



UNIVERSIDAD AUTÓNOMA METROPOLITANA
UNIDAD - IZTAPALAPA

DIVISIÓN DE CIENCIAS BÁSICAS E INGENIERIA
DEPARTAMENTO DE MATEMÁTICA

**INFERENCIA PARA MODELOS DE SUPERVIVENCIA
DE UN SOLO EVENTO Y EXTENSIONES PARA
MODELOS DE RIESGOS COMPETITIVOS**

T E S I S

QUE PARA OBTENER EL GRADO DE:

MAESTRO EN CIENCIAS

PRESENTA:

MARCO ANTONIO MATADAMAS SEGURA

Asesor de Tesis:

DR. GABRIEL ESCARELA PÉREZ

México, D.F., Abril de 2010

La Ley de la Atracción (Camilo Cruz)

Las personas no atraen hacia ellas aquello que quieren, sino aquello que son...FANTÁSTICO!

Camilo Cruz

El ser humano es el amo y señor de sus pensamientos, forjador de su carácter, creador y modelador de sus condiciones y de su entorno, y arquitecto de su propio destino. Cada persona esta donde esta por desición propia. Los pensamientos que han moldeado su carácter lo han llevado ahí, esto es válido tanto para aquellos que se sientes decepcionados con el mundo que los rodea, como para quienes están satisfechos con él. La persona que piensa que su vida es el resultado de condiciones externas suele ser victima de ellas. No obstante, cuando crea conciencia del poder creativo que reside dentro de ella, solo entonces se convierte en la dueña y señora de sus pensamientos. Siempre atraeremos aquello que ya se encuentra dentro de nosotros, tanto lo que amamos como lo que tenemos. Las circunstancias son simplemente los medios a través de los cuales recibimos aquello que merecemos o que creemos merecer.

James Allan

ÍNDICE GENERAL

Agradecimientos	1
Introducción	3
1. Conceptos Básicos	7
1.1. Características de los datos de supervivencia	8
1.1.1. Distribución de los datos	8
1.1.2. Datos Incompletos	8
1.2. Mecanismo de Censura	9
1.2.1. Censura por la derecha	9
1.3. Ejemplos de datos de supervivencia	14
1.3.1. Información Concomitante	14
1.3.2. Trasplante de Médula Osea por Leucemia	16
2. Análisis de Supervivencia Univariado	21
2.1. Función de supervivencia y función de riesgo	22
2.2. Modelos paramétricos: Distribuciones	30
2.2.1. Modelos de Mezclas Discretos (finitos)	34
2.3. Modelos paramétricos: Estimación	39
2.3.1. Método de Máxima Verosimilitud para datos censurados.	39
2.4. Ejemplo: Distribución Weibull	42
3. Estimación no-paramétrica de la función de supervivencia	47
3.1. La Distribución Empírica.	48
3.2. El Estimador Producto-Limite	48
3.2.1. El estimador-KM como un EMV	52
3.2.2. Intervalos de confianza del estimador Kaplan-Meier.	54
3.2.3. Integral producto y el estimador Nelson-Aalen	55
3.3. Comparación de distribuciones de supervivencia	60

4. Modelo de regresión para datos de supervivencia	65
4.1. Modelo de Riesgos Proporcionales	66
4.1.1. Función de supervivencia bajo riesgos proporcionales	67
4.2. El Modelo de Cox	69
4.2.1. Verificación de la suposición de riesgos proporcionales	70
4.2.2. Ajuste e interpretación del modelo de Cox	71
4.2.3. Verosimilitud Parcial: Un método de estimación.	74
4.2.4. Estimación de la función de supervivencia base.	78
4.2.5. Contraste de hipótesis para el Modelo de Cox ajustado	81
4.3. Ejemplo: Comparación de dos tratamientos médicos	82
5. Teoría de Riesgos Competitivos	89
5.1. Definición, suposiciones y algunos problemas básicos	89
5.2. Especificación del Modelo de Riesgos Competitivos	91
5.2.1. Función de Decremento Múltiple	91
5.2.2. La Función de Supervivencia Global	92
5.2.3. Distribución Conjunta de (T, J)	93
5.2.4. Riesgos Competitivos Independientes	99
5.3. Métodos Estadísticos de Inferencia	101
5.4. Identificabilidad y la Función de Decremento Múltiple	107
5.5. Razón de Riesgos Proporcionales	109
5.6. Modelado de Riesgos Competitivos	110
5.6.1. Modelo de Mezclas de Larson & Dinse (1985)	113
Conclusiones	119
A. Método de Máxima Verosimilitud	121
B. El Algoritmo EM	125

AGRADECIMIENTOS

A mis padres:

ESTELA SEGURA Y MARCO A. MATADAMAS,

Gracias por motivarme en la consecución de mis metas, así como el amor, la educación y el apoyo que me han dado: son unos padres maravillosos.

A mis hermanos:

NESTOR, MANUEL Y RAFAEL,

A la Licenciada:

ADRIANA MARGARITA RIVERA RENDON,

Por el inmenso amor y el apoyo incondicional que me brindó durante mis estudios de Posgrado. Gracias por ser una pareja maravillosa y por compartir conmigo momentos tan hermosos: TE AMO con letras mayúsculas.

A la que ha sido mi casa de estudios superiores y de posgrado:

UNIVERSIDAD AUTÓNOMA METROPOLITANA (Iztapalapa).

Gracias a los trabajadores de sus diferentes áreas, en particular al Depto. de Matemáticas y a la Asistente de Posgrado por su excelente labor, Mtra. Iseo González CH. También, gracias por los amigos que me permitiste hacer.

A mis sinodales:

Doctora SILVIA RUIZ VELASCO ACOSTA

Doctor JOSÉ AURELIO VILLASEÑOR ALVA,

que con sus comentarios y correcciones me ayudaron a mejorar esta Tesis.

Doctor RUSSELL JAMES BOWATER,

gracias por su apoyo.

Doctor GABRIEL ESCARELA PÉREZ

siempre tenías una palabra de aliento para que siguiera con mis estudios de posgrado y, a pesar de las circunstancias, confiaste en que acabaría esta tesis.

Al Honorable

CONACyT,

por el apoyo económico durante la realización de mis estudios de Posgrado.

INTRODUCCIÓN

El Análisis de Supervivencia es un conjunto de técnicas que permiten estudiar la variables tiempo Y hasta la ocurrencia de cierto evento de interés, dentro de un *periodo de seguimiento o de estudio* establecido. El evento de interés a observar puede ser de diferente índole, dependiendo del contexto que se considere, por ejemplo, Y puede medir el tiempo hasta que ocurra la muerte de un enfermo, tiempo hasta la recurrencia de la enfermedad, tiempo hasta que la pieza de una máquina falle, etc. Al periodo de tiempo que tarda en ocurrir el evento de interés comúnmente se le llama *tiempo de supervivencia o tiempo de fallo*. Hoy en día, tales técnicas estadísticas son una pieza fundamental en aplicaciones de Ingeniería (estudios de fiabilidad de equipos) y Economía, así como en Ciencias Biológicas y de la Salud (investigaciones medicas).

Cuando el evento de interés no es observado en un individuo dentro del periodo de estudio, se dice que su tiempo de supervivencia esta *censurado*. Otras de las características de la variable tiempo Y es que, en general, su distribución es asimétrica. Por eso, se requieren de técnicas especiales para estudiar la variable tiempo Y , ya que las técnicas convencionales utilizadas en estadística descriptiva e inferencial no son adecuadas para datos de tiempos de supervivencia.

Ahora bien, ¿Cómo se resume la información de datos de supervivencia? En general, se resumen mediante funciones: la función de distribución de supervivencia (FDS) que describe la proporción de individuos que han sobrevivido al evento de interés en un tiempo dado y , denotada como $S_Y(y) = \Pr(Y > y)$, y la función de riesgo condicional (*hazard function*), denotada como $h_Y(y)$, que nos da información del cambio en la probabilidad de que ocurra el evento en un instante de tiempo t , dado que no ha ocurrido. Entonces, asumiendo un modelo paramétrico, semi-paramétrico o no-paramétrico, para esas funciones, la metodología empleada en Análisis de Supervivencia permite estimar esas cantidades, en base a un conjunto de datos de tiempos de supervivencia que puede venir acompañado de información concomitante, con la particularidad de que tal metodología usa la *información parcial* proporcionada por las observaciones censuradas. Un supuesto importante en que se basan esas metodologías es que la causa de censura de un individuo es independiente del tiempo de supervivencia (mecanismo de censura *no-informativo*).

Una situación más general es cuando el evento de interés no ha podido ser obser-

vado (o su probabilidad de ocurrencia ha sido alterada) debido a algún otro evento, llamado *evento de riesgo competitivo*. El conjunto de datos de este tipo surge cuando una población esta expuesta a J tipos o modos de falla (mutuamente exclusivos), y el tiempo de supervivencia o falla de un individuo es debido una de esas causas. Entonces, puede haber varios riesgos competitivos, tanto *dependientes como independientes*. Al conjunto de tiempos de supervivencia resultantes se le ha llamado *datos de riesgos competitivos*. Debido a la variedad de disciplinas que dan lugar a este tipo de datos, tales como epidemiología, oncología e ingeniería, entre otras, su análisis a llegado ha ser una de las áreas de mayor interés para los investigadores en Análisis de Supervivencia.

Para el estudio de la ocurrencia de un evento de interés en presencia de eventos de riesgos competitivos se requiere de técnicas más elaboradas que las empleadas en Análisis de Supervivencia estándar. Sin embargo, algunos de esos métodos estándar (mediante ciertas adaptaciones) pueden ser usadas, mientras que otros son de uso limitado. Por tal motivo, si uno esta interesado en adentrarse por primera vez en el análisis de datos de riesgos competitivos, inclusive si se esta familiarizado con este tema, es necesario tener un conocimiento de los supuestos en que se basan esos métodos estándar y saber los conceptos básico de la Teoría de Riesgos Competitivos, así como los problemas fundamentales que hay en la interpretación de los resultados de un análisis de riesgos competitivos, por que de lo contrario existirá incertidumbre en el planteamiento de la metodología empleada.

Existe una gran variedad de información en Análisis de Supervivencia univariado, así como varios programas estadísticos en los que ya vienen implementadas las metodologías para hacer un análisis de datos de tiempos de falla.

Por otro lado, hay libros de análisis de supervivencia en los que se incluye un Capítulo para Riesgos Competitivos (e.g. [16],[24]) y libros, como el de Crowder [5], en los que se proporciona exclusivamente tanto la teoría como la metodología para el análisis de riesgos competitivos. También hay varios artículos para el análisis y modelado de este tipo de datos.

En cualquiera de los dos casos, es importante entender, hasta cierto punto, tanto las metodologías empleadas en Análisis de Supervivencia univariada como las usadas en Riesgos Competitivos, con la finalidad de hacer un buen uso de los programas estadísticos en los que vienen implementadas e interpretar adecuadamente los resultados obtenidos mediante estos.

El objetivo del presente trabajo es unificar, desde un punto de vista global, lo más relevante de la teoría de análisis de supervivencia univariado y la teoría básica de riesgos competitivos, mediante un planteamiento probabilístico y sin entrar en un rigor matemático muy formal, permitiendo al lector un entendimiento global de los métodos de análisis de datos de eventos que ocurren con el tiempo. Otro de los objetivos es el de incitar al lector a interactuar con algún programa estadístico, ya que en la actualidad es de gran importancia. En particular, se usa el paquete estadístico **R**, el cual es de distribución libre y puede ser descargado desde su sitio oficial de internet

<http://cran.r-project.org/>. En ese mismo sitio se pueden descargar los paquetes necesarios. Por ejemplo, el paquete `cmprsk` para una análisis de riesgos competitivos por sub-distribuciones. De esta manera, junto con los libros y artículos citados, la presente tesis puede servir como referencia en el estudio de Análisis de Supervivencia e introducción a la Teoría de Riesgos Competitivos.

El material de esta tesis esta organizado como sigue: en el Capítulo 1 se dan los conceptos básicos de Análisis de Supervivencia, las características y representación de un conjunto de datos de este tipo, así como los ejemplos que se usaran para ilustrar el contenido de los capítulos restantes. En el Capítulo 2 se revisan las cantidades básicas para el análisis de datos de tiempos de falla, la función de supervivencia y la función de riesgo. También se describen brevemente algunos modelos de supervivencia que pueden ser adoptados para esas cantidades, y el Método de Máxima Verosimilitud para este tipo de modelos. Un método no-paramétrico para estimar la función de supervivencia se expone en el Capítulo 3: el estimador Kaplan-Meier. Debido a que los datos de supervivencia pueden venir acompañados de información concomitante, es necesario la adopción de un modelo de regresión. En el Capítulo 4 se presenta el modelo de regresión semi-paramétrico de Cox, el cual esta basado en una técnica estadística para explorar la relación entre el tiempo de supervivencia de un individuo y un conjunto de variables explicativas. En el Capítulo 5 se revisan las cantidades básicas para analizar datos de riesgos competitivos: la función de riesgo de causa-específica y la función de incidencia acumulada de causa-específica. También se exponen los problemas básicos que pueden surgir en el planteamiento de un análisis de riesgos competitivos. Por ejemplo, un enfoque clásico para el modelado de riesgos competitivos ha sido considerar la función de supervivencia conjunta de m tiempos de falla latentes, Y_1, Y_2, \dots, Y_m , correspondientes a m distintos modos de falla.

CAPÍTULO 1

CONCEPTOS BÁSICOS

El Análisis de Supervivencia es un conjunto de técnicas, las cuales consisten en la medición del tiempo, mediante una variable respuesta, de la ocurrencia de algún evento de interés, el cual puede depender de información concomitante.

Dependiendo del contexto, el *evento de interés* puede ser de diferente índole. Por ejemplo, en tratamientos médicos, el investigador puede considerar como evento a estudiar la recaída, recuperación o muerte del paciente a causa de la enfermedad que padece; en aplicaciones de Ingeniería, el evento de interés puede ser la falla de un componente físico, mecánico o eléctrico de ciertos equipos industriales, etc.

El tiempo (años, meses, semanas o días) de estudio de la ocurrencia de cierto evento de interés comienza en un punto inicial de observación bien definido, hasta establecer un punto final, por ejemplo, la fecha final de cierto tratamiento al que fue sometido un grupo de pacientes, o cuando cierto número de aparatos eléctricos han presentado la falla de interés. Nos referiremos a ese tiempo como *período de seguimiento* o tiempo de estudio¹.

El tiempo de la ocurrencia del evento de interés comúnmente es llamado *tiempo de supervivencia* (*survival time, en inglés*), y representa el período desde el comienzo de observación de una individuo hasta que experimente el evento de interés².

La *variable respuesta*, la cual mide el tiempo de supervivencia de un individuo, es una variable aleatoria Y con valores reales positivos y definida sobre un espacio de probabilidad $(\Omega, \mathcal{S}, \mathbb{P})$, donde Ω es el espacio muestral, \mathcal{S} es la σ -álgebra de eventos y \mathbb{P} es una medida en la σ -álgebra \mathcal{S} de subconjuntos de Ω , es decir, \mathbb{P} es la medida de probabilidad en (Ω, \mathcal{S}) . Nos referiremos a este tipo de variables aleatorias como *variables aleatorias de supervivencia (v.a.s.)*.

Hoy en día, tales técnicas estadísticas son una pieza fundamental en aplicaciones de Ingeniería (Estudios de fiabilidad de equipos, así como en Ciencias Biológicas y de la Salud (investigaciones medicas). Por ejemplo, en aplicaciones de Ingeniería, uno podría estar interesado en saber el tiempo que tarda en fallar un componente

¹ El período de seguimiento puede ser asignado tanto a la muestra completa de individuos correspondiente a cierta población como individualmente.

² Generalmente, a este evento final también se le llama **falla**. Por esta razón, el tiempo de supervivencia también es llamado *tiempo de falla*.

físico, mecánico o eléctrico (*failure time*). En Ciencias Biológicas y de la Salud se interesan por el tiempo de supervivencia de un ser vivo, como es el caso de un paciente con alguna enfermedad crónica. En este caso, la variable respuesta mide el tiempo transcurrido (tiempo de supervivencia) entre el inicio del tratamiento y la consecución de un cierto evento (*v.g.* su muerte)

El término *supervivencia* se debe a que en las primeras aplicaciones de este método de análisis se utilizaba como evento la muerte de un paciente (Tablas de Vida), aunque cabe mencionar que a mediados del siglo pasado ya se comenzaba con este tipo de análisis en aplicaciones de ingeniería, fiabilidad de equipos militares (Segunda Guerra Mundial) y productos comerciales.

Ya que la aplicación práctica de toda la teoría dada en el presente trabajo será el área de Ciencias Biológicas y de la Salud, en particular, investigaciones médicas relacionadas con grupos de pacientes sometidos a períodos de observación, el evento de interés estará relacionado con la salud de estos, inclusive con su vida. El tiempo de supervivencia verdaderamente representará el tiempo de vida de ese paciente, desde un punto inicial de observación, hasta experimentar el evento de interés.

1.1. Características de los datos de supervivencia

Comencemos con varias cuestiones relativas al análisis de datos de supervivencia. ¿Por qué no se debe analizar este tipo de datos con los procedimientos tradicionales que se utilizan en estadística descriptiva e inferencial (procedimientos paramétricos)? ¿Por qué son los métodos no-paramétricos y semiparamétricos tan populares en análisis de supervivencia?, ¿Por qué necesitamos una teoría estadística especial para este tipo de datos?. Las respuestas a estas cuestiones tienen que ver con problemas en la medición de la variable tiempo de supervivencia y con la distribución de los datos recogidos.

1.1.1. Distribución de los datos

En gran parte del análisis estadístico se tiene la hipótesis de que los datos provienen de una población normal, sin embargo, los datos de supervivencia—por lo general—no son distribuidos simétricamente. Por eso es que surge la necesidad de considerar métodos *no-paramétricos y semiparamétricos*, es decir, aquellos métodos que no presuponen que los datos se ajustan a una distribución específica para el análisis de datos de supervivencia.

1.1.2. Datos Incompletos

Suponga que se establece un período de seguimiento de un grupo de pacientes sometidos a un estudio o tratamiento, comenzando en un tiempo origen bien definido, $t = t_0$ y concluyendo en una fecha premeditada, $t = t_f$. En el transcurso del estudio, puede ocurrir que algunos pacientes lo abandonen al tiempo c o lleguen al final de este sin experimentar el evento de interés, registrándose información parcial

(*censurada*) de su tiempo de supervivencia, a saber, que su tiempo de supervivencia es mayor que c o que t_f . Por otro lado, los pacientes que experimentaron el evento de interés dentro del período de observación proporcionaron información completa, *no-censurada*, de su tiempo de supervivencia. Uno desearía obtener siempre datos de tiempos de supervivencia completos, pero en la mayoría de los estudios, la realidad de las cosas es lo contrario.

Definición 1.1. *El tiempo de supervivencia de una persona está censurado cuando el evento de interés no ha sido observado durante el período de seguimiento.*

De esta manera, en el caso del tratamiento médico, si Y es la v.a.s. correspondiente a cierto paciente, su tiempo de supervivencia se considera como censurado cuando $Y > c$, donde c es su *tiempo de censura* asociado, y se considera como no-censurado cuando $Y \leq c$. Entonces, otra característica distintiva del Análisis de Supervivencia es que aprovecha esa información parcial proporcionada por la censura.

1.2. Mecanismo de Censura

Hay escenarios en los que se presentan diferentes mecanismos de censura:

- *censura por la derecha (o por la izquierda);*
- *censura por intervalos.*

Para propósitos del presente trabajo nos enfocaremos a datos *censurados por la derecha*, y asumiremos que el mecanismo de este tipo de censura es *no-informativo* en el sentido de que si el tiempo de supervivencia Y de un individuo está censurado al tiempo C , todo lo que se sabe es que $Y > C$, y el mecanismo de censura no proporciona alguna otra información acerca del tipo de falla eventual de ese individuo (Lagakos, 1979a).

Una información más amplia de los diferentes mecanismos de censura puede ser consultada en cualquier libro de Análisis de Supervivencia, por ejemplo, el libro de Klein et. al., 1997.

1.2.1. Censura por la derecha

Este mecanismo de censura es el caso más común de datos incompletos y se caracteriza de los demás en el hecho de que el tiempo de censura observado, digamos C , proporciona solamente una cota inferior para el tiempo de falla Y , y del cual, no se sabe cuando tendría lugar. Algunos tipos de censura por la derecha son: *Tipo I*, *Tipo II* y *censura aleatoria*.

1.2.1.1. Censura de Tipo I

En este tipo de censura, el período de seguimiento tiene una fecha de inicio, $t = 0$, y un tiempo final, $t = t_c$, que es el tiempo máximo de observación para la ocurrencia del evento de interés.

Notación 1. Sea Y_1, Y_2, \dots, Y_n una muestra de v.a.s. i.i.d. Sea t_c el tiempo de censura fijo. En vez de observar las variables aleatorias de interés, Y_i 's, solamente observamos los pares

$$(T_1, \delta_1), (T_2, \delta_2) \dots, (T_n, \delta_n),$$

donde

$$\delta_i = \mathbf{1}\{Y_i \leq t_c\} = \begin{cases} 1, & \text{si } Y_i \leq t_c \text{ (no-censurado);} \\ 0, & \text{si } Y_i > t_c \text{ (censurada).} \end{cases}$$

$$T_i = \min(Y_i, t_c) = \begin{cases} Y_i & \text{si } \delta_i = 1; \\ t_c & \text{si } \delta_i = 0. \end{cases}$$

Ejemplo 1.1. Sea Y_1, Y_2, \dots, Y_6 v.a.s. i.i.d. correspondientes a un grupo de 6 pacientes, los cuales presentan una enfermedad mortal y son sometidos a un período de observación, simultáneamente, comenzando en un tiempo t_0 y finalizando en un tiempo t_f . Suponga que el evento de interés es la muerte. El período de seguimiento de esos seis pacientes se ilustra gráficamente en la figura 1.1. El tiempo de supervivencia de los pacientes que murieron (D, death) pudo ser registrado, en cambio, el tiempo de supervivencia de los que llegaron con vida (A, alive) al final del estudio, fue censurado por t_f . Así, observamos los pares

$$(T_1, \delta_1), (T_2, \delta_2), \dots, (T_6, \delta_6).$$

donde

$$(T_i = t_f, \delta_i = 0) \quad \text{para } i = 3, 4, 6$$

son observaciones censuradas, y

$$(T_i = y_i, \delta_i = 1) \quad \text{para } i = 1, 2, 5$$

son observaciones no-censuradas. □

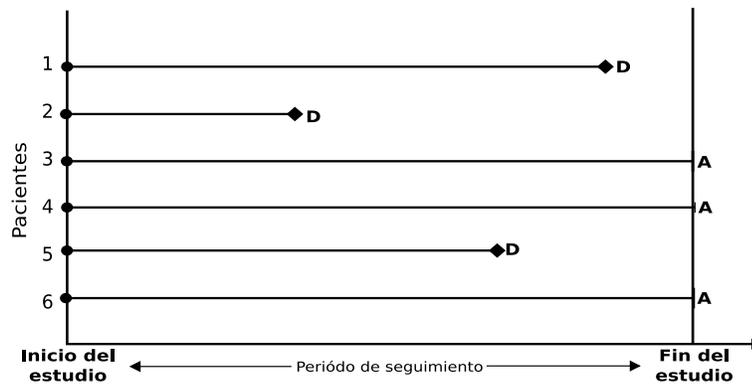


Figura 1.1: Censura de Tipo I. El número de censuras es aleatorio, pero la duración total del estudio es fijo.

Observación 1.1. Una generalización de la censura de Tipo I es cuando los pacientes son reclutados al estudio en tiempos diferentes y se les asigna a cada uno un tiempo máximo de observación, llamado su *tiempo de censura fijo*. Este puede diferir de un caso a otro y ser redefinido en el transcurso del estudio.

1.2.1.2. Censura de Tipo II

Suponga que, para una muestra de v.a.s. Y_1, Y_2, \dots, Y_n , el período de observación se termina cuando un número predeterminado, digamos $r (< n)$ de individuos experimentan el evento de interés, por lo que el tiempo de supervivencia de los r pacientes es conocido y los restantes $(n - r)$ pacientes son censurados a la derecha por el tiempo de supervivencia de la r -ésima observación.

Lo que se observa son los r estadísticos de orden más pequeños:

$$Y_{(1)} \leq Y_{(2)} \leq Y_{(3)} \leq \dots \leq Y_{(r-1)} \leq Y_{(r)}.$$

Ejemplo 1.2. Suponga que se tiene un lote de n transistores o tubos y se ponen a prueba en $t = 0$, para posteriormente registrar sus tiempos de falla, es decir, el tiempo que tardan en quemarse. Ya que algunos transistores pueden tardar un período muy largo en quemarse y, debido a ciertas circunstancias, no es posible esperar todo ese tiempo para finalizar el experimento, por lo que se toma la decisión de esperar hasta una preespecificada fracción r/n de transistores quemados, teniendo una censura de Tipo II. \square

Observación 1.2. En la censura de Tipo II, la duración total del estudio es aleatorio, pero el número de observaciones censuradas y no-censuradas es fijo.

1.2.1.3. Censura aleatoria

En este caso, el investigador no sabe cuando se presentara la censura, por lo que no es un tiempo fijo del todo y puede variar de un individuo a otro. Nos podemos preguntar el motivo de la presencia de este tipo de *censura aleatoria* en los datos recogidos y se nos pueden ocurrir muchos. Por eso, considere tres razones generales que causarían el individuo a estar censurado aleatoriamente por la derecha, con respecto al evento de interés:

1. *Pérdida del seguimiento.* El individuo bajo observación decide marcharse a otro lugar y la única información que se tiene es cuando se le vió libre del evento por última vez.
2. *Removido del estudio.* El individuo es retirado del estudio debido a que experimentó algún otro evento ajeno al de interés (riesgos competitivos). En este caso, hay que verificar que tal suceso no este relacionado con el evento que se desea observar. Otro motivo puede ser que el individuo presento reacciones adversas al tratamiento.
3. *Terminación del estudio.* Al final del estudio, el individuo no experimentó el evento de interés, dando una observación censurada.

Distribución de (T, δ)

Sean $F(y)$ y $G(t)$ las funciones de distribución acumuladas (FDA) de Y y C , respectivamente, donde Y es una v.a.s. y C otra variable aleatoria (v.a.) positiva denotando el tiempo de censura asociado a Y . Suponga que las variables aleatorias Y y C son independientes y continuas, con función de densidad $f(\cdot)$ y $g(\cdot)$, respectivamente. También supongase que $G(t)$ no depende de ninguno de los parámetros de $F(t)$. Así, mientras que la distribución $F(t)$ se considera de primordial interés, $G(t)$ se considera como un tedioso parámetro desconocido. En este modelo de censura, lo que se observa por unidad muestral es el par aleatorio (T, δ) definido como antes por

$$T = \min(Y, C) \quad \text{y} \quad \delta = I\{Y \leq C\}.$$

La distribución del par aleatorio (T, δ) se obtiene de la siguiente manera: la función de distribución de una observación censurada es

$$\begin{aligned} F_1(t) = \Pr(\{T \leq t\} \cap \{\delta = 0\}) &= \Pr(\{C \leq t\} \cap \{C \leq Y\}) \\ &= \int_0^t \left[\int_u^\infty dF(v) \right] dG(u) \\ &= \int_0^t [1 - F(u)]g(u) d(u), \end{aligned} \quad (1.1)$$

y derivando con respecto a t en (1.1) se obtiene la función de densidad

$$\begin{aligned} f_1(t) = \Pr(T = t, \delta = 0) &= \Pr(\{C = t\} \cap \{C \leq Y\}) \\ &= \frac{d}{dt} F_1(t) \\ &= [1 - F(t)]g(t). \end{aligned} \quad (1.2)$$

Análogamente, la función de distribución de una observación no-censurada es

$$\begin{aligned} F_2(t) = \Pr(T \leq t, \delta = 1) &= \Pr(\{Y \leq t\} \cap \{Y \leq C\}) \\ &= \int_0^t \left[\int_u^\infty dG(v) \right] dF(u) \\ &= \int_0^t [1 - G(u)]dF(u). \end{aligned} \quad (1.3)$$

La función de densidad asociada a (1.3) es

$$\begin{aligned} f_2(t) = \Pr(\{T = t\} \cap \{\delta = 1\}) &= \Pr(\{Y = t\} \cap \{Y \leq C\}) \\ &= \frac{d}{dt} F_2(t) \\ &= [1 - G(t)]f(t). \end{aligned} \quad (1.4)$$

Bajo la suposición de que todos los tiempos de supervivencia y de censura son mutuamente independientes, una muestra observada de tamaño n del par aleatorio (T, δ) consiste de los pares

$$(t_1, \delta_1), (t_2, \delta_2), \dots, (t_{n-1}, \delta_{n-1}), (t_n, \delta_n),$$

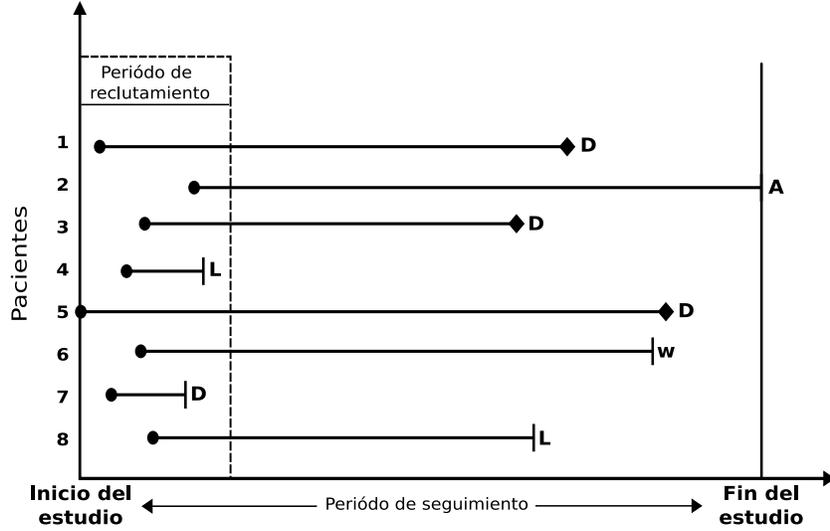


Figura 1.2: Censura aleatoria por la derecha.

Por lo tanto, la función de densidad de probabilidad (f.d.p.) del par observado (t_i, δ_i) es

$$\Pr(T_i = t_i, \delta_i) = [f(t_i)G(t_i)]^{\delta_i} [g(t_i)(1 - F(t_i))]^{1-\delta_i}. \quad (1.5)$$

La generalización de todas esas ideas se presentan en la *Teoría de Riesgos Competitivos (Competing Risks Theory)*.

Ejemplo 1.3. Sea Y_1, Y_2, \dots, Y_8 v.a.s. i.i.d. correspondientes a un grupo de ocho pacientes, los cuales presentar una enfermedad mortal y son sometidos a un período de observación, comenzando en un tiempo t_0 y finalizando en un tiempo t_f . El evento que se desea observar es la muerte del paciente como resultado de su enfermedad. Suponga que los pacientes pueden entrar en tiempos diferentes al estudio, por lo que se considera un período de reclutamiento. El período de seguimiento de esos ocho pacientes se ilustra gráficamente en la Figura 1.2. En el transcurso del período de estudio, se presentan observaciones censuradas por parte de pacientes a los cuales se les perdió el seguimiento (L , lost to follow-up) o se les retiró del estudio (W , withdrawals) debido a que experimentaron otra causa ajena e independiente a la de interés.

Otros pacientes murieron debido a su enfermedad mortal (D , death) a lo largo del período de seguimiento, por lo que sus tiempos de supervivencia pudieron ser registrados.

Al final del estudio, el tiempo de supervivencia de los pacientes que permanecieron con vida (A , alive) son censuradas por el tiempo en el que concluyó este, que es t_f . Así, observamos los pares

$$(t_1, \delta_1), (t_2, \delta_2), \dots, (t_8, \delta_8).$$

□

En aplicaciones de Ciencias de la Salud, en contraste con aplicaciones de Ingeniería, *la censura es más una parte de situaciones experimentales que un asunto de diseño premeditado.*

1.3. Ejemplos de datos de supervivencia

En datos de supervivencia con censura, por lo regular se acostumbra a escribir t para una observación no censurada y t^+ para una observación censurada, donde t y t^+ son valores específicos de interés para la *v.a.s.* T . Así, nuestros datos consisten en observaciones de la forma $1^+, 8^+, 6, 13, 25, 14^+, 2, 5, 16^+$, que al representarlos como pares ordenados $(t_{(i)}, \delta_{(i)})$, se obtiene

$$(1, 0), (2, 1), (5, 1), (6, 1), (8, 0), (13, 1), (16, 0), (25, 1),$$

donde $t_{(i)}$ es una realización de $T_i = \min(Y_i, C_i)$ y δ_i su indicadora de censura.

El principal paquete del programa estadístico **R** empleado para Análisis de Supervivencia es `survival`. En el paquete `KMsurv` del mismo programa se incluyen todos los conjuntos de datos del libro de Klein y Moeschberger [19], y en la página de la UCLA [35] se proporciona el aspecto computacional mediante diferentes paquetes estadísticos. Más adelante se dan ejemplos de datos de supervivencia, los cuales vienen incluidos en tales paquetes, que serán empleados para ilustrar el contenido de los Capítulos 1,2,3 y 4, mediante el uso de **R**. Para usar esos paquetes se escriben los siguientes comandos

```
> install.packages('KMsurv') #si aun no esta instalado;
> library(survival) #carga las librerias del paquete 'survival';
> data(leukemia) #lee el conjunto de datos "leukemia"
```

Cada paquete viene documentado. No obstante, para una consulta rápida se pueden usar los comandos `library(help=survival)` y `help(nombrefuncion)`.

1.3.1. Información Concomitante

En muchos estudios médicos que dan lugar a datos de supervivencia, se registra información adicional por cada individuo y de la cual se cree que depende el tiempo de supervivencia Y . Las variables mediante las cuales se observa este tipo de información se les llama **variables explicativas** y son denotadas por X . Debido a la influencia que puede tener cierto conjunto de covariantes $\{X_j : j = 1, 2, \dots, p\}$ sobre el tiempo de supervivencia, se usa el término de **variable respuesta**³ o **dependiente** para Y , mientras que para las X_j el de **variables predictivas** o **independientes**. Por ejemplo, en un trasplante de médula ósea, X puede denotar la edad o el sexo del paciente/donador, el tipo de tratamiento al que se sometió el paciente, etc. Todas esas variables explicativas se pueden considerar como un vector de covariantes:

$$\mathbf{X}^T = (X_1, X_2, X_3) = (\text{edad}, \text{sexo}, \text{tipo de tratamiento}),$$

³ El término respuesta se usa en el sentido de que la medida Y varía en respuesta de los valores observados del conjunto de variables explicativas.

donde \mathbf{X}^T denota el *vector transpuesto* de \mathbf{X} .

Entonces, para un individuo en particular, se observa el vector $\mathbf{X} = \mathbf{x}$, donde

$$\mathbf{x}^T = (x_1, x_2, x_3).$$

Las variables respuesta y las variables explicativas se pueden clasificar, según su escala de medición, como

1. **Variables Cuantitativas:** Son variables que se pueden medir expresándose numéricamente (cuantificar). Estas variables pueden ser de dos tipos:

- *Continuas*, cuando admiten tomar cualquier valor dentro de un rango numérico determinado. Por ejemplo, el tiempo de supervivencia de un individuo en un estudio médico, la edad, el peso, la talla, etc.
- *Discretas*, si solamente toman valores enteros, por lo que no admiten todos los valores intermedios en un rango dado (e.g. el número de hijos o de hermanos).

2. **Variables Cualitativas.** Son variables que representan distintas cualidades que posee un individuo de cierta población, o alguna cosa (no cuantificables). Cada una de esas cualidades presentes se denomina atributo o categoría, y la medición de éstas variables consiste en la clasificación de dichos atributos. Dependiendo de los valores que tome una variable cualitativa, ésta puede ser *dicotómica* (e.g. la variable SEXO sólo puede adoptar uno de los dos atributos que posee: varón o mujer), o bien, *politómica*. En el proceso de medición de las variables cualitativas, se pueden utilizar dos escalas

- *Escala ordinal:* La clasificación de las categorías correspondientes a la variable cualitativa presentan un orden natural, por ejemplo, la variable EDAD agrupada⁴ en tres categorías: joven, adulto (maduro) y anciano.
- *Escalas nominales.* A las categorías de la variable cualitativa no se les puede asignar un criterio de orden tanto inherente como jerárquico. Por ejemplo, la escala de colores, sexo, profesión, etc.

Una variable explicativa cualitativa es llamada **factor** y a sus categorías se les denomina *niveles* para el factor. Por ejemplo, la variable cualitativa EDAD es un factor clasificado en tres niveles: joven, adulto y anciano; el factor GÉNERO con los niveles varón y mujer, etc.

En general, se tiene un muestra de v.a.s. i.i.d, Y_1, Y_2, \dots, Y_n , censuradas por la derecha por tiempos de censura C_1, C_2, \dots, C_n y vector de variables explicativas $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$, respectivamente, donde

$$\mathbf{X}_i^T = (X_{i1}, X_{i2}, \dots, X_{ij}, \dots, X_{ip}), \quad i = 1, 2, \dots, n,$$

⁴ Las variables de medida continua, como lo es la edad de una persona, pueden ser objeto de categorización mediante intervalos.

son vectores de p variables explicativas. De esta manera, lo que observamos son ternas de la forma

$$(T_i, \delta_i, \mathbf{X}_i), \quad \text{para } i = 1, 2, \dots, n.$$

Los datos observados se resumen en la Tabla 1.1.

i	$T_i = t_i$	δ_i	\mathbf{x}_i^\top (vector de covariantes)					
1	t_1	δ_1	x_{11}	x_{12}	\cdots	x_{1j}	\cdots	x_{1p}
2	t_2	δ_2	x_{21}	x_{22}	\cdots	x_{2j}	\cdots	x_{2p}
\vdots						\vdots		\vdots
i	t_i	δ_i	x_{i1}	x_{i2}	\cdots	x_{ij}	\cdots	x_{ip}
\vdots						\vdots		\vdots
n	t_n	δ_n	x_{n1}	x_{n2}	\cdots	x_{nj}	\cdots	x_{np}

Tabla 1.1: Representación más general de la terna $(T_i, \delta_i, \mathbf{X}_i)$, con $\mathbf{X}_i \in \mathbb{R}^p$.

1.3.2. Trasplante de Médula Osea por Leucemia

El trasplante de células madre hematopoyéticas⁵ (H SCT, por sus siglas en inglés) consiste en el trasplante de células madre provenientes de la médula ósea. Este tipo de trasplante es un procedimiento médico que tiene lugar en los campos de la hematología y la oncología, y que por lo general, es realizado a personas con enfermedades en la sangre, médula ósea o ciertos tipos de cáncer. La mayoría de los receptores de H SCT son pacientes con mieloma múltiple o leucemia que no se beneficiarán con un tratamiento prolongado o que ya son resistentes a la quimioterapia.

Tipos de injerto

- **Autólogo.** En este caso, el donador y el receptor del injerto son la misma persona, por lo que la probabilidad de incidencia de una paciente de experimentar GvHD⁶ es casi nula. El H SCT autólogo tiene un menor riesgo de rechazo del injerto e infección, ya que la recuperación del sistema inmunológico es rápida.
- **Alogénico.** En este caso, hay dos personas involucradas: el donante (sano) y el receptor (paciente). Los donadores de H SCT alogénico deben tener un tipo de tejido (HLA) que coincida con el receptor.

⁵ La hepatopoyesis consiste en la generación de todas las células sanguíneas a partir de la célula troncal hematopoyética (CTH) o *célula madre*.

⁶ Enfermedad injerto-contra- intruso (Graft-vs-Host Disease). Es una complicación común de H SCT alogénico como resultado del trasplante de médula ósea, en la cual las células inmune-funcionales reconocen al receptor como una agente extraño y montan un ataque inmunológico.

Fuentes de HSC y almacenamiento

Para limitar el riesgo de rechazo del trasplante de células madre o de experimentar una GvHD aguda, en el caso de HSCT alogénico, el donador preferentemente deberá tener el mismo HLA que el destinatario. Las células madre pueden permanecer congeladas durante largos períodos de tiempo. En cuanto a las fuentes de HSCT esta la sangre periférica de células madre y la sangre del cordón umbilical.

Acondicionamiento del tratamiento

La administración de quimioterapia o radioterapia inmediatamente antes de un trasplante se le denomina acondicionamiento o régimen preparatorio, cuyo objetivo es contribuir a la erradicación de la enfermedad del paciente, antes de la infusión de HSC y suprimir las reacciones inmunes. Hay dos tipos de acondicionamiento:

- *Trasplante mieloablatoivo*. Tratamiento con alta dosis de quimioterapia.
- *Trasplante no-mieloablatoivo*. Tratamiento con intensidad reducida de quimioterapia y radioterapia.

Estado de remisión de un paciente con Leucemia

Se considera que un paciente está en *remisión completa (o parcial)* cuando todos (o varios de) los signos y síntomas de cáncer han desaparecido en respuesta a cierto tratamiento, es decir, es el período durante el cual la enfermedad del paciente está "bajo control", pero no curado de ella, por lo que puede recaer debido a esta.

Especificaciones conceptuales de análisis de supervivencia

- Después del trasplante, sigue un proceso de recuperación en el que el paciente esta en remisión (completa o parcial). Entonces, considere los eventos de interés a observar : la recaída del paciente debido a que su leucemia regreso(fuera de remisión) y muerte relacionada con el tratamiento (GvHD).
- El *tiempo de supervivencia* representa la longitud de tiempo en el que los pacientes están vivos y libres de la enfermedad después del trasplantes (duración del período de remisión).
- *Información concomitante*. En HSCT, el pronóstico de recuperación puede depender de *factores de riesgo* como pueden ser: etapa inicial de la enfermedad, edad y sexo del paciente y/o donador, la cantidad de medicamento suministrado, el tipo de trasplante que se le practico (Autólogo o Alogénico), etc.

Ejemplo 1.4. *Trasplante de Médula Osea Autólogo y Alogénico.*

(*Paquete: KMsurv; nombre: alloauto; variables: time, type y delta*). Este conjunto de datos contiene el registro de las longitudes de remisión de 101 pacientes con avanzada Leucemia Mielógena Aguda (LMA), reportado por el Registro Internacional de Trasplantes de Médula Osea. Cincuenta y uno de esos pacientes recibieron

Grupo A (referencia): no mantuvo quimioterapia						
$(t_{(i)}, \delta_{(i)}) :$	(5,1)	(5,1)	(8,1)	(8,1)	(12,1)	(16,0)
	(23,1)	(27,1)	(30,1)	(33,1)	(43,1)	(45,1)
Grupo B: mantuvo quimioterapia						
$(t_{(i)}, \delta_{(i)}) :$	(9,1)	(13,1)	(13,0)	(18,1)	(23,1)	(28,0)
	(31,1)	(34,1)	(45,0)	(48,1)	(161,0)	

Tabla 1.2: (Fuente: Rupert G. Miller, 1997). Longitud de remisión (completa), en semanas, de los pacientes con LMC.

un HSCT autólogo mieloablativo ($type = 2$), y cincuenta pacientes tuvieron un HSCT alogénico ($type = 1$).

Uno de los objetivos en estudios de personas que experimentaron un trasplante de médula ósea es la medición, a través de la longitud de su remisión, de la efectividad de esos dos métodos de trasplante. Los pacientes fueron seguidos desde su trasplante hasta que salieron de remisión o hasta que terminó el período de seguimiento. \square

Ejemplo 1.5. Estudio del mantenimiento de quimioterapia para pacientes con leucemia.

(Paquete: *survival*; nombre: *leukemia*; variables: *time*, *status*, *x*). Los datos de la Tabla 1.2 se refieren a un tratamiento clínico para evaluar la eficiencia de seguir recibiendo quimioterapia para combatir la Leucemia Mielógena Crónica⁷ (LMC). Después de investigar el estado de remisión inducido por tratamientos con quimioterapias, los pacientes quienes entraron al estudio fueron aleatoriamente divididos en el grupo A y el grupo B. El grupo B siguió manteniendo el tratamiento con quimioterapia. El grupo A, llamado grupo de control o referencia, ya no mantuvo el tratamiento con quimioterapia. El objetivo de la prueba fue ver si mantener la quimioterapia prolonga la ausencia de la enfermedad (longitud de remisión) hasta experimentar una recaída debido a esta. \square

Ejemplo 1.6. Duración de Remisión en una prueba médica para pacientes con leucemia. En la Tabla 1.3 se presentan una parte del conjunto de datos correspondiente a los tiempos de supervivencia (longitud de remisión) de 42 pacientes con leucemia como resultado de una prueba médica para comparar un tratamiento (droga 6-mercaptopurina) contra placebo. También se presenta información adicional:

- Rx: el tratamiento al que se somete el paciente. Rx=0 si es tratamiento con 6-MP; Rx=1 si es tratamiento con placebo.
- sexo: 1 = masculino, 0 = femenino.

⁷ La LMC es un cáncer que afecta a la sangre. Se le llama mielógeno porque afecta a un tipo particular de glóbulo blanco llamado mieloblasto, y se le considera una forma crónica de la leucemia porque el cáncer tiende a crecer lentamente durante un período prolongado.

- $\log WBC$: cantidad de glóbulos blancos (white bloodcell count), en escala logarítmica, para cada persona en el estudio⁸.

Estos pacientes fueron seleccionados de aquellos quienes tenían una remisión completa o parcial (de su leucemia) inducida por un tratamiento previo (droga prednisona⁹). Después fueron aleatorizados en dos grupos: el Grupo 1, llamado grupo de referencia o control, conformado con 21 pacientes, los cuales recibieron placebo¹⁰; otro grupo de 21 pacientes, con un tratamiento de 6-MP. Los pacientes estuvieron bajo observación hasta que salieron de remisión o hasta que finalizó el estudio.

Un paciente experimenta una falla cuando sale de remisión, es decir, cuando su leucemia regresa (recaída). Entonces, la variable respuesta tiempo, Y , mide el período desde que el paciente entro al estudio, estando en remisión, hasta que experimento una recaída debido a su enfermedad.

