

**BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE  
PUEBLA**

*Facultad de Ciencias Físico Matemáticas  
Postgrado en Ciencias Matemáticas*

**ANÁLISIS DE SUPERVIVENCIA CON PUNTOS DE  
CAMBIO EN LA FUNCIÓN DE RIESGO**

**T E S I S**

*Que para obtener el grado de:*

**Maestro en Ciencias**

*Presenta:*

**GUADALUPE YOANNA ARENAS MARTINEZ**

*Director de tesis:*

*DR. FRANCISCO SOLANO TAJONAR  
DRA. HORTENSIA REYES CERVANTES*

*PUEBLA, PUE.*

*JULIO 2014*

*Dedico esta tesis con todo mi cariño a:  
Esther, Eloy y Andres*

# ***Agradecimientos***

Le doy gracias a Dios por haberme dado la oportunidad de llegar a este momento tan importante de mi vida.

Agradezco a mis padres, Eloy y Esther porque siempre han estado a mi lado apoyándome en cada paso de mi vida, pero sobre todo, por el amor que siempre me han brindado, sin ellos este logro no sería posible. Gracias papis los quiero mucho.

Andrés, eres el mejor hermano que dios me pudo mandar; gracias porque con tu inmensa alegría haces que mi vida sea muy feliz, por ese cariño tan especial, la confianza y ser tan lindo conmigo. Te quiero.

En esta etapa de mi vida me acompañaron ángeles muy especiales a los cuales puedo llamar amigos; muchas gracias, porque ustedes fueron parte importante para poder culminar esta etapa, gracias por los consejos, la confianza, el apoyo incondicional y por estar siempre conmigo.

Gracias al Dr. Francisco Tajonar y Dra. Hortensia Reyes Cervantes, por haber aceptado dirigir este trabajo de tesis; pero sobre todo gracias por sus consejos, apoyo, paciencia, tolerancia, y por brindarme su amistad. Muchas gracias.

Gracias al Dr. Hugo Adán, Dra. Gladys Linares, Dr.Fernando Velasco y Dr. Oscar Palmeros, por haber aceptado revisar este trabajo. Gracias a sus comentarios, aportaciones y correcciones, porque fueron de vital importancia para poder realizar esta tesis.

Quiero agradecer al Consejo de Nacional de Ciencia y Tecnología (CONACYT), por el apoyo económico brindado para poder realizar mis estudios de maestría.

Finalmente agradezco a todas las personas que lean este trabajo, porque con ese hecho, ya forman parte de él.

Yoanna Arenas M.

# Introducción

El análisis de supervivencia es una línea de investigación de la estadística inferencial que nace de una teoría paralela llamada *teoría de confiabilidad*, la cual designa la probabilidad de que un sistema cumpla satisfactoriamente con la función para el que fue diseñado, durante un determinado período y en condiciones específicas de operación. Así, un evento que interrumpa este funcionamiento se denomina falla.

El desarrollo de las técnicas para el análisis de confiabilidad de componentes, equipos y sistemas ha estado asociado con el desarrollo de tecnologías complejas y de alto riesgo, tales como el campo de la aeronáutica, la militar y el nuclear, (ver [14], [4]).

El estudio del análisis de supervivencia juega un papel muy importante en la medicina; los datos proporcionados por los resultados de estudios clínicos se expresan en ciertas ocasiones en términos de supervivencia, también es aplicable para poder determinar los patrones de riesgo de la enfermedad, u otros eventos en el tiempo. Esta técnica no queda limitada al estudio de *vida* o *muerte*, sino también estudia situaciones en las que el interés es analizar el tiempo que transcurre desde un cierto origen y hasta que sucede un evento de interés, (ver [2]).

En general, los estudios que evalúan la supervivencia de una determinada situación, presentan características particulares que determinan la metodología que puede usarse en cada caso. El problema más común corresponde al

## II

análisis de la supervivencia de un paciente, registrada o medida a partir de un determinado instante de tiempo  $t \geq 0$ .

El análisis de supervivencia implica el seguimiento de los individuos a lo largo del tiempo  $t$ . Es común que durante el periodo de observación se presenten condiciones que compliquen el estado de los individuos. De esta manera, un evento favorable es observar de manera exacta el tiempo de ocurrencia del suceso de interés. En términos generales el análisis de supervivencia tiene como objetivo estudiar el tiempo de vida (o falla) del evento que se está analizando.