El tiempo de supervivencia Y puede estar censurado por la derecha, al tiempo C , debido a que permaneció en remisión hasta la fecha final del estudio, se le perdió el seguimiento, o se le retiró antes de concluir el estudio. Así, la muestra observada consiste de vectores de la forma

$$(t_i, \delta_i, \mathbf{X}_i), \quad i = 1, 2, \dots, 42,$$

donde t_i es una realización de $T_i = \min(Y_i, C_i)$, δ_i su indicadora de censura y \mathbf{X}_i es el vector de covariantes correspondiente a la i -ésima observación. \square

⁸ Las razones epidemiológicas para añadir esa información al conjunto de datos es que usualmente esa variable es considerada un importante predictor de supervivencia en pacientes con leucemia.

⁹ La prednisona es un fármaco corticosteroide sintético que se toma usualmente en forma oral, pero puede ser administrado por vía intramuscular (inyección) y es usado para un gran número de afecciones, en particular para prevenir y tratar rechazo de órganos en trasplantes.

¹⁰ Un placebo es una sustancia farmacológicamente inerte que se utiliza como control en la investigación clínica. Se le suministra a los pacientes para producir un efecto psicológico

i (paciente)	t (semanas)	δ (estatus)	X_{i1} (sexo)	X_{i2} (log WBC)	X_{i3} Rx
1	35	0	1	1.45	0
2	34	0	1	1.47	0
3	32	0	1	2.20	0
4	32	0	1	2.53	0
5	25	0	1	1.78	0
6	23	1	1	2.57	0
7	22	1	1	2.32	0
8	20	0	1	2.01	0
9	19	0	0	2.05	0
10	17	0	0	2.16	0
11	16	1	1	3.60	0
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
34	5	1	1	3.49	1
35	5	1	0	3.97	1
36	4	1	1	4.36	1
37	4	1	1	2.42	1
38	3	1	1	4.01	1
39	2	1	1	4.91	1
40	2	1	1	4.48	1
41	1	1	1	2.80	1
42	1	1	1	5.00	1

Tabla 1.3: (Fuente: Freireich et al., 1963.) Duración de la remisión de 6-MP contra placebo en niños con leucemia aguda. Este conjunto de datos aparece en internet con el nombre de "anderson".

ANÁLISIS DE SUPERVIVENCIA UNIVARIADO

El análisis de supervivencia univariado se refiere al análisis de muestras aleatorias independientes (*datos de supervivencia*) provenientes de poblaciones expuestas a un sólo tipo de falla.

Para introducirnos al análisis de datos de supervivencia se comenzará con el modelado de *poblaciones homogéneas*¹ a través de una simple representación matemática (modelo de supervivencia) para la distribución de probabilidad de la v.a.s. Y . Esa distribución se puede especificar a partir de diferentes cantidades, de las cuales, dos son de particular interés: una es mediante la **función de riesgo**, también llamada **tasa de mortalidad o fuerza de mortalidad**, denotada como $h(y)$, y la otra es la **función de supervivencia**, denotada como $S(y)$.

Considere la siguiente familia paramétrica de distribuciones

$$\mathcal{P} = \{F_{\theta} : \theta \in \Theta\}$$

donde θ es un parámetro finito-dimensional y Θ el espacio de valores de θ . Cuando se da una "forma específica" de la distribución de Y mediante un elemento de \mathcal{P} , entonces la función de supervivencia $S(y; \theta)$ y la función de riesgo $h(y; \theta)$ también presentan una forma específica.

Por otro lado, se puede adoptar un modelo no-paramétrico para la distribución de Y , por lo que es necesario el uso de métodos no-paramétricos para $S(y)$ y $h(y)$. Por ejemplo, en presencia de censura en los datos, un estimador no-paramétrico para la función de supervivencia de Y es el Estimador Kaplan-Meier.

¹ No hay distinción alguna entre los individuos bajo estudio, todos son susceptibles a experimentar el evento de interés, i.e. ausencia de variables explicativas.

2.1. Función de supervivencia y función de riesgo

Sea $(\Omega, \mathcal{S}, \mathbb{P})$ un espacio de probabilidad y $Y : \Omega \rightarrow \mathbb{R}^+$ una *v.a.s.* definida sobre este. Sea

$$F_Y(y) = \Pr(Y \leq y) \equiv \mathbb{P}(\{\omega \in \Omega : Y(\omega) \leq y\}),$$

la función de distribución acumulada de Y , donde $\{\omega \in \Omega : Y(\omega) \leq y\} \in \mathcal{S}$ para toda $y \in \mathbb{R}^+$. La cantidad cuantitativa que se emplea para describir la distribución del tiempo de supervivencia Y es la función de supervivencia.

Definición 2.1. *La función de supervivencia (FS), denotada por $S_Y(y)$, se define como*

$$\begin{aligned} S_Y(y) &= 1 - F_Y(y) \\ &= \Pr(Y > y) \equiv \mathbb{P}(\{\omega \in \Omega : Y(\omega) > y\}) \quad \forall y \in \mathbb{R}^+. \end{aligned}$$

La FS $S_Y(y)$ representa la probabilidad de que un individuo sobreviva desde un tiempo origen, digamos $y_0 = 0$, hasta algún punto mayor que el tiempo y . Como consecuencia de su definición, S_Y posee ciertas propiedades que la caracterizan, las cuales son resumidas en la siguiente proposición.

Proposición 2.1.1. *La función de supervivencia $S_Y(y)$ tiene las siguientes propiedades.*

- (i) $0 \leq S_Y(y) \leq 1 \quad \forall y \in \mathbb{R}^+$.
- (ii) $S_Y(y)$ es monótona decreciente: $y_1 \leq y_2$ implica $S_Y(y_1) \geq S_Y(y_2)$.
- (iii) $S_Y(0) = 1$ y $\lim_{y \rightarrow \infty} S_Y(y) = 0$.
- (iv) $S_Y(y)$ es continua por la derecha, es decir, para cada $y \in \mathbb{R}^+$ se tiene que

$$\lim_{\epsilon \rightarrow 0} S_Y(y + \epsilon) = S_Y(y).$$

Demostración. La demostración es similar a la que se da para las propiedades de $F_Y(y)$ (ver [36]). ■

Observación 2.1. En la Proposición 2.1.1, ya que Y es una v.a. positiva, en vez de considerar a $\lim_{y \rightarrow -\infty} S_Y(y)$ en la propiedad (iii), se adopta la convención $S_Y(0) = 1$.

Observación 2.2. Si se define la función de supervivencia como $S_Y(y) = \Pr\{Y \geq y\}$, entonces esta será una función continua por la izquierda.

En el contexto de Análisis de Supervivencia habrá distribuciones que posiblemente no cumplan con (iii) en la Proposición 2.1.1 (e.g. si los datos de supervivencia presentan censura), por lo que se dará una definición para distinguir entre ese tipo de funciones de supervivencia.

Definición 2.2. Sea Y una v.a.s. y sea $S_Y(y)$ su FS asociada. Se dice que $S_Y(y)$ es una FS **propia** en el sentido de que

$$S_Y(0) = 1 \quad \text{y} \quad \lim_{y \rightarrow \infty} S_Y(y) = 0.$$

Observación 2.3. La FS² $S_Y(y)$ es decreciente y acotada inferiormente. Por eso, si para $\tau \in \mathbb{R}^+$ dado, tenemos que $y \uparrow \tau$, se sigue que $S_Y(y)$ decrece hacia algún valor límite, llamémosle $\tau^* = \inf\{S_Y(y) : y < \tau\}$. Ya que S_Y es decreciente, y por definición de infimo, se tiene que $\tau^* \geq S_Y(\tau)$. Entonces, la FS S_Y posee un límite izquierdo τ^* para cada punto $\tau \in \mathbb{R}^+$, pero se puede dar el caso de que $\tau^* > S_Y(\tau)$.

Notación 2. De la Observación 2.3, denotemos por $S_Y(y-)$ el límite de $S_Y(\varepsilon)$ cuando ε tiende a y por la izquierda (o crece hacia y), eso es, $S_Y(y-) = \lim_{\varepsilon \uparrow y} S_Y(\varepsilon)$. Entonces, $S_Y(y-) \geq S_Y(y)$. Otra notación que también se usa es $S_Y(y^-) = \lim_{\varepsilon \rightarrow 0} S_Y(y - \varepsilon)$, con $\varepsilon > 0$.

Definición 2.3. El salto de S_Y en y es la diferencia $S_Y(y-) - S_Y(y)$. Decimos que y es un punto de continuidad de S_Y si S_Y es continua en y , en cuyo caso $S_Y(y-) = S_Y(y)$, por lo que el salto es cero.

Proposición 2.1.2. Para cualquier $y \in \mathbb{R}^+$ el salto de S_Y en y es igual a $\Pr(Y = y)$.

Demostración. Por definición, $S_Y(y) = \Pr(Y > y)$. Sea $A_n = \{\omega : Y(\omega) > y - \frac{1}{n}\}$ para $n \in \mathbb{N}$. Se puede verificar fácilmente que

$$A_1 \supseteq A_2 \supseteq \dots \quad \text{y} \quad \bigcap_{n=1}^{\infty} A_n = \{\omega \in \Omega : Y(\omega) \geq y\}.$$

Se sigue que $\Pr(A_n) \downarrow \Pr(Y \geq y)$ cuando $n \rightarrow \infty$. Pero

$$\lim_{n \rightarrow \infty} \Pr(A_n) = \lim_{n \rightarrow \infty} S_Y(y - \frac{1}{n}) = S_Y(y-),$$

de manera que $S_Y(y-) = \Pr(Y \geq y)$. Por lo tanto, el salto de S_Y en y es

$$S_Y(y-) - S_Y(y) = \Pr(Y \geq y) - \Pr(Y > y) = \Pr(Y = y).$$

■

Para especificar a la función de riesgo de Y , y su relación con la función de supervivencia es conveniente proceder por separado para v.a.s. continuas y discretas.

² Esta observación es una adaptación para la función de supervivencia $S_Y(y)$ de la que Wilde [36], en sus notas, da para la función de distribución acumulada $F_Y(y)$.

Y (absolutamente) continua

Sea Y una v.a.s. con función de densidad de probabilidad $f_Y(y)$ y función de distribución acumulada

$$F_Y(y) = \Pr(Y \leq y) = \int_0^y dF_Y(u) = \int_0^y f_Y(u) du, \quad (2.1)$$

que representa la probabilidad de que el evento de interés ocurra antes del tiempo y . Entonces, la función de supervivencia es

$$S_Y(y) = 1 - F_Y(y) = \Pr(Y > y) = \int_y^\infty f_Y(u) du, \quad (2.2)$$

Haciendo uso del Teorema Fundamental del Cálculo y de la ecuación (2.2) se cumple que

$$f_Y(y) = \frac{dF_Y(y)}{dy} = -\frac{d}{dy}[1 - F_Y(y)] = -\frac{d}{dt}S_Y(y). \quad (2.3)$$

Por otra parte, la función de riesgo (*hazard rate*) corresponde a una tasa instantánea, la cual nos da una idea de como va cambiando la probabilidad de presentar el evento de interés en el siguiente instante de tiempo, condicionado a que no se haya presentado al inicio de éste.

Ejemplo 2.1. *Suponga que el modelo para la supervivencia de un grupo de pacientes sometidos a un trasplante de corazón tiene función de riesgo en forma de joroba: al principio hay un gran riesgo de muerte, y éste decrece a medida que el paciente se recupera.*

Otro ejemplo es la función de riesgo en forma de bañera, como se observa en la Fig 2.1, y describe el tiempo de supervivencia de seres humanos: para recién nacidos, el inminente riesgo de muerte decrece con el tiempo; en una edad madura, el riesgo es relativamente constante; finalmente, este se incrementa en la vejez. \square

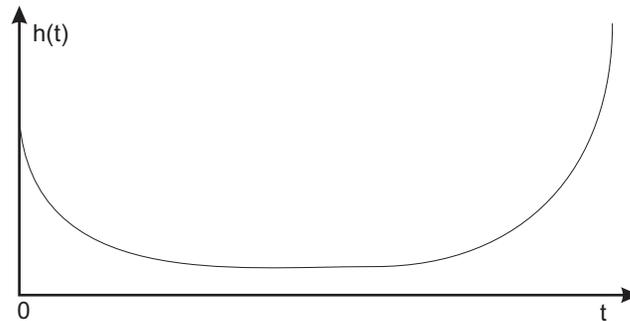


Figura 2.1: *Diagrama que muestra la gráfica de una fuerza de mortalidad en forma de bañera, característica del tiempo de vida de seres humanos.*

Para obtener una definición más formal de la función de riesgo, sea Δt un tiempo infinitesimal y considere la probabilidad condicional de que un individuo experimente una falla (e.g. la muerte) en el intervalo de tiempo $[y, y + \Delta y)$, dado que ha sobrevivido hasta el tiempo y , la cual se puede expresar como

$$\begin{aligned} \Pr(y < Y \leq y + \Delta y \mid Y > y) &= \frac{\Pr(\{y < Y \leq y + \Delta y\} \cap \{Y > y\})}{\Pr(Y > y)} \\ &= \frac{\Pr(y < Y \leq y + \Delta y)}{\Pr(Y > y)}. \end{aligned} \quad (2.4)$$

Al dividir (2.4) entre la longitud del intervalo de tiempo actual, Δy , se obtiene la **tasa media de mortalidad** a lo largo de dicho intervalo:

$$\frac{1}{\Delta y} \Pr(y < Y \leq y + \Delta y \mid Y > y) = \frac{\Pr(y < Y \leq y + \Delta y)}{\Delta y} \frac{1}{\Pr(Y > y)}. \quad (2.5)$$

En el límite, cuando $\Delta y \rightarrow 0^+$, la tasa media de mortalidad se convierte en una **tasa de mortalidad instantánea**,

$$\lim_{\Delta y \rightarrow 0^+} \frac{\Pr(y < Y \leq y + \Delta y \mid Y > y)}{\Delta y}. \quad (2.6)$$

Ya que

$$\Pr(y < Y \leq y + \Delta y) = \int_y^{y+\Delta y} f_Y(u) du = F_Y(y + \Delta y) - F_Y(y),$$

al substituir esa expresión en (2.6) y recordando la definición de la derivada de una función, se obtiene

$$\begin{aligned} \lim_{\Delta y \rightarrow 0^+} \frac{\Pr(y < Y \leq y + \Delta y)}{(\Delta y) \Pr(Y > y)} &= \lim_{\Delta y \rightarrow 0^+} \frac{F_Y(y + \Delta y) - F_Y(y)}{\Delta y} \frac{1}{S_Y(y^-)} \\ &= \frac{F'_Y(y)}{S_Y(y)} = \frac{f_Y(y)}{S_Y(y)}. \end{aligned}$$

donde $S_Y(y^-) \equiv \lim_{y \uparrow t} S_Y(y)$, que en el caso de una v.a.s. continua $S_Y(y^-) = S_Y(y)$.

El resultado anterior motiva la siguiente definición:

Definición 2.4. La **función de riesgo** de una v.a.s. Y denotada como $h(y)$, se define como

$$h_Y(y) = \frac{f_Y(y)}{S_Y(y)}, \quad \forall y > 0. \quad (2.7)$$

Esta representa la probabilidad condicional de presentar el evento de interés en un instante de tiempo dado, condicionada a que no se ha presentado al inicio de ese instante.

La función de supervivencia se expresa en términos de la fuerza de mortalidad mediante el siguiente teorema.

Teorema 2.1.1.

$$S_Y(y) = \exp \left\{ - \int_0^y h_Y(u) du \right\}, \quad \forall y \geq 0.$$

Demostración. Ya que $f_Y(u) = -\frac{d}{du}S_Y(u)$, se sigue que

$$h_Y(u) = -\frac{d}{du}[\log S_Y(u)].$$

Al integrar en cada lado de la expresión anterior entre 0 y t se obtiene

$$\log S_Y(u)|_0^y = - \int_0^y h_Y(u) du.$$

Tomando en cuenta que $S_Y(0) = 1$, el resultado del Teorema se sigue de la expresión anterior. ■

Del Teorema 2.1.1, cualquier función de riesgo de una v.a.s continua Y debe satisfacer las propiedades resumidas en el siguiente corolario.

Corolario 2.1.2. Sea Y una v.a.s. con función de densidad $f_Y(y)$ y sea $h_Y(y)$ la función de riesgo correspondiente. Entonces

1. $h_Y(y) \geq 0, \quad \forall y > 0$
2. $\int_0^y h_Y(u) du < \infty, \quad \text{para algún } y > 0,$
3. $\int_0^\infty h_Y(u) du = \infty$

Demostración. La condición uno se sigue de la ecuación (2.6) debido a que las cantidades en el numerador y el denominador son no negativas. La segunda y tercera propiedad se siguen del Teorema 2.1.1. En efecto, para $y > 0$,

$$\int_0^y h_Y(u) du = -\log S_Y(y) < \infty,$$

y ya que $h_Y(y) = -d \log S_Y(y)/dy$, se tiene que

$$\int_0^\infty h_Y(u) du = \int_0^\infty -d[\log S_Y(u)] = \infty.$$

■

Observación 2.4. Del Corolario anterior, la función de riesgo puede no ser monótona creciente o decreciente, no estar acotada cuando $y \rightarrow \infty$, y no es una probabilidad del todo. Esta nos describe la probabilidad condicional, por unidad de tiempo, de que un individuo experimente el evento de interés al tiempo y , dado que ha sobrevivido hasta ese tiempo. Las unidades de tiempo dependerán de la selección de medición del tiempo de supervivencia.

Definición 2.5. La *función de riesgo acumulada*, denotada como $H_Y(y)$, correspondiente a la v.a.s. Y con función de riesgo asociada $h_Y(y)$, se define como

$$H_Y(y) = \int_0^y h_Y(u) \, du = -\log S_Y(y). \quad (2.8)$$

Así, otra expresión para la función de supervivencia es

$$S_Y(y) = \exp\{-H_Y(y)\}. \quad (2.9)$$

Nótese que el límite inferior igual a 0 en la integral de (2.8) es consistente con $S_Y(0) = 1$ y, para una FS propia, $H_Y(y)$ debe tender a ∞ cuando $y \rightarrow \infty$.

Y variable aleatoria discreta.

Sea Y una v.a.s. discreta, la cual toma los valores $0 = \tau_0 < \tau_1 < \dots < \tau_n$. Sea

$$f_Y(y) = \Pr(Y = y) \quad y \quad F_Y(y) = \Pr(Y \leq y) = \sum_{j | \tau_j \leq y} f_Y(\tau_j),$$

la función de probabilidad y la función de distribución acumulada de Y , respectivamente. Entonces, la correspondiente función de supervivencia de Y se define como

$$S_Y(y) = \Pr(Y > y) = \sum_{j | \tau_j > y} f_Y(\tau_j), \quad \text{con } S_Y(\tau_0) = 1.$$

Esas funciones también están relacionadas por

$$f_Y(y) = F_Y(y) - F_Y(y^-) = S_Y(y^-) - S_Y(y) \quad (2.10)$$

donde $S_Y(y^-) = \Pr(Y \geq y) = \lim_{\varepsilon \rightarrow 0} S_Y(y - \varepsilon)$, con $\varepsilon > 0$.

Nótese que, para $\tau_{i-1} < y < \tau_i$, $f_Y(y) = 0$, por lo que

$$S_Y(\tau_i^-) = \lim_{\varepsilon \rightarrow 0} S(\tau_i - \varepsilon) \equiv S(\tau_{i-1}).$$

En la figura 2.2 se describe la FS $S(y)$ correspondiente a la v.a.s discreta Y , la cual toma los valores $0 = \tau_0 < \tau_1 < \tau_2 < \tau_3$, con $S(0) = 1$ y $S(\tau_3) = 0$.

Para obtener una expresión de la función de riesgo de Y , la cual es análoga a la ecuación (2.7), considere la siguiente probabilidad condicional:

$$\begin{aligned} \Pr(Y = \tau_i | Y \geq \tau_i) &= \frac{\Pr(\{Y = \tau_i\} \cap \{Y \geq \tau_i\})}{\Pr(Y \geq \tau_i)} \\ &= \frac{\Pr(Y = \tau_i)}{\Pr(Y \geq \tau_i)} \\ &= \frac{f(\tau_i)}{S(\tau_i^-)} \quad \text{para } i = 0, 1, 2, \dots, n. \end{aligned}$$

La expresión anterior da lugar a la definición de la función de riesgo para una v.a.s. discreta:

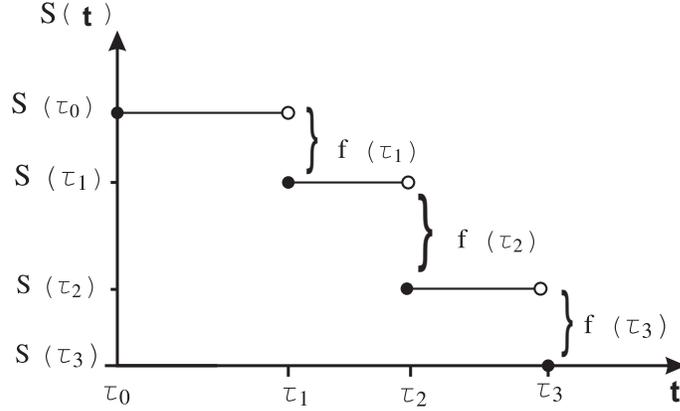


Figura 2.2: Ya que $S_Y(y)$ es una función monótona decreciente y continua por la derecha, se tiene que $f_Y(\tau_0) = 0$, y para $y \neq \tau_i$, $f_Y(y) = S_Y(y^-) - S_Y(y) = 0$; por otro lado, debido a que $S_Y(y)$ es discontinua por la izquierda, con puntos de discontinuidad en cada τ_i , entonces $f_Y(\tau_i) = S_Y(\tau_i^-) - S_Y(\tau_i) > 0$, para $i = 1, 2, 3$.

Definición 2.6. Sea Y una v.a.s. discreta, la cual toma los distintos valores en $0 = \tau_0, \tau_1, \dots, \tau_n$, y sea $S_Y(y)$ su FS. La función de riesgo de Y se define como

$$h_Y(y) = \frac{f_Y(y)}{S_Y(y^-)}, \quad \forall y \geq 0. \quad (2.11)$$

A partir de la definición anterior, se puede verificar que:

- $0 \leq h_Y(y) \leq 1, \quad \forall y > 0$;
- $h_Y(y) = 0$, excepto en los puntos τ_i , para $i = 1, 2, \dots, n$;
- $h_Y(0) = 0$; y $h_Y(y) = 1$ solamente en el punto y donde $S_Y(y^-) = f(y)$, que es el punto extremo superior $y = \tau_n$.

También se puede obtener una expresión de la función de supervivencia en términos de la función de riesgo. Para eso, primero considere el siguiente Lema:

Lema 2.1.1. Para $k \neq 0$ ($k \leq n$), $y \in [\tau_k, \tau_{k+1})$, la FS de Y se puede expresar como un producto de probabilidades condicionales

$$S_Y(y) = \prod_{j=1}^k \frac{S_Y(\tau_j)}{S_Y(\tau_j^-)}.$$

Demostración. Considere la siguiente probabilidad de supervivencia condicional:

$$\Pr(Y > \tau_i | Y > \tau_{i-1}) = \frac{\Pr(\{Y > \tau_i\} \cap \{Y > \tau_{i-1}\})}{\Pr(Y > \tau_{i-1})} = \frac{\Pr(Y > \tau_i)}{\Pr(Y > \tau_{i-1})} = \frac{S_Y(\tau_i)}{S_Y(\tau_{i-1})}.$$

Luego, para $\tau_1 < \dots < \tau_k \leq y$ se tiene que

$$\begin{aligned} S_Y(\tau_k) &= \frac{S_Y(\tau_k)}{S_Y(\tau_{k-1})} * \frac{S_Y(\tau_{k-1})}{S_Y(\tau_{k-2})} \cdots \frac{S_Y(\tau_2)}{S_Y(\tau_1)} * \frac{S_Y(\tau_1)}{S_Y(\tau_0)} \\ &= \prod_{j=1}^k \frac{S_Y(\tau_j)}{S_Y(\tau_{j-1})} = \prod_{j|\tau_j \leq t} \frac{S_Y(\tau_j)}{S_Y(\tau_j^-)}. \end{aligned}$$

■

Teorema 2.1.3. Para $y \in [\tau_k, \tau_{k+1})$, con $k \leq n$,

$$S_Y(y) = \prod_{j|\tau_j \leq y} \{1 - h_Y(\tau_j)\}, \quad (2.12)$$

con la convención $S_Y(\tau_0) = 1$.

Demostración. Sea Y una v.a.s con distribución de probabilidad $f_Y(y)$. Ya que $f_Y(y) = S_Y(y^-) - S_Y(y)$, la función de riesgo queda expresada como

$$h_Y(y) = \frac{f_Y(y)}{S_Y(y^-)} = \frac{S_Y(y^-) - S_Y(y)}{S_Y(y^-)} = 1 - \frac{S_Y(y)}{S_Y(y^-)},$$

o de manera equivalente

$$\frac{S_Y(y)}{S_Y(y^-)} = 1 - h_Y(y).$$

El resultado se sigue por el lema anterior. ■

Como consecuencia del Teorema anterior, otra expresión para la función de probabilidad de Y se da en el siguiente corolario.

Corolario 2.1.4. Para $i = 0$ $f_Y(\tau_0) = 0$, para $i = 1$ $f_Y(\tau_1) = h_Y(\tau_1)$, y para $i \geq 2$, con $\tau_i \leq y$,

$$f_Y(y) = h_Y(y) \prod_{j=1}^{i-1} \{1 - h_Y(\tau_j)\}. \quad (2.13)$$

Demostración. La convención $S_Y(\tau_0) = 1$ implica que $f_Y(\tau_0) = 0$. Luego, de (2.12), $S_Y(\tau_1) = 1 - h_Y(\tau_1)$, lo cual implica que $f_Y(\tau_1) = S_Y(\tau_1^-) - S_Y(\tau_1) = h_Y(\tau_1)$. Finalmente, de la definición de $h_Y(y)$ en (2.11), $f_Y(y) = h_Y(y)S_Y(y^-)$, por lo que, para $i \geq 2$

$$\begin{aligned} f_Y(\tau_i) &= h_Y(\tau_i)S_Y(\tau_i^-) \\ &= h_Y(\tau_i)S_Y(\tau_{i-1}) \\ &= h_Y(\tau_i) \prod_{j=1}^{i-1} \{1 - h_Y(\tau_j)\}. \end{aligned}$$

■

Escribiendo h_i para $h_Y(\tau_i)$, otras expresiones para (2.12) y (2.13) son:

$$\text{i) } S_Y(\tau_i) = \prod_{j=1}^i (1 - h_j) \text{ para } i \geq 1, \text{ con } S_Y(\tau_0) = 1 \text{ y } S_Y(\tau_n) = 0.$$

$$\text{ii) } f_Y(\tau_i) = h_i S_Y(\tau_i^-) = h_i \prod_{j=1}^{i-1} (1 - h_j) \text{ para } i \geq 2, \text{ con } f(\tau_0) = 0 = h_0 \text{ y } f(\tau_1) = h_1.$$

Finalmente, de la ecuación (2.8) se tiene

$$H_Y(y) = -\log S_Y(y) = -\sum_j \log(1 - h_j). \quad (2.14)$$

Si h_j en la suma de (2.14) es pequeña, entonces $\log(1 - h_j) \approx -h_j$ y

$$H_Y(y) \approx -\sum_j h_j, \quad (2.15)$$

la cual puede ser interpretada como la **función de riesgo acumulada**.

2.2. Modelos paramétricos: Distribuciones

En la Sección 2.1 se vio que la distribución de supervivencia de Y puede ser especificada, de manera equivalente, por: la función de densidad de probabilidad, $f_Y(y)$, la función de distribución acumulada, $F_Y(y)$, la función de supervivencia, $S_Y(y)$, o la función de riesgo, $h_Y(y)$. Esto significa que si se especifica alguna de esas expresiones, entonces las otras pueden ser derivadas de esta. Por lo tanto, al hacer un análisis de supervivencia de ciertos datos, se puede considerar a cualquiera de esas representaciones que mejor nos convenga, para posteriormente basar la inferencia de esos datos en esta.

Frecuentemente uno está interesado en la función de riesgo ya que su directa interpretación es un riesgo inminente, aunque esta selección dependerá del interés que se tenga.

Se han usado varias familias paramétricas para el modelado de datos de tiempos de supervivencia, pero entre esos modelos, algunos sobresalen debido a su flexibilidad y gran utilidad en ciertas situaciones. A continuación se mencionan algunas de esas familias para el modelado de la distribución del tiempo de supervivencia Y .

Notación 3. Sea

$$\mathcal{P} = \{f_\theta : \theta \in \Theta\}$$

una familia de funciones de densidad de probabilidad (o función de masa de probabilidad en el caso de distribuciones discretas) indexada por θ en el espacio de parámetros Θ . Entonces

$$f(t; \theta), \quad f_\theta(t) \quad \text{o} \quad f(t | \theta)$$

denotan a un elemento de \mathcal{P} como función de t manteniendo a θ fijo.

Modelo Exponencial

Suponga que la v.a.s. Y se distribuye exponencialmente con parámetro $\theta (> 0)$, eso es $Y \sim Exp(\theta)$. Este modelo supone un riesgo constante

$$h_Y(y; \theta) = \frac{1}{\theta}, \quad y \geq 0.$$

Por el Teorema 2.1.1 la función de supervivencia para Y es

$$S_Y(y; \theta) = \exp(-y/\theta), \quad \forall y \geq 0$$

Otra expresión para la distribución de Y que frecuentemente se emplea es con la reparametrización $\omega = 1/\theta$, en cuyo caso la f.d.p. es

$$f_Y(y) = \omega \exp(\omega y).$$

Modelo Weibull

Suponga que $Y \sim WEI(\theta, \alpha)$ con $\theta = (\theta, \alpha)$, donde $\theta > 0$ es el parámetro de escala³ y $\alpha > 0$ es el parámetro de forma. La función de riesgo de este modelo es

$$h_Y(y; \theta) = \frac{\alpha}{\theta} \left(\frac{y}{\theta} \right)^{\alpha-1}, \quad \forall y \geq 0. \quad (2.16)$$

Por el Teorema 2.1.1, la función de supervivencia y la función de densidad para (2.16) son

$$S_Y(y; \theta) = \exp(-[y/\theta]^\alpha), \quad y \quad f_Y(y; \theta) = \frac{\alpha}{\theta} \left(\frac{y}{\theta} \right)^{\alpha-1} \exp(-[y/\theta]^\alpha).$$

La función $h(y)$ es monótona creciente si $\alpha > 1$, monótona decreciente si $\alpha < 1$, y para $\alpha = 1$ se produce el modelo exponencial. Como se muestra en la Figura 2.3, debido a la forma de $h_Y(y; \theta)$ y $S_Y(y; \theta)$, hacen del Modelo Weibull una conveniente herramienta en análisis de supervivencia.

Al considerar la reparametrización

$$Y \sim WEI(\theta, \alpha), \quad \text{con} \quad \theta = (1/\omega)^{(1/\alpha)},$$

las funciones de riesgo y de supervivencia adoptan la siguiente forma

$$h_Y(y; \omega, \alpha) = \omega \alpha y^{\alpha-1} \quad \text{y} \quad S_Y(y; \omega, \alpha) = \exp(-\omega y^\alpha).$$

Nos referiremos ha esa reparametrización como *Weibull*(ω, α).

³ Si una familia de distribución de probabilidades es tal que existe un parámetro θ (y otro parámetro α) para el cual la función de distribución acumulada satisface

$$F(y; \theta, \alpha) = F(y/\theta; 1, \alpha)$$

entonces a θ se le llama *parámetro de escala*, ya que su valor determina la escala o dispersión estadística de la distribución de probabilidades.

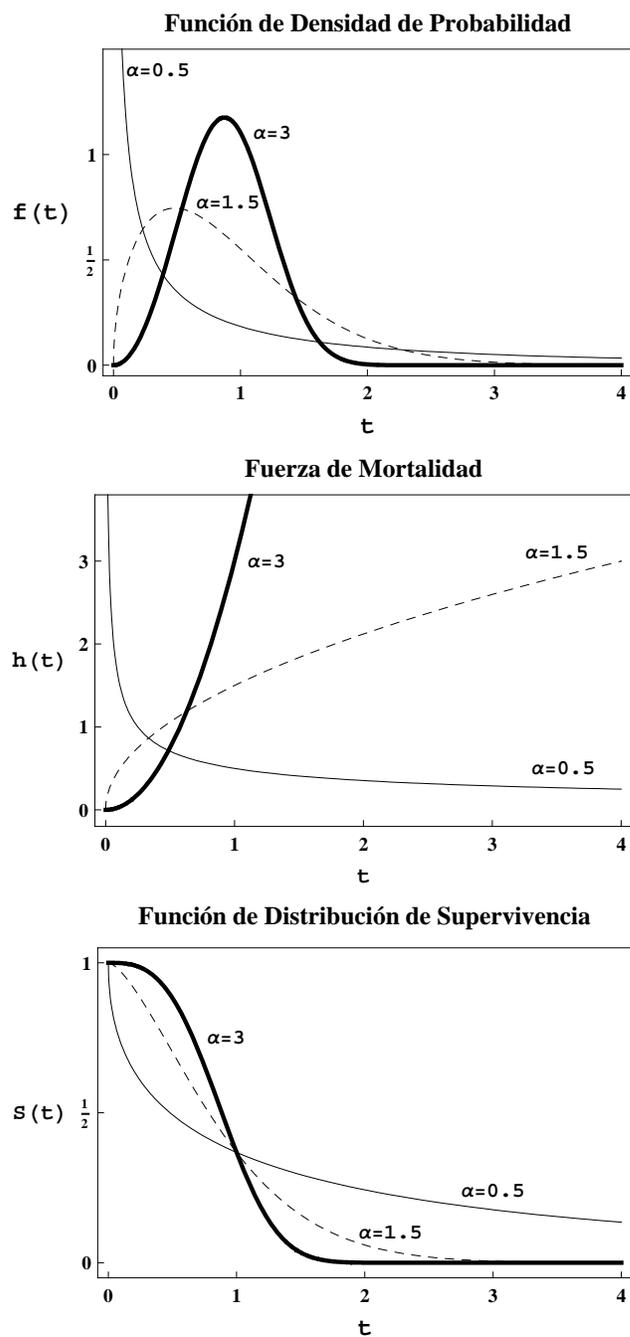


Figura 2.3: Funciones asociadas al Modelo Weibull con parámetro de escala $\theta^{-1} = 1$ y parámetro de forma $\alpha = 0.5, 1.5$, y 3 .

Modelo del Valor Extremo

Sea Y una v.a.s. y suponga que $Y \sim Weibull(\theta, \alpha)$. Entonces, la función de supervivencia es

$$S_Y(y; \theta, \alpha) = \exp(\theta y^\alpha), \quad \forall y > 0.$$

Aplicando la transformación logaritmo natural (denotada como \log) al tiempo de supervivencia Y , la función de supervivencia para la nueva v.a. $W = \log Y$ esta dada por

$$\begin{aligned} S_W(w; \theta, \alpha) &= \Pr(W > w) \\ &= \Pr(\log Y > w) \\ &= \Pr(Y > \exp(w)) \\ &= \exp(-\theta \exp\{\alpha w\}). \end{aligned} \tag{2.17}$$

Si redefinimos los parámetros en la expresión anterior como

$$\theta = \exp(-\mu/\sigma) \quad \text{y} \quad \alpha = 1/\sigma,$$

la función de supervivencia para W es

$$S_W(w; \mu, \sigma) = \exp \left\{ - \exp \left(\frac{w - \mu}{\sigma} \right) \right\}, \tag{2.18}$$

la cual corresponde a la función de supervivencia de una variable aleatoria con distribución Valor Extremo.

Por lo tanto, $W \sim ValorExtremo(\mu, \sigma)$, cuya función de densidad es

$$f_W(w; \mu, \sigma) = \frac{1}{\sigma} \exp \left(\frac{w - \mu}{\sigma} - \exp([w - \mu]/\sigma) \right), \quad \forall w \in (-\infty, \infty).$$

Así, la función de riesgo para W es

$$h_W(w; \mu, \sigma) = \frac{1}{\sigma} \exp \left\{ \frac{w - \mu}{\sigma} \right\}, \quad \forall w \in (-\infty, \infty).$$

Observación 2.5. Cuando $\mu = 0$ y $\sigma = 1$, se obtiene la distribución *Valor Extremo Estándar*, eso es, $W^* \sim ValorExtremo(0, 1)$.

Modelo de pedazos exponencial

En ocasiones, el patrón de dependencia del tiempo en los modelos especificados de forma completamente paramétrica, tales como el Modelo Weibull, no siempre dan un resultado satisfactorio en el ajuste de los datos. En tal situación, un *Modelo de pedazos (o escalones) exponencial (piece-wise exponential model* en inglés) bien puede constituir un esquema más apropiado.

Sea $a_0 < a_1 < \dots < a_M$ valores especificados, con $a_0 = 0$ y $a_{M+1} = \infty$. La función de riesgo de este modelo es

$$h_Y(y) = \exp(\alpha_\ell), \quad \text{si } y \in I_\ell, \quad (2.19)$$

donde

$$I_\ell = [a_{\ell-1}, a_\ell), \quad \ell = 1, 2, \dots, M,$$

son intervalos mutuamente excluyentes. De esta forma, $h_Y(y)$ es constante sobre cada intervalo y α_ℓ representa el log del riesgo en el intervalo I_ℓ .

La función de riesgo acumulada esta dada por

$$H_Y(y) = \begin{cases} y \exp(\alpha_1), & \text{si } y \in I_1; \\ \sum_{\kappa=1}^{\ell-1} (a_\kappa - a_{\kappa-1}) e^{\alpha_\kappa} + (y - a_{\ell-1}) e^{\alpha_\ell}, & \text{si } y \in I_\ell, \ell \geq 2. \end{cases}$$

Ya que la expresión anterior es algo tediosa, otro tipo de notación ha sido adoptada. Por ejemplo, introduciendo la notación (Lawless, 2003)

$$\Delta_\ell(y) = \int_{a_{\ell-1}}^{a_\ell} \mathbf{1}\{u \leq y\} du \quad \ell = 1, 2, \dots, M.$$

la función de riesgo acumulada es

$$H_Y(y) = \sum_{\ell=1}^M \exp(\alpha_\ell) \Delta_\ell(y). \quad (2.20)$$

Así, de la ecuación (2.9), la FS de (2.19) es

$$S_Y(y) = \exp \left\{ - \sum_{\ell=1}^M \exp(\alpha_\ell) \Delta_\ell(y) \right\},$$

y para $y \in I_\ell$, $f_Y(y) = \exp(\alpha_\ell) S_Y(y)$ es la función de densidad de probabilidad correspondiente.

2.2.1. Modelos de Mezclas Discretos (finitos)

Un modelo de mezclas discreto es una distribución de probabilidad que es una combinación convexa⁴ de otras distribuciones de probabilidad.

⁴ Dado un conjunto finito $\mathcal{A} = \{y_1, y_2, \dots, y_m\}$ en un espacio vectorial real V , una *combinación convexa* de esos puntos es otro punto de la forma

$$y = a_1 y_1 + a_2 y_2 + \dots + a_m y_m, \quad a_j \in \mathbb{R},$$

donde los a_j 's satisfacen $a_j \geq 0$ y $\sum_{j=1}^m a_j = 1$. La **corteza convexa** de \mathcal{A} es el conjunto de combinaciones convexas de subconjuntos finitos de puntos de \mathcal{A} .

Definición básica

Sea Y_1, Y_2, \dots, Y_n una muestra⁵ aleatoria de tamaño n y y_i el valor observado de Y_i . Supongamos que Y_i es una mezcla de m v.a. con función de distribución acumulada $F_1(y_i), F_2(y_i), \dots, F_m(y_i)$, respectivamente. Entonces, la distribución de probabilidad de Y_i está dada por

$$F(y_i) = \pi_1 F_1(y_i) + \pi_2 F_2(y_i) + \dots + \pi_m F_m(y_i), \quad (2.21)$$

donde las $F_j(y_i)$'s son llamadas las *distribuciones componentes* de la mezcla $F(y_i)$ y las cantidades $\pi_1, \pi_2, \dots, \pi_m$, llamadas las *proporciones mezcladas* o pesos, satisfacen

$$0 \leq \pi_j \leq 1, \quad j = 1, 2, \dots, m \quad \text{y} \quad \sum_{j=1}^m \pi_j = 1. \quad (2.22)$$

Las propiedades de los modelos de mezclas se siguen de las propiedades de las m distribuciones componentes: Suponga que Y_i es una mezcla de m v.a.s. *continuas* con función de densidad de probabilidad (f.d.p.)

$$f_1(y_i) = F_1'(y_i), \quad f_2(y_i) = F_2'(y_i), \dots, \quad f_m(y_i) = F_m'(y_i),$$

respectivamente. Entonces, de la expresión (2.21), T_i es una v.a.s. continua con f.d.p.

$$f(y_i) = \pi_1 f_1(y_i) + \pi_2 f_2(y_i) + \dots + \pi_m f_m(y_i). \quad (2.23)$$

Nota 2.1. Si Y_i es una mezcla de m v.a.s. *discretas*, la función de masa de probabilidad de Y_i es similar a la que se da en (2.23), pero los componentes de la mezcla son funciones de masa de probabilidad, correspondientes a las m v.a.s. discretas.

Modelos de mezclas paramétricos

Cuando los componentes de la mezcla en (2.23) pertenecen a una familia paramétrica de distribuciones, con parámetro desconocido ϕ_j , estos son llamados *Modelos de mezclas paramétricos*. En este caso, la f.d.p. de Y_i está dada por

$$f(y_i; \theta) = \sum_{j=1}^m \pi_j f_j(y_i; \phi_j). \quad (2.24)$$

donde $\theta = (\pi_1, \dots, \pi_{j-1}, \phi_1, \phi_2, \dots, \phi_m)^T \in \Theta$ es el vector de todos los parámetros desconocidos de la mezcla y Θ es el espacio de valores de θ .

Observación 2.6. Ya que la suma de las proporciones mezcladas π_j suman 1, una de ellas es redundante, por eso es que en el vector de parámetros desconocidos, θ , se ha omitido de manera arbitraria el m -ésimo peso π_m .

⁵ En la práctica, Y_i se toma como un vector aleatorio p -dimensional, \mathbf{Y}_i , cuyos componentes son variables aleatorias correspondientes a p medidas hechas en la i -ésima observación de ciertas características del fenómeno en estudio.

Por lo regular las $f_j(y; \phi_j)$ en (2.24) son especificadas como distribuciones de la misma familia paramétrica. Entonces, si

$$\mathcal{F} = \{f_Y(\cdot; \phi) : \phi \in \Phi\},$$

denota la clase de todas las distribuciones componentes y Φ el espacio de valores para ϕ , la densidad de mezclas de Y_i está dada por

$$f(y_i; \theta) = \sum_{j=1}^m \pi_j f_j(y_i; \phi_j) \quad \text{con} \quad f_j(y_i; \phi_j) = f_Y(y_i; \phi_j) \quad (j = 1, 2, \dots, m),$$

donde $f_Y(\cdot; \phi)$ es un miembro genérico de \mathcal{F} y θ es como antes, el vector de parámetros desconocidos de la mezcla con espacio de valores Θ .

Nota 2.2. Los **modelos de mezclas continuos** también pueden ser considerados y son definidos de manera similar: Sea $g(\cdot)$ una función del parámetro θ tal que

$$g(\theta) \geq 0 \quad \forall \theta \in \Theta \quad \text{y} \quad \int_{\Theta} g(\theta) d\theta = 1,$$

donde Θ es el espacio de parámetros de θ . Entonces la f.d.p. de Y se define como

$$f_Y(y) = \int_{\Theta} g(\theta) f_Y(y; \theta) d\theta.$$

Interpretación y algunas aplicaciones de los modelos de mezcla

Sea $\mathcal{J} = \{1, 2, \dots, m\}$ el conjunto de etiquetas de las distribuciones componentes (categorías) del modelo de mezclas en (2.21) y J una variable aleatoria asumiendo los posibles valores $1, 2, \dots, m$ con probabilidad $\pi_1, \pi_2, \dots, \pi_m$, respectivamente. Supongamos que la función de distribución *condicional* de Y_i dado $J_i = j$ es $F_j(y_i)$. Entonces, la función de distribución *incondicional*⁶ de Y_i está dada por $F(y_i)$.

Sea Γ_i ($i = 1, 2, \dots, n$) un vector m -dimensional, donde el j -ésimo elemento de Γ_i , Γ_{ij} , se define como

$$\Gamma_{ij} = I\{J_i = j\} = \begin{cases} 1, & \text{si el componente origen de } Y_i \text{ es } j \ (j \in \mathcal{J}); \\ 0, & \text{en otro caso.} \end{cases}$$

Bajo la suposición de que cada Y_i ($i = 1, 2, \dots, n$) proviene exactamente de una de las m distribuciones componentes, se tiene que

$$\sum_{j=1}^m \Gamma_{ij} = 1.$$

⁶ Esta puede verse como una distribución marginal. Considere al par aleatorio (Y, J) donde Y es una v.a.s. continua y J una v.a. asumiendo los posibles valores $1, 2, \dots, m$. Entonces, la distribución marginal de Y es

$$F_Y(y) = \Pr(Y \leq y) = \sum_{j=1}^m \Pr(\{Y \leq y\} \cap \{J = j\}) = \sum_{j=1}^m \Pr(J = j) \Pr(Y \leq y | J = j).$$

Entonces $\mathbf{\Gamma}_i$ se distribuye multinomial con m posibles resultados, los cuales tienen probabilidades $\pi_1, \pi_2, \dots, \pi_m$, respectivamente, es decir,

$$\mathbf{\Gamma}_i \sim Mult_m(1, \pi_1, \pi_2, \dots, \pi_m),$$

cuya función de masa de probabilidad es

$$\begin{aligned} \Pr(\mathbf{\Gamma}_i = \boldsymbol{\gamma}_i) &= \Pr(\{\Gamma_{i1} = \gamma_{i1}\} \cap \{\Gamma_{i2} = \gamma_{i2}\} \cap \dots \cap \{\Gamma_{im} = \gamma_{im}\}) \\ &= \pi_1^{\gamma_{i1}} \pi_2^{\gamma_{i2}} \dots \pi_m^{\gamma_{im}}, \end{aligned}$$

siendo γ_{ij} uno o cero, dependiendo del valor observado para el i -ésimo individuo. Más aún, se puede verificar fácilmente⁷ que para cada $j \in \mathcal{J}$, Γ_{ij} se distribuye Binomial con parámetros $r = 1$ y π_j , o de manera equivalente,

$$\Gamma_{ij} \sim Bernoulli(\pi_j),$$

con función de masas de probabilidad

$$\Pr(\Gamma_{ij} = \gamma_{ij}) = \pi_j^{\gamma_{ij}} (1 - \pi_j)^{1 - \gamma_{ij}}.$$

Una situación en la que el modelo de mezclas (2.21) es directamente aplicable es cuando Y_i proviene de una población G que consiste de m grupos G_1, G_2, \dots, G_m en proporciones $\pi_1, \pi_2, \dots, \pi_m$, respectivamente. Si la distribución de Y_i en el grupo G_j está dada por $F_j(y_i)$ para $j = 1, 2, \dots, m$, entonces la distribución de Y_i tiene la forma de la mezcla (2.21). De esta manera, los m componentes de la mezcla pueden ser físicamente identificados por los m grupos externamente existentes G_1, G_2, \dots, G_m . Así, cuando Y_i puede ser físicamente identificado que proviene del j -ésimo componente ($j \in \mathcal{J}$), el vector de etiquetas-componente correspondiente es

$$\mathbf{\Gamma}_i = \{0, \dots, 0, \Gamma_{ji} = 1, 0, \dots, 0\},$$

en otro caso, $\mathbf{\Gamma}_i$ será desconocido.

En algunos casos, la adopción de una simple distribución componente para el modelado de poblaciones heterogéneas puede resultar inadecuado. Una forma de lidiar con este problema es adoptar un modelo de mezclas de distribuciones para obtener una mayor flexibilidad en el modelado de ese tipo de poblaciones. Por ejemplo, Farewell (1982) y Larson y Dinse (1985) estuvieron entre los primeros en usar modelos de mezclas finitas para el análisis de datos de Riesgos Competitivos. En este contexto, si t_1, t_2, \dots, t_n denota una muestra aleatoria observada de tiempos de supervivencia, posiblemente censurados por la derecha, proveniente de una población expuesta a $m > 1$ modos de falla, debido a la censura en los datos, estos son parcialmente clasificados, es decir, si los datos se ordenan de tal manera que $t_{(1)}, t_{(2)}, \dots, t_{(r)}$ son las observaciones no-censuradas y las restantes $n-r$ están censuradas, los vectores de etiquetas-componentes m -dimensionales $\mathbf{\Gamma}_{(1)}, \mathbf{\Gamma}_{(2)}, \dots, \mathbf{\Gamma}_{(r)}$ están disponibles solamente para $t_{(1)}, t_{(2)}, \dots, t_{(r)}$.

⁷ Recuerde que la distribución Binomial es en realidad una Multinomial con $m=2$ posibles resultados: sea $r > 0$, $0 < p < 1$ y Γ una v.a. tal que $\Gamma \sim Mult_2(r, p, 1-p)$, entonces $\Gamma \sim Binomial(r, p)$.

2.2.1.1. Mezcla de Distribuciones de Supervivencia

Sea \tilde{T} una v.a.s. correspondiente a una población G en la que están actuando m distintos tipos o causas de falla (**Riesgos Competitivos**). Supongamos que el tiempo de falla \tilde{T} puede estar censurado a la derecha por una variable aleatoria positiva C . Sea J la v.a. que denota la causa de falla, tomando posibles valores en $\mathcal{J} = \{1, 2, \dots, m\}$. Para $j \in \mathcal{J}$, definamos la variable indicadora $\Gamma^j = I\{J = j\}$. En un escenario de riesgos competitivos, tanto el tipo de falla J como la indicadora Γ^j son observados solamente si el tiempo de supervivencia (o de falla) no está censurado. Por eso, se define la indicadora de censura como $\delta = 1\{\tilde{T} \leq C\}$, siendo uno si \tilde{T} no está censurada y cero en otro caso. Supongamos que se tiene una muestra aleatoria de tamaño n , cuyos datos consisten de vectores de la forma

$$\mathbf{T}_i = (T_i, \delta_i, \delta_i J_i), \quad i = 1, 2, \dots, n,$$

donde $T_i = \min\{\tilde{T}_i, C_i\}$.

El enfoque tradicional para el modelado de la distribución del tiempo de falla en un escenario de riesgos competitivos es suponer la existencia de los llamados tiempos de falla latentes, Y_1, Y_2, \dots, Y_m , correspondientes a las m causas de falla, y proceder con el modelado del vector aleatorio

$$\mathbf{T}_i = (T_i, \delta_i, \delta_i J_i), \quad i = 1, 2, \dots, n$$

donde $T_i = \min\{\tilde{T}_i, C_i\}$ y $\tilde{T}_i = \min(Y_{1i}, Y_{2i}, \dots, Y_{mi})$.

Un enfoque alternativo es adoptar un modelo de mezclas con m componentes, en el cual la Función de Supervivencia de T_i es modelada como

$$S(t_i) = \pi_1 S_1(t_i) + \pi_2 S_2(t_i) + \dots + \pi_m S_m(t_i) \quad (2.25)$$

donde la j -ésima FS componente

$$S_j(t_i) = \Pr(T_i > t_i \mid J = j)$$

denota la función de supervivencia condicional (*propia*), dado que la falla del i -ésimo individuo es por la j -ésima causa, y $\pi_j = \Pr(J = j)$ es la *probabilidad a priori* del tipo de falla j ($j = 1, 2, \dots, m$). Este enfoque de mezclas *asume que la eventual causa de falla es determinada al tiempo cero (i.e. al inicio del estudio) por algún mecanismo estocástico*. Por ejemplo, después de la terapia para pacientes con cáncer de pulmón un paciente que no está curado debe estar destinado a morir de la enfermedad; un paciente curado eventualmente morirá de otras causas.

La introducción del vector de etiquetas-componentes $\mathbf{\Gamma} = (\Gamma^1, \Gamma^2, \dots, \Gamma^m)^\top$ provee las bases para usar el Algoritmo EM en el proceso de estimación del vector de parámetros del modelo de mezclas.

Ejemplo 2.2 (Mezcla de exponenciales). *Supóngase que los componentes del modelo de mezclas en (2.25) proviene de la misma familia exponencial parametrizada por $\omega > 0$:*

$$\mathcal{P}_\omega = \{F(t; \omega) = 1 - \exp(-\omega t) \mid \omega \in (0, \infty)\}.$$

Entonces, la función de supervivencia de T_i es

$$S(t_i; \boldsymbol{\theta}) = \pi_1 S_1(t_i; \omega_1) + \pi_2 S_2(t_i; \omega_2) + \cdots + \pi_m S_m(t_i; \omega_m)$$

donde $\boldsymbol{\theta} = (\pi_1, \dots, \pi_{m-1}; \omega_1, \dots, \omega_m)$ es el vector de parámetros desconocidos de la mezcla con

$$\pi_j > 0, \quad y \quad \sum_{j=1}^m \pi_j = 1,$$

y la j -ésima distribución componente, la cual denota la función de supervivencia (propia) condicionada a que la falla es de la j -ésima causa, está dada por

$$S_j(t_i; \omega_j) = \exp(-\omega_j t_i).$$

□

2.3. Modelos paramétricos: Estimación

2.3.1. Método de Máxima Verosimilitud para datos censurados.

La aplicación del Método de Máxima Verosimilitud en Análisis de Supervivencia con *datos sin censura* se describe en el Apéndice A. En el caso de datos de supervivencia que presentan censura, hay que establecer la contribución de cada observación (censurada o no-censurada) a la función de verosimilitud, por lo que la construcción de esta difiere ligeramente. Para fines del presente trabajo, solamente se especificará la función de verosimilitud para la censura aleatoria y de Tipo I.

Sin pérdida de generalidad, se procederá en el caso de variables (absolutamente) continuas, pero se obtienen resultados similares en el caso de v.a. discretas.

Verosimilitud para datos con censura de Tipo I

Suponga un modelo de censura por la derecha (Tipo I). Sea Y_1, Y_2, \dots, Y_n una muestra de v.a.s. i.i.d. con función de densidad común $f(t; \boldsymbol{\theta})$ y *constantes de censuras* asociadas C_1, C_2, \dots, C_n , respectivamente. Recordando la notación para datos censurados por la derecha, estos se pueden escribir como

$$(T_1, \delta_1), (T_2, \delta_2), \dots, (T_n, \delta_n)$$

donde $T_i = \min(Y_i, C_i)$ y $\delta_i = I\{Y_i \leq C_i\}$ su indicador de censura. Sea t_1, t_2, \dots, t_n una realización, con censura presente, de T_1, T_2, \dots, T_n .

El siguiente teorema establece la función de verosimilitud para datos censurados por la derecha (de Tipo I).

Teorema 2.3.1. *Bajo la censura por la derecha de Tipo I con tiempos de censura fijos, la función de verosimilitud, $L(\boldsymbol{\theta})$, de los datos observados (t_i, δ_i) , $i = 1, 2, \dots, n$, está dada por*

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n [f(t_i; \boldsymbol{\theta})]^{\delta_i} [S(t_i; \boldsymbol{\theta})]^{1-\delta_i}, \quad (2.26)$$

Demostración. Comencemos por encontrar una expresión para

$$\Pr(\{T = t\} \cap \{\delta = 0\}) \quad \text{y} \quad \Pr(\{T = t\} \cap \{\delta = 1\}).$$

Si un individuo tiene un tiempo de supervivencia censurado en $C = t_c$, todo lo que sabemos es que su tiempo de supervivencia Y excede a t_c . La probabilidad de este evento es $S(t_c; \boldsymbol{\theta}) = \Pr(Y > t_c)$. En efecto, cuando $\delta = 0$, $T = t_c$ con probabilidad 1, de manera que

$$\begin{aligned} \Pr(\{T = t_c\} \cap \{\delta = 0\}) &= \Pr(T = t_c \mid \delta = 0) \Pr(\delta = 0) \\ &= \Pr(\delta = 0) = \Pr(Y > t_c) \\ &= S(t_c; \boldsymbol{\theta}). \end{aligned}$$

Luego, si un individuo tiene tiempo de supervivencia no-censurado y registrado en $Y = t_u$, condicionalmente, cuando $\delta = 1$, $Y \leq C$, por lo que

$$\begin{aligned} \Pr(\{T = t_u\} \cap \{\delta = 1\}) &= \Pr(T = t_u \mid \delta = 1) \Pr(\delta = 1) \\ &= \Pr(Y = t_u \mid Y \leq C) \Pr(Y \leq C) \\ &= \left[\frac{f(t_u; \boldsymbol{\theta})}{F(C; \boldsymbol{\theta})} \right] [F(C; \boldsymbol{\theta})] = f(t_u; \boldsymbol{\theta}) \end{aligned}$$

Así, la contribución a la verosimilitud del par aleatorio (t, δ) es

$$\Pr(T = t, \delta) = f(t; \boldsymbol{\theta})^\delta S(t; \boldsymbol{\theta})^{1-\delta},$$

Por lo tanto, la función de verosimilitud para la muestra aleatoria observada $(t_1, \delta_1), (t_2, \delta_2), \dots, (t_n, \delta_n)$ es

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n f(t_i; \boldsymbol{\theta})^{\delta_i} S(t_i; \boldsymbol{\theta})^{1-\delta_i}. \quad (2.27)$$

■

Note que si consideramos la relación en (2.7), la función de verosimilitud se puede reescribir como

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n h(t_i; \boldsymbol{\theta})^{\delta_i} S(t_i; \boldsymbol{\theta}).$$

Verosimilitud para datos con censura aleatoria

En la notación de la Sección 1.2, Y es una v.a.s. continua con función de distribución $F(t)$, función de supervivencia $S(t)$ y f.d.p. $f(t)$. El tiempo de censura (aleatorio) asociado a Y es la v.a. continua C con función de distribución acumulada $G(t)$ y f.d.p. $g(t)$. Bajo un mecanismo de censura no-informativo, se supone que Y y C son independientes. También se asume que $G(t)$ no depende de ninguno de los parámetros de $S(t)$, por lo que no aporta información alguna para la distribución del tiempo de

supervivencia Y . En este modelo de censura, lo que se observa por unidad muestral es el par aleatorio (T, δ) definido como

$$T = \min(Y, C) \quad \text{y} \quad \delta = I\{Y \leq C\}.$$

Una muestra observada de tiempos de supervivencia para n individuos consiste de los pares

$$(t_1, \delta_1), (t_2, \delta_2), \dots, (t_n, \delta_n),$$

El siguiente Teorema determina la función de verosimilitud para la muestra aleatoria observada descrita anteriormente.

Teorema 2.3.2. *En un modelo con censura aleatoria por la derecha, la función de verosimilitud, L , de los datos observados (t_i, δ_i) , $i = 1, 2, \dots, n$, es*

$$L = \left[\prod_{i=1}^n f(t_i)^{\delta_i} S(t_i)^{1-\delta_i} \right] \left[\prod_{i=1}^n (1 - G(t_i))^{\delta_i} g(t_i)^{1-\delta_i} \right], \quad (2.28)$$

Demostración. *Suponga un modelo con censura aleatoria por la derecha cuyos datos observados son:*

$$(t_1, \delta_1), (t_2, \delta_2), \dots, (t_n, \delta_n).$$

De (1.5), la contribución a la verosimilitud del par (t_i, δ_i) es

$$\Pr\{T_i = t_i, \delta_i\} = \left[f(t_i) (1 - G(t_i)) \right]^{\delta_i} \left[g(t_i) S(t_i) \right]^{1-\delta_i},$$

por lo que función de verosimilitud para (t_i, δ_i) , $i = 1, 2, \dots, n$, es

$$L = \prod_{i=1}^n \left[f(t_i) (1 - G(t_i)) \right]^{\delta_i} \left[g(t_i) S(t_i) \right]^{1-\delta_i}.$$

■

Observación 2.7. Ya que $G(t)$ y $g(t)$ no involucran a ninguno de los parámetros en $f(t)$, el último término en (2.28) se tratar como constante al momento de maximizar a L , por lo que pueden ser ignorados, obteniendo la expresión:

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

Ejemplo 2.3 (Modelo Exponencial). *Supongamos que*

$$Y_1, Y_2, \dots, Y_n \stackrel{i.i.d.}{\sim} \text{Exp}(\theta) \quad (\theta > 0),$$

es una muestra (censurada por la derecha) de tamaño n , donde Y_i ($i = 1, 2, \dots, n$) es una v.a.s. con f.d.p. y función de supervivencia

$$f(y_i; \theta) = \theta^{-1} \exp(-y_i/\theta) \quad \text{y} \quad S(y_i; \theta) = \exp(-y_i/\theta) \quad \forall y \in (0, \infty),$$

respectivamente. Denotemos por

$$(t_1, \delta_1)^\top, (t_2, \delta_2)^\top, \dots, (t_n, \delta_n)^\top$$

el conjunto de datos de tiempos de supervivencia observados. De esa manera, si Y_i no está censurado ($\delta_i = 1$), con tiempo de supervivencia observado en y_i , el valore realizado t_i de T_i es igual a y_i , cuando por el contrario, si Y_i está censurado en c_i ($\delta_i = 0$), entonces $t_i = c_i$ y $Y_i > c_i$.

En este ejemplo, el vector de parámetros desconocido θ es un escalar, siendo igual a θ . Supongamos ahora que las observaciones han sido reordenadas de manera que t_1, t_2, \dots, t_r denotan las r observaciones no censuradas y $t_{r+1}, t_{r+2}, \dots, t_n$ las $n-r$ observaciones censuradas. Entonces de (2.27), la función de verosimilitud para θ formada en base a \mathbf{y} , es

$$\begin{aligned} L(\theta | \mathbf{y}) &= \prod_{i=1}^n [\theta^{-1} \exp(-t_i/\theta)]^{\delta_i} [\exp(-t_i/\theta)]^{1-\delta_i} \\ &= \theta^{-r} \exp\left(\frac{1}{\theta} \sum_{i=1}^n t_i\right). \end{aligned}$$

Así, la función log-verosimilitud es

$$l(\theta) = -r \log \theta - \frac{1}{\theta} \sum_{i=1}^n t_i.$$

Se sigue que el EMV de θ es la solución de $\partial l(\hat{\theta})/\partial \hat{\theta} = 0$, que es

$$\hat{\theta} = \sum_{i=1}^n \frac{t_i}{r}. \quad (2.29)$$

En este caso, $\hat{\theta}$ se pudo obtener de manera explícita, por lo que no hubo necesidad de un cálculo iterativo. \square

2.4. Ejemplo: Distribución Weibull

Sea Y_1, Y_2, \dots, Y_n una muestra de v.a.s., posiblemente censurada a la derecha, tal que

$$Y_1, Y_2, \dots, Y_n \stackrel{i.i.d.}{\sim} Weibull(\theta, \alpha).$$

La función de supervivencia correspondiente de la distribución Weibull es

$$S(y) = \exp(-\theta y^\alpha) \quad \forall y > 0, \quad (2.30)$$

y la correspondiente función de riesgo es

$$h(y) = \theta \alpha y^{\alpha-1} \quad \forall y > 0.$$

Al tomar la transformación logaritmo natural del tiempo Y , lo que observamos son pares de la forma

$$(T_1, \delta_1), (T_2, \delta_2), \dots, (T_n, \delta_n),$$

donde $T_i = \min(W, \log C_i)$, $W = \log Y_i$, y δ_i su indicador de censura.

De (2.17) se tiene que la v.a. W tiene función de supervivencia

$$S_w(w) = \exp\{-\theta \exp(\alpha w)\},$$

Redefiniendo los parámetros como

$$\theta = \exp\left(-\frac{\mu}{\sigma}\right) \quad \text{y} \quad \sigma = \frac{1}{\alpha} \quad \text{con} \quad -\infty < \mu < \infty, \quad \sigma > 0,$$

se obtiene el siguiente modelo log-lineal de Tiempo de Falla Acelerado

$$W = \mu + \sigma W^*, \tag{2.31}$$

donde μ es el *parámetro de localización* (o intercepción), σ es el *parámetro de escala* y $W^* \sim \text{ValorExtremo}(0, 1)$, cuya función de supervivencia es

$$S_{w^*}(w^*) = \exp(-\exp^{w^*}), \quad \forall w^* \in (-\infty, \infty).$$

El modelo de tiempo de falla acelerado dado en (2.31) corresponde a un modelo de regresión de una población homogénea (ausencia de variables explicativas).

De esta forma las funciones de densidad de probabilidad y de supervivencia, respectivamente, para W son

$$f_w(t) = \frac{1}{\sigma} \exp\left\{\frac{w - \mu}{\sigma} - \exp\left[\frac{w - \mu}{\sigma}\right]\right\}$$

y

$$S_w(w) = \exp\left(-\exp[(w - \mu)/\sigma]\right).$$

Así, la función log-verosimilitud para datos de supervivencia con censura por la derecha esta dada por

$$l(\boldsymbol{\theta}) = \sum_{i=1}^n \delta_i \log f_{\tau}(t_i) + \sum_{i=1}^n (1 - \delta_i) \log S_{\tau}(t_i),$$

donde $\boldsymbol{\theta} = (\mu, \sigma)$. Para encontrar el EMV de $\boldsymbol{\theta}$ hay que resolver el sistema de ecuaciones:

$$\begin{cases} \frac{\partial l(\hat{\boldsymbol{\theta}})}{\partial \hat{\mu}} = 0 \\ \frac{\partial l(\hat{\boldsymbol{\theta}})}{\partial \hat{\sigma}} = 0 \end{cases} \tag{2.32}$$

Ya que las ecuaciones resultantes del sistema anterior son no lineales, se procede a resolverlas mediante algún método iterativo.

Una vez que se ha obtenido el EMV $\hat{\theta} = (\hat{\mu}, \hat{\sigma})$ de $\theta = (\mu, \sigma)$ se procede a calcular la matriz información observada

$$\mathbf{i}(\hat{\theta}) = -\frac{\partial^2 l(\theta)}{\partial \theta^2} \Big|_{\theta=\hat{\theta}}. \quad (2.33)$$

De acuerdo a (A.6) la inversa de (2.33) es un estimador de la matriz de varianza-covarianza de θ :

$$\mathbf{i}^{-1}(\hat{\theta}) = \begin{pmatrix} \widehat{Var}[\hat{\mu}] & \widehat{Cov}[\hat{\mu}, \hat{\sigma}] \\ \widehat{Cov}[\hat{\mu}, \hat{\sigma}] & \widehat{Var}[\hat{\sigma}] \end{pmatrix} \approx \begin{pmatrix} Var[\mu] & Cov[\mu, \sigma] \\ Cov[\mu, \sigma] & Var[\sigma] \end{pmatrix}$$

Tanto métodos iterativos para resolver el sistema (2.32) como la aproximación de la matriz de varianza-covarianza de θ están disponibles en varios paquetes estadísticos⁸.

Usando la propiedad de invarianza de los EMV de μ y σ se obtienen los EMV de θ y α dados por

$$\hat{\theta} = \exp(-\hat{\mu}/\hat{\sigma}) \quad \text{y} \quad \hat{\alpha} = 1/\hat{\sigma}.$$

Para obtener la varianza-covarianza de $\hat{\theta}$ y $\hat{\alpha}$ aplicamos el Método Delta, el cual consiste en una aproximación de Serie de Taylor de segundo orden:

$$Var[g(\hat{\theta}_1, \hat{\theta}_2)] \approx \left(\frac{\partial g}{\partial \hat{\theta}_1} \right)^2 Var[\hat{\theta}_1] + \left(\frac{\partial g}{\partial \hat{\theta}_2} \right)^2 Var[\hat{\theta}_2] + 2 \left(\frac{\partial g}{\partial \hat{\theta}_1} \frac{\partial g}{\partial \hat{\theta}_2} \right) Cov[\hat{\theta}_1, \hat{\theta}_2].$$

y

$$\begin{aligned} Cov[g_1(\hat{\theta}_1, \hat{\theta}_2), g_2(\hat{\theta}_1, \hat{\theta}_2)] &\approx \left(\frac{\partial g_1}{\partial \hat{\theta}_1} \frac{\partial g_2}{\partial \hat{\theta}_1} \right) Var[\hat{\theta}_1] + \left(\frac{\partial g_1}{\partial \hat{\theta}_2} \frac{\partial g_2}{\partial \hat{\theta}_2} \right) Var[\hat{\theta}_2] \\ &+ \left(\frac{\partial g_1}{\partial \hat{\theta}_1} \frac{\partial g_2}{\partial \hat{\theta}_2} + \frac{\partial g_1}{\partial \hat{\theta}_2} \frac{\partial g_2}{\partial \hat{\theta}_1} \right) Cov[\hat{\theta}_1, \hat{\theta}_2]. \end{aligned}$$

Ya que se han obtenido los EMV de θ y α y su correspondiente varianza, se procede a calcular intervalos de confianza y a realizar pruebas de hipótesis. Por ejemplo, si el intervalo de confianza para el EMV del parámetro de forma α contiene el valor 1, es posible que el modelo exponencial ajuste mejor nuestro conjunto de datos que el modelo Weibull (ver [31]. pag 135,136).

Ejemplo 2.4. *Considere los tiempos de falla del Ejemplo 1.4 de aquellos pacientes que recibieron un trasplante autólogo. Con la finalidad de ilustrar el proceso de inferencia descrito anteriormente, suponga que el Modelo Weibull se ajusta⁹ bien a ese conjunto de datos de tiempos de falla. El código en **R** es el siguiente:*

⁸ R, S-Plus, Minitab, S.A.S., SPSS, etc.

⁹ Para checar si un cierto conjunto de datos sigue una distribución *Weibull*(θ, α) se pueden emplea una gráfica de probabilidad. Ver [31], pág. 61, para más detalles.

Código en R. Ejemplo 2.4.

```

> library(KMsurv); data(alloauto);
> attach(alloauto) #se usan las variables de "alloauto" como objetos
> tiempo<-time[type==2] ; estatus<-delta[type==2];
> auto<-data.frame(tiempo,estatus)
> detach(alloauto)
# Realiza la regresion para el Modelo Weibull;
> auto.wei<-survreg(Surv(tempo,estatus)~1,data=auto, dist='weibull')
> summary(auto.wei)

```

	Value	Std. Error	z	valor-p
(Intercept)	3.452	0.218	15.821	2.24e-56
Log(scale)	0.105	0.158	0.666	5.05e-01

```

Scale= 1.11
Weibull distribution
Loglik(model)= -123.441  Loglik(intercept only)= -123.441

```

Ya que el paquete estadístico **R** nos da un estimador de $\log \sigma$, los EMV junto con su error estándar obtenidos, respectivamente, son

$$\hat{\mu} = 3.452, \quad s.e.[\hat{\mu}] = 0.218$$

y

$$\log \hat{\sigma} = 0.105, \quad s.e.[\log \hat{\sigma}] = 0.158.$$

Los límites de un intervalo de confianza al 95 % para $\log \hat{\sigma}$ son

$$\log \hat{\sigma} \pm 1.96 s.e.[\log \hat{\sigma}],$$

de manera que, un intervalo de confianza para $\log \hat{\sigma}$ es $(-2.20468, 0.41468)$. Como este intervalo incluye al cero y ya que

$$\alpha = 1 \quad \text{si y sólo si} \quad \log \sigma = 0,$$

es plausible que el modelo Exponencial proporcione un mejor ajuste a los datos que el modelo Weibull. Para eso, se prueba la hipótesis $H_0 : \alpha = 1$ (equivalentemente que $\log \sigma = 0$) mediante algún tipo de prueba, por ejemplo, la prueba de la razón de verosimilitud.

El resumen proporcionado por **R** para el conjunto de datos bajo un modelo Exponencial se obtiene con las mismas instrucciones, pero cambiando la distribución "weibull" por "exponential", dando como resultado:

```


```

	Value	Std. Error	z	p
(Intercept)	3.42	0.189	18.1	4.52e-73

```

Scale fixed at 1

Exponential distribution
Loglik(model)= -123.674  Loglik(intercept only)= -123.674

```

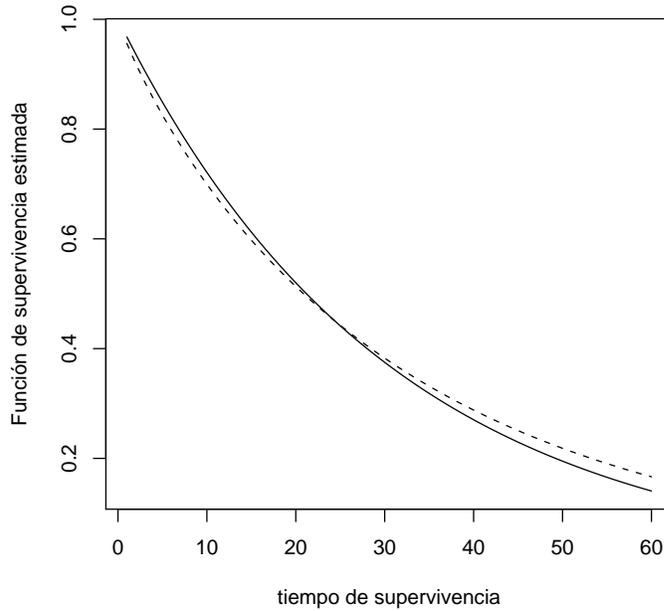


Figura 2.4: Función de supervivencia estimada con el ajuste de la distribución Weibull(0.045,0.90) y de la distribución Exponencial(0.0327) (línea continua).

Para el modelo Weibull la función log-verosimilitud es $\log L(\hat{\alpha}, \hat{\omega}) = -123.441$ y para el modelo exponencial es $\log L(\hat{\omega}_0) = -123.674$. Rechazamos la hipótesis nula si $\chi_{obs}^2 \geq 3.842$, donde

$$\chi_{obs}^2 = -2 \log \left[\frac{L(\hat{\omega}_0)}{L(\hat{\alpha}, \hat{\omega})} \right] = -2 \left[\log L(\hat{\omega}_0) - \log L(\hat{\alpha}, \hat{\omega}) \right], \quad (2.34)$$

y 3.842 es el valor crítico del percentil $\chi_{0.05,1}^2$. Al substituir los valores correspondientes en (2.34) se tiene que $\chi_{obs}^2 = 0.467$. El valor p es

$$\text{valor } p = \Pr(X > \chi_{obs}^2) = 0.494, \quad \text{donde } X \sim \chi_1^2,$$

el cual es mayor que 0.05. Por lo tanto, no tenemos evidencia alguna para rechazar la hipótesis nula, lo cual sugiere que una distribución exponencial puede proporcionar un mejor ajuste que la distribución Weibull para ese conjunto de datos.

Los correspondientes EMV del Modelo Weibull son $\hat{\theta} = 0.045$, $\hat{\alpha} = 0.900$. El EMV para el modelo exponencial es $\hat{\theta} = 0.0327$. Las curvas de las funciones de supervivencia se obtiene al substituir los EMV en cada modelo, respectivamente (ver Figura 2.4).

□

ESTIMACIÓN NO-PARAMÉTRICA DE LA FUNCIÓN DE SUPERVIVENCIA

Cuando se quiere hacer inferencia en una muestra aleatoria proveniente de una población en la que se desea estudiar alguna característica, la selección del método de inferencia (e.g. paramétrico, no-paramétrico) es importante, ya que una mala selección de este conducirá a conclusiones erróneas. Uno de los aspectos que se puede tomar en cuenta para la selección del método de inferencia es la manera en la que se distribuyen los datos. Sin embargo, algunos métodos de inferencia estadísticos se basan en suposiciones que no siempre son verificadas por la muestra. Por ejemplo, un histograma construido a partir de una muestra de tiempos de falla con censura a la derecha, por lo regular, no es simétrico. En consecuencia, no es razonable suponer que este tipo de datos tiene una distribución normal. Cuando la ausencia de simetría es obvia, puede aplicarse alguna transformación a la muestra para que su distribución sea aproximadamente simétrica, e.g. la transformación logaritmo natural, y de esta manera emplear métodos de inferencia basados en la suposición de normalidad.

Por otra lado, se optaría por un modelo paramétrico (e.g. Exponencial, Weibull) si se estuviese seguro de que tal modelo se ajusta bien a nuestros datos. A pesar de que existen métodos para evaluar la bondad de ajuste de un modelo paramétrico, podemos no estar completamente seguros de que este sea el adecuado. Entonces, cuando hay incertidumbre en la selección de un modelo paramétrico para cierto conjunto de datos, se pueden utilizar métodos de inferencia *no-paramétricos*, los cuales no requieren de supuestos sobre la forma funcional de la distribución de probabilidad de los datos. Por esta razón también son llamados *libres de distribución*.

El método de inferencia no-paramétrico que se describe en este Capítulo proporciona un estimador de la función de supervivencia de una muestra aleatoria independiente de tiempos de falla con censura a la derecha, llamado *estimador producto-límite* (PL) o estimador¹ Kaplan-Meier (KM).

¹ Kaplan y Meier fueron los primeros en discutir las propiedades del estimador PL en su artículo de 1958, de aquí que este es también llamado el estimador KM.

3.1. La Distribución Empírica.

Sea y_1, y_2, \dots, y_n realizaciones de la muestra de v.a.s. Y_1, Y_2, \dots, Y_n en la que *ninguna de las observaciones presenta censura*, además de no hacer suposición alguna de su distribución. A partir de eso, ¿Como se puede estimar la función de supervivencia, $S(y) = \Pr(Y > y)$? El método más simple es a través de la *función de supervivencia empírica*.

Definición 3.1. Si no hay observaciones censuradas en la muestra de tamaño n y_1, y_2, \dots, y_n , la función de supervivencia empírica (FSE) se define como

$$\widehat{S}(y) = \frac{\text{número de observaciones } > y}{n} \quad y \geq 0, \quad (3.1)$$

la cual proporciona un estimador no-paramétrico de la función de supervivencia S_Y .

Cuando todas las observaciones son distintas, la FSE es una función escalonada (decreciente) con saltos cuya longitud es igual a $1/n$, solamente después de cada tiempo de supervivencia observado. En el caso de que se tengan d tiempos de supervivencia empatados e iguales a y , la FSE decrece con longitud de salto igual a d/n . Además, si $y_{(1)}$ es la observación más pequeña y $y_{(n)}$ la más grande, entonces $\widehat{S}(y) = 1$ para $y < y_{(1)}$ y $\widehat{S}(y) = 0$ para $y > y_{(n)}$.

Debido a las limitaciones de la FSE en presencia de censura por la derecha en los datos, es necesario hacer algunas modificaciones de (3.1), las cuales se describen a continuación.

3.2. El Estimador Producto-Limite

Bajo la notación de la Sección 1.2, considere la muestra aleatoria homogénea observada (posiblemente censurada por la derecha) de tiempos de supervivencia continuos

$$(t_1, \delta_1), (t_2, \delta_2), \dots, (t_n, \delta_n),$$

generada a partir de alguna función de supervivencia (fundamental) desconocida $S(t)$. Como se mencionó al principio, no se supone una forma paramétrica para $S(t)$. El proceso de estimación aquí es similar al que se da con el Estimador Actuarial ([1]) y en el caso de tiempos de supervivencia discretos (Sección 2.1). La diferencia es que en el Método Kaplan-Meier los diferentes tiempos de supervivencia observados, digamos $t_{(i)}$, $i=1,2,\dots,r$, toman el lugar de los tiempos fijos τ_j haciendo que la longitud de cada intervalo de la forma $I_j = [t_{(j)}, t_{(j+1)})$ sea variable.

Supongamos que inicialmente *no hay empates en las n observaciones*, y que de estas, r ($\leq n$) son tiempos de falla no-censurados y los restantes $n-r$ son observaciones censuradas. Sean $0 < t_{(1)} < t_{(2)} < \dots < t_{(r)} < \infty$ esos distintos tiempos de falla observados y ordenados. En cuanto al orden del indicador de censura, se tiene que $\delta_{(j)} = \delta_j$ cuando $t_{(j)} = t_j$, pero $\delta_{(1)}, \delta_{(2)}, \dots, \delta_{(n)}$ no están ordenados como los $t_{(j)}$'s. Además, *suponga que el mecanismo de censura es no-informativo*, es decir, que

el tiempo de supervivencia es independiente del mecanismo que causaría a éste su censura. Usando esos r distintos tiempos de falla observados, dividamos \mathbb{R}^+ en una sucesión de intervalos disjuntos, eso es,

$$I_0 = [t_0, t_{(1)}), I_1 = [t_{(1)}, t_{(2)}), \dots, I_r = [t_{(r)}, t_{(r+1)}), \quad \text{con } t_0 = 0 \text{ y } t_{r+1} = +\infty.$$

Definición 3.2. *El conjunto de riesgo al tiempo t , denotado por $R(t^-)$, es el conjunto de índices de los sujetos que permanecen vivos y bajo observación, "justo antes" del tiempo t , incluyendo los que están por morir en ese tiempo.*

Defina a lo siguiente:

- $n \equiv$ el número de individuos bajo estudio, es decir, el número de individuos inicialmente en riesgo;
- $d_j \equiv$ número de individuos quienes fallan por el evento de interés al tiempo $t_{(j)}$;
- $n_j \equiv$ el número de elementos en $R(t_{(j)}^-)$;
- $p_j \equiv \Pr(T > t_{(j)} \mid T > t_{(j-1)})$, la probabilidad condicional de sobrevivir a lo largo del j -ésimo intervalo, dado que a sobrevivido al inicio de éste.

Observación 3.1. Note que si m_j individuos están censurados en el intervalo I_j al tiempo $t_{(j)1}, t_{(j)2}, \dots, t_{(j)m_j}$, respectivamente, las etiquetas de sus tiempos de falla censurados pertenecen al conjunto de riesgo $R(t_{(j)}^-)$, es decir, están incluidos en n_j .

Deseamos encontrar un estimador de $S(t) = \Pr(T > t)$, mediante estimadores de los p_j 's. Por el momento, sólo se dará un breve argumento de la derivación de éstos (ver [1]), pero más adelante se verá que el estimador de cada p_j es en realidad el EMV no-paramétrico de $1 - h(t)$ en $t = t_{(j)}$.

Sea $\varepsilon > 0$ un tiempo infinitesimal y considere el j -ésimo intervalo $I_j = [t_{(j)}, t_{(j+1)})$. Note que el intervalo de tiempo que va de $t_{(j)} - \varepsilon$ a $t_{(j)}$ solamente incluye un tiempo de supervivencia, mientras que el intervalo que va de $t_{(j)}$ a $t_{(j+1)} - \varepsilon$ no contiene ninguna falla, ya que éste se presenta hasta $t = t_{(j+1)}$. Como n_j es el número de individuos vivos poco antes de $t_{(j)}$ y d_j las fallas ocurridas en ese mismo tiempo, la probabilidad de que un individuo presente una falla en el intervalo que va de $t_{(j)} - \varepsilon$ a $t_{(j)}$ se estima con d_j/n_j . Así, la probabilidad condicional de que un individuo sobreviva más allá de $t_{(j)}$ dado que ha sobrevivido hasta $t_{(j)} - \varepsilon$ es estimada por $1 - d_j/n_j$, y la probabilidad de sobrevivir de $t_{(j)}$ a $t_{(j+1)} - \varepsilon$ es uno. De esta manera, la probabilidad conjunta de sobrevivir de $t_{(j)} - \varepsilon$ a $t_{(j)}$ y de $t_{(j)}$ a $t_{(j+1)} - \varepsilon$ se puede estimar con $(1 - \frac{d_j}{n_j}) * (1)$. Por lo tanto, en el límite cuando $\varepsilon \rightarrow 0$,

$$1 - \frac{d_j}{n_j}$$

llega a ser un estimador de la probabilidad condicional p_j correspondiente al intervalo I_j . Entonces, para un tiempo fijo $t \in [t_{(i)}, t_{(i+1)})$ el producto de esos estimadores,

\hat{p}_j ($j = 1, 2, \dots, i$), da una estimación de la función de supervivencia al tiempo t , es es

$$\begin{aligned} S(t) &= \Pr(T > t) \\ &\approx \prod_{j=1}^i \hat{p}_j = \prod_{j=1}^i \left(1 - \frac{d_j}{n_j}\right) \\ &= \prod_{j | t_{(j)} \leq t} \left(1 - \frac{1}{n_j}\right). \end{aligned}$$

Definición 3.3. Denotemos por $0 < t_{(1)} < t_{(2)} < \dots < t_{(r)}$ los r ($\leq n$) distintos tiempos de falla observados en una muestra de tamaño n , es decir, no hay tiempos de supervivencia empatados. Para $t \in [t_{(i)}, t_{(i+1)})$ fijo, el **estimador producto-límite** $\hat{S}(t)$ de $S(t)$ se define como

$$\hat{S}(t) = \prod_{j | t_{(j)} \leq t} \left(1 - \frac{1}{n_j}\right) \quad \text{con} \quad \hat{S}(0) = 1. \quad (3.2)$$

De acuerdo a (3.2), la gráfica de $\hat{S}(t)$ es una función escalonada, en la cual las probabilidades de supervivencia son constantes entre los tiempos de falla adyacentes, es monótona decreciente, continua por la derecha y discontinua por la izquierda² con saltos de discontinuidad "únicamente" en las observaciones no-censuradas.

Al momento de recoger datos de supervivencia, estos pueden presentar empates, ya sea entre observaciones censuradas, no censuradas o una combinación de ambas, por lo que la definición del estimador Kaplan-Meier (KM) dado en (3.2) debe ser redefinida, con el objetivo de considerar los casos de observaciones empatadas. Otro caso que se debe tomar en cuenta es cuando la observación más grande, digamos $t_{(r)}$, es un tiempo de supervivencia censurado, ya que el estimador-KM solamente presenta saltos en observaciones no-censuradas.

Caso 1 *Observaciones no-censuradas empatadas.*

Supóngase que antes del tiempo t_j hay n_j individuos libres del evento (e.g. vivos), y al tiempo t_j ocurren d_j ($d_j > 1$) fallas (e.g. muertes). Imagine que las d_j fallas son separadas sobre un intervalo de tiempo infinitesimalmente pequeño. De esta manera se obtiene una expresión para el factor de las d_j fallas en el estimador KM, el cual está dado por

$$\begin{aligned} &\left(1 - \frac{1}{n_j}\right) \left(1 - \frac{1}{n_j - 1}\right) \dots \left(1 - \frac{1}{n_j - d_j + 1}\right) \\ &= \left(\frac{n_j - 1}{n_j}\right) \left(\frac{n_j - 2}{n_j - 1}\right) \dots \left(\frac{n_j - d_j}{n_j - d_j + 1}\right) \\ &= \frac{n_j - d_j}{n_j}. \end{aligned}$$

² Cuando se toma $\{j | t_{(j)} < t\}$ en (3.2), entonces $\hat{S}(t)$ es una función escalonada continua por la izquierda.

Caso 2 *Empates entre observaciones censuradas y no-censuradas.* Es posible que varios tiempos de falla censurados sean registrados con el mismo tiempo que una o más tiempos de fallas no-censurados. Entonces, suponga que los tiempos de falla censuradas ocurren inmediatamente después de los que no están censurados. De esta manera, las observaciones censuradas contribuyen al número de individuos en riesgo en ese tiempo.

Caso 3 *La observación más grande esta censurada.* Cuando la observación más grande, digamos τ , es un tiempo de supervivencia censurado, $\widehat{S}(t)$ no esta definido para $t > \tau$, es decir,

$$\lim_{t \rightarrow \infty} \widehat{S}(t) > 0 \quad \text{si } t > \tau.$$

Por otra parte, cuando τ es un tiempo no-censurado, entonces, $\widehat{S}(t) = 0$ para $t > \tau$.

La consideración de esos casos permite la generalización de la Definición 3.3 para observaciones empatadas en los datos.

Definición 3.4. Sean $t_{(1)} < t_{(2)} < \dots < t_{(r)}$ los r distintos tiempos de falla observados en una muestra aleatoria de tamaño n , posiblemente con censurada por la derecha. La posibilidad de que se de más de una falla en $t_{(j)}$ está permitido, por lo que denotaremos al número de fallas en ese tiempo por d_j (≥ 1). Para t fijo, el **estimador Kaplan-Meier** \widehat{S} de S se define como

$$\widehat{S}(t) = \prod_{j | t_{(j)} \leq t} \left(1 - \frac{d_j}{n_j} \right) \quad (3.3)$$

Se puede verificar fácilmente que si $t_{(j)}$ es una observación no-censurada, entonces

$$\widehat{S}(t_{(j)}^+) := \lim_{\varepsilon \rightarrow 0} \widehat{S}(t_{(j)} + \varepsilon) = \widehat{S}(t_{(j)}),$$

es decir, el estimador-PL es una función escalonada continua por la derecha, con puntos de discontinuidad presentes solamente en las observaciones no-censuradas. Más aún, si no hay empates en $t = t_{(j)}$, el tamaño de la discontinuidad en ese punto es

$$\widehat{S}(t_{(j)}^-) - \widehat{S}(t_{(j)}) = \frac{\widehat{S}(t_{(j-1)})}{n_j}.$$

En caso de que se presenten d_j observaciones no-censuradas en $t_{(j)}$, el tamaño de la discontinuidad es

$$\widehat{S}(t_{(j-1)}) \frac{d_j}{n_j}.$$

El estimador-KM puede ser determinado de manera recursiva en observaciones no-censuradas adyacentes como sigue:

Sean $t_{(j-1)} < t_{(j)}$ dos observaciones no-censuradas adyacentes. Por un lado

$$\Pr(\{T > t_{(j)}\} \cap \{T > t_{(j-1)}\}) = \Pr(T > t_{(j)}) = S(t_{(j)}). \quad (3.4)$$

Por el otro lado,

$$\begin{aligned} \Pr(\{T > t_{(j)}\} \cap \{T > t_{(j-1)}\}) &= \Pr(T > t_{(j)} \mid T > t_{(j-1)}) \Pr(T > t_{(j-1)}) \\ &= \Pr(T > t_{(j)} \mid T > t_{(j-1)}) S(t_{(j-1)}). \end{aligned} \quad (3.5)$$

Ya que $(1 - d_j/n_j)$ es un estimador de p_j , y considerando a (3.4) y (3.5) se tiene que

$$\widehat{S}(t_{(j)}) = \widehat{S}(t_{(j-1)}) \left(1 - \frac{d_j}{n_j}\right). \quad (3.6)$$

Otra forma de deducir (3.6) es considerando el tamaño de la discontinuidad en una observación no-censurada.

Observación 3.2. Nótese que en ausencia de censura en los datos observados, el estimador-PL, $\widehat{S}(t)$, se reduce a la función de supervivencia empírica

$$\widetilde{S}(t) = \frac{\text{número de observaciones } > t}{n}.$$

3.2.1. El estimador-KM como un EMV

De la manera como se describe en Kalbfleisch and Prentice (1980), el estimador Kaplan-Meier es el estimador no-paramétrico de máxima verosimilitud de la función de supervivencia.