Las funciones de riesgo son una herramienta muy importante en supervivencia, ya que estas describen la posibilidad de falla en un momento determinado, a través de ellas es posible determinar una distribución del tiempo de falla. La función de riesgo juega un papel importante en la supervivencia ya que por lo general esta función puede aumentar o disminuir, en intervalos de tiempo muy pequeños, estos tipos de cambio son comunes en los estudios médicos como en enfermedades de cáncer, trasplantes de corazón o riñón, donde los pacientes tienen un riesgo cada vez mayor durante el periodo de prácticas y una disminución de riesgos después de este periodo de adaptación. En un estudio de la recuperación de cáncer de mama, se ha observado que se produce la máxima mortalidad después de tres años y luego disminuye lentamente durante un periodo de tiempo fijo ver [10].

El tipo de cambios que se presentan en la función de riesgo son importantes pues si estos cambios son conocidos y se estima cuando pueden ocurrir se ayudará a mejorar el estudio realizado tomando decisiones adecuadas, uno de estos casos, es el punto de cambio en los modelos de duración, estos asumen una función con tasas de riesgo diferentes que cambian en algunos puntos en muy poco tiempo, estos puntos son llamados puntos de cambio, los cuales se requieren determinar y estimar. En la literatura los modelos más estudiados son aquellos que suponen un sólo punto de cambio (ver [18]), es importante

considerar a la función de riesgo y determinar el número de puntos de cambio ya que esta puede tener más cambios (ver [12], [11]).

De lo anterior, tenemos que el objetivo principal del trabajo de tesis es analizar el modelo constante por tramos con múltiples puntos de cambio, que pueden ayudar a modelar los cambios en los tratamientos, por ejemplo el del cáncer.

Para poder realizar un buen análisis de tiempos de vida es necesario lo siguiente:

- \* Tener bien definido el evento de interés.
- \* Definir de forma apropiada el origen o inicio del estudio, es decir, indicar el comienzo del análisis.
- \* Definir la escala de tiempo.

Tomando en cuenta estas consideraciones se puede llegar a tener un buen resultado del análisis que se está efectuando, sin olvidar el hecho de que los datos tienen que ser representativos de la población bajo estudio.

En este trabajo se hará la selección de un modelo que se ajuste a un conjunto de datos, una vez realizada la selección, se resuelve el problema de la estimación, es decir, se estima a los puntos de cambio, y cada uno de los parámetros del modelo elegido.

La estructura de la tesis es la siguiente:

- \* En el Capítulo 1 se presentan los conceptos básicos del análisis de supervivencia en el caso continuo y discreto, se describen las principales características de los modelos clásicos. A la par, se describe el concepto de análisis secuencial. Finalmente se discute la prueba tipo Wald, la cual es utilizada para determinar los puntos de cambio del modelo elegido.

#### IV

- \* En el Capítulo 2 se desarrolla el análisis de supervivencia para un modelo constante y un modelo lineal por tramos con múltiples puntos de cambio. Se presentan los resultados de las pruebas de hipótesis realizadas, la característica importante es que en todas las pruebas se preserva el error tipo I.

Para poder trabajar con la función de densidad del modelo exponencial es necesario optimizar su función de *log-verosimilitud*, la cual está basada en el método simplex de Nelder Mead, que es descrito en este capítulo.

- \* En el Capítulo 3 se ilustra una simulación para el modelo constante por tramos con múltiples puntos de cambio, y se encuentra al modelo que mejor se ajusta al conjunto de datos obtenidos.
- \* Al final se presentan las conclusiones obtenidas a lo largo del desarrollo de este trabajo de tesis.