Sean $t_1 < t_2 < \dots < t_r$ los distintos tiempos de supervivencia observados y no-censurados en una muestra de tamaño n , correspondiente a una población homogénea con función de supervivencia $S(t)$. Suponga que d_j individuos fallan en t_j y m_j individuos están censurados en el intervalo $[t_j, t_{j+1})$ al tiempo $t_{j1}, t_{j2}, \dots, t_{jm_j}$ ($j = 0, 1, 2, \dots, r$), donde $t_0 = 0$ y $t_{r+1} = \infty$. Sea $n_j = (m_j + d_j) + \dots + (m_r + d_r)$ el número de individuos en riesgo justo antes del tiempo t_j . Suponga un modelo de supervivencia en tiempo discreto. La contribución a la verosimilitud de una observación no-censurada es

$$f(t_j) = \Pr(T = t_j) = S(t_j^-) - S(t_j),$$

y la contribución a la verosimilitud de un tiempo de supervivencia censurado en t_{jl} es

$$\Pr(T > t_{jl}) = S(t_{jl}),$$

Bajo un mecanismo de censura no-informativo, la función de verosimilitud formulada con base en los tiempos de falla observados es

$$L = \prod_{j=1}^r \left\{ [S(t_j^-) - S(t_j)]^{d_j} \prod_{l=1}^{m_j} S(t_{jl}) \right\} \quad (3.7)$$

El EMV $\hat{S}(t)$ de $S(t)$ debe ser discontinuo por la izquierda en los tiempos de falla observados, de lo contrario $S(t_j^-) = S(t_j)$ lo cual hace que $L=0$. Además, ya que $t_{jl} > t_j$, $\hat{S}(t_{jl}) \leq \hat{S}(t_j)$ y por eso L es maximizada tomando $S(t_{jl}) = S(t_j)$ ($j = 1, 2, \dots, r$, $l = 1, 2, \dots, m_j$) con $S(t_{0l}) = 1$ ($l = 1, 2, \dots, m_0$). Del Teorema 2.1.3, la función de supervivencia con función de riesgo $h(t_j) \equiv h_j$ está dada por

$$S(t_i^-) = \prod_{j=1}^{i-1} (1 - h_j), \quad (3.8)$$

y

$$S(t_i) = \prod_{j=1}^i (1 - h_j). \quad (3.9)$$

Substituyendo (3.8) y (3.9) en (3.7), la función de verosimilitud resultante es

$$L = \prod_{j=1}^r \left\{ \left[\prod_{l=1}^{j-1} (1 - h_l) - \prod_{l=1}^j (1 - h_l) \right]^{d_j} \prod_{l=1}^{m_j} S(t_{jl}) \right\} \quad (3.10)$$

El primer factor de (3.10) se puede simplificar de la siguiente manera

$$\prod_{l=1}^{j-1} (1 - h_l) - \prod_{l=1}^j (1 - h_l) = \prod_{l=1}^{j-1} (1 - h_l) (1 - (1 - h_j)) = \left[\prod_{l=1}^{j-1} (1 - h_l) \right] h_j.$$

El segundo factor de (3.10) se puede reescribir como

$$\prod_{l=1}^{m_j} S(t_{jl}) = \prod_{l=1}^{m_j} S(t_j) = [S(t_j)]^{m_j} = \prod_{l=1}^j (1 - h_l)^{m_j}.$$

Al substituir esas expresiones en (3.10) y considerando que

$$n_1 = \sum_{j=1}^r (d_j + m_j), n_2 = \sum_{j=2}^r (d_j + m_j), \dots, n_r = d_r + m_r$$

se tiene que

$$\begin{aligned}
L &= \prod_{j=1}^r \left\{ \prod_{l=1}^{j-1} (1-h_l)^{d_j} h_j^{d_j} \left[\prod_{l=1}^j (1-h_l)^{m_j} \right] \right\} \\
&= \prod_{j=1}^r \left\{ \prod_{l=1}^j \left[(1-h_l)^{d_j} \right] (1-h_j)^{-d_j} h_j^{d_j} \left[\prod_{l=1}^j (1-h_l)^{m_j} \right] \right\} \\
&= \left[\prod_{j=1}^r h_j^{d_j} (1-h_j)^{-d_j} \right] \prod_{j=1}^r \left[\prod_{l=1}^j (1-h_l)^{d_j} \prod_{l=1}^j (1-h_l)^{m_j} \right] \\
&= \left[\prod_{j=1}^r h_j^{d_j} (1-h_j)^{-d_j} \right] \prod_{j=1}^r \left[\prod_{l=1}^j (1-h_l)^{d_j+m_j} \right] \\
&= \left[\prod_{j=1}^r h_j^{d_j} (1-h_j)^{-d_j} \right] (1-h_1)^{n_1} (1-h_2)^{n_2} \cdots (1-h_r)^{n_r} \\
&= \left[\prod_{j=1}^r h_j^{d_j} (1-h_j)^{-d_j} \right] \prod_{j=1}^r (1-h_j)^{n_j} \\
&= \prod_{j=1}^r h_j^{d_j} (1-h_j)^{n_j-d_j} \tag{3.11}
\end{aligned}$$

Claramente, el estimador de máxima verosimilitud³ de h_j obtenido de (3.11) es

$$\hat{h}_j = \frac{d_j}{n_j}, \quad j = 1, 2, \dots, r.$$

Por lo tanto, el estimador *producto-límite* de la función de supervivencia está dado por

$$\hat{S}(t) = \prod_{j | t_j < t} (1 - \hat{h}_j) = \prod_{j | t_j < t} \left(1 - \frac{d_j}{n_j} \right).$$

3.2.2. Intervalos de confianza del estimador Kaplan-Meier.

La varianza del estimador Kaplan-Meier es aproximada por la Fórmula de Greenwood

$$Var[\hat{S}(t)] \approx \hat{\sigma}^2(t) = [\hat{S}(t)]^2 \sum_{t_j \leq t} \frac{d_j}{n_j(n_j - d_j)}. \tag{3.12}$$

Si no hay observaciones censuradas, (3.12) se reduce a $\hat{S}(t)[1 - \hat{S}(t)]/n$, que es el estimador de la varianza binomial estándar. Ver la Sección 2.1.3 de [1] para un mayor detalle.

³ Considere el vector $\mathbf{h} = (h_1, h_2, \dots, h_r)$ como los parámetros desconocidos y de interés en la distribución del tiempo de supervivencia. En este caso, la definición de máxima verosimilitud es una generalización del concepto usual empleado en modelos paramétricos (ver [28], pág 57 y 58: El estimador Kaplan-Meier da un Estimador de Máxima Verosimilitud Generalizado).

En el caso de muestras grandes el estimador Kaplan-Meier, en un tiempo fijo t , se distribuye aproximadamente Normal. Entonces, un intervalo de confianza al $100(1 - \alpha)\%$ de $\widehat{S}(t)$ para $S(t)$ está dado por

$$\widehat{S}(t) \pm z_{1-\alpha/2} \widehat{\sigma}(t),$$

donde $z_{1-\alpha/2}$ denota el percentil de una distribución normal estándar al nivel $1 - \alpha/2$, eso es, $\Pr(Z < z_{1-\alpha/2}) = 1 - \alpha/2$, con $Z \sim N(0, 1)$. Sin embargo, algunas veces los intervalos estandar así construidos pueden incluir valores fuera del rango $[0, 1]$. Esto puede ser evitado aplicando la distribución normal asintótica a una transformación de $S(t)$ para la cual el rango no está restringido. Por ejemplo, la transformación⁴

$$\log\{-\log \widehat{S}(t)\},$$

permite determinar límites de intervalos de confianza asintóticos al $100(1 - \alpha)\%$ dados por

$$\widehat{S}(t)^{\exp\{\pm z_{1-\alpha/2} \widehat{\sigma}(t) / [\widehat{S}(t) \log \widehat{S}(t)]\}},$$

los cuales toman valores en $[0, 1]$.

3.2.3. Integral producto y el estimador Nelson-Aalen

Sea T una v.a.s. con función de supervivencia $S_T(t) = \Pr(T > t)$ definida para todo $t \geq 0$, con $S(0) = 1$. Supongamos que T es continua, con función de densidad $f_T(t)$ y función de riesgo $h_T(t)$. Entonces, de la Sección 2.1, se cumple la siguiente relación

$$S_T(t) = \exp \left\{ - \int_0^t h_T(u) du \right\}. \quad (3.13)$$

Si T es una v.a.s. cuya distribución es discreta, tomando los posibles valores en $0 < t_1 < t_2 < \dots$, con función de masa de probabilidad $f_T(t) = \Pr(T = t)$ y función de riesgo $h_T(t) = \Pr(T = t \mid T \geq t) = f_T(t) / S_T(t^-)$, la función de supervivencia se expresa como

$$S_T(t) = \prod_{j|t_j < t} [1 - h_T(t_j)]. \quad (3.14)$$

Al comparar (3.13) y (3.14) debe notarse que la expresión de $S_T(t)$ en el caso continuo no proporciona una buena generalización del caso discreto, además de la falta de una intuitiva interpretación, en contraste con (3.14).

Gill *et. al* (1990) describe la manera en la que ambas fórmulas pueden ser unificadas e intuitivamente interpretables usando la *integral producto*, denotada comúnmente por el símbolo \prod , el cual está definida como el límite de aproximaciones de productos finitos⁵.

⁴ Ver [1] y [16] para detalles.

⁵ Esta definición es similar a la que se da para la ordinaria integral, denotada por el símbolo \int , la cual es el límite de aproximaciones de sumas finitas.

Comencemos por definir la fórmula

$$\Lambda_T(t) = \int_0^t \frac{dS_T(u)}{S_T(u^-)}, \quad (3.15)$$

que es una expresión general para la función de riesgo acumulada $H_T(t)$ ya que, en el caso continuo (3.15) da $\Lambda_T(t) = \int_0^t h_T(u) du$, y en el caso de una distribución discreta esta es $\Lambda_T(t) = \sum_{j|t_j \leq t} h_T(t_j)$. Sea du un tiempo infinitesimal; el riesgo

$$\Delta\Lambda_T(u) = \Lambda_T(u + du) - \Lambda_T(u) = \Pr(T \leq u + du \mid T > u)$$

se puede considerar como la probabilidad condicional de presentar una falla en el intervalo que va de u a $u + du$, dado que se ha sobrevivido al inicio de ese pequeño intervalo de tiempo, entonces

$$1 - \Delta\Lambda_T(u) = \Pr(T > u + du \mid T > u)$$

es interpretada como la probabilidad condicional de sobrevivir a través de este pequeño intervalo de tiempo, dado que se ha sobrevivido desde su inicio. De esta manera, la probabilidad incondicional $\Pr(T > t)$ será el límite de aproximaciones de productos finitos de la forma

$$\prod_{j=1}^k \Pr(T > t_j \mid T > t_{j-1}) = \prod_{j=1}^k [1 - \Pr(T \leq t_j \mid T > t_{j-1})],$$

donde $0 = t_0 < t_1 < \dots < t_k = t$ es una partición del intervalo de tiempo $[0, t]$. Por eso es que, usando la integral producto, la función de supervivencia de T , en ambos casos tanto continuo como discreto, se puede escribir como

$$S_T(t) = \prod_0^t [1 - d\Lambda_T(u)], \quad (3.16)$$

donde el *producto integral* \prod está definido por

$$\prod_0^t [1 - d\Lambda_T(u)] = \lim_{k \rightarrow \infty} \prod_{j=1}^k (1 - [\Lambda_T(u_j) - \Lambda_T(u_{j-1})]),$$

tomando el límite cuando $k \rightarrow \infty$ y $\max(t_i - t_{i-1}) \rightarrow 0$. Así, cuando T sea una v.a.s. continua se tiene

$$S_T(t) = \prod_0^t [1 - d\Lambda_T(u)] = \prod_0^t [1 - h_T(u)du] = \exp \left\{ - \int_0^t h_T(u)du \right\}.$$

En el caso discreto, (3.16) se reduce ha

$$S_T(t) = \prod_0^t [1 - d\Lambda_T(u)] = \prod_{j|t_j \leq t} [1 - h_T(t_j)].$$

Ahora veamos como se relaciona el estimador Nelson-Aalen para la función de riesgo acumulada $\Lambda_T(t)$ y el estimador Kaplan-Meier para la función de supervivencia $S_T(t)$. Para eso, considere una muestra independiente de tiempos de supervivencia censurados. Denote por $0 = t_0 < t_1 < t_2 < \dots < t_r$ los distintos tiempos en los que se presenta una falla, i.e. las observaciones no-censuradas; sea n_j y d_j definidos como antes. Entonces, el *estimador Nelson-Aalen* para la función de riesgo acumulada correspondiente a $S_T(t)$ es

$$\widehat{\Lambda}(t) = \sum_{j|t_j \leq t} \frac{d_j}{n_j}. \quad (3.17)$$

Este es una función escalonada continua por la derecha cuyos incrementos son los riesgos empíricos estimados, $\widehat{h}(t_j) = d_j/n_j$. Debido a esto, (3.17) corresponde a una distribución con toda la masa de probabilidad concentrada únicamente en los tiempos de falla observados. De la ecuación (3.16), la integral producto de $\widehat{\Lambda}(t)$ es entonces

$$\widehat{S}(t) = \prod_0^t (1 - d\widehat{\Lambda}(t)) = \prod_{j|t_j \leq t} \left(1 - \frac{d_j}{n_j}\right),$$

que no es más que el *estimador Kaplan-Meier*. De esta manera, el estimador Kaplan-Meier y el estimador Nelson-Aalen quedan relacionados del mismo modo como están relacionadas $S_T(t)$ y $\Lambda_T(t)$. Sin embargo, cuando se consideran las relaciones

$$H_T(t) = -\log S_T(t) \quad \text{y} \quad S_T(t) = \exp \{ -H_T(t) \},$$

ese hecho se pierde de vista ya que éstas solamente son válidas en el caso continuo.

Estimador de la Función de Riesgo Acumulada

Si $\widehat{S}(t)$ es el estimador KM de la función de supervivencia, de (3.13) el estimador de la función de riesgo acumulada es $\widehat{H}_T(t) = -\log \widehat{S}_T(t)$. Por otro lado, un segundo método para estimar $H(t)$ es usando el estimador Nelson-Aalen dado en (3.17). Se puede mostrar fácilmente, considerando la aproximación de primer orden de la serie de Taylor de $\log(1 - x)$ alrededor de $a = 1$, que

$$\widehat{H}_T(t) \approx \widehat{\Lambda}_T(t) = \sum_{i:t_i \leq t} \frac{d_i}{n_i}.$$

La varianza del estimador Nelson-Aalen esta dada por

$$\text{Var}[\widehat{\Lambda}(t)] = \sum_{i:t_i \leq t} \frac{d_i}{n_i^2}$$

Ejemplo 3.1. Considere la longitud de los tiempos de remisión de leucemia mielógena crónica del Ejemplo 1.5 correspondientes a los pacientes que no mantuvieron el tratamiento con quimioterapia (Grupo A). El estimador KM de la función de supervivencia para ese grupo de pacientes se observa en la Figura 3.1, y los calculos de la función de supervivencia estimada se muestran en la Tabla 3.1. \square

t_j (tiempo)	n_j (en riesgo)	d_j (fallas)	w_j (censurados)	$\hat{S}(t_j)$ (estimador-KM)	I. de C. al 95%
5	12	2	0	0.8333	(0.6470, 1.000)
8	10	2	0	0.6667	(0.4468, 0.995)
12	8	1	1	0.5833	(0.3616, 0.941)
23	6	1	0	0.4861	(0.2675, 0.883)
27	5	1	0	0.3889	(0.1854, 0.816)
30	4	1	0	0.2917	(0.1148, 0.741)
33	3	1	0	0.1944	(0.0569, 0.664)
43	2	1	0	0.0972	(0.0153, 0.620)
45	1	1	0	0.0000	NA

Tabla 3.1: Error estandar e intervalos de confianza (puntuales) para los valores de $\hat{S}(t)$ correspondientes a los datos de LMC. Note que $n_j = n_{j-1} - (d_j + w_j)$, con $n_0 = n$, $d_0 = w_0 = 0$, $\hat{S}(t_0) = 1$, y $\hat{S}(t_j) = \hat{S}(t_{j-1}) * \hat{p}_j$, donde $\hat{p}_j = 1 - (d_j/n_j)$.

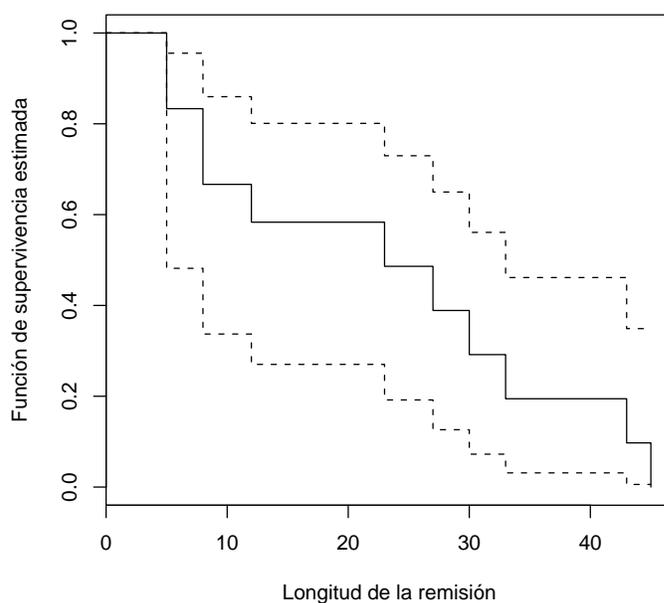


Figura 3.1: Gráfica del estimador Kaplan-Meier de la función de supervivencia para el grupo de pacientes con leucemia que no mantuvo el tratamiento con quimioterapia y su correspondiente banda de confianza.

Código en R, Ejemplo 3.1.

```

# Los datos se guardado en el objeto "lmc" con formato "data.frame".
> lmc<-data.frame( tiempo=c(9,13,13,18,23,28,31,34,45,48,161,5,5,8,8,
      12,16,23, 27,30,33,43,45), censura=c(1,1,0,1,1,0,1,1,0,1,1,0,1,0,1,
      1,1,1,1,0,1,1,1,1,1,1), X=c( 1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,
      2,2,2,2,2,2 ))
# Se crea el objeto "survfit"
> lmc.survfit<-survfit(Surv(tiempo,censura),data=lmc,
  error="greenwood", conf.type="log-log")
> summary(lmc.survfit)
# Grafica el estimador-KM con sus bandas de confianza.
> plot(lmc.survfit,conf.int=TRUE,mark.time=FALSE)
> title(xlab='Longitud de la remisi'on', ylab='Funcion de
  supervivencia estimada')

```

Ejemplo 3.1 (Continuación). En la Figura 3.2 se comparan los estimadores de la función de riesgo para el Grupo A, obtenidos mediante el estimador Kaplan-Meier, $\hat{H}_T(t) = -\log \hat{S}_T(t)$, y el estimador Nelson-Aalen, respectivamente. \square

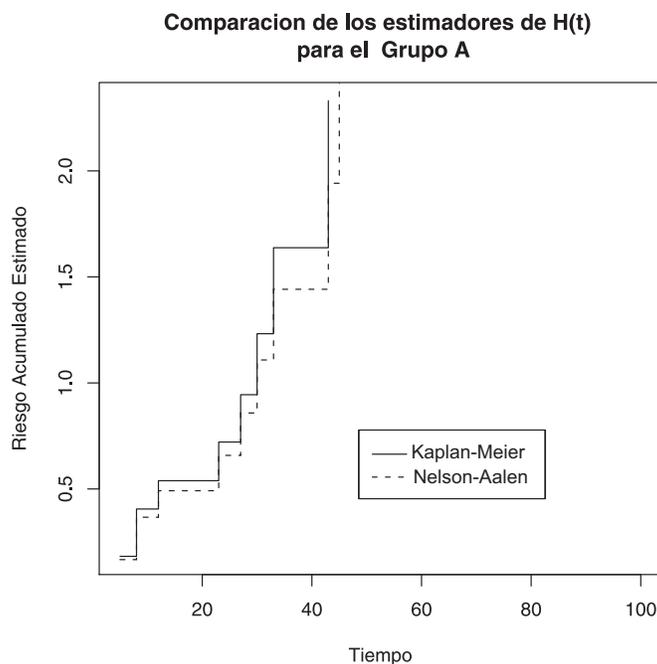


Figura 3.2: Estimadores de la Función de Riesgo Acumulada. El comportamiento de los estimadores de $H_T(t)$ se debe a que la última observación no está censurada.

3.3. Comparación de distribuciones de supervivencia

Hay situaciones que dan lugar a la necesidad de comparar la supervivencia de dos o más grupos de individuos, e.g. a través de sus correspondientes curvas de supervivencia. Una de la formas más fácil de llevar a cabo eso es mediante una visión gráfica.

Ejemplo 3.2. Considere nuevamente el conjunto de datos de LMC descritos en el Ejemplo 1.5. El estimador-KM para cada grupo de pacientes se muestra en la Figura 3.3. Al hacer una "comparación visual" de las curvas de supervivencia para los dos grupos se observa que los pacientes que mantuvieron el tratamiento con quimioterapia (B) tuvieron un mejor tiempo de remisión⁶ hasta experimentar una recaída (falla), en comparación con el grupo de referencia (A). Eso nos sugiere que si el paciente mantiene la quimioterapia tardara más tiempo en experimentar una recaída. Nótese que el estimador-KM de $S(t)$ correspondiente al Grupo B no está definido después de la observación más grande ($t = 161$), debido a que es un tiempo de supervivencia censurado. □

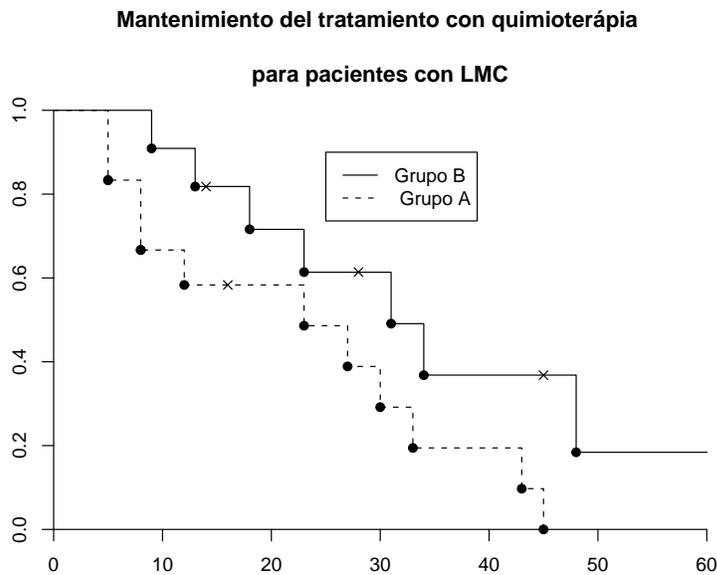


Figura 3.3: Gráficas de los estimadores-KM para el Grupo de control (A) y el Grupo que mantuvo el tratamiento con quimioterapia (B).

⁶ El hecho de que la curva de supervivencia del grupo B sea mayor que la del grupo A en algún punto en el tiempo, digamos t_i , quiere decir que en ese tiempo la proporción de individuos en remisión (i.e. libres de recaída) del grupo B es mayor que la del grupo A.

Prueba de hipótesis para dos o más muestras

Como recordara, una *prueba de hipótesis* es un procedimiento que nos permite evaluar hasta que punto un conjunto de datos observados es consistente con una hipótesis en particular, conocida como *hipótesis nula*. La Figura 3.3 del Ejemplo 3.2 nos sugiere, mediante una visión gráfica, que hay una diferencia entre las dos curvas de supervivencia estimadas, pero *¿Que tan significativa es esa diferencia observada?* Si consultamos el ejemplo de la comparación de las funciones de supervivencia para dos grupos en [15], se daría cuenta de que este no es el mejor camino para resolver nuestro dilema. Existe una gran variedad de pruebas estadísticas de significancia (pruebas de hipótesis) que nos ayudarían a responder a esa pregunta. Sin embargo, no todas las pruebas de hipótesis conocidas se aplican en este caso. Esto se debe, de nueva cuenta, a la censura en los datos de supervivencia o a que los datos no presentan una distribución simétrica. Por ejemplo, si no hubiese censura en los datos, la prueba no-paramétrica de suma de rangos de Wilcoxon podría ser apropiada para comparar dos muestras independientes de datos de supervivencia; la prueba t – *Student* tampoco es apropiada debido a la distribución en este tipo de datos.

Ya que en la mayoría de los casos se presentan muestras con observaciones censuradas, se debe de utilizar otro tipo de pruebas de hipótesis tales como:

- *Prueba log-rank (o prueba de Mantel-Haenszel)*. Por lo regular, la prueba log-rank es la más usada para la comparación de dos (o más) distribuciones de supervivencia. Esta asume que las funciones de riesgo correspondientes a los grupos son paralelas. Sin embargo, si las curvas de supervivencia se cortan, la prueba log-rank presenta problemas. Esta da el mismo peso tanto a los primeros como a los últimos tiempos de falla. A continuación se dará un bosquejo de la construcción de esta prueba para dos Grupos, pero se puede consultar [1] para una descripción más detallada.

Suponga que existen r distintos tiempos de falla, $t_{(1)} < t_{(2)} < \dots < t_{(r)}$, por los dos grupos. Se considera cada tiempo de falla $t_{(i)}$ ($i = 1, 2, \dots, r$) por separado y se crea una tabla de contingencia de 2×2 en la que se muestra d_{ji} , el número de fallas en el j -ésimo grupo, y n_{ji} , el número de sujetos en riesgo en el grupo j antes del tiempo $t_{(i)}$, para $j = 1, 2$. Consecuentemente, al tiempo $t_{(i)}$ hay $d_i = d_{1i} + d_{2i}$ fallas en total de $n_i = n_{1i} + n_{2i}$ individuos en riesgo de falla. Ahora consideremos la hipótesis nula de que no hay diferencia entre los tiempos de falla de los dos grupos. Si los totales marginales en la tabla de contingencia se toman como fijos, bajo la hipótesis nula, sus cuatro entradas están determinadas por el valor de d_{1i} , el número de fallas del Grupo 1 en $t_{(i)}$. De esta manera, d_{1i} se puede considerar como una v.a. cuya distribución es *HIPERGEOMETRICA*($d_{1i}; n_{1j}, n_j, d_j$). Por eso, el número esperado de fallas en el Grupo 1 es

$$e_{1i} = n_{1i} \frac{d_j}{n_j},$$

y su varianza esta dada por

$$v_{1i} = \frac{n_{1i}n_{2i}d_i(n_i - d_i)}{n_i^2(n_i - 1)}.$$

Finalmente, el estadístico de contraste se define de la siguiente manera:

$$W = \frac{\left[\sum_{i=1}^r (d_{1i} - e_{1i}) \right]^2}{\sum_{i=1}^r v_{1i}}. \quad (3.18)$$

Cuando el número de fallas es relativamente grande, W se distribuye aproximadamente χ^2 con un grado de libertad. Este estadístico resume las desviaciones que hay entre los tiempos de falla observados en los dos grupos y los tiempos de aquellos esperados bajo la hipótesis nula de no diferencia en los dos grupos. Entre más grande sean los valores del estadístico W mayor será la evidencia en contra de la hipótesis nula.

- *Prueba de Wilcoxon (modificación de Peto y Peto de la prueba Wilcoxon-Gehan).* Esta es más potente que la prueba log-rank cuando las fuerzas de mortalidad no son paralelas (i.e. cuando se cortan en algún punto) y se tengan pocas observaciones censuradas. Otra de las características de esta prueba es que da más peso a los primeros tiempos de falla (ver [1] para una descripción detallada).

Ejemplo 3.3. Retomemos el conjunto de tiempos de remisión de los pacientes con LMC del ejemplo anterior. Se desea probar la hipótesis nula de que no hay diferencia en los tiempos de falla de los dos grupos mediante la prueba log-rank, con un nivel de significancia de $\alpha = 0.05$. Para llevar a cabo esta prueba en \mathbf{R} nos auxiliaremos de la variable explicativa X , la cual toma el valor de 0 si el paciente pertenece al Grupo A, y de uno si es del Grupo B. Nuestro conjunto de datos consta de 23 observaciones de la forma

$$(t_1, \delta_1, x_1), (t_2, \delta_2, x_2), \dots, (t_{23}, \delta_{23}, x_{23}),$$

Esta prueba se realiza mediante la función `survdif(formula, rho)`. El valor de `rho=0` es el que inicialmente tome \mathbf{R} y corresponde a la prueba log-rank; un valor de `rho=1` es la prueba de Wilcoxon. En general, el peso de la curva de supervivencia, al inicio o al final de esta, se asigna dependiendo del valor que se le de a `rho`. Bajo la hipótesis nula de que no hay diferencia entre los tiempos de supervivencia de los dos grupos, nuestro estadístico de prueba se distribuye asintóticamente χ^2 con un grado de libertad, eso es, $W \stackrel{a}{\sim} \chi_{(1)}^2$.

El valor observado del estadístico W obtenido de \mathbf{R} es $\chi_{obs} = 3.396$ y el valor- P correspondiente es 0.0654. Puesto que valor- p es mayor que $\alpha = 0.05$, no se puede rechazar la hipótesis nula; esto nos sugiere que no hay evidencia real de que el mantener el tratamiento con quimioterapia prolonge el tiempo de remisión de pacientes con LMC. \square

Código en R, Ejemplo 3.3.

```
>library(survival)
> lmc.survfit<- survfit(Surv(tiempo,censura$)~X, data=lmc)
> plot(lmc.survfit,xlim=c(0,60),axes=FALSE,lty=1:2, mark.time=FALSE)
> title(xlab='Longitud de la remisión', ylab='Función de
  supervivencia estimada', font.lab=4)
> legend(25,0.9,legend=c( "Grupo B"," Grupo A"),lty=1:2)
> axis(1,pos=c(0,0)); axis(2,pos=c(0,0))
> title("Mantenimiento del tratamiento para LMC.")
> lmc.logrank<-survdiff(Surv(tiempo,censura)~X,data=lmc,rho=0)
> lmc.logrank$chisq
  3.396389
```

En contraste con el ejemplo anterior, hay casos en donde no es tan claro tomar una decisión respecto al tiempo de supervivencia de dos grupos de pacientes debido a ciertos factores, como el hecho de que las curvas de supervivencia de los dos grupos se crucen.

Ejemplo 3.4. *Los datos del Ejemplo 1.4 corresponden a la longitud de remisión de 101 pacientes a los cuales se les realizó un trasplante de médula osea. Cincuenta y uno de esos pacientes recibieron un trasplante autólogo, mientras que los restantes cincuenta recibieron un trasplante alogénico. La gráfica del estimador Kaplan-Meier de la función de supervivencia para cada tratamiento se muestra en la Figura 3.4.*

*Uno de los objetivos en este tipo de estudios es el de evaluar la eficiencia de los dos tratamientos mediante la longitud de remisión de los pacientes (tiempo de supervivencia). Nótese que las curvas de supervivencia se cruzan aproximadamente en $t = 12$ meses, además de que el grupo de trasplante alogénico tiene un mayor número progresivo de censuras que el otro grupo, ambos en los últimos tiempos de supervivencia. Al realizar la prueba de Wilcoxon en **R** se obtiene un valor $-p = 0.977$. Esto nos sugiere que no hay una diferencia significativa entre los dos trasplantes de médula osea.*

Código en R, Ejemplo 3.4.

```
>library(KMsurv)
> data(alloauto)
> attach(alloauto)
> survdiff(Surv(time,delta)~type,rho=1)
```

□

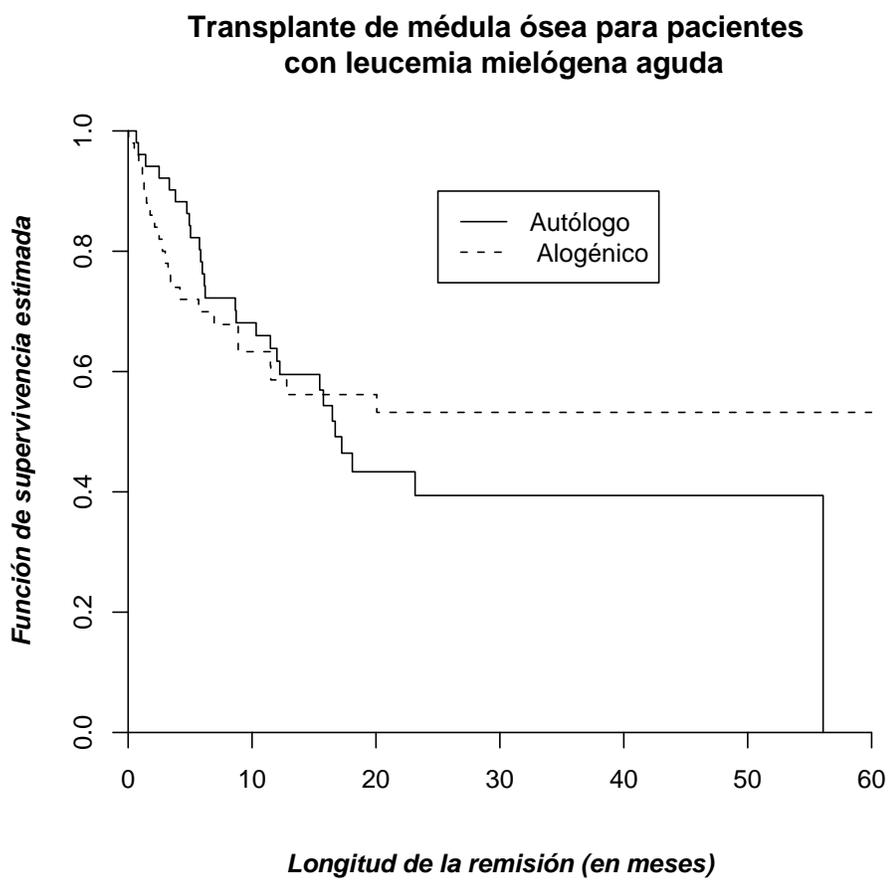


Figura 3.4: Gráfica del estimador Kaplan-Meier para cada tipo de transplante realizado a los pacientes con LMA.

MODELO DE REGRESIÓN PARA DATOS DE SUPERVIVENCIA

Hasta el momento se han considerado modelos y métodos estadísticos para proceder con el Análisis de Supervivencia para datos provenientes de poblaciones homogéneas. Sin embargo, en muchos estudios médicos que dan lugar a datos de este tipo, se registra información adicional por cada individuo y de la cual se cree que puede depender el tiempo de supervivencia. Por eso, surge la necesidad de considerar generalizaciones tanto de esos modelos como de los métodos empleados para tomar en cuenta ese tipo de *información concomitante*.

Los modelos de regresión, los cuales son aún más generales que los vistos anteriormente, pueden usarse tanto con el objetivo de *predecir* el valor esperado de la respuesta Y usando un conjunto de variables explicativas fijas $\{X_1, X_2, \dots, X_p\}$, así como el de *estimar* el efecto de una o más variables explicativas X_j sobre la variable respuesta Y . El modelo de regresión descrito en el presente Capítulo, el cual aparece en la literatura estadística bajo el nombre de *Modelo de Regresión de Cox* o *Modelo de Riesgos Proporcionales de Cox*, permite estimar la relación que hay entre un conjunto de variables explicativas fijas X_1, X_2, \dots, X_p y la respuesta o tiempo de falla Y , o más bien, con la función de riesgo $h(y; \mathbf{x})$, que es la tasa instantánea del suceso de interés.

El objetivo principal del presente Capítulo es la exploración de la dependencia de Y respecto de \mathbf{X} mediante un modelo de regresión adecuado. Esencialmente, son dos las clases generales de modelos que comúnmente se consideran:

- *Los Modelos de Riesgos Proporcionales.*

- *Los Modelos de Tiempos de Vida Acelerados.*

Para fines del presente trabajo nos enfocaremos en los Modelos de Riesgos Proporcionales.

4.1. Modelo de Riesgos Proporcionales

Sea Y una v.a.s. que depende, en la respuesta, del vector observado de valores de la covariable $\mathbf{x}^\top = (x_1, x_2, \dots, x_p)$. Por un **modelo de riesgos proporcionales** para Y nos referimos al modelo

$$h(y; \mathbf{x}) = h_0(y) \psi(\mathbf{x}) \quad (4.1)$$

donde $h(y; \mathbf{x})$ es la función de riesgo condicional de Y dado \mathbf{x} , $\psi(\mathbf{x})$ es una función positiva en \mathbf{x} tal que $\psi(\mathbf{0}) = 1$, y $h_0(y)$ es la **función de riesgo base**, la cual representa la función de riesgo para una observación cuyo vector de covariables es cero, es decir, $h(y; \mathbf{x} = \mathbf{0}) = h_0(y)$. La selección de $\psi(\cdot)$ dependerá de los datos a considerar (ver [16]), la forma de $\psi(\mathbf{x})$ en (4.1) más conveniente y flexible para muchos propósitos es (Cox, 1972a):

$$\psi(\mathbf{x}) = \psi(\mathbf{x}; \boldsymbol{\beta}) = \exp(\boldsymbol{\beta}^\top \mathbf{x}),$$

donde $\boldsymbol{\beta}^\top = (\beta_1, \beta_2, \dots, \beta_p)$ es un vector de parámetros desconocidos. De esta manera, el modelo de riesgos proporcionales en (4.1) toma la forma

$$h(y; \mathbf{x}) = h_0(y) \exp(\boldsymbol{\beta}^\top \mathbf{x}). \quad (4.2)$$

Esta clase de modelos se caracteriza por el hecho de que las variables explicativas tienen un *efecto multiplicativo* a través del producto con $h_0(y)$. Más aún, si \mathbf{x} y \mathbf{x}^* representan los vectores de covariables observados de dos individuos diferentes, y denotamos como HR (por sus siglas en inglés *hazard ratio*) a la *razón* de sus funciones de riesgo, la cual está dada por

$$HR \equiv \frac{h(y; \mathbf{x})}{h(y; \mathbf{x}^*)} = \frac{\psi(\mathbf{x})h_0(y)}{\psi(\mathbf{x}^*)h_0(y)} = \exp \left\{ \boldsymbol{\beta}^\top (\mathbf{x} - \mathbf{x}^*) \right\}, \quad (4.3)$$

se tiene que no depende del tiempo de falla y , y sólo depende de la diferencia en el valor de la variable explicativa. De aquí el nombre de *riesgos proporcionales*.

Cuando se ajusta un conjunto de datos bajo la suposición de riesgos proporcionales usando el Modelo de Regresión (4.2) y se supone una distribución "base" para $h_0(y)$, digamos $F_0(y; \boldsymbol{\theta})$, el proceso involucra la estimación de los vectores de parámetros desconocidos $\boldsymbol{\theta}$ y $\boldsymbol{\beta}$, empleando las respuestas observadas Y junto con el vector de valores observados \mathbf{x} correspondiente a \mathbf{X} .

Ejemplo 4.1 (Modelo para la comparación de dos grupos). *Considere a un grupo de n pacientes con cierta enfermedad, los cuales son aleatorizados en dos grupos disjuntos para recibir un nuevo tratamiento (factor de exposición) o seguir con el tratamiento estándar. Denotemos por Grupo I a todos los pacientes que reciben el tratamiento estándar, y por Grupo II a los que reciben un nuevo tratamiento. Supongamos que las fuerzas de mortalidad de los dos grupos son proporcionales en cualquier instante de tiempo y , y tomemos al Grupo I como referencia del riesgo inicial, es decir, aquellos pacientes que no están expuestos a la droga del nuevo tratamiento.*

Del modelo de riesgos proporcionales en (4.2), la función de riesgo al tiempo y para cualquier paciente i ($i = 1, 2, \dots, n$) en la muestra formada por los dos grupos es

$$h_i(y) = \exp(\beta x_i) h_0(y), \quad \forall y > 0 \text{ y } \forall \beta \in (-\infty, \infty), \quad (4.4)$$

donde x_i es el valor que toma la variable indicadora X_i , la cual está definida como

$$X = \begin{cases} 1, & \text{paciente del Grupo II (expuestos al factor "nuevo tratamiento");} \\ 0, & \text{paciente del Grupo I.} \end{cases}$$

De esta manera, la función de riesgo al tiempo y para un individuo en el Grupo I es $h_0(y)$, y la función de riesgo al tiempo y para un individuo del Grupo II es $h_0(y) \exp(\beta)$. De (4.3) con $p = 1$, la HR en cualquier tiempo y para un individuo que recibe el nuevo tratamiento relativo a un individuo con el tratamiento estándar es

$$HR = \frac{h(y; \mathbf{x} = 1)}{h_0(y)} = \exp(\beta)$$

En esta situación, la interpretación de la HR es como sigue: si $\exp(\beta) < 1$, el riesgo de presentar el evento de interés al tiempo y es menor para un paciente con la nueva droga, relativo a un paciente con el tratamiento estándar, lo cual nos sugiere que el nuevo tratamiento es mejor que el estándar (factor protector). Por otra parte, si $\exp(\beta) > 1$, el riesgo de presentar el evento de interés al tiempo y es más grande para un paciente en el Grupo II relativo a otro paciente en el Grupo I, que es el grupo de referencia (i.e. que no está expuesto al factor). El tratamiento estándar se muestra superior ante la nueva droga (factor de riesgo). \square

Si suponemos que el tiempo de supervivencia del Grupo I en el ejemplo anterior tiene un distribución *Weibull*(λ, α), la función de riesgo para los individuos en ese grupo es

$$h(y; x = 0) = h_0(y) = \lambda \alpha y^{\alpha-1},$$

En consecuencia, la función de riesgo para los individuos del Grupo II es

$$h(y; x = 1) = \exp(\beta) h_0(y) = \exp(\beta) \lambda \alpha y^{\alpha-1},$$

la cual corresponde a la función de riesgo de una distribución *Weibull* con parámetro de escala $\lambda \exp(\beta)$ y parámetro de forma α . Nótese que tal modelo es completamente paramétrico. En este caso, se dice que la distribución *Weibull* tiene la *propiedad de riesgos proporcionales* [1].

Collet(1994) discute la generalización del modelo de riesgos proporcionales dado en (4.4) a una situación en la que el riesgo de muerte al tiempo y depende de los valores x_1, x_2, \dots, x_p de p variables explicativas X_1, X_2, \dots, X_p .

4.1.1. Función de supervivencia bajo riesgos proporcionales

Sea Y una v.a.s. continua con vector observado de variables explicativas $\mathbf{x} \in \mathbb{R}^p$. Ya se mencionó que bajo un modelo de riesgos proporcionales las covariables tienen

un efecto multiplicativo en la función de riesgo condicional de Y . Entonces, es natural preguntarse sobre el efecto de las covariables en la distribución del tiempo de supervivencia de Y . Si definimos $S(y; \mathbf{x})$ como la función de supervivencia condicional de Y , cuya función de riesgo condicional es $h(y; \mathbf{x})$, se tendrá que el vector de covariables tiene un efecto potencia en la función de supervivencia. En efecto, del Teorema 2.1.1,

$$\begin{aligned} S(y; \mathbf{x}) &= \exp \left\{ - \int_0^y h(u; \mathbf{x}) du \right\} = \exp \left\{ - \int_0^y h_0(u) \psi(\mathbf{x}) du \right\} \\ &= \exp \left\{ - \psi(\mathbf{x}) \int_0^y h_0(u) du \right\} = \left[\exp \left\{ - \int_0^y h_0(u) du \right\} \right]^{\psi(\mathbf{x})} \\ &= [S_0(y)]^{\psi(\mathbf{x})} \end{aligned} \quad (4.5)$$

donde

$$S_0(y) \equiv S(y; \mathbf{x} = 0) = \exp \left\{ - \int_0^y h_0(u) du \right\},$$

es la **función de supervivencia base**.

De manera análoga, definimos la **función de riesgo acumulada base**

$$H_0(y) = \int_0^y h_0(u) du,$$

cuando $\mathbf{x} = \mathbf{0}$.

Modelo de tiempos de vida acelerado

Sea Y una v.a.s. que depende, en la respuesta, del vector observado de covariables $\mathbf{X}^T = (X_1, X_2, \dots, X_p)$. En el modelo de tiempos de vida acelerado, en contraste con el de riesgos proporcionales, se supone que las covariables actúan directamente sobre el tiempo de supervivencia, de tal manera que el efecto de estas se ve reflejado en un aumento o disminución de su progreso.

Considere nuevamente el Ejemplo 4.1 en el que se tiene una muestra de n individuos con una sola covariable X definida como antes. Collet (1994) también describe la generalización de este ejemplo bajo un modelo de tiempos de vida acelerado en el que la función de riesgo condicional para el i -ésimo individuo dado el vector de valores observado de p covariables, $\mathbf{x}_i \in \mathbb{R}^p$, es

$$h_i(y; \mathbf{x}_i) = \exp\{\boldsymbol{\beta}^T \mathbf{x}_i\} h_0(y \exp\{\boldsymbol{\beta}^T \mathbf{x}_i\}) \quad (4.6)$$

donde

$$\boldsymbol{\beta}^T \mathbf{x}_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}$$

es el *componente lineal* del modelo, en el cual x_{ij} es el valor de la j -ésima covariable correspondiente al i -ésimo individuo ($i = 1, 2, \dots, n$, $j = 1, 2, \dots, p$), $\boldsymbol{\beta} \in \mathbb{R}^p$ es un vector de parámetros desconocidos, y la función de riesgo base, $h_0(y)$, que es

el riesgo de muerte al tiempo y para un individuo con vector de covariables igual a cero. Análogamente, la función de supervivencia condicional del i -ésimo individuo con función de riesgo como en (4.6) es

$$S_i(y; \mathbf{x}_i) = S_0(y \exp\{\boldsymbol{\beta}^\top \mathbf{x}_i\}), \quad (4.7)$$

donde $S_0(y)$ es la *función de supervivencia base* y representa la función de supervivencia de un individuo con vector de covariables igual a cero ($\mathbf{x} = \mathbf{0}$).

Ejemplo 4.2. *Considere nuevamente los dos grupos de pacientes del Ejemplo 4.1, pero ahora bajo la suposición de un modelo de tiempos de vida acelerado. También suponga que el tiempo de supervivencia del Grupo I se distribuye Weibull(λ, α). Entonces, la función de riesgo base ($x = 0$) para un individuo de ese grupo es*

$$h_0(y) = \lambda \alpha y^{\alpha-1}.$$

De la ecuación (4.6), la fuerza de mortalidad del i -ésimo individuo en la muestra formada por los dos grupos es

$$\begin{aligned} h_i(y) &= \exp(\boldsymbol{\beta}^\top \mathbf{x}_i) \lambda \alpha [y \exp(\boldsymbol{\beta}^\top \mathbf{x}_i)]^{\alpha-1} \\ &= [\lambda \exp(\alpha \boldsymbol{\beta}^\top \mathbf{x}_i)] \alpha y^{\alpha-1}, \end{aligned} \quad (4.8)$$

que corresponde a la función de riesgo de una distribución Weibull con parámetro de forma α y parámetro de escala $\lambda \exp(\alpha \boldsymbol{\beta}^\top \mathbf{x}_i)$. Por lo tanto, se dice que la distribución Weibull tiene la propiedad de tiempos de vida acelerados. \square

Observación 4.1. La distribución Weibull es la única distribución que posee tanto la propiedad de riesgos proporcionales como la propiedad de tiempos de vida acelerado. Debido a esto, un enfoque alternativo al modelo de riesgos proporcionales Weibull es un modelo de tiempos de vida acelerado (ver [1],[15]).

4.2. El Modelo de Cox

Cuando se supone una forma paramétrica de la función de riesgo base $h_0(y)$ en (4.1), el problema consiste en la estimación de los parámetros involucrados a partir de la información contenida en los datos. Aunque existen métodos para evaluar la Bondad del Ajuste del modelo supuesto para cierto conjunto de datos, siempre habrá incertidumbre en su adecuación. Una manera de relajar esa situación es establecer un modelo *semi-paramétrico* en el que no se supone una forma específica para $h_0(y)$, como lo es el Modelo de Cox. Este tipo de modelos también son llamados *libres de distribución* en el sentido de que su validez y ciertas propiedades no dependen de la forma de $h_0(y)$.

Bajo la notación de la Sección 1.2, supongamos que (t_i, δ_i) , $i = 1, 2, \dots, n$ denotan una muestra aleatoria independiente de tiempos de supervivencia continuos, presentándose en orden ascendente y posiblemente censurados por la derecha bajo un mecanismo de censura no-informativo. Considere el vector $\mathbf{x}^\top = (x_1, \dots, x_p)$ de

variables explicativas "fijas", es decir, covariables independientes del tiempo¹, para cada individuo en la muestra. El tiempo de supervivencia de cada individuo tiene una distribución, dependiente en \mathbf{x} , que puede ser caracterizada ya sea o por la función de supervivencia $S(t; \mathbf{x})$ o por la función de riesgo $h(t; \mathbf{x})$. El modelo de Cox (1972a) especifica la función de riesgo para el i -ésimo individuo al tiempo t como

$$h_i(t) \equiv h(t; \mathbf{x}_i) = h_0(t) \exp\{\boldsymbol{\beta}^T \mathbf{x}_i\}. \quad (4.9)$$

donde $\boldsymbol{\beta} \in \mathbb{R}^p$ es un vector de coeficientes de regresión desconocido, $h_0(t)$ es alguna **función de riesgo base** no especificada, que representa el riesgo de falla al tiempo t en un conjunto de condiciones estándar, es decir, cuando $\mathbf{x}=\mathbf{0}$, y $\psi(\mathbf{x}_i; \boldsymbol{\beta}) = \exp\{\boldsymbol{\beta}^T \mathbf{x}_i\}$ es una función (positiva) específica del vector de variables explicativas \mathbf{x}_i . De esta manera, el modelo de riesgos proporcionales en (4.9) es **semi-paramétrico** en que $\exp(\boldsymbol{\beta}^T \mathbf{x}_i)$ es paramétricamente especificada, pero $h_0(t)$ no. Esto a su vez significa que el Modelo de Cox no supone una forma paramétrica específica de la distribución del tiempo de supervivencia, haciendo este método considerablemente **robusto**, ya que buenos estimadores de los coeficientes de regresión, y en consecuencia de la función de riesgo y de la función de supervivencia ajustada, pueden ser obtenidos para una gran variedad de conjuntos de datos.

Como ocurre en los modelos lineales y logísticos, surge la necesidad de contar con una medida del efecto que describa la relación entre el tiempo de supervivencia (respuesta) y algún factor de exposición. Tal efecto estará ajustado por ciertas variables explicativas relevantes incluidas en el modelo de regresión. En los modelos logísticos, la medida de ese efecto es un *cociente de momios*, mientras que en análisis de supervivencia la medida de ese efecto es llamado **razón de riesgos (instantáneos)**, denotado como HR. Pues resulta que la simple interpretación de la razón de riesgo como un tipo de *riesgo relativo* es otra de las atractivas características del Modelo de Cox (ver Ejemplo 4.1).

4.2.1. Verificación de la suposición de riesgos proporcionales

La forma en (4.9) es una suposición fuerte y requiere de una cuidadosa verificación al momento de su aplicación, por lo que métodos para la evaluación de su validez son necesarios. En el libro de Kleinbaum [21] se explican tres enfoques que se pueden usar para verificar la suposición de riesgos proporcionales: *métodos gráficos*, *pruebas de bondad de ajuste*, y *variables dependientes del tiempo*. Cuando este último método es usado para evaluar la suposición de riesgos proporcionales de la ℓ -ésima covariable

¹ El modelo de Cox tiene la ventaja de que puede ser extendido para permitir modelar covariables que dependen del tiempo, eso es, el riesgo para el i -ésimo individuo está dado por

$$h(t_i; \mathbf{X}_i(t)) = h_0(t_i) \exp(\boldsymbol{\beta}^T \mathbf{X}_i(t))$$

donde $\mathbf{X}_i(t)$ es el vector de variables explicativas para el i -ésimo individuo en el tiempo. Sin embargo, el **Modelo de Cox Extendido** no está en el alcance del presente trabajo, pero los procedimientos descritos en este Capítulo igualmente pueden ser extendidos para tratar con variables explicativas dependientes del tiempo. El libro de Kleinbaum (1996) es una buena opción para familiarizarse con este tema.

fija, el modelo de Cox es extendido para incluir una interacción de la forma $X_{i\ell} \times v(t)$, donde $v(t)$ es alguna función del tiempo (e.g. $v_\ell(t) = t$). De esta manera, el modelo de Cox Extendido para verificar la suposición de riesgos proporcionales (HR) de la ℓ -ésima covariable es

$$h(t, \mathbf{x}_i) = h_0(t) \exp \left\{ \sum_{k=1}^p \beta_k x_{ik} + \eta [x_{i\ell} \times v(t)] \right\} \quad (4.10)$$

donde η es el coeficiente de interacción.

Usando el modelo (4.10), verificamos la suposición HR probando la significancia de η . Por lo tanto, la hipótesis nula es $H_0 : \eta = 0$. Note que si la hipótesis nula es verdadera, (4.10) se reduce a (4.9). También se puede realizar una prueba de hipótesis desde un punto de vista GLOBAL, es decir, evaluar la suposición HR para todas las covariables incluidas en el modelo, de manera simultanea. Por eso, el Modelo de Cox extendido toma la forma

$$h(t, \mathbf{x}_i) = h_0(t) \exp [\boldsymbol{\beta}^\top \mathbf{x}_i^\top + \boldsymbol{\eta}^\top \mathbf{z}_i(t)]$$

donde $\mathbf{z}_i(t) = (x_{i1}v_1(t), \dots, x_{ip}v_p(t))^\top$, y $\boldsymbol{\eta} = (\eta_1, \dots, \eta_p)^\top$. En este caso, la hipótesis nula es $H_0 : \boldsymbol{\eta} = \mathbf{0}$. Estas pruebas son llevadas a cabo mediante el estadístico de la razón de verosimilitud.

Modelo de Cox Estratificado

Este modelo es un modificación de (4.9), el cual permite modelar la función de riesgo para distintos grupos disjuntos o estratos. El modelo de Cox para el k -ésimo estrato se define como

$$h(t, \mathbf{X}) = h_{0k}(t) \exp (\boldsymbol{\beta}^\top \mathbf{Z}(t)), \quad (4.11)$$

En el caso de que alguna de las covariables incluidas en el Modelo de Cox no cumple con la suposición de riesgos proporcionales, el modelo (4.11) puede ser empleado para controlar el efecto de tal covariables mediante la estratificación de esta (ver [21]).

4.2.2. Ajuste e interpretación del modelo de Cox

Cuando se usa el modelo de Cox con el objetivo de estimar la relación que hay entre el tiempo de supervivencia y una o más variables explicativas, surgen dos conceptos importantes: **confusión** e **interacción**. Para entender el concepto de confusión, supongamos que un modelo de riesgos proporcionales es considerado para un conjunto de datos de supervivencia, digamos

$$\text{Modelo(1): } h(t) = h_0(t) \exp\{\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3\}$$

donde X_1 , X_2 y X_3 son las $p = 3$ variables explicativas de posible relevancia, y β_j es el *efecto principal* correspondiente a X_j ($j = 1, 2, 3$). El efecto de cualquier término en el Modelo(1) no se puede estudiar independientemente de los otros, por lo que

existe una dependencia entre ellos, por ejemplo, el efecto β_1 de X_1 en el Modelo(1) depende de X_2 y X_3 . En este caso, se dice que β_1 es **ajustado** por X_2 y X_3 . Así, $\exp(\beta_1)$ representa la *HR* debida al factor X_1 , ajustado por X_2 y X_3 .

Ahora considere un segundo modelo en el que se han ajustado las $p = 3$ variables del Modelo(1) más una variable adicional X_4 , de la cual se sospecha que es una buena indicadora del pronóstico de supervivencia, eso es,

$$\text{Modelo(2)} : h(t) = h_0(t) \exp\{\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4\}.$$

Se dice que X_4 es una *factor de confusión* para cualquiera de los efectos β_j en el Modelo(1), digamos β_1 , si el Modelo(2) produce una estimación para β_1 diferente a la obtenida en el Modelo(1). Similarmente, el efecto de X_4 en el Modelo(2), que es β_4 , se dice que es ajustado por las $p = 3$ variables explicativas que ya han sido ajustadas, X_1, X_2, X_3 .

Por otra parte, cuando la relación entre alguno de los factores de estudio en el Modelo(2), digamos X_1 , y la variable del tiempo de supervivencia (respuesta) es modificado según el valor de una tercera, digamos X_3 , se dice que hay una *interacción* entre X_1 y X_3 . Entonces, si queremos evaluar el posible efecto de interacción de X_3 sobre X_2 , es necesario incluir una quinta variable, eso es, el producto de X_1 y X_3 .

Ya que el modelo en (4.9) se puede re-expresar en la forma

$$\log \left\{ \frac{h_i(t)}{h_0} \right\} = \boldsymbol{\beta}^T \mathbf{x} = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}, \quad (4.12)$$

el modelo de riesgos proporcionales también puede ser visto como un modelo lineal para el logaritmo natural del cociente de los riesgos. La manera de incluir información concomitante en el modelo de riesgos proporcionales es a través del *componente lineal* $\boldsymbol{\beta}^T \mathbf{x}$. Collett (1994) presenta una extensa discusión del ajuste del modelo de Cox mediante $\boldsymbol{\beta}^T \mathbf{x}$, además de la interpretación del vector de parámetros estimado $\boldsymbol{\beta}$ con distintos modelos ajustados. Esta parte también puede ser discutida desde el punto de vista de los Modelos Lineales Generalizados (MLG); una ventaja de este enfoque es que surge el concepto de **matriz diseño** (ver [7]). Por eso, las técnicas estándar para incluir covariables en modelos de regresión (e.g. logísticos, lineales) pueden ser usadas en este contexto. En cuanto a su escala de medición, ya se mencionó de manera breve en la Subsección 1.3.1.

Ejemplo 4.3. *Considere a un grupo de pacientes con cáncer sometidos a un tratamiento médico. Supongamos que se desea modelar la dependencia de la función de riesgo sobre el factor A ="grado del tumor", al cual se le ha clasificado en tres niveles descriptivos. Así, los individuos son también clasificados en los grupos I, II y III, respectivamente. Si β_j representa el efecto de A en el nivel j ($j = 1, 2, 3$), podemos definir variables indicadoras X_1, X_2, X_3 para incluir esos términos en la parte lineal del modelo de riesgos proporcionales, con la restricción de que alguno de los β_j sea cero para ser consistentes con la definición de $h_0(t)$, por ejemplo, tomando $\beta_1 = 0$*

podemos adoptar como **referencia** del riesgo de falla (i.e todas las variables explicativas iguales a cero) a los pacientes cuyo factor A está en el primer nivel, es decir, el Grupo I. Así, los valores que toman las variables indicadoras, junto con la función de riesgo de los pacientes cuyo factor A está en el nivel j quedan como se muestra en la Tabla 4.1. \square

Nivel del Factor A	\mathbf{X}_2	\mathbf{X}_3	función de riesgo
1	0	0	$h(t) = h_0(t)$
2	1	0	$h(t) = h_0(t) \exp(\beta_2)$
3	0	1	$h(t) = h_0(t) \exp(\beta_3)$

Tabla 4.1: Valores de \mathbf{X}_j para indicar el nivel en el que se encuentra el factor A.

Una expresión más general para la función de riesgo condicional usada para identificar los tres niveles del factor cáncer de un individuo con vector de covariables $\mathbf{x} = (1, x_2, x_3)^\top$ es

$$h(t; \mathbf{x}) = h_0(t) e^{\beta_2 x_2 + \beta_3 x_3}. \quad (4.13)$$

Tomando el riesgo relativo, HR, de dos individuos cuyos factores se encuentran en niveles diferentes, la interpretación de los términos del componente lineal en (4.13) son más explícitos: $\exp(\beta_2)$ es el riesgo de falla de un paciente con el factor A en el nivel 2 (Grupo II) *relativo* al riesgo de falla de un paciente con A en el nivel 1, que es el grupo de referencia; y $\exp(\beta_3)$ es el riesgo de falla de un paciente en el Grupo III, relativo al riesgo de falla de otro paciente en el Grupo I. De esta manera, si $\exp(\beta_2) < 1$, digamos 0.5, el riesgo de falla en el Grupo I es la mitad del riesgo para el Grupo II, y si $\exp(\beta_2) > 1$, digamos 2, el riesgo de falla en el Grupo II es el doble que el del Grupo I.

Supongamos que se tiene un modelo de riesgos proporcionales en el cual se incluye una sola covariable continua X . De este modo, la función de riesgo del i -ésimo de los n individuos, para $X = x_i$, es

$$h(t; x_i) = h_0(t) \exp(\beta x_i).$$

Ahora considere la HR para un individuo cuyo valor registrado de X es $x + 1$, relativo a un individuo para el cual se obtuvo un valor x , eso es,

$$HR = \frac{h(t; x + 1)}{h(t; x)} = \frac{\exp\{\beta(x + 1)\}}{\exp\{\beta x\}} = \exp(\beta)$$

Así, en el modelo de riesgos proporcionales ajustado, $\hat{\beta}$ es el cambio estimado del logaritmo de la razón de riesgo cuando el valor de X es incrementado en una unidad.

Usando un argumento similar, el cambio estimado en el log HR cuando el valor de la covariable X se incrementa en k unidades es $k\hat{\beta}$, y el correspondiente estimador de la HR es $\exp(k\hat{\beta})$ (ver [1]).

4.2.3. Verosimilitud Parcial: Un método de estimación.

Cox (1972a) introdujo un método para estimar β en (4.9) sin la especificación previa de la función de riesgo base $h_0(t)$ mediante una verosimilitud, que posteriormente llamó *función de verosimilitud parcial* (Cox, 1975). Así, una vez obtenido el estimador $\hat{\beta}$ de β , se procede a estimar la función de riesgo base $h_0(t)$. Equivalentemente uno puede considerar a la función de riesgo acumulada base $H_0(t) = \int_0^t h_0(u)du$ o a la función de supervivencia base $S_0(t) = \exp\{-H_0(t)\}$ en lugar de $h_0(t)$.

Supongamos ahora que las fallas observadas ocurren en los distintos tiempos $t_{(1)} < t_{(2)} < \dots < t_{(r)}$; los restantes $n-r$ tiempos se toman como censurados por un mecanismo no-informativo. Por el momento ignoremos el caso de tiempos de falla empatados. Se puede verificar fácilmente que la función usual de verosimilitud para ese conjunto de datos correspondiente al modelo en (4.9) es

$$L[\beta, h_0(t)] = \prod_{j=1}^n h_0(t_j)^{\delta_j} [\exp(\beta^T \mathbf{x}_j)]^{\delta_j} \exp\{-H_0(t_j) \exp(\beta^T \mathbf{x}_j)\}, \quad (4.14)$$

Ya que el modelo de Cox no supone una forma específica de $h_0(t)$, no es posible emplear directamente métodos estándar con (4.14) para obtener un estimador del vector de parámetros desconocido β . No obstante, Cox (1972a) sugirió la siguiente función de verosimilitud para estimar β en la Ecuación 4.9:

$$L(\beta) = \prod_{j=1}^r \frac{\exp\{\beta^T \mathbf{x}_{(j)}\}}{\sum_{\ell \in R_j} \exp\{\beta^T \mathbf{x}_\ell\}} \quad (4.15)$$

donde $\mathbf{x}_{(j)} \in \mathbb{R}^p$ es el vector de covariables asociado con el individuo cuyo tiempo de falla es $t_{(j)}$ y $R_j \equiv R(t_{(j)})$ es definido como antes: el conjunto de riesgo al tiempo t , que consta de los individuos que no han presentado una falla y que permanecen bajo observación (no-censurados) al tiempo t^- , es decir, justo antes del tiempo t .

Un argumento para la obtención del j -ésimo factor en (4.15) es como sigue: dado R_j y dado que ocurre una falla al tiempo $t_{(j)}$, la probabilidad condicional de que un miembro de R_j con vector de covariables $\mathbf{x}_{(j)}$ falle en $t_{(j)}$ se puede expresar de manera equivalente² como

$$\frac{\Pr(\text{el individuo de } R_j \text{ con vector de covariables } \mathbf{x}_{(j)} \text{ falle en } t_{(j)})}{\Pr(\text{un miembro de } R_j \text{ falle en } t_{(j)})}. \quad (4.16)$$

El numerador en la expresión anterior es el riesgo de falla al tiempo $t_{(j)}$ correspondiente al individuo con vector de covariables $\mathbf{x}_{(j)}$. Si éste es el i -ésimo individuo de R_j quien presenta la falla en $t_{(j)}$, entonces $\mathbf{x}_{(j)} = \mathbf{x}_i$ y su función de riesgo se puede escribir como $h_i(t_{(j)}) = h_0(t_{(j)}) \exp\{\beta^T \mathbf{x}_{(j)}\}$. El denominador es la suma de los riesgos de falla al tiempo $t_{(j)}$ sobre todos los individuos quienes están en riesgo de falla en ese tiempo. Así, la probabilidad condicional en (4.16) se puede aproximar por

$$\frac{h_i(t_{(j)})\Delta t_{(j)}}{\sum_{\ell \in R_j} h_\ell(t_{(j)})\Delta t_{(j)}} = \frac{\exp\{\beta^T \mathbf{x}_{(j)}\}}{\sum_{\ell \in R_j} \exp\{\beta^T \mathbf{x}_\ell\}}, \quad (4.17)$$

² $\Pr(A | B) = \Pr(A \cap B) / \Pr(B)$.

Tomando el producto sobre $i = 1, 2, \dots, r$ de (4.17) se obtiene la verosimilitud $L(\boldsymbol{\beta})$ dada en (4.15). Cox (1972a) se refirió a ésta como una **verosimilitud condicional**, un concepto bastante engañoso ya que la verosimilitud marginal y condicional son casos especiales del concepto más general de **verosimilitud parcial** (Cox, 1975). Más adelante se dará un esbozo de la justificación de $L(\boldsymbol{\beta})$ como una verosimilitud parcial.

Nótese que $L(\boldsymbol{\beta})$ en (4.15) no es una verosimilitud en el sentido de la Definición A.1, ya que los tiempos de supervivencia observados no figuran numéricamente en su cálculo. Sin embargo, la maximización de (4.17) proporciona un estimador $\hat{\boldsymbol{\beta}}$ que es consistente, asintóticamente normal bajo apropiadas condiciones de regularidad (Cox, 1975; Tsiatis 1981) y eficiente (Efron, 1977), y estadísticos como la razón de verosimilitud, matriz información y puntaje basados en $L(\boldsymbol{\beta})$ se comportan como si se tratase de una verosimilitud ordinaria. De esta manera, el proceso de estimación de $\boldsymbol{\beta}$ requiere la solución del sistema de ecuaciones

$$\mathbf{u}(\boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\beta}} l(\boldsymbol{\beta}) \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = \mathbf{0}, \quad (4.18)$$

donde $l(\boldsymbol{\beta}) \equiv \log L(\boldsymbol{\beta})$ y $\mathbf{u}(\boldsymbol{\beta}) = \partial l(\boldsymbol{\beta}) / \partial \boldsymbol{\beta}$ es la **función score o puntaje**. Esto se puede llevar a cabo mediante algún método iterativo, tal como el Método Newton-Raphson.

Bajo apropiadas condiciones de regularidad, $\mathbf{i}^{-1}(\hat{\boldsymbol{\beta}}) = -\partial^2 l(\boldsymbol{\beta}) / \partial \boldsymbol{\beta}^2$ proporciona un estimador para la matriz-varianza de $\hat{\boldsymbol{\beta}}$, por lo que

$$\hat{\boldsymbol{\beta}} \stackrel{a}{\sim} N(\boldsymbol{\beta}, \mathbf{i}^{-1}(\hat{\boldsymbol{\beta}})). \quad (4.19)$$

La mayoría del software en estadística incluye paquetes para obtener el estimador de máxima verosimilitud $\hat{\boldsymbol{\beta}}$ de $\boldsymbol{\beta}$, así como errores estándar, pruebas de hipótesis o intervalos de confianza basados en la aproximación asintótica normal estándar dada en (4.19).

Verosimilitud parcial en presencia de empates.

Cuando los datos contienen tiempos observados **empataados**³, la verosimilitud parcial (4.15) tiene que ser modificada de alguna forma. Se han propuesto varias aproximaciones para la función de verosimilitud parcial en esta situación, por ejemplo, Breslow (1974), Efron (1977), Cox (1972) (ver [1], [16] para más detalles).

Sea $t_{(1)} < t_{(2)} < \dots < t_{(r)}$ los r distintos tiempos observados y ordenados. Sea d_j el número de fallas ocurridas en $t_{(j)}$ y $\mathcal{D}_j \equiv \mathcal{D}(t_{(j)}) = \{j_1, j_2, \dots, j_{d_j}\}$ el conjunto de etiquetas de los individuos que fallan en $t_{(j)}$. Sea $\mathbf{s}_j = \sum_{\ell \in \mathcal{D}_j} \mathbf{x}_\ell$ y R_j el conjunto de subíndices de individuos en riesgo al tiempo $t_{(j)}$.

³ El caso de datos de supervivencia sin tiempos empatados es más realista que ocurra cuando la variable del tiempo de supervivencia tiene una distribución continua. En contraste con distribuciones discretas del tiempo de supervivencia, estas permiten la presencia de empates en los datos, dependiendo también de la escala de medición que se considere (semanas, meses, años).

La aproximación de la verosimilitud parcial sugerida por Breslow (1974) considera que las d_j fallas al tiempo $t_{(j)}$ son distintos y ocurriendo secuencialmente. Cuando se tienen pocos empates, esta proporciona una muy buena aproximación de la función de verosimilitud parcial. La verosimilitud debida a Breslow (1974), en el caso de empates, es

$$\prod_{j=1}^r \frac{\exp(\boldsymbol{\beta}^\top \mathbf{s}_j)}{[\sum_{\ell \in R_j} \exp(\boldsymbol{\beta}^\top \mathbf{s}_\ell)]^{d_j}}.$$

Observación 4.2. Ya que la expresión para la verosimilitud sugerida por Breslow es la más simple, esta viene implementada en la mayoría de los paquetes estadísticos.

Una aproximación alternativa, sugerida por Efron (1977), es

$$\prod_{j=1}^r \frac{\exp(\boldsymbol{\beta}^\top \mathbf{s}_j)}{\prod_{k=1}^{d_j} [\sum_{\ell \in R_j} \exp(\boldsymbol{\beta}^\top \mathbf{s}_\ell) - (k-1)d_k^{-1} \sum_{\ell \in \mathcal{D}_j} \exp(\boldsymbol{\beta}^\top \mathbf{x}_\ell)]}.$$

Cox (1972) sugirió la aproximación

$$\prod_{j=1}^r \frac{\exp(\boldsymbol{\beta}^\top \mathbf{s}_j)}{\sum_{\ell \in R(t_{(j)}; d_j)} \exp(\boldsymbol{\beta}^\top \mathbf{s}_\ell)}.$$

donde $R(t_{(j)}; d_j)$ denota el conjunto de todos los subconjuntos de d_j individuos seleccionados del conjunto de riesgo $R(t_j)$ sin reemplazo. De este modo, si $\ell \in R(t_{(j)}; d_j)$, éste es de la forma $\{\ell_1, \ell_2, \dots, \ell_{d_j}\}$. La verosimilitud parcial anterior es computacionalmente difícil si el número de empates es grande.

4.2.3.1. Justificación de $L(\boldsymbol{\beta})$ como una verosimilitud parcial

Crowder (2001) proporciona una justificación informal de la inferencia de $\boldsymbol{\beta}$ basada en la verosimilitud parcial (4.15) siguiendo Cox (1975), además de referencias adicionales en las que se presenta de manera más formal una teoría general de este tema. En Kalbfleisch & Prentice (2002) también se presenta una estrecha relación con el trabajo de Cox (1972a, 1975). Aquí solamente se dará un esbozo de la justificación de $L(\boldsymbol{\beta})$ como una verosimilitud parcial.

En términos generales, la función de verosimilitud parcial del modelo de Cox está basada en un argumento de probabilidad condicional como sigue. Supongamos que $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\phi})$, donde $\boldsymbol{\beta}$ es el parámetro de interés y $\boldsymbol{\phi}$ es considerado como un parámetro de ruido o de estorbo de dimensión infinita; en el caso del modelo de Cox dado en (4.9), la función de riesgo base, $h_0(t)$, toma el papel de $\boldsymbol{\phi}$. Supongamos que el vector observado \mathbf{t} , cuya f.d.p. es $f(\mathbf{t}; \boldsymbol{\theta})$, puede ser transformado en la secuencia de pares de eventos $(A_1, B_1), \dots, (A_r, B_r)$. Sea $A^{(m)} = (A_1, A_2, \dots, A_m)$ y $B^{(m)} = (B_1, B_2, \dots, B_m)$, donde el número de términos m puede ser aleatorio o fijo. Supongamos que

$$Pr(A^{(r)}, B^{(r)} | \boldsymbol{\theta}) = \prod_{j=1}^r Pr(B_j | B^{(j-1)}, A^{(j-1)}; \boldsymbol{\theta}) \prod_{j=1}^r Pr(A_j | B^{(j)}, A^{(j-1)}; \boldsymbol{\beta}),$$

por lo que la función de verosimilitud (completa) del conjunto de datos observado \mathbf{t} se puede escribir como

$$L(\boldsymbol{\theta}; \mathbf{t}) = \prod_{j=1}^r f(b_j | b^{(j-1)}, a^{(j-1)}; \boldsymbol{\beta}, \phi) \prod_{j=1}^r f(a_j | b^{(j)}, a^{(j-1)}; \boldsymbol{\beta}). \quad (4.20)$$

El segundo termino en (4.20) se llama la **verosimilitud parcial** de $\boldsymbol{\beta}$ basado en $\{A_j\}$ en la sucesión $\{A_j, B_j\}$ (Cox, 1975). Por lo regular existirá alguna perdida de información al usar solamente la verosimilitud parcial

$$\prod_{j=1}^r f(a_j | b^{(j)}, a^{(j-1)}; \boldsymbol{\beta}) \quad (4.21)$$

para hacer inferencia sobre $\boldsymbol{\beta}$, ya que el primer término en (4.20), es decir,

$$\prod_{j=1}^r f(b_j | b^{(j-1)}, a^{(j-1)}; \boldsymbol{\beta}, \phi) \quad (4.22)$$

dependerá hasta cierto grado del parámetro de interés, $\boldsymbol{\beta}$, conteniendo alguna información residual de éste.

En el caso del modelo de riesgos proporcionales, Cox (1975) argumentó que no se cuenta con información alguna de $\boldsymbol{\beta}$ proveniente de (4.22), ya que esa información residual está inextricablemente⁴ ligada con la información sobre el parámetro de ruido ϕ , es decir, con la función de riesgo base $h_0(\cdot)$. Además, ya que no se supone una forma específica para $h_0(\cdot)$, es concebible que ésta tome un valor muy cercano a cero en aquellos intervalos de tiempo en que no hay falla alguna. En esta situación, se puede elegir a (4.21) para hacer inferencia sobre $\boldsymbol{\beta}$ e ignorar a (4.22) sin perder mucha información del parámetro de interés (ver también Collett, 1994, pag. 63).

Para ver a $L(\boldsymbol{\beta})$ en (4.15) como una verosimilitud parcial empleando la factorización de la Expresión 4.20, retomemos nuevamente la suposición de que hay r distintos tiempos de supervivencia ordenados y no-censurados $t_{(1)} < t_{(2)} < \dots < t_{(r)}$. Los restantes $n-r$ tiempos de falla están censurados por la derecha. Luego, tomemos a B_j en (4.20) para especificar la información de los tiempos de falla censurados en el intervalo $[t_{(j-1)}, t_{(j)})$ y la información de que se ha observado una falla en el intervalo $[t_{(j)}, t_{(j)} + \Delta t_{(j)}]$; A_j especifica la información del individuo en particular, digamos $i_{(j)}$, quien presenta la falla en $[t_{(j)}, t_{(j)} + \Delta t_{(j)})$. De esta manera, bajo un mecanismo de censura no informativo, el j -ésimo factor de la verosimilitud parcial en (4.21), que es $L_j(\boldsymbol{\beta}) = f(a_j | b^{(j)}, a^{(j-1)}; \boldsymbol{\beta})$, es justo la probabilidad condicional de la Ecuación 4.17, es decir,

$$L_j(\boldsymbol{\beta}) = \frac{h_{i_{(j)}}(t_{(j)})\Delta t_{(j)}}{\sum_{\ell \in R_j} h_{\ell}(t_{(j)})\Delta t_{(j)}} = \frac{\exp\{\boldsymbol{\beta}^T \mathbf{x}_{(j)}\}}{\sum_{\ell \in R_j} \exp\{\boldsymbol{\beta}^T \mathbf{x}_{\ell}\}}. \quad (4.23)$$

⁴ **Adj.** Que no se puede desenredar; muy intrincado y confuso.

Bajo el modelo de regresión (4.9), (4.23) se simplifica ya que el riesgo base $h_0(t_{(j)})\Delta t_{(j)}$ se cancela tanto en el numerador como en el denominador. Entonces, tomando el producto sobre j ($j=1,2,\dots,r$) se obtiene la verosimilitud parcial para β ,

$$L(\beta) = \prod_{j=1}^r \frac{\exp\{\beta^\top \mathbf{x}_{(j)}\}}{\sum_{\ell \in R_j} \exp\{\beta^\top \mathbf{x}_\ell\}}.$$

Note que el numerador del j -ésimo termino en $L(\beta)$ depende meramente de la información (en el vector de covariables $\mathbf{x}_{(j)}$) proporcionada por el individuo que falla en $t_{(j)}$, mientras que la suma en el denominados utiliza la información de aquellos individuos en riesgo al tiempo $t_{(j)}^-$. Sin embargo, no se hace un uso directo tanto de los tiempos de supervivencia no-censurados como de aquellos que están censurados, por lo que $L(\beta)$ no es la usual función de verosimilitud (completa) para el vector de datos \mathbf{t} . De aquí su nombre de *función de verosimilitud parcial*.

4.2.4. Estimación de la función de supervivencia base.

Una vez que se ha obtenido el estimador $\hat{\beta}$ de β en el modelo de Cox, se procede a la estimación de la función de riesgo base $h_0(t)$, o de manera equivalente, a la función de supervivencia base $S_0(t)$. De esta manera, la función de supervivencia para el individuo con vector de variables explicativas fijas, \mathbf{x} , se estima mediante

$$\hat{S}(t; \mathbf{x}) = \hat{S}_0(t)^{\exp(\hat{\beta}^\top \mathbf{x})}. \quad (4.24)$$

El proceso de estimación para $S_0(t)$ es análogo al que se uso con el estimador Kaplan-Meier (Sección 3.2). Como antes, sean $t_{(1)}, t_{(2)}, \dots, t_{(r)}$ los distintos tiempos de supervivencia no-censurados, $R_j = R(t_{(j)})$ el conjunto de índices o etiquetas asociados con los individuos en riesgo al tiempo $t_{(j)}^-$, D_j el conjunto de índices asociados con los d_j individuo que fallan en $t_{(j)}$ y C_j el conjunto de índices asociados con los individuos censurados en el intervalo $[t_{(j)}, t_{(j+1)})$ ($j = 1, 2, \dots, r$), donde $t_{(0)} = 0$ y $t_{(r+1)} = \infty$. Bajo un mecanismo de censura no-informativo, la contribución a la verosimilitud de un individuo con vector de covariables \mathbf{x} quien falla en $t_{(j)}$ es $S(t_{(j)}^-; \mathbf{x}) - S(t_{(j)}; \mathbf{x})$ y la contribución de una observación no-censurada al tiempo t es $S(t; \mathbf{x})$. La función de verosimilitud se puede escribir como

$$L = \prod_{i=1}^r \left\{ \prod_{\ell \in D_i} [S_0(t_{(i)}^-)^{\exp(\beta^\top \mathbf{x}_\ell)} - S_0(t_{(i)})^{\exp(\beta^\top \mathbf{x}_\ell)}] \prod_{\ell \in C_i} S_0(t_{(i)})^{\exp(\beta^\top \mathbf{x}_\ell)} \right\}, \quad (4.25)$$

donde $D_0 = \emptyset$.

Es claro que L es maximizada tomando $S_0(t) = S_0(t_i)$ para $t_{(i)} \leq t < t_{(i+1)}$ y permitiendo masa de probabilidad solamente en los tiempos de falla $t_{(1)}, t_{(2)}, \dots, t_{(r)}$. Como con el estimador Kaplan-Meier, esas observaciones conducen a considerar un

modelo discreto con función de supervivencia base⁵

$$S_0(t_{(i)}) = \prod_{j=0}^i \xi_j \quad \text{con } \xi_0 = 1, \quad (4.26)$$

para $i = 1, 2, \dots, r$. Así, $h_0(t_{(j)}) = 1 - \xi_j$ es el riesgo base en cada tiempo de falla observado $t_{(j)}$. A manera de simplificar un poco la notación sea $\theta_\ell = \exp(\boldsymbol{\beta}^\top \mathbf{x}_\ell)$. Substituyendo (4.26) en (4.25), obtenemos

$$\begin{aligned} L &= \prod_{i=1}^r \left\{ \prod_{\ell \in D_i} \left[\prod_{j=1}^{i-1} \xi_j^{\theta_\ell} - \prod_{j=1}^i \xi_j^{\theta_\ell} \right] \prod_{\ell \in C_i} \prod_{j=1}^{i-1} \xi_j^{\theta_\ell} \right\} \\ &= \prod_{i=1}^r \left\{ \prod_{j \in D_i} \left[(1 - \xi_i^{\theta_j}) \prod_{j=0}^{i-1} \xi_j^{\theta_\ell} \right] \prod_{\ell \in C_i} \prod_{j=1}^{i-1} \xi_j^{\theta_\ell} \right\} \\ &= \left[\prod_{i=1}^r \prod_{j \in D_i} (1 - \xi_i^{\theta_j}) \right] \prod_{i=1}^r \left[\prod_{\ell \in D_i} \prod_{j=0}^{i-1} \xi_j^{\theta_\ell} \prod_{\ell \in C_i} \prod_{j=0}^i \xi_j^{\theta_\ell} \right] \\ &= \left[\prod_{i=1}^r \prod_{j \in D_i} (1 - \xi_i^{\theta_j}) \right] \left[\prod_{\ell \in R_1 - D_1} \xi_1^{\theta_\ell} \prod_{\ell \in R_2 - D_2} \xi_2^{\theta_\ell} \dots \prod_{\ell \in R_r - D_r} \xi_r^{\theta_\ell} \right] \\ &= \left[\prod_{i=1}^r \prod_{j \in D_i} (1 - \xi_i^{\theta_j}) \right] \left[\prod_{i=1}^r \prod_{\ell \in R_i - D_i} \xi_i^{\theta_\ell} \right] \\ &= \prod_{i=1}^r \left[\prod_{j \in D_i} (1 - \xi_i^{\exp(\boldsymbol{\beta}^\top \mathbf{x}_j)}) \prod_{\ell \in R_i - D_i} \xi_i^{\exp(\boldsymbol{\beta}^\top \mathbf{x}_\ell)} \right] \end{aligned}$$

De la expresión anterior, la función de verosimilitud a ser maximizada es

$$L(\boldsymbol{\beta}, \xi_1, \dots, \xi_r) = \prod_{i=1}^r \left[\prod_{j \in D_i} (1 - \xi_i^{\exp(\boldsymbol{\beta}^\top \mathbf{x}_j)}) \prod_{\ell \in R_i - D_i} \xi_i^{\exp(\boldsymbol{\beta}^\top \mathbf{x}_\ell)} \right] \quad (4.27)$$

Tomando $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$ como el estimador de la función de verosimilitud parcial de Cox, maximizamos (4.27) con respecto a $\xi_1, \xi_2, \dots, \xi_r$. Diferenciando el logaritmo de (4.27)

⁵ Este es el caso particular de covariables fijas, para el caso más general de covariables dependientes del tiempo, $\mathbf{x}(t)$, ver [16]

con respecto a ξ_i obtenemos

$$\begin{aligned}
\frac{\partial \log L}{\partial \xi_i} &= - \sum_{j \in D_i} \frac{e^{\beta^\top \mathbf{x}_j} \xi_i^{\exp(\beta^\top \mathbf{x}_j) - 1}}{1 - \xi_i^{\exp(\beta^\top \mathbf{x}_j)}} + \sum_{\ell \in R_i - D_i} \frac{e^{\beta^\top \mathbf{x}_\ell}}{\xi_i} \\
&= \left[- \sum_{j \in D_i} \frac{e^{\beta^\top \mathbf{x}_j} \xi_i^{\exp(\beta^\top \mathbf{x}_j) - 1}}{1 - \xi_i^{\exp(\beta^\top \mathbf{x}_j)}} - \sum_{j \in D_i} \frac{e^{\beta^\top \mathbf{x}_j}}{\xi_i} \right] + \left[\sum_{\ell \in R_i - D_i} \frac{e^{\beta^\top \mathbf{x}_\ell}}{\xi_i} + \sum_{\ell \in D_i} \frac{e^{\beta^\top \mathbf{x}_\ell}}{\xi_i} \right] \\
&= - \sum_{j \in D_i} \left[\frac{e^{\beta^\top \mathbf{x}_j} \xi_i^{\exp(\beta^\top \mathbf{x}_j) - 1}}{1 - \xi_i^{\exp(\beta^\top \mathbf{x}_j)}} + \frac{\exp(\beta^\top \mathbf{x}_j)}{\xi_i} \right] + \sum_{\ell \in R_i} \frac{e^{\beta^\top \mathbf{x}_\ell}}{\xi_i} \\
&= - \sum_{j \in D_i} \frac{\exp(\beta^\top \mathbf{x}_j)}{(1 - \xi_i^{\exp(\beta^\top \mathbf{x}_j)}) \xi_i} + \sum_{\ell \in R_i} \frac{\exp(\beta^\top \mathbf{x}_\ell)}{\xi_i} \tag{4.28}
\end{aligned}$$

Igualando a cero en (4.28) el EMV de ξ_i ($i = 1, 2, \dots, r$) se obtiene como una solución de la ecuación

$$\sum_{j \in D_i} \frac{\exp(\hat{\beta}^\top \mathbf{x}_j)}{1 - \hat{\xi}_i^{\exp(\hat{\beta}^\top \mathbf{x}_j)}} = \sum_{\ell \in R_i} \exp(\hat{\beta}^\top \mathbf{x}_\ell), \tag{4.29}$$

la cual se resuelve, por lo regular, mediante algún método iterativo.

En el caso de que exista una sola falla en $t_{(i)}$, i.e. $d_i = 1$ ($i=1, 2, \dots, r$), (4.29) puede resolverse de manera explícita para ξ_i , obteniendo

$$\hat{\xi}_i = \left(1 - \frac{\exp(\hat{\beta}^\top \mathbf{x}_i)}{\sum_{\ell \in R_i} \exp(\hat{\beta}^\top \mathbf{x}_\ell)} \right)^{\exp(-\hat{\beta}^\top \mathbf{x}_i)}.$$

El estimador de máxima verosimilitud de la función de supervivencia base es

$$\hat{S}_0(t) = \prod_{i | t_{(i)} < t} \hat{\xi}_i, \tag{4.30}$$

que, al igual que el estimador-KM, es una función escalonada con discontinuidad en cada tiempo de supervivencia observado $t_{(i)}$. El correspondiente estimador de la función de riesgo acumulada base es

$$\Lambda_0(t) = \sum_{j | t_{(j)} \leq t} (1 - \hat{\xi}_j)$$

Finalmente, en el caso de covariables fijas, la función de supervivencia estimada para un individuo con vector de variables explicativas \mathbf{x} es

$$\hat{F}(t; \mathbf{x}) = \prod_0^t [1 - d\hat{\Lambda}_0]^{\exp[\hat{\beta}^\top \mathbf{x}]},$$

el cual se reduce ha $\hat{S}(t; \mathbf{x}) = [\hat{S}_0(t)]^{\exp(\hat{\beta}^\top \mathbf{x})}$ (ver [16] para más detalles).

Se han propuesto otros estimadores para la función de supervivencia base $S_0(t)$, o de manera equivalente, para la función de riesgo acumulada base $H_0(t) = \int_0^t h_0(u) du$. Por ejemplo, en el caso de una sola muerte en cada tiempo de supervivencia $t_{(i)}$, el *estimador Breslow* de la función de riesgo acumulada base es

$$\hat{H}_0(t) = \sum_{i|t_{(i)} < t} \frac{1}{\sum_{\ell \in R_i} \exp(\hat{\boldsymbol{\beta}}^\top \mathbf{x}_\ell)} \quad (4.31)$$

Este y otros estimadores de la función de supervivencia, junto con una variedad de ejemplos se discuten con mayor detalle en [19].

4.2.5. Contraste de hipótesis para el Modelo de Cox ajustado

Hay otros modelos formados a partir de la inclusión de dos o más factores, interacciones entre estos, o también de la inclusión de un factor y una variable explicativa numérica (ver [1]). El conocimiento del contexto en el cual se obtuvieron los datos (e.g. los resultados de un análisis exploratorio en los datos, relaciones teóricas entre las variables, el diseño del estudio, etc.) pueden ayudar para la formulación de un modelo adecuado.

Una vez que el Modelo de Cox a sido ajustado, existen tres contrastes de hipótesis para verificar la significancia del modelo:

- Prueba del cociente de verosimilitud;
- Prueba de Wald;
- Prueba de Puntajes.

Más adelante se da un ejemplo en el que se utiliza el **método de eliminación hacia atrás** para seleccionar el modelo más parsimonioso, empleando la prueba del cociente de verosimilitud⁶.

Este comienza con la comparación de modelos alternativos: el llamado **Modelo Saturado** (MS), el cual contiene a todas las covariables fijas consideradas como predictoras del tiempo de supervivencia, y el llamado **Modelo Actual** (MA), el cual esta *anidado* en el (MS), ambos ajustados al mismo conjunto observado de datos de supervivencia. Cuando se comparan modelos de regresión alternativos, el mayor interés está centrado en la hipótesis de que alguno de los parámetros de regresión en el modelo es igual a cero. Para probar esa hipótesis se empleará la prueba del cociente de verosimilitud.

Supongamos que el $MS(p+q)$ contiene $p+q$ parámetros, $\beta_1, \beta_2, \dots, \beta_p, \beta_{p+1}, \dots, \beta_{p+q}$, y se desea comparar con el $MA(p)$ que solamente contiene los p parámetros β_1, \dots, β_p . Esto equivale a probar la hipótesis nula de que los q parámetros desconocidos $\beta_{p+1}, \beta_{p+2}, \dots, \beta_{p+q}$ en $MS(p+q)$ son todos iguales a cero. Denotemos por $\hat{\boldsymbol{\beta}}_1$ el vector de estimadores bajo $MS(p+q)$, y $\hat{\boldsymbol{\beta}}_2$ para $MA(p)$. La prueba del cociente de verosimilitud

⁶ Collet (1994) sugieren este contraste de hipótesis debido a que presenta una mayor confiabilidad

de la hipótesis nula de que $\beta_{p+1} = \beta_{p+2} = \dots = \beta_{p+q} = 0$ en $MS(p+q)$ se basa en el estadístico

$$2\{\log L(\hat{\beta}_1) - \log L(\hat{\beta}_2)\}.$$

Bajo la hipótesis nula, ese estadístico se distribuye χ^2 con q grados de libertad. Si el valor observado de ese estadístico no es significativamente grande, los dos modelos serán igualmente adecuados. Entonces, uno se inclinaría por el modelo que incluye menos términos. Por el otro lado, si los valores $-2 \log L(\hat{\beta})$ para los dos modelos es significativamente diferente, eso nos sugiere que los términos adicionales son necesarios, por lo que se adoptaría el modelo más complejo, que es $MS(p+q)$.

4.3. Ejemplo: Comparación de dos tratamientos médicos

Para ilustrar el ajuste del modelo de Cox, se utilizaran los datos de la Tabla 1.3. Kleinbaum (1996), Lawles (2003) y otros han discutido ese conjunto de datos, los cuales tratan de una prueba médica para examinar los tiempos de remisión inducidos por esteroides (en semanas) de 42 pacientes con leucemia. La variable dicotómica Rx (1= placebo, 0=medicamento 6-MP), y la variable $\log WBC$ son incluidas en el modelo, siendo la primera de primordial interés ya que el objetivo del estudio es cuantificar el efecto del tipo de tratamiento, ajustado por el posible efecto de confusión de $\log WBC$, que es una variable pronóstico de importancia⁷ en el estudio sobre la supervivencia de esos pacientes. Debido a eso, también se considera el posible efecto de interacción de $\log WBC$ sobre la variable explicativa Rx, es decir, se toma en cuenta la posibilidad de que el efecto del tipo de tratamiento sobre el tiempo de supervivencia de los pacientes dependa de los niveles de $\log WBC$. Es obvio que primero debe contrastarse la interacción y después, en caso de que esta no existiera, la confusión. Para llevar a cabo esto, considere los siguientes modelos de regresión:

- **Modelo Nulo** : $S_0(t)$
- **Modelo(1)** : $S(t; \mathbf{x}) = S_0(t)^{\exp(\beta_1 * Rx)}$
- **Modelo(2)** : $S(t; \mathbf{x}) = S_0(t)^{\exp(\beta_1 * Rx + \beta_2 * \log WBC)}$
- **Modelo(3)** : $S(t; \mathbf{x}) = S_0(t)^{\exp(\beta_1 * Rx + \beta_2 * \log WBC + \beta_3 * [Rx * \log WBC])}$

Los resultados del ajuste para los tres diferentes modelos de riesgos proporcionales obtenidos del paquete estadístico **R** se llevan a cabo mediante la función `coxph`. En caso de "empate" en los datos, **R** usa por default la aproximación de la función de verosimilitud parcial de Efron (1977), pero hay otros métodos alternativos disponibles: Método de Breslow y Método Exacto. Si no hay empates en los tiempos de falla, los tres métodos son equivalentes. Para este ejemplo se usa el Método de Breslow.

⁷ La variable explicativa $\log WBC$ se considera un buen indicador del pronóstico de supervivencia para pacientes con leucemia en el sentido de que un alto WBC tiene un mal pronóstico de supervivencia

Veamos como usar esos resultados para evaluar el posible efecto del tipo de tratamiento (Rx), que es la covariable de primordial interés, sobre el tiempo de remisión T ajustado por el efecto de interacción y confusión de la variable explicativa $\log WBC$. La mayoría del software que se usa en análisis de supervivencia, en particular **R**, proporcionan los resultados de la prueba de hipótesis $H_0 : \beta_j = 0$, usando el *estadístico de Wald* $\beta_j/s.e.(\beta_j)$, junto con su correspondiente *valor-P*. En este ejemplo se usará la prueba del cociente de verosimilitud para la significancia del modelo.

Denotemos por \hat{L}_1 , \hat{L}_2 y \hat{L}_3 el valor de la función de máxima verosimilitud para cada modelo, respectivamente, y por \hat{L}_0 el correspondiente al Modelo Nulo. Entonces, los estadísticos de prueba de la razón de verosimilitud para el Modelo(2) y el Modelo(3) son

$$-2 \log \frac{\hat{L}_0}{\hat{L}_2} \quad \text{y} \quad -2 \log \frac{\hat{L}_0}{\hat{L}_3},$$

los cuales se distribuyen asintóticamente $\chi_{(2)}^2$ y $\chi_{(3)}^2$, respectivamente, bajo la hipótesis nula de que todos los parámetros β_j incluidos en cada modelo son iguales a cero. Al tomar la diferencia entre esas dos razones de verosimilitud se obtiene el estadístico

$$-2 \log \frac{\hat{L}_2}{\hat{L}_3} = \left(-2 \log \frac{\hat{L}_0}{\hat{L}_3} \right) - \left(-2 \log \frac{\hat{L}_0}{\hat{L}_2} \right) \quad (4.32)$$

el cual tienen una distribución asintótica $\chi_{(3-2)}^2 = \chi_{(1)}^2$ y puede usarse para probar la significancia del término interacción en el Modelo(3). Las instrucciones en **R** son las siguientes:

Resumen de R para el Modelo(3).

```
# Supongase que los datos se encuentran en el objeto "remision" en
# formato "data.frame".
> library(survival)
> mod3.ph<-coxph(Surv(tiempo,censura)~Rx+logWBC+Rx*logWBC,remision)
> summary(mod3.ph)
```

	coef	exp(coef)	se(coef)	z	Pr(> z)
Rx	2.3749	10.7500	1.7055	1.393	0.164
logWBC	1.8724	6.5040	0.4514	4.148	3.35e-05
Rx:logWBC	-0.3175	0.7280	0.5258	-0.604	0.546

```

Rsquare= 0.674 (max possible= 0.988 )
Likelihood ratio test= 47.07 on 3 df, p=3.356e-10
Wald test = 32.39 on 3 df, p=4.326e-07
Score(logrank) test = 49.86 on 3 df, p=8.539e-11
```

Analógamente, para el Modelo(2) se obtiene:

Resumen de R para el Modelo(2).

```

> mod2.ph<-coxph( Surv(tiempo,censura) ~ Rx + logWBC, remision)
> summary(mod2.ph)

```

	coef	exp(coef)	se(coef)	z	Pr(> z)
Rx	1.3861	3.9991	0.4248	3.263	0.00110
logWBC	1.6909	5.4243	0.3359	5.034	4.80e-07

```

Rsquare= 0.671 (max possible= 0.988 )
Likelihood ratio test= 46.71 on 2 df, p=7.187e-11
Wald test = 33.6 on 2 df, p=5.061e-08
Score (logrank) test = 46.07 on 2 df, p=9.92e-11

```

Nótese primero que el valor de $-2\log(\widehat{L}_0/\widehat{L}_3)$ para el Modelo(3) es 47.07, mientras que el de $-2\log(\widehat{L}_0/\widehat{L}_2)$ correspondiente al Modelo(2) es 46.71. Así, bajo la hipótesis nula de que no hay efecto de interacción ($\widehat{\beta}_3 = 0$), calculamos la diferencia de esas dos cantidades obteniendo 0.360. Este es el valor observado, basado en los datos, del estadístico (4.32). El *valor-P* correspondiente es de 0.555. Entonces, no hay evidencia alguna para rechazar la hipótesis nula de que $\beta_3 = 0$, es decir, no hay una interacción significativa entre *Rx* y *log WBC*.

Ahora nos enfocaremos en evaluar el efecto de la covariable *Rx* ajustado por *log WBC*, usando los resultados obtenidos del Modelo(1).

Resumen de R para el Modelo(1).

```

> mod1.ph<-coxph( Surv(tiempo,censura) ~ Rx, data=remision)
> summary(mod1.ph)

```

	coef	exp(coef)	se(coef)	z	Pr(> z)
Rx	1.5721	4.8169	0.4124	3.812	0.000138

```

Rsquare= 0.322 (max possible= 0.988 )
Likelihood ratio test= 16.35 on 1 df, p=5.261e-05
Wald test = 14.53 on 1 df, p=0.0001378
Score (logrank) test = 17.25 on 1 df, p=3.283e-05

```

Procediendo de manera análoga como con el Modelo(3), la diferencia entre las razones de verosimilitud obtenidas de **R** para el Modelo(1) y el Modelo (2) es

$$\left(-2\log\frac{\widehat{L}_0}{\widehat{L}_2} \right) - \left(-2\log\frac{\widehat{L}_0}{\widehat{L}_1} \right) = 47.61 - 16.35 = 31.26,$$

el cual tiene un *valor-P* < 0.05 . Por lo tanto, los resultados de esa prueba muestran que usando el Modelo(2), el efecto de *log WBC* es significativo, después de ser ajustado por la covariable *Rx* (tipo de tratamiento) al nivel del 5%.

Los resultados anteriores nos sugieren que las covariables *Rx* y *log WBC* tienen un efecto significativo en el tiempo de remisión, al nivel del 5%, por eso es que el

Modelo(2) proporciona un buen ajuste para ese conjunto de datos.

El Modelo de Cox supone que los riesgos entre dos individuos distintos con covariables \mathbf{X} y \mathbf{X}^* son proporcionales, es decir, que la HR no depende del tiempo. Por eso, la verificación de tal supuesto es importante. Si en el modelo hay una o más covariables cuyo coeficiente varía con el tiempo, o si hay covariables que dependen del tiempo, la suposición de riesgos proporcionales no se cumplirá. Se puede verificar que las covariables incluidas en el Modelo(2) son independientes del tiempo mediante los métodos descritos en el libro de Kleinbaum([21]). Por otra parte, para verificar que el coeficiente de regresión de cada covariable en el Modelo(2) no depende del tiempo, se puede usar el comando de **R** `cox.zph`, el cual usa Residuos de Schoenfeld (ver [1]). Los comandos en **R** son los siguientes:

```
> plot(cox.zph(mod2.ph),var=1,main="Gráfica de Betas para Rx")
> plot(cox.zph(mod2.ph),var=2,main="Gráfica de Betas para log WBC")
```

Otra información importante obtenida a través de la salida del Modelo(2) es la estimación de la HR para el efecto de la covariable Rx ajustada por la covariable $\log WBC$, que es $\exp(\hat{\beta}_1) = 3.999$, y la estimación de la HR para el efecto de la covariable $\log WBC$ ajustada por la variable indicadora Rx , $\exp(\hat{\beta}_2) = 5.424$. De lo anterior podemos decir que un paciente tratado con placebo tiene 3.999 veces el riesgo de presentar una recaída relativo a un paciente tratado con 6-MP, para un valor fijo de la covariable $\log WBC$. Ahora considere a $\exp(\hat{\beta}_2) = 5.424$, que es la estimación de la razón de riesgo para un paciente con un tratamiento en particular (Rx fija) y un valor específico de $\log WBC$, relativo a un paciente en el mismo tratamiento, pero con valor de $\log WBC$ menor que el dado anteriormente. El valor de la HR estimada nos dice que a mayor cantidad de glóbulos blancos en la sangre, se corre más riesgo de experimentar una recaída en cualquier instante de tiempo ($\widehat{HR} > 1$).

Habiendo ajustado un modelo de Cox para este conjunto de datos, también es de interés examinar la distribución estimada de los tiempos de supervivencia, llamada *curva de supervivencia ajustada*. Sea $\overline{\mathbf{X}} = (\overline{Rx}, \overline{\log WBC})$. Bajo el Modelo(2), la gráfica del estimador de la función de supervivencia, ajustada por las covariables Rx y $\log WBC$ es

$$\begin{aligned} \widehat{S}(t; \overline{\mathbf{X}}) &= [\widehat{S}_0(t)]^{\exp\{\hat{\beta}_1 \overline{Rx} + \hat{\beta}_2 \overline{\log WBC}\}} \\ &= [\widehat{S}_0(t)]^{\exp\{1.3861(0.5) + 1.6909(2.93)\}} \\ &= [\widehat{S}_0(t)]^{284.3} \end{aligned} \tag{4.33}$$

Sin embargo, el objetivo es describir el efecto del tipo de tratamiento (Rx) en el tiempo de supervivencia, ajustado por $\log WBC$, por lo que el uso de (4.33) no es apropiado. Entonces, procedemos a calcular una curva de supervivencia para cada valor de Rx ; la covariable $\log WBC$ es fijada con el valor promedio de todos los individuos en el estudio, que es $\overline{\log WBC} = 2.93$. De esta manera, la curva de supervivencia estimada

para cada valor de Rx y $\log WBC = 2.93$ es

$$\begin{aligned}\widehat{S}(t; Rx = 0, \log WBC = 2.93) &= [\widehat{S}_0(t)]^{\exp\{\widehat{\beta}_1(0) + \widehat{\beta}_2(2.93)\}} \\ \widehat{S}(t; Rx = 1, \log WBC = 2.93) &= [\widehat{S}_0(t)]^{\exp\{\widehat{\beta}_1(1) + \widehat{\beta}_2(2.93)\}}\end{aligned}\quad (4.34)$$

En el software estadístico **R** se emplea la función `survfit` para estimar (4.33). El método `plot` usa el objeto regresado por `survfit` para graficar la función de supervivencia estimada, eso es,

```
> plot(survfit(mod2.coxph), xlab='Tiempo de remisi'on',
      ylab='Curva de Supervivencia Ajustada')
```

Para obtener las curvas de supervivencia ajustadas (4.34), se debe construir un nuevo marco de datos con dos renglones, uno por cada valor de Rx , el cual es pasado a `survfit` mediante el argumento `newdata`:

```
> attach(anderson)
> anderson.rx <- data.frame(Rx=c(0,1), logWBC=rep(mean(logWBC), 2))
> detach()
> plot(survfit(mod2.coxph, newdata=anderson.rx), lty=c(1,2))
> legend(locator(1), legend=c('Rx=0', 'Rx=1'), lty=c(1,2))
> title(main="Curvas de Supervivencia Ajustadas por log WBC")
```

Usualmente, las gráficas de las curvas de supervivencia ajustadas por un Modelo de Cox ajustado son funciones escalonadas, con salto de discontinuidad en los tiempos de falla observados (i.e. no-censurados).

Hay otros paquetes estadísticos alternativos, como es **SAS**. En [21] se proporciona el código para obtener las curvas de supervivencia ajustadas (4.34) mediante **SAS**. Las gráficas resultantes en **SAS**, las cuales aparecen en la Figura 4.1, muestran que el grupo de tratamiento ($Rx=0$) consistentemente tiene una probabilidad más alta de supervivencia que el grupo de placebo ($Rx=1$), después de ajustarse por $\log WBC$.

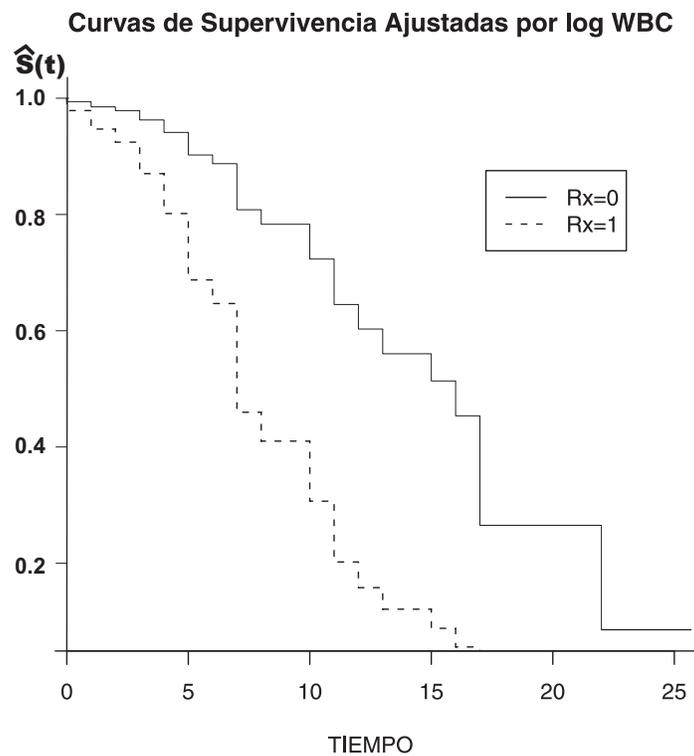


Figura 4.1: *Modelo(2)*. Curvas de supervivencia estimadas para los Grupos de tratamiento y placebo, correspondientes a los datos de Remisión.

TEORÍA DE RIESGOS COMPETITIVOS

Los datos de supervivencia de un solo evento pueden ser más generales en dos importantes aspectos. Primero, un individuo puede experimentar uno de $m (> 1)$ distintos tipos de falla $\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_m$; tales datos son comúnmente llamados datos de *riesgos competitivos* o múltiples modos de falla. Segundo, a cada individuo le puede corresponder más de un tiempo de falla, los cuales pueden presentarse cuando $m (> 1)$ v.a.s. Y_1, Y_2, \dots, Y_m son de interés, e.g. la ocurrencia sobre el tiempo de estudio de un mismo evento varias veces, en cuyo caso se tiene una restricción adicional. Por eso, surge la necesidad de considerar distribuciones de tiempos de supervivencia multivariadas, las cuales pueden especificarse en términos de funciones de supervivencia conjuntas de la forma

$$S_{1\dots m}(y_1, y_2, \dots, y_m) = \Pr(\{Y_1 > y_1\} \cap \{Y_2 > y_2\} \cap \dots \cap \{Y_m > y_m\}), \quad \forall y_j > 0.$$

Para fines del presente trabajo, se discutirán brevemente las cantidades básicas para el análisis de datos de riesgo competitivos, enfocándonos en problemas donde los métodos basados en tiempos de supervivencia univariados pueden ser adoptados. Por otra parte, si se desea un conocimiento más amplio de este tema, se puede consultar la mayoría de los libros estándar de Análisis de Supervivencia: Kalbfleisch y Prentice [16], Capítulo 8: Competing Risks and Multistate Models; En Crowder [5] se proporciona un buen resumen tanto de la teoría como de métodos para el análisis de datos de riesgos competitivos. Entre los artículos escritos en los que se da una revisión del presente tema está el de Gail [10] y el de Lindqvist [25].

5.1. Definición, suposiciones y algunos problemas básicos

Considere una población de individuos (por el momento homogénea) sujeta a $m (> 1)$ causas de falla $\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_m$, las cuales están operando simultáneamente entre ellos. Supongámonos un mecanismo de censura (por la derecha) independiente del tiempo de supervivencia. Entonces, cuando un miembro de esta población experimenta una falla, su tiempo de supervivencia Y y la causa (fundamental) \mathcal{C}_j son registradas. Sin embargo, cuando se tiene que Y es un tiempo de supervivencia censurado, la causa de falla será desconocida. También supóngase que

- **A1:** Cuando la falla ocurre, este puede ser de uno de los m distintos tipos o causas.
- **A2:** Cada individuo en una población dada es susceptible a experimentar una falla de cualquiera de las causas que están actuando en esa población.

Una manera clásica e intuitiva de describir un escenario de riesgos competitivos con m causas de falla¹ es suponer que a cada causa \mathcal{C}_j se le asocia una v.a.s. Y_j continua, la cual describe el *tiempo de falla latente*² de un individuo³ en condiciones hipotéticas donde \mathcal{C}_j es el único riesgo de falla ($j = 1, \dots, m$). De esta forma, cada individuo está asociado con un vector de v.a.s. $\mathbf{Y} = (Y_1, Y_2, \dots, Y_m)^\top$. Sin embargo, ya que todas las causas están actuando simultáneamente, y en vista de la suposición **A1**, no podemos observar Y_1, Y_2, \dots, Y_m conjuntamente. En lugar de eso, cada individuo está caracterizado por el par aleatorio observado (T, J) , donde

$$T = \min\{Y_1, Y_2, \dots, Y_m\}$$

es la v.a.s. (continua) que denota el tiempo de falla, y J la variable del tipo o modo de falla, tomando posibles valores en $\mathcal{J} = \{1, 2, \dots, m\}$. Por eso, se requiere de un modelo conjunto para T y J . Esto se puede llevar a cabo especificando modelos para $\Pr(\{T \leq t\} \cap \{J = j\})$ o especificando *funciones de riesgo condicionales de causa-específica*

$$h_j(t) = \lim_{\Delta t \rightarrow 0^+} \frac{\Pr(\{T \leq t + \Delta t\} \cap \{J = j\} \mid T > t)}{\Delta t}$$

para los distintos tipos de falla $j = 1, 2, \dots, m$ y $t > 0$. Como se verá más adelante, el análisis de este tipo de modelos está estrechamente relacionado con el análisis de datos de supervivencia univariado.

En cuanto a los problemas que surgen en el análisis de datos de riesgos competitivos, tres de ellos son de relevancia:

1. Inferencia en la relación entre un conjunto de variables explicativas y la tasa instantánea de falla de causa-específica, $h_j(t)$.
2. El estudio de la relación entre las funciones de riesgo de causa-específica bajo un conjunto específico de condiciones de estudio.
3. La estimación de la tasa instantánea de falla para alguna causa dada la eliminación de alguna o de todas las otras causas.

¹ Se le ha llamado riesgo antes de la falla, y causa después de esta: *el riesgo compete por ser la causa de falla* [5].

² En estadística, las variables latentes son variables que no son observables directamente pero son más bien inferidas (mediante un modelo matemático) a partir de otras variables que sí son observables y directamente medibles. A estas variables también se las conoce como **variables hipotéticas**.

³ Por individuos nos estaremos refiriendo a miembros de una población dada.

Por lo regular esos problemas son planteados en términos de los tiempos de falla latentes, Y_j 's, para cada tipo de falla, C_j , $j = 1, 2, \dots, m$. Sin embargo, este enfoque es criticado en base a suposiciones no garantizadas (riesgos independientes), falta de interpretación física (una vez que la falla se presenta al tiempo Y_j por causa C_j , los tiempos de supervivencia restantes ya no son observados) y *problemas de identificabilidad*.

Un enfoque alternativo se basa en la *función de riesgo de causa-específica*, mostrándose está una básica cantidad estimable en el esquema de riesgos competitivos [33].

Cabe mencionar que las probabilidades de falla dada la eliminación de alguna o de todas las otras causas (problema 3) ha sido un tema central y clásico en el análisis de riesgos competitivos. En artículos recientes esas probabilidades aparecen con otros nombres, pero anteriormente se referían a esas cantidades como:

- *Probabilidades netas*. A excepción de la j -ésima causa, todas las otras causas son eliminadas.
- *Probabilidades crudas (parciales)*. Todas las causas están presentes (algunas pero no todas las causas son eliminadas).

En el artículo de Prentice y Kalbfleisch [33] se encuentra una buena discusión de tales problemas.

5.2. Especificación del Modelo de Riesgos Competitivos

5.2.1. Función de Decremento Múltiple

Se define la *función de decremento múltiple* o *función de supervivencia conjunta* de los tiempos de falla latentes Y_1, Y_2, \dots, Y_m como ([16], pág. 259)

$$S_{1\dots m}(y_1, y_2, \dots, y_m) = \Pr \left(\bigcap_{j=1}^m \{Y_j > y_j\} \right) \quad (5.1)$$

para todo $y_j > 0$, donde $F_{y_1, \dots, y_m}(\cdot)$ es la función de distribución acumulada (FDA) conjunta correspondiente.

Es conveniente (aunque no esencial) suponer que (5.1) es una distribución **propia** en el sentido de que

$$S_{1\dots m}(0, \dots, 0) = 1 \quad \text{y} \quad S_{1\dots m}(\infty, \dots, \infty) = 0. \quad (5.2)$$

También se tiene una expresión para la función de riesgo multivariada con respecto a y_j en $\mathbf{y} = (y_1, \dots, y_m)$, definida como (Johnson y Kotz, 1972)

$$\begin{aligned}
h_j(\mathbf{y}) &= \lim_{\Delta y \rightarrow 0^+} \frac{1}{\Delta y} \Pr \left(\{Y_j \leq y_j + \Delta y\} \bigcap_{\substack{i=1 \\ i \neq j}}^m \{Y_i > y_i\} \mid \bigcap_{j=1}^m \{Y_j > y_j\} \right) \\
&= -\frac{1}{S_{1\dots m}(\cdot)} \cdot \frac{\partial S_{1\dots m}(\cdot)}{\partial y_j} = -\frac{\partial \log S_{1\dots m}(\cdot)}{\partial y_j}.
\end{aligned} \tag{5.3}$$

5.2.2. La Función de Supervivencia Global

Como se mencionó anteriormente, cada individuo es caracterizado por un vector de tiempos de falla latentes $\mathbf{Y} = (Y_1, Y_2, \dots, Y_m)$. Sin embargo, ya que la probabilidad de presentar dos o más fallas distintos de manera simultanea es cero (suposición **A1**), no podemos observar conjuntamente los componentes de \mathbf{Y} . Lo que se observa es el mínimo, digamos T , de esos tiempos de supervivencia latentes en que la falla ha ocurrido, eso es,

$$T = \min(Y_1, \dots, Y_m),$$

refiriendose a éste como el *tiempo de supervivencia real* [34]. De esta forma, se define la *función de supervivencia global* como

$$S_T(t) = \Pr(T > t) = \Pr \left(\bigcap_{j=1}^m \{Y_j > t\} \right) = S_{1\dots m}(t, t, \dots, t), \tag{5.4}$$

la cual se interpreta como la probabilidad de sobrevivir de cualquier causa hasta el tiempo t . Ya que (5.4) es función de una sola variable, del Capítulo 2, la *tasa de falla instantanea global* (i.e. de cualquier causa), o *función de riesgo del tiempo de supervivencia real*, es

$$\begin{aligned}
h_T(t) &= \lim_{\Delta t \rightarrow 0^+} \frac{1}{\Delta t} \Pr(T \leq t + \Delta t \mid T > t) \\
&= -\frac{d \log S_T(t)}{dt} = -\frac{d \log S_{1\dots m}(t, \dots, t)}{dt}.
\end{aligned} \tag{5.5}$$

Integrando con respecto a t produce la inversa relación

$$S_T(t) = \exp \left\{ - \int_0^t h_T(u) du \right\} = \exp \{-H_T(t)\} \tag{5.6}$$

donde $H_T(t) = \int_0^t h_T(u) du$ es la **función de riesgo acumulada**.

La distribución conjunta del par aleatorio (T, J) puede especificarse ya sea o por la *función de riesgo de causa-específica* o por la *función de incidencia acumulada*⁴ para la falla de tipo j .

⁴ También llamada Distribución de Probabilidad Cruda por Causa C_j .

5.2.3. Distribución Conjunta de (T, J)

Función de Riesgo de causa-específica

Para la especificación del modelo se define una tasa instantánea de falla de causa-específica.

Definición 5.1. *Se define la función de riesgo al tiempo t por causa \mathcal{C}_j en presencia de todas las otras causas actuando simultáneamente en la población como*

$$\begin{aligned} h_j(t) &= \lim_{\Delta t \rightarrow 0^+} \frac{1}{\Delta t} \Pr \left(\{Y_j \leq t + \Delta t\} \bigcap_{\substack{i=1 \\ i \neq j}}^m \{Y_i > t\} \mid \bigcap_{j=1}^m \{Y_j > t\} \right) \\ &= - \frac{\partial \log S_{1\dots m}(y_1, \dots, y_m)}{\partial y_j} \Big|_{y_1 = \dots = y_m = t} \end{aligned} \quad (5.7)$$

para $j = 1, 2, \dots, m$ y $t > 0$.

La función de riesgo de causa-específica representa la tasa instantánea para la falla de tipo j al tiempo t , y en presencia de todos los otros tipos de falla, condicionada a que no ha ocurrido ninguno de ellos hasta ese tiempo. Otra expresión que más a menudo se usa para (5.7) es

$$\begin{aligned} h_j(t) &= \lim_{\Delta t \rightarrow 0^+} \frac{1}{\Delta t} \Pr(\text{morir solamente de } \mathcal{C}_j \text{ en } (t, t + \Delta t) \text{ y todos los riesgos} \\ &\quad \text{actuando en } (t, t + \Delta t) \mid \text{sobrevivo de todas las causas hasta } t) \\ &= \lim_{\Delta t \rightarrow 0^+} \frac{\Pr(\{T \leq t + \Delta t\} \cap \{J = j\} \mid T > t)}{\Delta t} \end{aligned} \quad (5.8)$$

para $j = 1, 2, \dots, m$ y $t > 0$. En terminologías más antiguas (5.7) aparece con el nombre de *función de riesgo*, mientras que Crowder (1994) la llamó función de sub-riesgo.

Antes de mencionar la relación que guarda la función de riesgo global, $h_T(t)$, con las funciones de riesgo de causa-específica en (5.7), considere el siguiente Teorema.

Teorema 5.2.1. *Sea $z = f(y_1, \dots, y_m)$ una función de m variables. Supongase que cada variables y_j depende de t , es decir, $y_j = \sigma_j(t)$ para $j = 1, 2, \dots, m$. Si $z(\cdot)$ y $\sigma_1(t), \dots, \sigma_m(t)$ son diferenciables, también su composición y*

$$\frac{dz}{dt} = \frac{\partial f(\cdot)}{\partial y_1} \cdot \frac{d\sigma_1(t)}{dt} + \frac{\partial f(\cdot)}{\partial y_2} \cdot \frac{d\sigma_2(t)}{dt} + \dots + \frac{\partial f(\cdot)}{\partial y_m} \cdot \frac{d\sigma_m(t)}{dt}.$$

Tomando $\sigma_j(t) = t$, para $j = 1, 2, \dots, m$ en el Teorema anterior se tiene

$$\frac{d \log S_{1\dots m}(t, \dots, t)}{dt} = \sum_{j=1}^m \frac{\partial \log S_{1\dots m}(y_1, \dots, y_m)}{\partial y_m} \Big|_{y_1 = \dots = y_m = t}, \quad (5.9)$$

de manera que

$$h_T(t) = h_1(t) + h_2(t) + \dots + h_m(t). \quad (5.10)$$

Observación 5.1. La *propiedad aditiva* de las funciones de riesgo de causa-especifica establecida en (5.10) también se sigue de la definición probabilística de $h_T(t)$ y $h_j(t)$, para $j = 1, 2, \dots, m$, bajo la suposición **A1** (i.e. la probabilidad de morir simultáneamente de dos o más causas es cero) y aplicando la Ley de Alternativas (Ley de Probabilidad Total).

Una consecuencia de (5.10) es el siguiente Teorema que será de utilidad más adelante.

Teorema 5.2.2 (de Factorización). *Cualquier FDS conjunta, $S_{1\dots m}(y_1, \dots, y_m)$, puede ser factorizada en $y_1 = \dots = y_m = t$ como*

$$S_T(t) = S_{1\dots m}(t, \dots, t) = \prod_{j=1}^m G_j(t) \quad (5.11)$$

donde el j -ésimo factor está dado por

$$G_j(t) = \exp \left\{ - \int_0^t h_j(u) du \right\}. \quad (5.12)$$

Demostración. Al definir

$$H_j(t) = \int_0^t h_j(u) du,$$

$H_T(t) = \sum_{j=1}^m H_j(t)$ es la función de riesgo acumulado de T . Entonces, el resultado se sigue de (5.6). ■

Se puede verificar fácilmente que las funciones $G_j(t)$, $j = 1, 2, \dots, m$, tienen las propiedades matemáticas de las funciones de supervivencia continuas en el caso de un solo tipo de falla. No obstante, en el caso de riesgos competitivos dependientes, estas no son las funciones de supervivencia de cualesquiera variables aleatorias observables. Para $m > 1$ y sin suposiciones adicionales, $G_j(t)$ no tendrá la interpretación de una FDS, simplemente representa una distribución asociada con la causa \mathcal{C}_j , suponiendo que la correspondiente función de riesgo de causa-especifica es $h_j(t)$.

Modelos paramétricos para la distribución conjunta de (T, J) pueden ser obtenidos de diferentes maneras, por ejemplo, especificando paramétricamente la función de decremento múltiple en (5.1). Sin embargo, el planteamiento más común es mediante la especificación paramétrica de las $h_j(t)$'s.

Ejemplo 5.1. *Considere el caso de no covariables presentes. Supongamos una parametrización Weibull(α, γ) para las funciones de riesgo de causa-especifica en (5.7). Por eso, estas tienen la forma*

$$h_j(t; \boldsymbol{\theta}_j) = \frac{\alpha_j}{\gamma_j} \left(\frac{t}{\gamma_j} \right)^{\alpha_j - 1} \quad j = 1, 2, \dots, m, \quad (5.13)$$

donde $\boldsymbol{\theta}_j = (\alpha_j, \gamma_j)^\top$. Entonces, la función de riesgo (global) de T es

$$h_T(t; \boldsymbol{\theta}) = \sum_{j=1}^m \frac{\alpha_j}{\gamma_j} \left(\frac{t}{\gamma_j} \right)^{\alpha_j - 1}.$$

donde $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)$. De (5.6), la función de supervivencia de T es

$$S_T(t; \boldsymbol{\theta}) = \exp \left\{ - \int_0^t h_T(u; \boldsymbol{\theta}) du \right\} = \exp \left\{ - \sum_{j=1}^m \left(\frac{t}{\gamma_j} \right)^{\alpha_j} \right\} = \prod_{j=1}^m \exp \left\{ - \left(\frac{t}{\gamma_j} \right)^{\alpha_j} \right\}.$$

La función de sub-densidad $f_j(t)$ ($j = 1, 2, \dots, m$) se puede calcular como $h_j(t)S_T(t)$. Por otro lado, para incluir covariables en el modelo, se puede suponer un efecto multiplicativo de éstas en la función de riesgo de causa-específica, es decir,

$$h_j(t) = h_{0j}(t) \exp(\boldsymbol{\beta}_j^\top \mathbf{x}) \quad j = 1, 2, \dots, m.$$

□

La Función de Incidencia Acumulada de causa-específica

La distribución conjunta del par (T, J) también esta completamente especificada por las llamadas funciones de incidencia acumuladas de causa-específica.

De la expresión (5.8), la probabilidad condicional de experimentar una falla por causa \mathcal{C}_j en el intervalo $(t, t + \Delta t)$, condicionado a que se a sobrevivido (i.e. libre de cualquier evento) hasta el tiempo t , y en presencia de todas las causas actuando simultáneamente en la población, es aproximadamente $h_j(t)dt$, eso es,

$$\Pr(\{T < t + \Delta t\} \cap \{J = j\} \mid T > t) \approx h_j(t) \Delta t,$$

Entonces, la probabilidad incondicional de morir de causa \mathcal{C}_j en $(t, t + \Delta t)$ es aproximadamente $h_j(t)S_T(t) dt$. De esta manera, se define la *función de sub-distribución o de incidencia acumulada* al tiempo t por causa \mathcal{C}_j y en presencia de todas las causas actuando simultáneamente en una población como

$$Q_j(t) = \Pr(\{T \leq t\} \cap \{J = j\}) = \int_0^t h_j(u)S_T(u)du \quad t > 0, \quad (5.14)$$

para $j = 1, 2, \dots, m$. De manera análoga, se define la *función de sub-supervivencia* como

$$P_j(t) = \Pr(\{T > t\} \cap \{J = j\}) = \int_t^\infty h_j(u)S_T(u)du, \quad (5.15)$$

la cual nos describe la probabilidad de eventualmente morir de causa \mathcal{C}_j en un tiempo mayor que t .

Tsiatis (1975) demostró que para cualquier función conjunta de tiempos de falla latentes caracterizada por (5.1), se tiene que

$$\frac{dP_j(t)}{dt} = \frac{\partial S_{1\dots m}(y_1, \dots, y_m)}{\partial y_j} \Big|_{y_1 = \dots = y_m = t}, \quad (5.16)$$

Observación 5.2. El resultado de Tsiatis (1975) en (5.16) indica que cualquier función de decremento múltiple $S_{12\dots m}(y_1, y_2, \dots, y_m)$ determina de manera única al conjunto de funciones de sub-supervivencia $\{P_j(t) : j = 1, 2, \dots, m\}$.

Así, la *función de sub-densidad* asociada con la función de sub-supervivencia $P_j(t)$, cuando existe, puede ser calculada como

$$f_j(t) = \frac{\partial S_{1\dots m}(y_1, \dots, y_m)}{\partial y_j} \Big|_{y_1=\dots=y_m=t},$$

la cual representa el riesgo incondicional de que un individuo experimente una falla al tiempo t de causa \mathcal{C}_j . De esta manera se obtiene la usual relación

$$h_j(t) = -\frac{\partial \log S_{1\dots m}(y_1, \dots, y_m)}{\partial y_j} \Big|_{y_1=\dots=y_m=t} = \frac{f_j(t)}{S_{\mathcal{T}}(t)}. \quad (5.17)$$

Nótese que despejando $f_j(t)$ en (5.17) y tomando la suma sobre $j = 1, 2, \dots, m$, se obtiene la función de densidad del tiempo de falla T , eso es,

$$\sum_{j=1}^m f_j(t) = S_{\mathcal{T}}(t) \sum_{j=1}^m h_j(t) = S_{\mathcal{T}}(t) h_{\mathcal{T}}(t) = f_{\mathcal{T}}(t) = -dS_{\mathcal{T}}(t)/dt.$$

Una vez definidas esas cantidades, se tiene que

- La *distribución marginal* de T está dada por la **función de distribución acumulada**:

$$\begin{aligned} \sum_{j=1}^m Q_j(t) &= \sum_{j=1}^m \int_0^t h_j(u) S_{\mathcal{T}}(u) du = \int_0^t \sum_{j=1}^m h_j(u) S_{\mathcal{T}}(u) du \\ &= \int_0^t h_{\mathcal{T}}(u) S_{\mathcal{T}}(u) du = -\int_0^t dS_{\mathcal{T}}(u) \\ &= 1 - S_{\mathcal{T}}(t) = F_{\mathcal{T}}(t) \end{aligned} \quad (5.18)$$

donde

$$S_{\mathcal{T}}(t) = \Pr(T > t) = \sum_{j=1}^m P_j(t) \quad (5.19)$$

es la FDS marginal de T .

- La proporción esperada de fallas por causa \mathcal{C}_j , denotada por π_j ($j \in \mathcal{J}$), está dada por la *distribución marginal de J* ,

$$\begin{aligned} \pi_j &= \Pr(J = j) = \Pr(\{T < \infty\} \cap \{J = j\}) \\ &= \int_0^{\infty} h_j(u) S_{\mathcal{T}}(u) du \\ &= Q_j(\infty) = P_j(0) \end{aligned} \quad (5.20)$$

con

$$\pi_j > 0 \quad \text{y} \quad \sum_{j=1}^m \pi_j = 1. \quad (5.21)$$

Observéese que las sub-distribuciones antes mencionadas se relacionan mediante

$$Q_j(t) + P_j(t) = \pi_j \quad \text{para } j = 1, 2, \dots, m.$$

Por otro lado, de (5.18) se tiene que

$$F_{\mathcal{T}}(t) + S_{\mathcal{T}}(t) = 1,$$

De lo anterior se concluye que las funciones $Q_j(t)$ y $P_j(t)$ no son distribuciones *propias*, ya que cada π_j es menor que uno. Sin embargo, se pueden definir *distribuciones condicionales propias* en términos de esas sub-distribuciones. Por ejemplo, la FDS condicional propia (en presencia de todas las causas) asociada con la causa \mathcal{C}_j , y denotada por $S_j^*(t)$, es

$$\begin{aligned} S_j^*(t) &= \Pr(T > t \mid J = j) = \frac{P_j(t)}{P_j(0)} = \frac{1}{\pi_j} P_j(t) \\ &= \frac{1}{\pi_j} \int_t^{\infty} h_j(u) S_{\mathcal{T}}(u) du \quad j = 1, 2, \dots, m. \end{aligned} \quad (5.22)$$

Entonces

$$F_j^*(t) = 1 - S_j^*(t) = \Pr(T \leq t \mid J = j) \quad (5.23)$$

denota la probabilidad condicional de presentar una falla antes del tiempo t y en presencia de todas las causas, dado que el individuo tendrá una falla por causa \mathcal{C}_j .

Para obtener una expresión de la probabilidad de tener una falla por causa j después del tiempo t , dado que se ha sobrevivido a todas las causas de falla hasta ese tiempo, procedemos como sigue: la correspondiente función de distribución de probabilidades de (5.22) es

$$f_j^*(t) = -\frac{dS_j^*(t)}{dt} = -\frac{1}{\pi_j} \frac{dP_j(t)}{dt} = \frac{1}{\pi_j} h_j(t) S_{\mathcal{T}}(t). \quad (5.24)$$

Por eso

$$h_j(t) = -\frac{1}{S_{\mathcal{T}}(t)} \frac{dP_j(t)}{dt} = -\frac{dP_j(t)/dt}{\sum_{k=1}^m P_k(t)} \quad (5.25)$$

Por otro lado, la función de riesgo de $S_j^*(t)$ es

$$\begin{aligned} h_j^*(t) &= \lim_{\Delta t \rightarrow 0^+} \frac{1}{\Delta t} \Pr(t < T \leq t + \Delta t \mid T > t, J = j) \\ &= -\frac{d \log S_j^*(t)}{dt} = \frac{f_j^*(t)}{S_j^*(t)} \\ &= \frac{h_j(t) S_{\mathcal{T}}(t)}{\int_t^{\infty} h_j(u) S_{\mathcal{T}}(u) du} = -\frac{1}{P_j(t)} \frac{dP_j(t)}{dt} \end{aligned} \quad (5.26)$$

Tomando la razón de (5.25) y (5.26), obtenemos

$$\frac{h_j(t)}{h_j^*(t)} = \frac{P_j(t)}{S_T(t)} = \frac{P_j(t)}{\sum_{j=1}^m P_j(t)} = \Pr(J = j \mid T > t) = \pi_j(t), \quad (5.27)$$

Esta representa la probabilidad condicional de que un individuo falle por causa \mathcal{C}_j después del tiempo t , dado que ha sobrevivido a todas las causas hasta ese tiempo.

Otro enfoque para la distribución de T y J sería especificar las distribuciones $S_j^*(t)$ en (5.22) paramétricamente, y tratar las π_j 's como parámetros adicionales. Obviamente, estos modelos tiene más parámetros que los obtenidos a partir de (5.13).

Ejemplo 5.2. Considere el caso de no covariables presentes y supongamos una distribución (reparametrizada) Weibull(α, γ) para la función de riesgo de $S_j^*(t)$, eso es,

$$h_j^*(t) = \frac{-dP_j(t)/dt}{P_j(t)} = \alpha_j \gamma_j^{\alpha_j} t^{\alpha_j-1} \quad j = 1, 2, \dots, m.$$

Entonces, la función de supervivencia condicional (propia) asociada a la causa \mathcal{C}_j es

$$S_j^*(t) = \frac{1}{\pi_j} P_j(t) = \exp \left\{ - \int_0^t h_j^*(u) du \right\} = \exp \left\{ - (\gamma_j t)^{\alpha_j} \right\}. \quad (5.28)$$

y la distribución de probabilidad del tiempo de falla por causa j es

$$f_j^*(t) = - \frac{1}{\pi_j} \frac{dP_j(t)}{dt} = h_j^*(t) S_j^*(t) \quad j = 1, 2, \dots, m.$$

De (5.28) se sigue que la función de supervivencia marginal de T esta dada por

$$S_T(t) = \sum_{j=1}^m P_j(t) = \sum_{j=1}^m \pi_j S_j^*(t) = \sum_{j=1}^m \pi_j \exp \left\{ - (\gamma_j t)^{\alpha_j} \right\} \quad (5.29)$$

y la función de riesgo por la j -ésima causa es

$$h_j(t) = \frac{f_j(t)}{S_T(t)} = \frac{\pi_j f_j^*(t)}{\sum_{\ell=1}^m \pi_\ell S_\ell^*(t)} \quad j = 1, 2, \dots, m. \quad (5.30)$$

□

De la expresión (5.30) se ve que las funciones de riesgo de causa-específica tienen una forma más compleja en contraste con los modelos basados en (5.13). Sin embargo, este modelo se puede interpretar como una mezcla de distribuciones Weibull (Subsección 2.2.1): las proporciones mezcladas son las π_j 's y, para un individuo en el j -ésimo grupo, la distribución del tiempo de falla es Weibull con parámetro de forma α_j y parámetro de escala γ_j .

5.2.4. Riesgos Competitivos Independientes

Una situación de particular interés en la literatura de riesgos competitivos es la situación en que los tiempos de falla latentes Y_1, Y_2, \dots, Y_m son *mutuamente independientes*. Una primera consecuencia de tal suposición es que la función de supervivencia conjunta de los tiempos de falla latentes está determinada por las *funciones de supervivencia marginales*⁵ $S_j(y_j) = \Pr(Y_j > y_j)$, $j = 1, 2, \dots, m$, como

$$S_{1\dots m}(y_1, \dots, y_m) = \prod_{j=1}^m S_j(y_j), \quad \forall y_j \in (0, \infty). \quad (5.31)$$

La función de riesgo marginal de Y_j , también llamada *función de riesgo neta*, se define como

$$\lambda_j(y) = \lim_{dy \rightarrow 0^+} \frac{1}{dy} \Pr\{y < Y_j \leq y + dy \mid Y_j > y\} = -\frac{d \log S_j(y)}{dy}$$

y representa la tasa de mortalidad instantánea de presentar una falla al tiempo t en condiciones hipotéticas donde los modos de falla, salvo j , han sido eliminados, condicionada a que no ha ocurrido una falla por causa \mathcal{C}_j hasta ese tiempo. Nótese que sin suposiciones adicionales, y en general,

$$h_j(t) \neq \lambda_j(t) \quad \text{y} \quad h_{\tau}(t) \neq \sum_{j=1}^m \lambda_j(t),$$

donde $h_j(t)$ es la función de riesgo de causa-específica para el tipo de falla j y $h_{\tau}(t)$ es la función de riesgo global.

El siguiente teorema da algunas otras implicaciones como resultado de suponer riesgos competitivos independientes (*Gail, 1975; Crowder, 2001*).

Teorema 5.2.3. *Las implicaciones (i) \Rightarrow (ii) \Rightarrow (iii) \Rightarrow (iv) se sostienen para las siguientes proposiciones:*

- (i) *se tienen riesgos competitivos independientes;*
- (ii) *Suposición de Makeham⁶: $h_j(t) = \lambda_j(t)$ para cada j y $t > 0$;*
- (iii) *el conjunto de funciones de sub-supervivencia $P_j(t)$ determinan el conjunto de FDS marginales $S_j(t)$, explícitamente,*

$$S_j(t) = \Pr(Y_j > t) = \exp \left\{ - \int_0^t h_j(u) du \right\}; \quad (5.32)$$

$$(iv) \quad S_{1\dots m}(t, \dots, t) = \prod_{j=1}^m S_j(t).$$

⁵ La función $S_j(t) = \Pr\{Y_j > t\}$ se puede pensar como la FDS de una población que esta bajo condiciones hipotéticas en donde solamente la falla de tipo j está actuando.

⁶ Gail (1975) le llamó a ésta condición la "suposición de Makeham".

Demostración. Bajo la proposición (i), aplicando logaritmo en (5.31), se obtiene

$$\log S_{1\dots m}(y_1, \dots, y_m) = \sum_{j=1}^m \log S_j(y_j).$$

Luego, diferenciando con respecto a y_j en ambos lados de la igualdad anterior y evaluando en $y_1 = \dots = y_m = t$, se tiene que

$$\left[-\frac{\partial \log S_{1\dots m}(y_1, \dots, y_m)}{\partial y_j} \right]_{y_1=\dots=y_m=t} = \frac{-d \log S_j(t)}{dt}, \quad (5.33)$$

para $j = 1, 2, \dots, m$; recordando la definición de $h_j(t)$ y $\lambda_j(t)$, respectivamente, (5.33) es precisamente (ii).

Por un lado, de (5.16) y (5.19) se tiene

$$h_j(t) = \frac{f_j(t)}{S_{\mathbb{T}}(t)} = \frac{dP_j(t)/dt}{\sum_{\ell=1}^m P_{\ell}(t)}.$$

Por otro lado, $\lambda_j(t) = -d \log S_j(t)/dt$. Entonces, bajo la suposición de (ii) se obtiene

$$\frac{dP_j(t)/dt}{\sum_{\ell=1}^m P_{\ell}(t)} = \frac{-d \log S_j(t)}{dt},$$

o de manera equivalente,

$$S_j(t) = \exp \left\{ -\int_0^t \left[\frac{dP_j(u)/du}{\sum_{\ell=1}^m P_{\ell}(u)} \right] du \right\}$$

Por eso, el conjunto de FDS marginales $\{S_j(t) : j = 1, 2, \dots, m\}$ queda determinado a partir del conjunto de funciones de sub-supervivencia $\{P_j(t) : j = 1, 2, \dots, m\}$, y (iii) se mantiene. Luego, integrando la Ecuación 5.33 con respecto a t se obtiene (5.32). De la Ecuación 5.32,

$$\prod_{j=1}^m S_j(t) = \exp \left\{ -\sum_{j=1}^m \int_0^t h_j(u) du \right\} = \exp \left\{ -\int_0^t h_{\mathbb{T}}(u) du \right\} = S_{\mathbb{T}}(t),$$

que es (iv). ■

Observación 5.3. Retomando la proposición (ii) bajo (i) del teorema anterior, la función de riesgo de causa-específica (i.e. en presencia de todas las causas), $h_j(t)$, y la función de riesgo marginal (solamente en presencia de la causa \mathcal{C}_j), $\lambda_j(t)$, son la misma para cada j y $t > 0$. Sin embargo, esto no se cumple para las funciones de supervivencia crudas y netas, debido a que la FDS (propia) en presencia de todas las causas es como se dio en (5.22), que es,

$$S_j^*(t) = \frac{1}{\pi} \int_t^{\infty} h_j(u) S_{\mathbb{T}}(u) du,$$

mientras que la FDS marginal es

$$S_j(t) = \exp \left\{ - \int_0^t \lambda_j(u) du \right\} = \exp \left\{ - \int_0^t h_j(u) du \right\}.$$

Claramente,

$$S_j^*(t) \neq S_j(t).$$

Observación 5.4. El significado en la práctica de la proposición (iii) bajo (ii) es que las distribuciones marginales se pueden estimar de una manera consistente a partir del conjunto de datos de riesgos competitivos.

Observación 5.5. El Teorema 5.2.3 nos indica que (i) es más restrictiva que (ii), o que (iv) es una condición algo más débil que (i), ya que, bajo la suposición de independencia estadística entre los tiempos de falla latentes Y_j 's se tiene

$$S_{1\dots m}(y_1, \dots, y_m) = \prod_{j=1}^m S_j(y_j), \quad \forall (y_1, y_2, \dots, y_m) \in \mathbb{R}_+^m. \quad (5.34)$$

En particular, para $y_1 = \dots = y_m = t$, la Ecuación 5.34 implica la proposición (ii) y de aquí (iv). Un contraejemplo ha sido dado por Hakulinen y Rahiala (1977).

Para concluir esta parte, nótese que del Teorema 5.2.2 (de Factorización), la distribución asociada con $h_j(t)$ es

$$G_j(t) = \exp \left\{ - \int_0^t h_j(u) du \right\}, \quad \forall t \in \mathbb{R}_+.$$

Entonces, en el caso de riesgos independientes, la $G_j(t)$ es de hecho la FDS marginal de Y_j . Así

$$S_{1\dots m}(t, \dots, t) = \prod_{j=1}^m G_j(t) = \prod_{j=1}^m S_j(t), \quad \forall t \in \mathbb{R}_+.$$

5.3. Métodos Estadísticos de Inferencia

De manera similar como en la Sección 1.2, pero ahora en un escenario de Riesgos Competitivos donde existen $m > 1$ distintos tipos o modos de falla, sea T una v.a.s y C su variable de censura. Sea $\tilde{T} = \min(T, C)$ y $\delta = \mathbf{1}\{T \leq C\}$ su indicadora de censura. Si \tilde{T} es un tiempo de supervivencia censurado, el tipo de falla $J \in \mathcal{J}$ no es observado.

Supongamos que se tiene un conjunto de n observaciones correspondiente a una muestra aleatoria, posiblemente censurada por la derecha, de individuos provenientes de una población en la que existen m tipos o modos de falla. Sea \mathbf{X}_i el p -vector de variables explicativas correspondiente al i -ésimo individuo en la muestra. Entonces, si el i -ésimo tiempo de supervivencia está censurado en $C_i = t_i$, se observa el vector $(\tilde{T}_i = t_i, \delta_i = 0, \mathbf{x}_i) \equiv (T_i > t_i, \delta_i = 0, \mathbf{x}_i)$, y su contribución a la verosimilitud es

$S_T(t_i; \mathbf{x}_i) = \Pr(T_i > t_i \mid \mathbf{x}_i)$. Por otro lado, el i -ésimo tiempo de supervivencia no-censurado $T_i = t_i$, cuya terna observada es $(\tilde{T}_i = t_i, \delta_i = 1, \mathbf{x}_i, J_i)$, contribuye a la verosimilitud con $f_{J_i}(t_i; \mathbf{x}_i) = h_{J_i}(t_i; \mathbf{x}_i)S_T(t_i; \mathbf{x}_i)$. Por lo tanto, bajo un mecanismo de censura no-informativo, la función de verosimilitud es proporcional a

$$\begin{aligned} L &= \prod_{i=1}^n \left\{ [h_{J_i}(t_i; \mathbf{x}_i)S_T(t_i; \mathbf{x}_i)]^{\delta_i} [S_T(t_i; \mathbf{x}_i)]^{1-\delta_i} \right\} \\ &= \prod_{i=1}^n \left\{ [h_{J_i}(t_i; \mathbf{x}_i)]^{\delta_i} S_T(t_i; \mathbf{x}_i) \right\}. \end{aligned} \quad (5.35)$$

La verosimilitud anterior se puede expresar de una forma más conveniente. Para esto, note del Teorema 5.2.2 que

$$\begin{aligned} S_T(t_i) &= \exp \left\{ - \int_0^{t_i} \sum_{j=1}^m h_j(u) du \right\} \\ &= \exp \left\{ - \sum_{j=1}^m H_j(t_i) \right\} \\ &= \prod_{k=1}^m G_k(t_i), \end{aligned}$$

donde cada $G_k(t_i) = \exp\{-H_k(t_i)\}$ ($k = 1, 2, \dots, m$) posee las propiedades matemáticas de una FDS continua en el caso de un solo tipo de falla, más no representan la FDS de ninguna variable aleatoria específica. No obstante, definiendo la variable indicadora de la falla de tipo j para el i -ésimo individuo como $\delta_{ij} = \mathbf{1}\{J_i = j, \delta_i = 1\}$, la verosimilitud (5.35) se puede reescribir como

$$\begin{aligned} L &= \prod_{i=1}^n \left\{ h_{J_i}(t_i; \mathbf{x}_i)^{\delta_i} S_T(t_i; \mathbf{x}_i) \right\} = \prod_{i=1}^n \left\{ \left[\prod_{j=1}^m h_j(t_i; \mathbf{x}_i)^{\delta_{ij}} \right] S_T(t_i; \mathbf{x}_i) \right\} \\ &= \prod_{i=1}^n \left\{ \left[\prod_{j=1}^m h_j(t_i; \mathbf{x}_i)^{\delta_{ij}} \right] \left[\prod_{k=1}^m G_k(t_i; \mathbf{x}_i) \right] \right\} \\ &= \prod_{i=1}^n \left\{ \prod_{j=1}^m \left[h_j(t_i; \mathbf{x}_i)^{\delta_{ij}} G_j(t_i; \mathbf{x}_i) \right] \right\} \\ &= \prod_{i=1}^n \left\{ \prod_{j=1}^m [g_j(t_i; \mathbf{x}_i)]^{\delta_{ij}} [G_j(t_i; \mathbf{x}_i)]^{1-\delta_{ij}} \right\}, \end{aligned} \quad (5.36)$$

donde

$$g_j(t; \mathbf{x}_i) = h_j(t; \mathbf{x}_i)G_j(t; \mathbf{x}_i) = -\frac{dG_j(t; \mathbf{x}_i)}{dt}.$$

Nótese que la verosimilitud L en (5.36) está completamente en términos de las $h_j(t)$'s y debido a la forma de L , esas funciones son estimables en base a la información

de los datos $(\tilde{T}_i, \delta_{ij}, \mathbf{x}_i, J_i)$ $i = 1, 2, \dots, n$. Más aún, al invertir el orden de los productos en (5.36), vemos que L toma la forma de un producto con m factores, uno por cada tipo de falla. A pesar de que $G_j(t_i; \mathbf{x}_i)$ no tienen una interpretación probabilística dentro de los modelos de riesgos competitivos, a menos que se consideren suposiciones adicionales, si en el j -ésimo factor

$$L_j = \prod_{i=1}^n g_j(t_i; \mathbf{x}_i)^{\delta_{ij}} G_j(t_i; \mathbf{x}_i)^{1-\delta_{ij}}, \quad (5.37)$$

interpretamos a $g_j(\cdot)$ y a $G_j(\cdot)$ como la función de densidad y función de supervivencia, respectivamente, para cierta distribución de tiempos de supervivencia, la expresión (5.37) tiene la misma forma de una función de verosimilitud ordinaria. Esta representa una muestra aleatoria en la que todos los tipos de fallas, salvo el modo de falla j , son considerados como censurados en el tiempo de falla correspondiente. Así, se pueden aplicar las técnicas de datos de supervivencia vistas en Capítulos anteriores para hacer inferencia sobre las $h_j(t)$'s (y por eso igualmente para las $g_j(t)$'s y $G_j(t)$'s) a partir de los datos $(\tilde{T}_i, \delta_i, \mathbf{x}_i, J_i)$. En particular, si se supone un *modelo paramétrico* $h_j(t; \mathbf{x}_i, \boldsymbol{\theta}_j)$ para la función de riesgo de tipo j ($j \in \mathcal{J}$), de tal manera que no se tengan parámetros en común entre esas funciones, la inferencia para $\boldsymbol{\theta}_j$ puede estar basada en (5.37).

También se pueden implementar las técnicas de estimación semi-paramétricas y no-paramétricas vistas en capítulos anteriores para modelos basados en las $h_j(t)$'s. Por ejemplo, métodos tales como el Estimador Kaplan-Meier y Nelson-Aalen pueden ser generalizados para datos de riesgos competitivos. Para eso, considere el conjunto de datos independientes censurados por la derecha de un modelo de riesgos competitivos homogéneo (no covariables). Ignorando la asociación entre los tipos de falla, supongamos que se tienen r distintos tiempos de falla ordenados (i.e. todos los distintos tiempos de falla combinados) y denotados por $t_1 < t_2 < \dots < t_r$. Sea d_{ji} el número de fallas de tipo j al tiempo t_i y n_i el número de individuos en riesgo al tiempo t_i , $i = 1, 2, \dots, r$; $j = 1, 2, \dots, m$. Argumentos similares a los que se dan para el estimador-KM conducen a un modelo de riesgos competitivos discreto⁷, con función de riesgo discreta $h_j(t_i) \equiv h_{ji}$ para el j -ésimo tipo de falla al tiempo t_i . Entonces, la función de verosimilitud, bajo un modelo discreto de riesgos competitivos, se puede expresar como (ver Kalbfleisch y Prentice, 2002)

$$L = \prod_{i=1}^n \left[\prod_{j=1}^m \left(h_{ji}^{d_{ij}} \right) (1 - h_i)^{n_i - d_i} \right], \quad (5.38)$$

donde $d_i = \sum_{j=1}^m d_{ij}$ es el número de fallas al tiempo t_i y $h_i \equiv h_{\mathcal{T}}(t_i) = \sum_{j=1}^m h_{ij}$ es la función de riesgo discreta global (i.e de cualquier causa) al tiempo t_i . El estimador de máxima verosimilitud de h_{ij} se obtiene al maximizar (5.38), dando $\hat{h}_{ij} = d_{ij}/n_i$. Entonces, el estimador Nelson-Aalen de la función de riesgo acumulada para la falla

⁷ Ver [5] para más detalles sobre los modelo de riesgos competitivos discretos.

de tipo j es

$$\widehat{H}_j(t) = \sum_{i|t_i \leq t} \frac{d_{ij}}{n_i} \quad t > 0. \quad (5.39)$$

para $j = 1, 2, \dots, m$. De esta manera, el Estimador Nelson-Aalen para la función de riesgo global $H_T(t)$ esta dado por

$$\widehat{H}_T(t) = \sum_{j=1}^m \widehat{H}_j(t) \quad t > 0, \quad (5.40)$$

y el estimadro Kaplan-Meier de la función de supervivencia global, $S_T(t)$,

$$\begin{aligned} \widehat{S}_T(t) &= \prod_{\ell|t_\ell \leq t} \left[1 - \widehat{h}_T(t_\ell) \right] \\ &= \prod_{\ell|t_\ell \leq t} \left[1 - \sum_{j=1}^m \widehat{h}_j(t_\ell) \right] \\ &= \prod_{\ell|t_\ell \leq t} \left[1 - \sum_{j=1}^m \frac{d_{\ell j}}{n_\ell} \right] \\ &= \prod_{\ell|t_\ell \leq t} \left[1 - \frac{d_{\ell \cdot}}{n_\ell} \right], \end{aligned}$$

para $t > 0$, donde $d_{\ell \cdot} = \sum_{j=1}^m d_{\ell j}$ denota el número total de fallas de todas las causa al tiempo t_ℓ . El estimador Nelson-Aalen en (5.40) proporciona un estimador alternativo de $S_T(t)$, que es

$$\widehat{S}_T(t) = \exp\{-\widehat{H}_T(t)\} = \exp\left\{-\sum_{j=1}^m \widehat{H}_j(t)\right\}.$$

Así, un estimador razonable para la función de sub-distribución o de incidencia acumulada de causa-específica $Q_j(t)$ en (5.14) es

$$\widehat{Q}_j(t) = \sum_{i|t_i \leq t} \widehat{S}_T(t_i^-) \frac{d_{ij}}{n_i} \quad j = 1, 2, \dots, m. \quad (5.41)$$

donde $\widehat{S}_T(t_i^-) = \lim_{\varepsilon \rightarrow 0^+} \widehat{S}_T(t_i - \varepsilon) = \widehat{S}_T(t_{i-1})$.

Sea t_r el último tiempo de falla (i.e. observación no-censurada). De (5.20), la probabilidad de finalmente presentar una falla por causa j se puede estimar como

$$\widehat{\pi}_j = \widehat{Q}_j(t_r) \quad j = 1, 2, \dots, m.$$

De esta forma, el EMV de $F_j^*(t)$ en (5.23) es

$$\widehat{F}_j^*(t) = \frac{\widehat{Q}_j(t)}{\widehat{\pi}_j},$$

que denota la probabilidad de falla de un individuo por causa j , y en presencia de todos los otros tipos de falla, condicionada a que presentará una falla de esa causa en t_r . De manera similar como en (5.18), se tiene

$$\sum_{j=1}^m \hat{Q}_j(t) = 1 - \hat{S}_T(t).$$

Entonces

$$\sum_{j=1}^m \hat{\pi}_j \leq 1, \quad (5.42)$$

y se tiene la igualdad cuando el tiempo observado más grande en los datos, digamos τ , es un tiempo de falla. Cuando se desea obtener estimadores que sumen 1 en (5.42), se puede reemplazar $\hat{\pi}_j$ con

$$\hat{\pi}'_j = \frac{\hat{\pi}_j}{\sum_{l=1}^m \hat{\pi}_l},$$

Ese ajuste se basa en la suposición de que los $\hat{\pi}'_j$ ($j = 1, 2, \dots, m$) no cambiarán después del último tiempo de falla observado. Gaynor *et.al* (1993) da un ejemplo en el que no se cumple tal suposición. Para τ grande, una alternativa a $\hat{\pi}_j$ es el estimador de $\Pr(J = j \mid T \leq t)$, el cual puede obtenerse a partir del estimador de (5.27).

Se puede demostrar ([11]) que en ausencia de censura en los datos, aunque en general no es así, las π'_j s se estiman como

$$\hat{\pi}_j = \frac{d_{.j}}{n} \quad j = 1, 2, \dots, m,$$

que es la proporción observada de individuos que presentan una falla por causa j , n es el tamaño total de la muestra y $d_{.j} = \sum_{\ell=1}^r d_{\ell j}$ es el número total de fallas observados en la muestra por causa j . Así,

$$\sum_{j=1}^m \hat{\pi}_j = \frac{\sum_{j=1}^m d_{.j}}{n} = 1,$$

y

$$\hat{F}_j^*(t) = \sum_{i|t_i \leq t} \frac{d_{ij}}{d_{.j}},$$

que es la *función de sub-distribución empírica* para la falla de tipo j .

El Estimador Kaplan-Meier y la Función de Incidencia Acumulada de causa-específica.

Suponga un modelo de riesgos competitivos con $m > 1$ tipos de falla. Sea Y_j el tiempo de falla latente correspondiente a la j -ésima causa de falla, para $j = 1, 2, \dots, m$. Bajo la suposición **A1** y por el Teorema 5.2.2, en cualquiera de los dos casos, ya sea

que se tengan riesgos competitivos dependientes o independientes, la FDS conjunta de los Y_j 's se puede escribir como

$$S_{1\dots m}(y_1, \dots, y_m) = \prod_{j=1}^m G_j(y_j)$$

donde

$$G_j(y) = \exp \left\{ - \int_0^y h_j(u) du \right\} = \exp \left\{ - H_j(y) \right\},$$

y $h_j(\cdot)$ es la función de riesgo de la j -ésima causa. Note que

$$\begin{aligned} Q_j(t) &= \int_0^t S_T(u) h_j(u) du = \int_0^t \exp \left\{ - \sum_{l=1}^m H_l(u) \right\} dH_j(u) \\ &\leq \int_0^t \exp \left\{ - H_j(u) \right\} dH_j(u) = 1 - G_j(t) \end{aligned} \quad (5.43)$$

para toda t . De hecho Gaynor et. al. (1993) establecen que la igualdad entre $Q_j(t)$ y $1 - G_j(t)$ se da solamente cuando se reúnen dos suposiciones hipotéticas:

- (1) *la falla debido a otras causas a sido eliminado; eso es, $h_l(t) = 0$ para $l \neq j$;*
- (2) *la función de riesgo de causa-específica de interés, $h_j(t)$, permanece sin cambios dada la suposición (1).*

El método Kaplan-Meier (KM) es una herramienta ampliamente usada para estimar la función de supervivencia de datos que presentan un solo tipo falla (análisis de supervivencia univariado), debido a la simple interpretación que tiene y a la amplia variedad de paquetes estadísticos en los que está implementado. Por otro lado, la expresión en (5.37) tiene la forma matemática de la función de verosimilitud para datos censurados de una distribución de tiempos de falla con función de supervivencia, densidad y función de riesgo $G_j(t)$, $g_j(t)$ y $h_j(t)$, respectivamente. Consecuentemente, $G_j(t)$ podría ser estimada usando el estimador Kaplan-Meier basado en los datos (t_i, d_{ji}) , $i=1,2,\dots,r$, donde los tiempos de falla observados por las m causas, salvo la causa \mathcal{C}_j , son considerados como observaciones censuradas. De esta manera, el estimador-KM para la causa \mathcal{C}_j de $G_j(t)$ es

$$\widehat{G}_j^{\text{KM}}(t) = \prod_{i | t_i \leq t} \left(1 - \frac{d_{ij}}{n_i} \right),$$

donde d_{ij} denota el número de individuos que presentan una falla de modo j en t_i , y n_i denota el número de individuos en riesgo de falla al tiempo t_i .

Comunmente $1 - \widehat{G}_j^{\text{KM}}(t)$ ha sido erróneamente empleado para estimar a $Q_j(t)$ en un escenario en el que existe más de un tipo de falla (riesgos competitivos) y bajo la clásica suposición de riesgos independientes, el cual es una de las hipótesis en la que se basa el método-KM. Esto conduce a una estimación inflada de la proporción

de individuos quienes están en riesgo de falla al tiempo t ; esto a su vez causa que $1 - \widehat{G}_j^{\text{KM}}(t)$ sobreestime a $Q_j(t)$, es decir,

$$1 - \widehat{G}_j^{\text{KM}}(t) \geq \widehat{Q}_j(t), \quad \forall t > 0. \quad (5.44)$$

La expresión (5.44), la cual es análoga a la desigualdad en (5.43), puede ser verificada por inducción matemática. En el caso de una sola causa de falla presente ($m = 1$) se cumple que

$$1 - \widehat{G}_1^{\text{KM}}(t) = \widehat{Q}_1(t).$$

Varios autores han criticado el uso erróneo del estimador KM en un escenario de riesgos competitivos (ver [11], [13], [14] y [30]).

5.4. Identificabilidad y la Función de Decremento Múltiple

Ya se mencionó que el enfoque clásico de la teoría de riesgos competitivos es considerar tiempos de falla latentes Y_1, \dots, Y_m correspondientes a las m tipos de falla. Bajo la suposición **A1**, lo que se observa (no-censura) es el tiempo de falla real $T = \min(Y_1, \dots, Y_m)$ y el tipo de falla J , donde $Y_J = T$. Sea

$$S_{1\dots m}(y_1, \dots, y_m)$$

la función de decremento múltiple (FDM) o función de supervivencia conjunta de los tiempos de falla latentes. Entonces, los problemas mencionados en la Sección 5.1 se pueden plantear en términos de los tiempos de falla latentes y de su FDM como sigue:

- (1) La clásica suposición de independencia entre los $m > 1$ riesgos competitivos se identifica con la independencia estadística entre los m tiempos de falla latentes.
- (2) La distribución de Y_j bajo condiciones hipotéticas en las que se han eliminado todos los modos de falla, salvo j , se supone como la *distribución marginal* de Y_j , $S_j(y_j) = \Pr(Y_j > y_j)$.
- (3) Por lo regular, el interés de un análisis de riesgos competitivos está en la función de supervivencia conjunta y distribuciones marginales de Y_1, Y_2, \dots, Y_m . Sin embargo, el problema al que uno se enfrenta al considerar un enfoque de riesgos competitivos vía un modelo de tiempos de falla latentes es que, una vez que el tiempo de falla por causa C_j es observado, es decir, se tiene el *mínimo identificable* (T, J) , surge una falta de significado físico para los restantes tiempos de falla latentes en términos de cantidades estimables (Sección 5.3). Entonces, no será sorpresa toparse con problemas al querer *identificar* a la función de supervivencia conjunta de los tiempos de falla latentes, $S_{1\dots m}(\cdot)$ y las funciones de supervivencia marginales, $S_j(y_j)$ ($j = 1, 2, \dots, m$), basándose en los datos (T, J) y sin la introducción de suposiciones adicionales fuertes, tal como la independencia estadística entre los tiempos de falla latentes. Este hecho es llamado el problema de *no-identificabilidad*, basándose en los datos (T, J) , de la distribución de Y_1, Y_2, \dots, Y_m .

La distribución del par (T, J) está completamente especificada por las funciones de sub-supervivencia para la causa \mathcal{C}_j dadas en (5.15). Ya que esas funciones son estimables a partir del conjunto de datos (t_i, δ_i, J_i) , para $i = 1, 2, \dots, n$, se supondrá que el conjunto $\{P_j(t); j = 1, 2, \dots, m\}$ es dado, y (3) se planteará en términos de este. Por otra parte, cuando se supone una función de decremento múltiple $S_{1\dots m}(y_1, \dots, y_m)$, esta determina de manera única al conjunto de funciones de sub-supervivencia (Tsiatis, 1975). Otro enfoque alternativo para el clásico problema de riesgos competitivos en (3) se puede dar en términos de las funciones de riesgo de causa-específica (ver [16]).

El problema de identificabilidad en (3) se refiere a que el conjunto de $P_j(t)$'s no proporciona suficiente información para determinar, en general, la FDS conjunta de los tiempos de falla latentes. Eso es, existen ambos, un modelo con riesgo independientes y uno o más modelos con riesgos dependientes para Y_1, Y_2, \dots, Y_m que dan lugar al mismo conjunto de funciones de sub-supervivencia.

Para tener más claro este punto, considere dos vectores de v.a.s. (Y_1, Y_2, \dots, Y_m) y $(Y'_1, Y'_2, \dots, Y'_m)$, con FDS conjunta $S_{1\dots m}(y_1, \dots, y_m)$ y $S'_{1\dots m}(y_1, \dots, y_m)$, respectivamente. Sea

$$T = \min(Y_1, \dots, Y_m) \quad \text{y} \quad T' = \min(Y'_1, \dots, Y'_m).$$

Definición 5.2. *Dos modelos de supervivencia son equivalentes si*

$$S_T(t) = S_{T'}(t) \quad \text{y} \quad P_j(t) = P'_j(t). \quad (5.45)$$

Teorema 5.4.1 (Tsiatis, 1975). *Supongase que el conjunto de funciones de sub-supervivencia $P_j(t)$ ($j = 1, 2, \dots, m$) está dado por algún modelo con riesgos dependientes. Entonces, existe un único modelo con riesgos independientes produciendo idénticas $P_j(t)$'s, el cual está definido por*

$$S'_{1\dots m}(y_1, \dots, y_m) = \prod_{j=1}^m S'_j(y_j),$$

donde

$$S'_j(t) = \exp \left\{ - \int_0^t h_j(u) du \right\} \quad (5.46)$$

y la función de riesgo de causa-específica $h_j(u)$ se deriva a partir las $P_j(t)$ dadas.

Demostración. Ver Tsiatis [34], Crowder [5]. ■

El Teorema 5.4.1 solamente establece que cada modelo con riesgos dependientes es equivalente a un único modelo con riesgos independientes. Entonces, los datos de tipo (T, J) no permiten distinguir a un modelo con riesgos independientes y a uno o más modelos con riesgos dependientes, ya que ambos dan lugar al mismo conjunto de funciones de sub-supervivencia $P_j(t)$, $j=1, 2, \dots, m$.

Una manera de lidiar con el problema de no-identificabilidad es la clásica suposición de **riesgos independientes**, ya que, siguiendo el Teorema 5.4.1, las FDS

marginales de los Y_j ahora pueden ser calculadas a partir del conjunto de funciones de sub-supervivencia $\{P_j(t), j = 1, 2, \dots, m\}$. Sin embargo, esta restrictiva suposición ha sido fuertemente criticada por varios investigadores, entre ellos Prentice *et al.* (1978), debido a la falta de interpretación física de los tiempos de falla latentes "no observados" y a la hipótesis de independencia entre los riesgos de falla.

Otra forma de sobrellevar esta situación es derivar **cotas** para las funciones de supervivencia marginales $S_j(y_j)$, en términos de las funciones de sub-supervivencia observables, $P_j(t)$ (ver [29], [31]).

Hasta el momento el problema de no-identificabilidad de las FDS marginales ha sido discutido en un sentido no-paramétrico, debido a que el conjunto dado de $P_j(t)$ se obtiene mediante ese tipo de métodos (Sección 5.3). Otros autores han optado por estudiar este problema en un escenario **paramétrico**, es decir, especificando paramétricamente la función de supervivencia conjunta de los tiempos de falla latentes. En este caso, el problema al que uno se enfrenta es a la identificabilidad del conjunto finito de parámetros de la función de supervivencia conjunta a partir de la distribución de (T, J) . Crowder [5] y Moeschberger y Klein [29] revisan modelos para los cuales la identificabilidad se mantiene.

5.5. Razón de Riesgos Proporcionales

El riesgo relativo de falla por causa \mathcal{C}_j al tiempo t es $h_j(t)/h_{\mathcal{T}}(t)$. Supongamos que esa razón no depende del tiempo t para cada j , es decir,

$$h_j(t) = a_j h_{\mathcal{T}}(t) \quad j = 1, 2, \dots, m, \quad (5.47)$$

donde $0 < a_j < 1$ es una constante. En este caso se dice que se han obtenido **riesgos proporcionales**. Así, conforme transcurre el tiempo, el riesgo relativo correspondiente a cada modo de falla j permanece constante. Un primer resultado de la suposición de riesgos proporcionales es el siguiente:

Proposición 5.5.1. *Supongase que la función de riesgo de causa específica, $h_j(t)$, es proporcional a la función de riesgo global, $h_{\mathcal{T}}(t)$. Entonces, la FDS (condicional) cruda entre aquéllos quiénes mueren de causa \mathcal{C}_j , $S_j^*(t)$, es la misma que la FDS global, $S_{\mathcal{T}}(t)$.*

Demostración. De la Ecuación 5.15, tenemos

$$P_j(t) = \int_t^{\infty} h_j(u) S_{\mathcal{T}}(u) du = a_j \int_t^{\infty} h_{\mathcal{T}}(u) S_{\mathcal{T}}(u) du = a_j S_{\mathcal{T}}(t), \quad (5.48)$$

y así

$$P_j(0) = a_j = \pi_j.$$

Por eso

$$S_j^*(t) = \frac{1}{\pi_j} P_j(t) = S_{\mathcal{T}}(t).$$

■

Como ya se dijo, es de interés estudiar la razón en (5.27) dada por

$$\frac{P_j(t)}{S_{\mathbb{T}}(t)} = \Pr(J = j \mid T > t) = \pi_j(t)$$

Si esta probabilidad permanece bastante estable, entonces, de (5.48), la distribución del tiempo de falla por causa \mathcal{C}_j en presencia de todas las causas, $S_j^*(t)$, tiene el mismo modelo que la FDS global, $S_{\mathbb{T}}(t)$. Note que este resultado no requiere que los tiempos de falla latentes sean independientes (ver [5]).

Si **A1** se mantiene, al igual que (5.47) para cada $j = 1, 2, \dots, m$, también se tiene que

$$a_1 + a_2 + \dots + a_m = 1.$$

En este caso, de (5.25), es fácil ver que

$$\frac{h_j(t)}{h_k(t)} = \frac{a_j}{a_k} = a_{jk},$$

y por eso

$$\frac{P_j(t)}{P_k(t)} = a_{jk},$$

es decir, $P_j(t) = a_{jk}P_k(t)$, para cualquier $j, k \in \mathcal{J}$.

Otra propiedad interesante de la suposición de riesgos proporcionales tiene que ver con la FDS marginal, $S_j(t)'$, del Teorema 5.4.1. Si se cumple (5.47), de (5.46) se tiene

$$S_j'(t) = \exp \left[-a_j \int_0^t h_{\mathbb{T}}(u) du \right] = \left\{ \exp \left[- \int_0^t h_{\mathbb{T}}(u) du \right] \right\}^{a_j} = [S_{\mathbb{T}}(t)]^{a_j} \quad (5.49)$$

Sin la suposición de independencia $S_j'(t)$ no es observable, pero sí estimable. Si podemos encontrar empíricamente que las funciones de riesgo de causa-específica son proporcionales, entonces $S_j'(t)$ puede ser fácilmente estimada a partir de (5.49). Por otro lado, se puede demostrar que en el caso de riesgos competitivos independientes, en particular bajo la suposición de Makeham,

$$\text{se obtienen riesgos proporcionales} \iff S_j'(t) = [S_{\mathbb{T}}(t)]^{a_j}$$

En este caso, $S_{1,2,\dots,m}'(t_1, t_2, \dots, t_n) = \prod_{j=1}^m [S_{\mathbb{T}}(t_j)]^{a_j}$. Esto significa que solo tenemos que preocuparnos por la FDS $S_{\mathbb{T}}(t)$, en lugar del conjunto de funciones de supervivencia $\{P_j(t) : j = 1, 2, \dots, m\}$.

5.6. Modelado de Riesgos Competitivos

Modelado de Tiempos de Falla Latentes

El enfoque tradicional para el modelado de riesgos competitivos dependientes ha sido vía la especificación de la función de distribución conjunta de los tiempos de

falla latentes Y_1, Y_2, \dots, Y_m o de su FDS conjunta $S_{1\dots m}(y_1, \dots, y_m)$. Sin embargo, la consideración de este enfoque conduce a problemas de *identificabilidad* del modelo por parte de la distribución de (T, J) . Varios investigadores han estudiado modelos para los cuales la identificabilidad se mantiene, por ejemplo, la distribución Normal y Weibull bivariada considerada por Moeschberger (1974), o la distribución Exponencial de Gumbel (1960) (Ver [5]).

Modelado de la Función de Riesgo de causa-específica

Otra estructura de modelado e inferencia de riesgos competitivos se basa en las funciones de riesgo de causa-específica, $h_j(t)$ ($j = 1, 2, \dots, m$), definidas en (5.7) como

$$h_j(t) = \lim_{\Delta t \rightarrow 0^+} \frac{\Pr(\{t < T \leq t + \Delta t\} \cap \{J = j\} \mid T > t)}{\Delta t}$$

para $j = 1, 2, \dots, m$ y $t > 0$. Estas funciones son estimables a partir de los datos sin hacer suposiciones injustificables (e.g. independencia de riesgos competitivos). Actualmente se han desarrollado e implementado metodologías modernas en torno a las funciones de riesgo de causa-específica [5]. El enfoque común es suponer un modelo paramétrico para $h_j(t)$, por ejemplo, un Modelo Weibull, cuya función de riesgo por la causa j es

$$h_j(t; \alpha_j, \theta_j) = \frac{\alpha_j}{\theta_j} \frac{t^{1-\alpha_j}}{\theta_j}$$

para $j = 1, 2, \dots, m$ y $t > 0$. Como se menciona en el caso univariado, *i*) para $\alpha = 1$ se obtiene un riesgo constante conforme transcurre el tiempo, *ii*) el riesgo se incrementa para $\alpha > 1$, y *iii*) este decrece para $\alpha < 1$, mostrándose así un modelo flexible para describir el comportamiento del riesgo en diferentes situaciones.

Modelado de la Función de sub-distribución

Entre los primeros enfoques basados en la función de sub-distribución o de incidencia acumulada de causa-específica, definida en (5.14) como

$$Q_j(t) = \Pr(T \leq t, J = j) = \int_0^t h_j(u) \exp \left[- \int_0^u h_T(v) dv \right] du \quad (5.50)$$

se encuentra el de Prentice, *et. al.* (1978) y Kalbfleisch & Prentice (1980) (ver [16]-2002), donde argumentan fuertemente en contra del clásico planteamiento para el modelado de riesgos competitivos (i.e. vía un modelo de tiempos de falla latentes). Esto se debió, en parte, al problema de identificabilidad de la FDS conjunta de los tiempos de falla latentes. Por otro lado, de (5.50) se ve que el enfoque en cuestión se basa en cantidades observables, evitando tal problema de identificabilidad.

En el programa computacional estadístico **R** se puede obtener el estimador no-paramétrico de la función de incidencia acumulada dado en (5.41) con el paquete `cmprsk` de Robert Gray. Por otro lado, hay situaciones en donde se desea comparar

curvas de incidencia acumuladas de un tipo de falla en particular, entre diferentes grupos y en presencia de riesgos competitivos. Gray(1988) propuso una clase de pruebas de hipótesis para tal situación, y también puede ser obtenida con la misma librería.

Un primer enfoque del modelado de riesgos competitivos mediante (5.50) se basa en la función de sub-supervivencia $P_j(t) = \Pr(T > t, J = j) = \pi_j - Q_j(t)$ definida en (5.15), y en la FDS condicional (propia) dado que la falla es por causa \mathcal{C}_j , y en presencia de riesgos competitivos, definida en (5.22) como

$$S_j^*(t) = \Pr(T > t \mid J = j) = \frac{1}{\pi_j} P_j(t) \quad (5.51)$$

donde $\pi_j = \Pr(J = j)$. A partir de $S_j^*(t)$ se obtiene la función de densidad condicional dada en (5.24) y la función de riesgo condicional en (5.26) definidas, respectivamente, como

$$f_j^*(t) = \Pr(T = t \mid J = j) = \frac{f_j(t)}{\pi_j} \quad \text{y} \quad h_j^*(t) = -\frac{d \log S_j^*(t)}{dt} = \frac{f_j^*(t)}{S_j^*(t)}. \quad (5.52)$$

De (5.19), la FDS marginal del tiempo de falla real T es

$$S_T(t) = \sum_{j=1}^m P_j(t) = \sum_{j=1}^m \pi_j S_j^*(t) \quad (5.53)$$

donde $\sum_{j=1}^m \pi_j = 1$. De la Sección 2.2, $S_T(t)$ se ve como una mezcla de FDS con m componentes.

Debido a la amplia variedad de situaciones que dan lugar a datos de riesgos competitivos, habrá casos donde la suposición **A2** no se cumpla. Por ejemplo, una población heterogénea en la que están actuando $m = 2$ causa de falla, con una proporción de individuos π_1 determinada a presentar una falla solamente de la causa \mathcal{C}_1 , y otra proporción π_2 determinada a presentar una falla por la causa \mathcal{C}_2 ; tal población es considerada como una *mezcla* de individuos destinados a presentar una falla de dos causa diferentes. En este caso, la FDS para el tiempo de falla T correspondiente a esa población heterogénea es $S_T(t) = \pi_1 S_1^*(t) + \pi_2 S_2^*(t)$. Entonces, otro enfoque para el modelado de riesgos competitivos es suponer una forma paramétrica para $S_j^*(t)$ y considerar a (5.53) como una *mezcla* de FDS con m componentes, cuyos pesos están dados por las π_j 's. Este enfoque de mezclas supone que un individuo fallará de un riesgo en particular, escogido por un mecanismo estocástico (Larson & Dinse, 1985).

Modelo de Regresión

Crowder (2001) discute las adaptaciones que se dan de manera natural para los modelos de regresión en su versión univariada, al caso de Riesgos Competitivos: riesgos proporcionales, tiempos de vida acelerados, odds proporcionales y vida residual media. Aquí se mencionara riesgos proporcionales solamente.

Sea z_i el vector de variables explicativas para el i -ésimo individuo en la muestra. Un enfoque común para incluir variables explicativas en un modelo de regresión

para Riesgos Competitivos vía la función de riesgo de cuasa-específica es hacer una adaptación de la definición univariada de "riesgos proporcionales":

$$h_j(t_i; \mathbf{z}_i) = \psi_i(\mathbf{z}_i)h_{0j}(t_i) \quad t > 0,$$

donde $h_{0j}(t)$ es la función de riesgo base ($\mathbf{z}_i = 0$) para el modo de falla j , y $\psi_j(\mathbf{z}_i)$ es una función positiva del vector de covariables, por ejemplo, una conveniente selección es $\psi_j(\mathbf{z}_i) = \exp(\boldsymbol{\beta}_j^\top \mathbf{z}_i)$, para algún vector de parámetros $\boldsymbol{\beta}_j$.

5.6.1. Modelo de Mezclas de Larson & Dinse (1985)

El modelo de estos autores, el cual presenta una especificación completamente paramétrica, ha servido como punto de partida para la consideración de modelos de mezclas más generales en un escenario de riesgos competitivos. Ellos suponen que un individuo presentará una falla de un riesgo en particular escogido por un mecanismo estocástico.

Sean \mathbf{Z} y \mathbf{X} dos vectores de covariables p y q dimensionales, respectivamente (\mathbf{Z} y \mathbf{X}) pueden tener componentes en común). Supongamos que se tiene una muestra aleatoria independiente de datos de riesgos competitivos con m modos de falla, denotada como

$$\mathbf{y}_i = (t_i, \delta_i, \mathbf{x}_i, \mathbf{z}_i, \delta_i J_i)^\top \quad i = 1, 2, \dots, n$$

donde el tiempo de falla actual t_i es una realización de $T_i = \min(\tilde{T}_i, C_i)$, \tilde{T}_i es el tiempo de falla del i -ésimo individuo en la muestra, C_i la variable de censura asociado a \tilde{T}_i , $\delta_i = \mathbf{1}\{\tilde{T}_i \leq C_i\}$ la indicadora de censura, y J_i la variable indicadora del modo de falla para el i -ésimo individuo, tomando posibles valores en $\mathcal{J} = \{1, 2, \dots, m\}$.

Para relacionar la distribución de J con \mathbf{x}_i , se supone un modelo de regresión logística multinomial dado por

$$\pi_j(\mathbf{x}_i) = \Pr(J = j \mid \mathbf{x}_i) = \frac{\exp(\boldsymbol{\eta}_j^\top \mathbf{v}_i)}{\sum_{l=1}^m \exp(\boldsymbol{\eta}_l^\top \mathbf{v}_i)} \quad (j \in \mathcal{J}) \quad (5.54)$$

donde $\boldsymbol{\eta}_j = (\eta_{0j}, \eta_{1j}, \dots, \eta_{qj})^\top$ es un vector $(q+1)$ -dimensional de coeficientes logísticos desconocidos y $\mathbf{v}_i = (1, \mathbf{x}_i^\top)^\top$ es un $(q+1)$ -vector, el cual contiene a la ordenada. Por unicidad se requiere la restricción $\boldsymbol{\eta}_m = \mathbf{0}$. Para el efecto del vector de covariables \mathbf{z} sobre $h_j^*(t)$ se supone un modelo de riesgos proporcionales (Cox, 1972); eso es,

$$h_j^*(t; \mathbf{z}) = h_{0j}^*(t) \exp(\boldsymbol{\beta}_j^\top \mathbf{z}), \quad j \in \mathcal{J}$$

donde $\boldsymbol{\beta}_j$ es un vector p -dimensional de parámetros de regresión desconocidos y $h_{0j}^*(t)$ es la función de riesgo base para la falla de tipo j , la cual está especificada paramétricamente por un modelo de escalones exponenciales definido en la Sección 2.2 como:

$$h_{0j}(t) = \exp(\alpha_{jl}) \quad \text{si } t \in I_l = [a_{l-1}, a_l),$$

donde I_1, \dots, I_M son intervalos (pre-especificados) mutuamente excluyentes que abarcan toda la recta real positiva, y α_{jl} es un parámetro que representa el log del riesgo

base para la falla de tipo j en el intervalo I_l ($j = 1, 2, \dots, m; l = 1, 2, \dots, M$). El modelo paramétrico de estos autores se establece para el mismo conjunto de variables explicativas, i.e. $\mathbf{z} = \mathbf{x}$. Así, de (5.53) la función de supervivencia condicional para el tiempo de falla, dado la falla da causa j y el vector de covariables \mathbf{x} es

$$S_j^*(t; \mathbf{x}, \boldsymbol{\beta}_j) = [\exp\{-H_{0j}^*(t)\}]^{\exp(\boldsymbol{\beta}_j^\top \mathbf{x})} \quad j \in \mathcal{J}$$

donde

$$H_{0j}^*(t) = \int_0^t h_{0j}^*(u) du = \sum_{l=1}^M \exp(\alpha_{jl}) \Delta_l(t)$$

es la función de riesgo acumulada base para la falla de tipo j y

$$\Delta_l(t) = \int_{a_{l-1}}^{a_l} \mathbf{1}\{u \leq t\} \quad \text{para } t > 0.$$

Sea $\boldsymbol{\eta} = (\boldsymbol{\eta}_1^\top, \dots, \boldsymbol{\eta}_{m-1}^\top)^\top$, $\boldsymbol{\Lambda} = (H_{01}^*(t), \dots, H_{0m}^*(t))$ y $\boldsymbol{\theta} = (\boldsymbol{\eta}^\top, \boldsymbol{\beta}_1^\top, \dots, \boldsymbol{\beta}_m^\top, \boldsymbol{\Lambda})$. A partir de esas especificaciones se sigue que la FDS para el i -ésimo individuo es

$$S_T(t_i; \mathbf{x}_i, \boldsymbol{\theta}) = \sum_{j=1}^m \pi_j(\mathbf{x}_i; \boldsymbol{\eta}) S_j^*(t_i; \mathbf{x}_i, \boldsymbol{\beta}_j).$$

La función de verosimilitud de $\boldsymbol{\theta}$ dado $\mathbf{y} = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$ se construye de manera similar como en el caso univariado. La contribución a la verosimilitud de la i -ésima observación no-censurada es $\pi_j(\mathbf{x}_i; \boldsymbol{\eta}) f_j^*(t_i; \mathbf{x}_i, \boldsymbol{\beta}_j)$, mientras que $S_T(t_i; \mathbf{x}_i, \boldsymbol{\theta})$ es la contribución a la verosimilitud de una observación censurada. Entonces, usando la notación $\delta_{ij} = \mathbf{1}\{J_i = j\}$, de (5.36) la función de verosimilitud para $\boldsymbol{\theta}$ formulada con base en \mathbf{y} es

$$L(\boldsymbol{\theta} | \mathbf{y}) = \prod_{i=1}^n \left\{ \prod_{j=1}^m \left[\pi_j(\mathbf{x}_i; \boldsymbol{\eta}) f_j^*(t_i; \mathbf{x}_i, \boldsymbol{\beta}_j) \right]^{\delta_{ij}} \left[\sum_{j=1}^m \pi_j(\mathbf{x}_i; \boldsymbol{\eta}) S_j^*(t_i; \mathbf{x}_i, \boldsymbol{\beta}_j) \right]^{1-\delta_{i\cdot}} \right\} \quad (5.55)$$

donde $\delta_{i\cdot} = \sum_{j=1}^m \delta_{ij}$. Aplicando el Método de Máxima Verosimilitud a $L(\boldsymbol{\theta} | \mathbf{y})$ (Apendice A), la ecuación resultante no se puede resolver analíticamente. Larson y Dinse calcularon el EMV de $\boldsymbol{\theta}$ dado el conjunto de datos incompletos \mathbf{y} vía un Algoritmo⁸ EM.

5.6.1.1. Aplicación del Algoritmo EM

En vez de trabajar directamente con la función de verosimilitud (5.55), el EMV de $\boldsymbol{\theta}$ se puede encontrar aplicando el algoritmo EM. A cada individuo i ($i = 1, 2, \dots, n$) se

⁸ Dempster *et. al.* (1977) aplicaron el algoritmo EM en varios contextos, entre los que se incluyen mezclas de distribuciones.

le asigna un vector aleatorio m -dimensional de indicadoras de etiquetas-componentes $\mathbf{\Gamma}_i = (\Gamma_{i1}, \dots, \Gamma_{im})^\top$, donde $\Gamma_{ij} = \mathbf{1}\{J_i = j\} = \gamma_{ij}$ es cero o uno de acuerdo a si el individuo i presenta una falla por causa j o no ($j \in \mathcal{J}$). Se sigue que $\pi_j(\mathbf{x}_i; \boldsymbol{\eta}) = \Pr(\gamma_{ij} = 1 \mid \mathbf{X}_i)$. Para el i -ésimo individuos con tiempo de supervivencia no-censurado $\mathbf{\Gamma}_i$ es conocido, y por tanto $\gamma_{ij} = \delta_{ij} = 1$ para algún $j \in \mathcal{J}$. Por otro lado, para el i -ésimo individuo con tiempo de falla censurado el vector aleatorio de etiquetas-componentes $\mathbf{\Gamma}_i$ será desconocido. Entonces, podemos aplicar el algoritmo EM en un escenario donde

- $\mathbf{Y}_c = \{(\mathbf{y}_1, \mathbf{\Gamma}_1), (\mathbf{y}_2, \mathbf{\Gamma}_2), \dots, (\mathbf{y}_n, \mathbf{\Gamma}_n)\}$ es visto como el conjunto de datos completos;
- $\mathbf{Y}_{obs} = \{\mathbf{y}_i : i = 1, 2, \dots, n\} \cup \{\mathbf{\Gamma}_i : \delta_i = 1\}$ como el conjunto de datos observados (i.e. incompletos); y
- $\mathbf{Y}_{miss} = \{(\tilde{T}_i, \mathbf{\Gamma}_i) : \delta_i = 0, i = 1, 2, \dots, n\}$, como el conjunto de datos perdidos, donde \tilde{T}_i es el tiempo de falla real, pero desconocido, del i -ésimo individuo. Todo lo que se ha observado de esos individuos es que su tiempo de falla real es mayor que su tiempo de censura c_i , eso es, $\tilde{T}_i > c_i$.

De esta manera, la función de verosimilitud de datos-completos es

$$L_c(\boldsymbol{\theta} \mid \mathbf{Y}_c) = \prod_{i=1}^n \left\{ \prod_{j=1}^m [\pi_j(\mathbf{x}_i; \boldsymbol{\eta}) f_j^*(t_i; \mathbf{x}_i, \boldsymbol{\beta}_j)]^{\delta_{ij} \gamma_{ij}} \left[\prod_{j=1}^m \left(\pi_j(\mathbf{x}_i; \boldsymbol{\eta}) S_j^*(t_i; \mathbf{x}_i, \boldsymbol{\beta}_j) \right)^{\gamma_{ij}} \right]^{1-\delta_{ij}} \right\}.$$

y la correspondiente función log-verosimilitud de datos-completos queda expresado como

$$\begin{aligned} \ell_c(\boldsymbol{\theta} \mid \mathbf{Y}_c) = \sum_{i=1}^n \left\{ \sum_{j=1}^m \log [\pi_j(\mathbf{x}_i; \boldsymbol{\eta}) f_j^*(t_i; \mathbf{x}_i, \boldsymbol{\beta}_j)] \right. \\ \left. + \sum_{j=1}^m (1 - \delta_{i\cdot}) \gamma_{ij} \log [\pi_j(\mathbf{x}_i; \boldsymbol{\eta}) S_j^*(t_i; \mathbf{x}_i, \boldsymbol{\beta}_j)] \right\} \end{aligned} \quad (5.56)$$

Paso-E

Sea $\boldsymbol{\theta}^{(k)}$ el estimador actual de $\boldsymbol{\theta}$. En la $(k+1)$ -ésima iteración, el paso-E consiste en calcular la esperanza condicional de $\ell_c(\boldsymbol{\theta}; \mathbf{Y}_c)$, dado los datos observados \mathbf{Y}_{obs} y $\boldsymbol{\theta}^{(k)}$. Note que (5.56) es una función lineal del conjunto de datos perdidos \mathbf{Y}_{miss} , por lo que este paso se reduce a reemplazar γ_{ij} en $\ell_c(\boldsymbol{\theta}; \mathbf{Y}_c)$ por $w_{ij}^{(k)}$, que es su esperanza

condicional, dado \mathbf{Y}_{obs} y $\boldsymbol{\theta}^{(k)}$. Ahora

$$\begin{aligned} w_{ij}^{(k)} &= \mathbb{E}[\gamma_{ij} \mid \mathbf{Y}_{obs}, \boldsymbol{\theta} = \boldsymbol{\theta}^{(k)}] \\ &= \int_{\gamma_{ij} \in \{0,1\}} \gamma_{ij} \Pr(\gamma_{ij} \mid \mathbf{y}_i, \boldsymbol{\theta}^{(k)}) d\gamma_{ij} \\ &= (1) \Pr(\gamma_{ij} = 1 \mid T > t_i, \boldsymbol{\theta}^{(k)}) + (0) \Pr(\gamma_{ij} = 0 \mid T > t_i, \boldsymbol{\theta}^{(k)}) \\ &= \Pr(J_i = j \mid T > t_i, \boldsymbol{\theta}^{(k)}) \end{aligned}$$

De (5.27), $w_{ij}^{(k)}$ representa la probabilidad de que el i -ésimo individuo con tiempo de censura t_i eventualmente presentará una falla de causa j , dado que no ha ocurrido ninguna falla al tiempo t_i :

$$w_{ij}^{(k)} = \frac{\pi_j(\mathbf{x}_i, \boldsymbol{\eta}^{(k)}) S_j^*(t_i; \mathbf{x}_i, \boldsymbol{\beta}_j^{(k)})}{\sum_{l=1}^m \pi_l(\mathbf{x}_i, \boldsymbol{\eta}^{(k)}) S_l^*(t_i; \mathbf{x}_i, \boldsymbol{\beta}_l^{(k)})} \quad j = 1, 2, \dots, m. \quad (5.57)$$

Así, la función $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$ es

$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) &= \sum_{i=1}^n \left\{ \sum_{j=1}^m \log [\pi_j(\mathbf{x}_i; \boldsymbol{\eta}) f_j^*(t_i; \mathbf{x}_i, \boldsymbol{\beta}_j)] \right. \\ &\quad \left. + \sum_{j=1}^m (1 - \delta_{i\cdot}) w_{ij}^{(k)} \log [\pi_j(\mathbf{x}_i; \boldsymbol{\eta}) S_j^*(t_i; \mathbf{x}_i, \boldsymbol{\beta}_j)] \right\} \end{aligned} \quad (5.58)$$

Paso-M

Este paso involucra maximizar $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$ con respecto a $\boldsymbol{\theta}$ para obtener $\boldsymbol{\theta}^{(k+1)}$ como sigue. Se puede verificar fácilmente que (5.58) se puede expresar como una suma de $(m+1)$ componentes funcionalmente independientes, eso es

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = \ell(\boldsymbol{\pi}) + \sum_{j=1}^m \ell_j(S^*)$$

donde

$$\ell(\boldsymbol{\pi}) = \sum_{i=1}^n \sum_{j=1}^m g_{ij}^{(k)} \log \pi_j(\mathbf{x}_i; \boldsymbol{\eta}),$$

$g_{ij}^{(k)} = \delta_{ij} + (1 - \delta_{i\cdot}) w_{ij}^{(k)}$, y para $j = 1, 2, \dots, m$

$$\ell_j(S^*) = \sum_{i=1}^n \left\{ \delta_{ij} [\log h_{0j}^*(t_i) + \boldsymbol{\beta}_j^T \mathbf{x}_i] + g_{ij}^{(k)} \log S_j^*(t_i; \mathbf{x}_i, \boldsymbol{\beta}_j) \right\}. \quad (5.59)$$

Dado el valor actual $\boldsymbol{\theta}^{(k)}$, cada $g_{ij}^{(k)}$ se trata como una constante conocida y luego se usa por separado el procedimiento Newton-Raphson para encontrar los estimadores

de los coeficientes logísticos $\{\boldsymbol{\eta}_j : j = 1, 2, \dots, m - 1\}$ que maximizen $\ell(\boldsymbol{\pi})$ y los valores de $\boldsymbol{\beta}_j$ y $\{\alpha_{jl} : l = 1, 2, \dots, M\}$ que maximizen $\ell_j(S^*)$ para $j = 1, 2, \dots, m$.

En resumen, el algoritmo EM comienza con la selección de un conjunto de estimadores iniciales $w_{ij}^{(0)}$ de w_{ij} ($i=1, 2, \dots, n; j=1, 2, \dots, m$), los cuales pueden ser obtenidos a través de estimadores iniciales de $\pi_j(\mathbf{x}_i; \boldsymbol{\eta})$ y $S_j^*(t_i; \mathbf{x}_i, \boldsymbol{\beta}_j)$ (e.g. ignorando las observaciones censuradas). En la $(k+1)$ -ésimo iteración, el Paso-E trata los estimadores actuales $\pi_j(\mathbf{x}_i, \boldsymbol{\eta}^{(k)})$ y $S_j^*(t_i; \mathbf{x}_i, \boldsymbol{\beta}^{(k)})$ como conocidos a fin de actualizar cada estimador de w_{ij} , y a su vez de g_{ij} . En el Paso-M, los valores actuales $g_{ij}^{(k)}$ son tratados como conocidos, y actualizan los estimadores de $\pi_j(\mathbf{x}_i; \boldsymbol{\eta})$ y $S_j^*(t_i; \mathbf{x}_i, \boldsymbol{\beta}_j)$, obteniendo $\pi_j(\mathbf{x}_i; \boldsymbol{\eta}^{(k+1)})$ y $S_j^*(t_i; \mathbf{x}_i, \boldsymbol{\beta}_j^{(k+1)})$. Bajo apropiadas condiciones de regularidad ([8]), estos estimadores de $\pi_j(\mathbf{x}; \boldsymbol{\eta})$ y $S_j^*(t; \mathbf{x}, \boldsymbol{\beta}_j)$ eventualmente convergerán a los verdaderos EMV.

CONCLUSIONES

Cuando se estudia la variable tiempo hasta la ocurrencia del evento de interés en presencia de eventos de riesgos competitivos (RC) se distinguen tres situaciones: (1) analizar el evento de interés ignorando riesgos competitivos; (2) considerar conjuntamente los eventos de todos los modos de fallo, colapsándolos en un sólo evento y proceder con el análisis, y (3) emplear adecuadamente la metodología desarrollada en riesgos competitivos. El presente trabajo expone la metodología empleada en el análisis de supervivencia univariado, tanto desde el punto de vista teórico como computacionalmente práctico, de tal manera que uno debe notar que el primer planteamiento no es correcto (e.g. el uso del método KM en presencia de RC). El segundo enfoque trata de la extensión del método Kaplan-Meier para incluir RC. Sin embargo, este es de uso limitado ya que no permite explorar los efectos de variables explicativas sobre el tiempo de supervivencia. El tercer punto es el adecuado, siempre y cuando se considere una metodología igualmente adecuada, ya que de lo contrario uno se puede topar con los problemas básicos que surgen en el análisis de este tipo de datos. Por ejemplo, ya que los riesgos usualmente serán dependientes, el enfoque clásico de tiempo de fallo latentes para el modelado de RC conduce al problema de identificabilidad del modelo basándose únicamente en los datos de la forma (T, J) y sin suposiciones adicionales. Por otro lado, un análisis de datos de riesgos competitivos mediante sub-distribuciones evita tal problema de identificabilidad, ya que éstas son cantidades observables y estimables a partir de los datos (T, J) .

La metodología presentada en éste trabajo para el análisis de supervivencia univariado y riesgos competitivos son de rápida comprensión y sencilla aplicación. Lo que debe notarse es el esfuerzo por reunir los conceptos fundamentales para entender modelos mas generales y métodos más sofisticados de análisis para datos relacionados con la variable tiempo hasta la ocurrencia de un evento. Por ejemplo, el modelo paramétrico de mezclas exponenciales de Larson & Dinse (1985), permitido para co-variantes, en el cual se usa el Algoritmo EM para la estimación de los parámetros involucrados en este. Otro ejemplo es el modelado de *eventos secundarios* en presencia de riesgos competitivos; en este caso, el individuo presenta dos eventos de interés consecutivos, cada uno con su propio tiempo de fallo. La forma de realizar el análisis consiste en derivar a partir de dos modelos de riesgos competitivos, como los que se

han estudiado en el Capítulo 5, dos estimadores de las funciones de incidencia acumulada correspondientes al primero y segundo evento. Un ejemplo específico de este tipo de modelos ha sido estudiado por Katsahian *et. al.* (2004).

Recientemente se han propuesto e investigado generalizaciones semiparamétricas del modelos de Larson & Dinse. Por ejemplo, Escarela y Bowater (2008) y Ng y McLachlan (2003) consideran modelos de mezclas semi-paramétrico con variables explicativas presentes, en donde se propone e implementa un algoritmo ECM, que es una extensión del algoritmo EM.

Otra manera de analizar los datos sobre eventos observados con el tiempo, es presentar los métodos expuestos en esta tesis pero con la estructura de procesos contables. Existen varios libros que exploran el enfoque de martingalas para el análisis estadístico de procesos contables, enfatizando en las aplicaciones de esos métodos. Una de las ventajas de considerar esa estructura es que se puede adentrar al estudio de métodos de análisis de otro tipo de datos, por ejemplo, *datos de eventos recurrentes*. Este tipo de datos surgen en una amplia variedad de escenarios, incluyendo Salud Pública, Ingeniería, Economía y Sociología. En este caso, el evento de primordial interés es recurrente y puede suceder varias veces durante el período de seguimiento para un individuo bajo estudio (e.g. la repetida hospitalización de pacientes con enfermedades renales en etapa terminal).

Esperando que el presente trabajo sirva de introducción a los métodos estadísticos empleados en el análisis de datos de tiempos de supervivencia, y que sea usado para el entendimiento e interpretación de los resultados de tal análisis, invito a las nuevas generaciones de estudiantes relacionados con estadística a tener una experiencia practica-computacional, y no quedarse con un planteamiento teórico-abstracto.

APÉNDICE A

MÉTODO DE MÁXIMA VEROSIMILITUD

Sea Y_1, Y_2, \dots, Y_n una muestra de v.a.s. independientes e idénticamente distribuidas (i.i.d.) con densidad común $f(y_i)$, eso es

$$Y_1, Y_2, \dots, Y_n \stackrel{i.i.d.}{\sim} F,$$

donde $F(y_i)$ denota la función de distribución correspondiente a $f(y_i)$. Denotemos por $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^\top$ el vector que contiene a la muestra entera y por \mathbf{y} una realización¹ de \mathbf{Y} . Supongamos que $f(\cdot)$ pertenece a la familia paramétrica

$$\mathcal{P} = \{F_\theta \mid \theta = (\theta_1, \theta_2, \dots, \theta_p)^\top \in \Theta \subset \mathbb{R}^p\},$$

donde θ es un vector de parámetros desconocidos y Θ es el espacio de valores de θ .

Como ya se sabe, debido a la independencia de las Y_i 's, la función de densidad conjunta del vector aleatorio $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^\top$ es el producto de las densidades marginales

$$\prod_{i=1}^n f_{Y_i}(y_i; \theta) = \prod_{i=1}^n f(y_i; \theta).$$

Entonces, la función de densidad conjunta de las n observaciones independientes $\mathbf{y} = (y_1, y_2, \dots, y_n)$ es

$$\mathbf{f}_{\mathbf{Y}}(\mathbf{y}; \theta) = \mathbf{f}_{\mathbf{Y}}(y_1, \dots, y_n; \theta) = \prod_{i=1}^n f(y_i; \theta).$$

Nótese que $\mathbf{f}_{\mathbf{Y}}(\mathbf{y}; \theta)$ es una función de los datos observados \mathbf{y} , manteniendo fijo el vector de parámetros θ . Eso motiva la siguiente definición:

¹ Una realización de \mathbf{Y} se refiere a la muestra aleatoria observada (i.e. los datos)

$$\mathbf{y} = (y_1, y_2, \dots, y_n)^\top,$$

donde y_i es el valor observado de la v.a.s. Y_i ($i=1, 2, \dots, n$).

Definición A.1. Sea $\mathbf{Y} \in \mathbb{R}^n$ un vector de v.a.s. i.i.d. con función de densidad conjunta $f_{\mathbf{Y}}(y_1, \dots, y_n; \boldsymbol{\theta})$, la cual depende del vector de parámetros $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p$ ($p < n$). La **función de verosimilitud**, $L(\boldsymbol{\theta}; \mathbf{Y})$, se define como

$$L(\boldsymbol{\theta}) = L(\boldsymbol{\theta}; \mathbf{Y}) \propto f_{\mathbf{Y}}(y_1, \dots, y_n; \boldsymbol{\theta}) = \prod_{i=1}^n f(y_i; \boldsymbol{\theta}). \quad (\text{A.1})$$

El Método de Máxima Verosimilitud establece que busquemos como valor estimado de $\boldsymbol{\theta}$ aquel vector de valores, $\hat{\boldsymbol{\theta}} \in \Theta$, mediante la optimización de la función de verosimilitud, $L(\boldsymbol{\theta})$, de tal manera que nos de la mayor probabilidad de que ocurra el evento de interés, según lo que hemos observado.

Definición A.2. Sea $\Theta \subseteq \mathbb{R}^p$ el conjunto de todos los posibles valores del vector de parámetros $\boldsymbol{\theta}$. El **Estimador de Máxima Verosimilitud (EMV)** de $\boldsymbol{\theta}$ se define como el valor $\hat{\boldsymbol{\theta}} \in \Theta$ (si existe) el cual maximiza (A.1), eso es

$$L(\hat{\boldsymbol{\theta}}; \mathbf{Y}) = \max_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta}; \mathbf{Y}).$$

En muchos problemas, es un tanto conveniente trabajar con el logaritmo natural² de la función de verosimilitud.

Definición A.3. La **función log-verosimilitud**, $l(\boldsymbol{\theta}; \mathbf{Y})$ es el logaritmo natural de la función de verosimilitud:

$$l(\boldsymbol{\theta}; \mathbf{Y}) = \log L(\boldsymbol{\theta}; \mathbf{Y}).$$

Ya que la función logaritmo natural es monótona, el EMV, $\hat{\boldsymbol{\theta}}$, satisface

$$l(\hat{\boldsymbol{\theta}}; \mathbf{Y}) = \max_{\boldsymbol{\theta} \in \Theta} l(\boldsymbol{\theta}; \mathbf{Y}).$$

El EMV de $\boldsymbol{\theta}$ se obtiene resolviendo el sistemas de ecuaciones

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}, \quad (\text{A.2})$$

Método Newton-Raphson.

Típicamente, para resolver el sistema de ecuaciones (A.2), que por lo regular es no lineal, se recurre a computadoras y software adecuado para la implementación de algún método iterativo, como es el **Método Newton-Raphson**. Sin embargo, la estimación numérica de tales parámetros no es tan trivial como parece, salvo por algunos casos donde la expresión matemática de máxima verosimilitud es de hecho simple.

Considere la expansión en Series de Taylor de $l(\boldsymbol{\theta})$ alrededor de $\boldsymbol{\theta}^i$:

$$l(\boldsymbol{\theta}) = l(\boldsymbol{\theta}^i) + (\boldsymbol{\theta} - \boldsymbol{\theta}^i)' \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}^i} + \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\theta}^i)' \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2} \Big|_{\boldsymbol{\theta}^i} (\boldsymbol{\theta} - \boldsymbol{\theta}^i) + \dots$$

²Notación. $\log(x) := \log_e(x) = \ln(x)$

Cuando $\boldsymbol{\theta}$ esta cerca de $\boldsymbol{\theta}^i$, los términos restantes son insignificantes, por lo que se pueden ignorar, obteniendo una aproximación cuadrática de $l(\boldsymbol{\theta})$:

$$l(\boldsymbol{\theta}) \approx (\boldsymbol{\theta}^i) + (\boldsymbol{\theta} - \boldsymbol{\theta}^i) \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}^i} + \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\theta}^i)' \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2} \Big|_{\boldsymbol{\theta}^i} (\boldsymbol{\theta} - \boldsymbol{\theta}^i). \quad (\text{A.3})$$

Por lo tanto, el punto estacionario de (A.3) es

$$\boldsymbol{\theta} = \boldsymbol{\theta}^i + \left[- \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2} \Big|_{\boldsymbol{\theta}^i} \right]^{-1} \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}^i}. \quad (\text{A.4})$$

Este resultado proporciona las bases de una aproximación, de manera iterativa, para el cálculo del EMV. Dado un valor inicial fijo de $\boldsymbol{\theta}$, usamos la ecuación (A.4) para obtener una mejor estimación y repetimos el proceso hasta generar una sucesión de estimadores de $\boldsymbol{\theta}$ que converge al EMV $\hat{\boldsymbol{\theta}}$, bajo ligeras condiciones de regularidad. El proceso iterativo descrito anteriormente mediante (A.4) se le llama el **Método Newton-Raphson**. Al vector de primeras derivadas parciales de $l(\boldsymbol{\theta})$, denotado por

$$\mathbf{u}(\boldsymbol{\theta}) = \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}},$$

se le llama **función score**, a la matriz

$$\mathbf{i}(\boldsymbol{\theta}) = - \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2},$$

se le conoce como la **matriz información observada**, y a la esperanza de $\mathbf{i}(\boldsymbol{\theta})$, $E[\mathbf{i}(\boldsymbol{\theta})] = \mathbf{I}(\boldsymbol{\theta})$, se le llama **información de Fisher**. Substituyendo la información observada en (A.4) por la información de Fisher resulta

$$\boldsymbol{\theta} = \boldsymbol{\theta}^i + \left[- \mathbf{I}(\boldsymbol{\theta}^i) \right]^{-1} \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}^i}. \quad (\text{A.5})$$

El proceso iterativo dado en (A.5) se le conoce como **Método de Scoring**.

Pruebas de hipótesis

En presencia de censura por la derecha (aleatoria y de Tipo I), bajo ligeras condiciones de regularidad, se tiene que

$$\hat{\boldsymbol{\theta}} \sim_a N(\boldsymbol{\theta}, \mathbf{I}(\hat{\boldsymbol{\theta}})^{-1}) \quad (\text{A.6})$$

Por lo tanto, podemos estimar la varianza asintótica del EMV $\hat{\boldsymbol{\theta}}$ de $\boldsymbol{\theta}$ como

$$\text{Var}(\hat{\boldsymbol{\theta}}) \approx \mathbf{I}(\hat{\boldsymbol{\theta}})^{-1}.$$

Las pruebas de hipótesis para $\boldsymbol{\theta}$ se pueden basar en la varianza asintótica de $\hat{\boldsymbol{\theta}}$ (o de $\mathbf{u}(\boldsymbol{\theta})$). Para probar $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ contra adecuadas alternativas o construir intervalos de confianza, tenemos los siguientes tres procedimientos:

1. Prueba de Wald: $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^2 \mathbf{I}(\hat{\boldsymbol{\theta}}) \approx \left[\frac{\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0}{S.E.(\hat{\boldsymbol{\theta}})} \right]^2 \sim_a \chi_\nu^2;$
2. Prueba de la Razón de Verosimilitud: $-2 \log \left(\frac{L(\boldsymbol{\theta}_0)}{L(\hat{\boldsymbol{\theta}})} \right) \sim_a \chi_\nu^2;$
3. Prueba Score: $\frac{[\mathbf{u}(\boldsymbol{\theta}_0)]^2}{\mathbf{I}(\boldsymbol{\theta}_0)} \sim_a \chi_\nu^2.$

Observación A.1. Si p es el número de parámetros en el modelo 1 y $p + q$ en el modelo 2, entonces los grados de libertad ν dependerán de la diferencia de parámetros del modelo 2 respecto del modelo 1, que en este caso es $\nu = (p + q) - p = q$.

Sin embargo, en presencia de censura, puede que sea necesario reemplazar la información de Fisher por la información observada, ya que el cálculo de $\mathbf{I}(\boldsymbol{\theta})$ usualmente es complicado y no puede ser obtenido.

Ejemplo A.1 (Modelo Exponencial). Sea $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^\top$ una muestra aleatoria de tamaño n correspondiente a los tiempos de supervivencia de individuos provenientes de cierta población. Supongamos que

$$Y_1, Y_2, \dots, Y_n \stackrel{i.i.d.}{\sim} \text{Exponencial}(\theta) \quad (\theta > 0),$$

donde Y_1 es una v.a.s. con función de densidad de probabilidad (f.d.p.) dada por

$$f(y_i; \theta) = \theta^{-1} \exp(-y_i/\theta) \quad \forall t \in (0, \infty).$$

Denotemos por

$$\mathbf{y} = (y_1, y_2, \dots, y_n)^\top$$

los datos observados completos (es decir, tiempos de supervivencia no-censurados).

En este ejemplo, el vector de parámetros desconocidos $\boldsymbol{\theta}$ es un escalar, siendo igual a θ . De esta manera, la función de verosimilitud de θ formada en base a \mathbf{y} es

$$L(\theta) = \mathbf{f}_{\mathbf{Y}}(\mathbf{y}; \theta) = \prod_{i=1}^n f(y_i; \theta) = \frac{1}{\theta^n} \exp \left\{ - \sum_{i=1}^n \frac{y_i}{\theta} \right\},$$

y la función log-verosimilitud está dada por

$$l(\theta) = -n \log \theta - \frac{1}{\theta} \sum_{i=1}^n y_i. \quad (\text{A.7})$$

En este caso, el EMV, $\hat{\boldsymbol{\theta}}$, de θ puede deducirse de manera explícita igualando la derivada de (A.7) a cero para obtener

$$\hat{\boldsymbol{\theta}} = \sum_{i=1}^n \frac{y_i}{n}.$$

□

APÉNDICE B

EL ALGORITMO EM

El algoritmo EM (Dempster *et. al.* 1977) es un método iterativo que proporciona una alternativa del que se usa comúnmente en el Método de Máxima Verosimilitud (i.e. Algoritmo Newton-Raphson) para estimar el vector de parámetros $\boldsymbol{\theta} \in \Theta$ correspondiente a un modelo asumido $S(t; \boldsymbol{\theta})$ para un conjunto de datos dado \mathbf{Y} , en situaciones donde la solución analítica para la maximización de la función de verosimilitud de $\boldsymbol{\theta}$ dado \mathbf{Y} , $L(\boldsymbol{\theta}; \mathbf{Y})$, es difícil o imposible, así como en el caso de que se tenga un conjunto de datos incompletos (i.e. parcialmente observados).

Supongamos que se tiene un conjunto de datos $\mathbf{Y}_c = (\mathbf{Y}_{obs}, \mathbf{Y}_{miss})$ cuyas realizaciones pertenecen a un espacio muestral \mathcal{Y}_c , donde \mathbf{Y}_{obs} son realizaciones de otro espacio muestral \mathcal{Y}_{obs} y denotan la parte observada de \mathbf{Y}_c (también llamado el conjunto de datos *incompletos*), y \mathbf{Y}_{miss} representa la parte no-observada (missing) de \mathbf{Y}_c . Refirámonos a \mathbf{Y}_c como el conjunto de *datos-completos*, el cual no es directamente observable, pero sólo indirectamente a través de \mathbf{Y}_{obs} . Asumamos un modelo para \mathbf{Y}_c con distribución de probabilidad asociada $f_c(\mathbf{Y}_c | \boldsymbol{\theta})$ indexada por un vector de parámetros desconocido $\boldsymbol{\theta} \in \Theta$, siendo Θ el espacio de valores de $\boldsymbol{\theta}$. La función de densidad de \mathbf{Y}_{obs} , $f(\mathbf{Y}_{obs} | \boldsymbol{\theta})$, se relaciona con la de \mathbf{Y}_c mediante la ecuación

$$f(\mathbf{Y}_{obs} | \boldsymbol{\theta}) = \int f_c(\mathbf{Y}_{obs}, \mathbf{Y}_{miss} | \boldsymbol{\theta}) d\mathbf{Y}_{miss}.$$

El objetivo del algoritmo EM es maximizar la función de verosimilitud $L(\boldsymbol{\theta} | \mathbf{Y}_{obs}) = f(\mathbf{Y}_{obs} | \boldsymbol{\theta})$ con respecto a $\boldsymbol{\theta}$ dado \mathbf{Y}_{obs} , pero haciendo uso de $f_c(\mathbf{Y}_{obs}, \mathbf{Y}_{miss} | \boldsymbol{\theta})$. Así, la función log-verosimilitud de $\boldsymbol{\theta}$ dado \mathbf{Y}_c , llamada *log-verosimilitud de datos-completos*, es

$$\ell_c(\boldsymbol{\theta} | \mathbf{Y}_c) = \log L_c(\boldsymbol{\theta} | \mathbf{Y}_c) \tag{B.1}$$

Note que (B.1) es de hecho una variable aleatoria ya que \mathbf{Y}_{obs} y $\boldsymbol{\theta}$ son constantes, y la información perdida \mathbf{Y}_{miss} es desconocida (i.e. aleatoria).

Sea $\Pr(\mathbf{Y}_{miss} | \mathbf{Y}_{obs}, \boldsymbol{\theta}^{(k)})$ la distribución de probabilidad condicional de \mathbf{Y}_{miss} dado el conjunto de datos observados \mathbf{Y}_{obs} y el valor actual de los estimadores de los parámetros, $\boldsymbol{\theta}^{(k)}$. En términos generales, este algoritmo consta de dos pasos:

- **Paso-E (Esperanza):** Calcula $E[\ell_c(\boldsymbol{\theta} | \mathbf{Y}_c) | \mathbf{Y}_{obs}, \boldsymbol{\theta}^{(k)}]$, eso es, la esperanza condicional de la verosimilitud de datos-completos, $\ell_c(\boldsymbol{\theta} | \mathbf{Y}_c)$, respecto a los

datos perdidos \mathbf{Y}_{miss} , dado los datos observados \mathbf{Y}_{obs} y los estimadores de los parámetros actuales, $\boldsymbol{\theta}^{(k)}$. Esa esperanza comúnmente se denota como

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = \int \ell_c(\boldsymbol{\theta} | \mathbf{Y}_c) \Pr(\mathbf{Y}_{miss} | \mathbf{Y}_{obs}, \boldsymbol{\theta}^{(k)}) d\mathbf{Y}_{miss} \quad (\text{B.2})$$

- **Paso-M (Maximización):** La función Q en (B.2) es maximizada con respecto a $\boldsymbol{\theta}$ para obtener $\boldsymbol{\theta}^{(k+1)}$, eso es

$$\boldsymbol{\theta}^{(k+1)} = \arg \max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$$

Estos dos pasos son iterados hasta que $\|\boldsymbol{\theta}^{(k+1)} - \boldsymbol{\theta}^{(k)}\|$ sea suficientemente pequeño.

Una modificación del paso-M es la siguiente: en vez de maximizar $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$ con respecto a $\boldsymbol{\theta}$, se busca algún $\boldsymbol{\theta}^{(k+1)}$ tal que $Q(\boldsymbol{\theta}^{(k+1)}, \boldsymbol{\theta}^{(k)}) \geq Q(\boldsymbol{\theta}^{(k)}, \boldsymbol{\theta}^{(k)})$. Tal algoritmo es llamado EM Generalizado (EMG), y también es garantizado que converge.

El Algoritmo EM asume que es más fácil maximizar la log-verosimilitud completa, $\ell_c(\boldsymbol{\theta} | \mathbf{Y}_c)$, que la log-verosimilitud incompleta $\ell(\boldsymbol{\theta} | \mathbf{Y}_{obs})$; cuando $\ell_c(\boldsymbol{\theta} | \mathbf{Y}_c)$ en si es una función lineal de los datos perdidos, \mathbf{Y}_{miss} , el paso-E se simplifica aún más: este consiste en reemplazar los valores perdidos por sus esperanzas condicionales en la función log-verosimilitud completa $\ell_c(\boldsymbol{\theta} | \mathbf{Y}_c)$. Otra de las ventaja de este algoritmo es que cada iteración es garantizada que incrementa la log-verosimilitud $\ell(\boldsymbol{\theta} | \mathbf{Y}_{obs})$, eso es,

$$\ell(\boldsymbol{\theta}^{(k+1)} | \mathbf{Y}_{obs}) \geq \ell(\boldsymbol{\theta}^{(k)} | \mathbf{Y}_{obs}),$$

con igualdad si y sólo si $Q(\boldsymbol{\theta}^{(k+1)}, \boldsymbol{\theta}^{(k)}) = Q(\boldsymbol{\theta}^{(k)}, \boldsymbol{\theta}^{(k)})$. Si la sucesión $\{\boldsymbol{\theta}^{(k)}\}$ converge, esta converge a un máximo local o a un punto silla de $\ell(\boldsymbol{\theta} | \mathbf{Y}_{obs})$, dependiendo del valor inicial seleccionado $\boldsymbol{\theta}^{(0)}$. Este también tiene una interpretación particularmente simple y útil cuando los datos completos \mathbf{Y}_c tienen una distribución proveniente de la *familia exponencial regular* definida por

$$f(\mathbf{Y}_c | \boldsymbol{\theta}) = b(\mathbf{Y}_c) \exp[s(\mathbf{Y}_c)\boldsymbol{\theta}/a(\boldsymbol{\theta})]$$

donde $\boldsymbol{\theta}$ denota un vector de parámetros ($d \times 1$)-dimensional, $s(\mathbf{Y}_c)$ denota un vector de estadísticos suficientes de *datos-completos*, y a y b son funciones de $\boldsymbol{\theta}$ y \mathbf{Y}_c , respectivamente.

Por otro lado, una de las desventaja de este algoritmo es que la convergencia puede ser muy lenta si la información perdida es una fracción grande, ya que su convergencia es lineal con tasa de convergencia proporcional a la máxima fracción de información perdida. En algunos problemas el paso-M es difícil (e.g. no se cuenta con una forma cerrada) por lo que su simplicidad teórica no se ve reflejada en la práctica. No obstante, hay dos tipos de extensiones de este algoritmo que a menudo pueden evitar ese y otros problemas.

Para un interés más amplio y en detalle del Algoritmo EM puede consultar el libro de McLachlan y Krishnan [29], y por supuesto [8].

Ejemplo B.1 (Modelo Exponencial). Considere el Ejemplo 2.3 en el que se tiene una muestra censurada de datos de supervivencia provenientes de una distribución exponencial, denotada por

$$(t_1, \delta_1)^\top, (t_2, \delta_2)^\top, \dots, (t_n, \delta_n)^\top$$

donde $t_i = \min\{Y_i, C_i\}$ y $\delta_i = I\{Y_i \leq C_i\}$ su indicadora de censura. Como ya se vio, no hay necesidad de un cálculo iterativo para $\hat{\theta}$. No obstante, este simple caso de datos incompletos servirá para ilustrar la manera en que trabaja el algoritmo EM. Sea $\mathbf{Y}_c = (\mathbf{Y}_{obs}, \mathbf{Y}_{miss})$ el conjunto de datos-completos, donde

$$\mathbf{Y}_{obs} = \{(t_i, \delta_i) : i = 1, 2, \dots, n\}$$

denota la información observada de la muestra (datos incompletos), y

$$\mathbf{Y}_{miss} = \{y_i : i = r + 1, \dots, n\}$$

el conjunto de datos perdidos (i.e. los tiempos de supervivencia censurados). Nos referimos a \mathbf{Y}_c como el conjunto de datos completos en el sentido de que se asume que no hay tiempos de supervivencia censurados, por lo que \mathbf{Y}_{miss} se considera como un conjunto de observaciones no-censuradas, para posteriormente calcular su valor esperado. La función log-verosimilitud completa es

$$\ell(\theta | \mathbf{Y}_c) = -n \log \theta - \theta^{-1} \left\{ \sum_{i=1}^r y_i + \sum_{i=r+1}^n y_i \right\} \quad (\text{B.3})$$

Note que $\ell_c(\theta | \mathbf{Y}_c)$ pertenece a una familia exponencial regular, además de la propiedad de la pérdida de memoria de la distribución exponencial. No obstante, se procederá sin hacer uso de esas propiedades. Entonces, dado un valor inicial $\theta^{(0)}$ del estimador de θ , en la $(k+1)$ -ésima iteración:

- **Paso-E.** Sea $\theta^{(k)}$ el valor actual de θ . Ya que $\ell_c(\theta | \mathbf{Y}_c)$ en (B.3) es una función lineal del conjunto de datos \mathbf{Y}_{miss} , este paso se reduce a calcular la esperanza condicional de cada $y_i \in \mathbf{Y}_{miss}$, dado $\theta^{(k)}$ y la información observada de y_i que es $\delta_i = 0$, lo cual es equivalente a que $Y_i > c_i$ y $c_i = t_i$. Se puede verificar fácilmente que la función de densidad condicional de Y_i dado que $Y_i > c_i$ es

$$f(y_i | Y_i > c_i) = \frac{1}{\theta} \exp\{-(y_i - c_i)/\theta\}, \quad y_i \in (c_i, \infty).$$

Luego, para $i = r + 1, \dots, n$, se tiene que

$$\begin{aligned} \tau_i^{(k)} = E[Y_i | \{Y_i > c_i\}, \theta^{(k)}] &= \int_{c_i}^{\infty} y_i f(Y_i = y_i | Y_i > c_i, \theta = \theta^{(k)}) dy_i \\ &= c_i + \theta^{(k)}. \end{aligned}$$

Así,

$$\begin{aligned}
 Q(\theta, \theta^{(k)}) &= E[\ell_c(\theta | \mathbf{Y}_c) | \mathbf{Y}_{obs}, \theta^{(k)}] \\
 &= -n \log \theta - \theta^{-1} \left\{ \sum_{i=1}^r t_i + \sum_{i=r+1}^n E[Y_i | Y_i > c_i, \theta^{(k)}] \right\} \\
 &= -n \log \theta - \theta^{-1} \left\{ \sum_{i=1}^r t_i + \sum_{i=r+1}^n \tau_i^{(k)} \right\}. \tag{B.4}
 \end{aligned}$$

- **Paso-M.** Este paso consiste en maximizar a (B.4) con respecto a θ para obtener $\theta^{(k+1)}$. Resolviendo $dQ(\theta, \theta^{(k)})/d\theta = 0$ con respecto a θ , y poniendo $\theta = \theta^{(k+1)}$ se obtiene

$$\theta^{(k+1)} = \frac{1}{n} \left\{ \sum_{i=1}^r t_i + \sum_{i=r+1}^n \tau_i^{(k)} \right\}.$$

El algoritmo EM es iterado hasta que $|\theta^{(k+1)} - \theta^{(k)}|$ sea suficientemente pequeña. La sucesión obtenida $\{\theta^{(k)}\}$ converge al EMV $\hat{\theta} = \sum_{i=1}^n t_i/r$ dado en (2.29), para $r < n$.

□

Código en R. Algoritmo EM para un Modelo Exponencial.

```

EM<-function(T.i,delta.i,tolerancia,theta.inicial){
  n<-length(T.i)
  tau<-array(1:n)*0
  theta.actual<-theta.inicial
  theta<-theta.actual
  theta.k<-0
  iter<-0
  while(abs(theta.k-theta)>tolerancia){
    E.Ti<-PasoE(T.i,delta.i,theta.actual,n,tau)
    theta<-PasoM(T.i,delta.i,E.Ti,n)
    theta.k<-theta.actual
    theta.actual<-theta
    iter<-iter+1
  }# fin while
  cat("núm. iteraciones:",iter,"\nEMV de theta:",theta.actual)
  } #fin EM

PasoE<-function(Ti,deltai,theta.actual,n,tau){
  for(i in 1:n){ tau[i]<-Ti[i]+theta.actual
                 tau*(1-deltai) }
  }

PasoM<-function(Ti,deltai,E.Ti,n){
  theta.est<-(1/n)*(sum(Ti*deltai)+ sum((1-deltai)* E.Ti))
  }

```

BIBLIOGRAFÍA

- [1] Collett, D. (1994). *Modelling Survival Data in Medical Research*. Chapman and Hall.
- [2] Cox, D.R. (1959). The analysis of exponentially distributed lifetime with two types of failure. *Journal of the Royal Statistical Society, Series B*, **21**: 411-421.
- [3] Cox, D. R., (1972a). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B*, **34**: 187-220.
- [4] Cox, D. R. (1975). Partial likelihood. *Biometrika*, **62**:269-276.
- [5] Crowder, M.J. (2001). *Classical competing risks*. Chapman & Hall/CRC: Boca raton.
- [6] Escarela, G. y Bowater, R.J. (2008). Fitting a Semi-Parametric Mixture Model for Competing Risks in Survival Data. *Communications in Statistics: Theory and Methods*, **37(2)** : 277 – 293.
- [7] Escarela, P.G.(2006). Notas de docencia: *Modelos de Supervivencia*. Universidad Autónoma Metropolitana, Unidad Iztapalapa, División de Ciencias Básicas e Ingeniería, Departamento de Matemáticas.
- [8] Dempster, A.P., Laird, N.M., Rubin D.B. (1977), Maximum Likelihood from Incomplete Data via the EM Algorithm, *Journal of the Royal Statistical Society. Series B (Methodological)*, **39** : 1 – 38.
- [9] Farewell, V. T. (1982), The Use of Mixture Models for the Análisis of Survival Data whit Long-Term Survivors, *International Biometric Society*, **38-4**: 1041-1046.
- [10] Gail, M. (1975), A review and critique of some models used in competing risks analysis, *Biometrics*, **31**: 209-222.
- [11] Gaynor, J.J., Feuer, E.J., Tan, C.C., Wu, D.H., Little, C.R., Straus, D.J., Clarkson, B.D. y Brennan, M.F. (1993). On the use of cause-specific failure and conditional failure probabilities: Examples from clinical oncology data, *Journal of the American Statistical Association*, **88**: 400-409.

- [12] Gill, R. and S. Johansen (1990). A survey of product-integration with a view towards application in survival analysis. *Annals of Statistics*, **18**: 1501-1555.
- [13] Gooley, T. A., Leisenring, W., Crowley, J. y Storer B.E. (1999). Estimation of failure probabilities in the presence of competing risks: New presentations of old estimators. *Statistics in Medicine*, **18**: 695-706.
- [14] Haesook, T. K. (2007), Cumulative Incidence in Competing Risks Data and Competing Risks Regression Analysis. *American Journal of the Association for Cancer Research*.
- [15] Hosmer and Lemeshow (1999). *Applied Survival Analysis: regression modeling of time to event*. Wiley Series in Probability and Statistics.
- [16] Kalbfleisch J.D. & Prentice R.L. (2002). *The Statistical Analysis of Failure Time Data*. Second Edition, John Wiley & Sons: New York.
- [17] Kaplan E. L. y Meier P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* , **53**: 457-481.
- [18] Katsahian S., Porcher R., Jean-Yves M. y Chevret S. (2004), The graft-versus-leukaemia effect after allogeneic bone-marrow transplantation: assessment through competing risks approaches. *Statistics in Medicine*. **23** : 3851-3863.
- [19] Klein, John P. y Moeschberger, Melvin L. (1997). *SURVIVAL ANALYSIS: techniques for Censored and Truncated data*. Springer-Verlag New York.
- [20] Johnson, N.L. and Kotz, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions*. John Wiley & Sons, New York.
- [21] Kleinbaum, D.G.(1996). *Survival Analysis, A Self-Learning Text*. Springer-Verlag, Berlin -Heidelberg- New York.
- [22] Lagakos, S. (1979a), General right censoring and its impact on the analysis of survival data. *Biometrics*,**35**:139-156.
- [23] Larson, M. G. y Dinse, G. E. (1985). A mixture model for the regression analysis of competing risks data. *Applied Statistics*. Addison-Wesley, **34**: 201-211.
- [24] Lawless, J. F. (2003), *Statistical models and methods for lifetime data*, 1ra ed. Wiley-Interscience: Hoboken NJ.
- [25] Lindqvist H (2006), *A review of Competing Risks*, Department of Mathematical Sciences, Norwegian University of Science and Technology.
- [26] McLachlan G. y Peel D. (2000), *Finite Mixture Models*. John Wiley & Sons, Inc.
- [27] McLachlan G. y Krishnan T. (2008), *The EM algorithm and Extension*. 2nd ed., John Wiley & Sons, Inc.

- [28] Miller, Rupert G. (1997), *Survival Analysis*. John Wiley & Sons
- [29] Moeschberger M L and Klein J. P. (1995). Statistical methods for dependent competing risks. *Lifetime Data Analysis*. **1**:195-204.
- [30] Pepe, M. S. y Mori M. (1993). Kaplan-Meier, marginal or conditional probability in summarizing competing risks failure time data? *Statistic in Medicine*, **21**: 737-751
- [31] Peter J. Smith (2002). *Analysis of Failure and Survival Data*. Chapman & Hall/CRC.
- [32] Peterson A. V. (1976). Bounds for a joint distribution function with fixed sub-distribution functions: application to competing risk. *Proceedings of National Academy of Sciences USA*, **73**:11-13.
- [33] Prentice, R.L., Kalbfleisch, J.D., Peterson, A.V., Flournoy, N., Farewell, V.T. y Breslow, N.E. (1978). The analysis of failure times in the presence of competing risks, *Biometrics*, **34**: 541-554.
- [34] Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proceedings of National Academy of Sciences USA*, **72**: 20-22.
- [35] UCLA Academic Technology Services, Statistical Computing.
Sitio-Web: <http://www.ats.ucla.edu/stat/>.
- [36] Wilde, Ivan F., *Measure, Integration and Probability*, Mathematics Department, King's College London. Class note.
Sitio-Web: <http://homepage.ntlworld.com/ivan.wilde/notes/>.



UNIVERSIDAD AUTÓNOMA METROPOLITANA
UNIDAD - IZTAPALAPA

DIVISIÓN DE CIENCIAS BÁSICAS E INGENIERIA
DEPARTAMENTO DE MATEMÁTICA

INFERENCIA PARA MODELOS DE SUPERVIVENCIA
DE UN SOLO EVENTO Y EXTENSIONES PARA
MODELOS DE RIESGOS COMPETITIVOS

T E S I S

QUE PARA OBTENER EL GRADO DE:

MAESTRO EN CIENCIAS

PRESENTA:

MARCO ANTONIO MATADAMAS SEGURA

Gabriel Escarela Pérez

Asesor de Tesis:

DR. GABRIEL ESCARELA PÉREZ

México, D.F., Abril de 2010