# Índice

<b>1. Conceptos básicos</b>	<b>1</b>
1.1. Modelos discretos . . . . .	5
1.2. Modelos paramétricos clásicos . . . . .	7
1.2.1. Distribución Normal . . . . .	8
1.2.2. Distribución Exponencial . . . . .	9
1.2.3. Distribución Weibull . . . . .	11
1.2.4. Distribución de Valores Extremos . . . . .	12
1.2.5. Distribución Log-logística . . . . .	14
1.2.6. Distribución Gama . . . . .	15
1.2.7. Distribución Gama generalizada . . . . .	16
1.3. Censura . . . . .	17
1.3.1. Censura por la derecha . . . . .	18
1.3.2. Censura tipo I . . . . .	18
1.3.3. Censura aleatoria independiente . . . . .	19
1.3.4. Censura tipo II . . . . .	20
1.4. Análisis secuencial . . . . .	22
1.5. Prueba chi-cuadrada . . . . .	22
<b>2. Modelo constante y lineal por tramos</b>	<b>25</b>
2.1. Formulación de supervivencia . . . . .	25
2.2. Generalización de censura por la derecha . . . . .	32
2.3. Optimización . . . . .	34
2.3.1. Multiplicadores de Lagrange . . . . .	35
2.3.2. Método Simplex . . . . .	35

2.3.3. Método de Nelder-Mead . . . . .	37
2.4. Modelo constante por tramos con puntos de cambio . . . . .	40
2.4.1. Prueba de tipo Wald para el modelo constante por tramos .	46
2.4.2. Estimación para el modelo constante por tramos . . . . .	47
2.5. Modelo lineal por tramos con puntos de cambio . . . . .	47
2.5.1. Prueba de tipo Wald para el modelo lineal por tramos . . .	53
2.5.2. Estimación para el modelo lineal por tramos . . . . .	53
2.6. Preservando el error tipo I . . . . .	54
<b>3. Simulación</b>	<b>57</b>
<b>Conclusiones</b>	<b>67</b>
<b>A. Algoritmo</b>	<b>69</b>
<b>Bibliografía</b>	<b>73</b>

# Capítulo 1

## Conceptos básicos

Se considera que el tiempo de vida de un objeto, persona o grupo (de personas) se estudia a través de una variable aleatoria  $T$ . Más específicamente,  $T$  es una variable aleatoria continua, discreta o mixta, no negativa, que representa el tiempo de vida de un individuo que se encuentra en la población de estudio. Por lo que esta variable aleatoria está definida sobre el intervalo de tiempo  $[0, \infty)$ , y se denota por  $f(t)$  a la función de densidad de probabilidad de  $T$ .

**Definición 1.1.** *La probabilidad de que un individuo sobreviva hasta un tiempo  $t$  está definida como*

$$S(t) = P(T \geq t) = \int_t^{\infty} f(x)dx, \quad t \geq 0. \quad (1.1)$$

$S(t)$  es la función de supervivencia de  $T$  e indica la probabilidad de que el individuo sobreviva al menos hasta el tiempo  $t$ .

**Observación 1.2.** *En términos de la función de distribución de  $T$ , que es denotada por  $F(t)$ , se tiene que la función de supervivencia se expresa como*

$$S(t) = 1 - P(T \leq t) = 1 - F(t), \quad t \geq 0. \quad (1.2)$$

## 2

En los casos donde los sistemas están dados bajo ciclos o tiempos de vida de fabricación de objetos,  $S(t)$  representa a la función de confiabilidad, donde

\*  $S(t)$  es una función continua, monótona decreciente.

\*  $S(0) = 1$ .

\*  $S(\infty) = \lim_{t \rightarrow \infty} S(t) = 0$ .

**Nota 1.3.** En el caso donde  $S(\infty) > 0$  indicará que el individuo no muere o no falla, estos son tratados como casos especiales, los cuales no serán vistos en este trabajo, ver [8].

Otra función que es muy importante en el análisis de tiempos de vida es la *función de riesgo*, esta ayuda a determinar la probabilidad de falla en un determinado tiempo  $t$ , ver [8].

**Definición 1.4.** La función de riesgo asociada al tiempo de vida o falla de  $T$  está dada por

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)}, \quad t \geq 0. \quad (1.3)$$

con  $F(t) < 1$ .

La función de riesgo especifica la tasa de falla instantánea en el tiempo  $t$ , para comprender este resultado, observe que la función de riesgo no es otra cosa que una razón de cambio, para ver esto se hace uso de la probabilidad condicional, esto es

$$\begin{aligned} h(t) &= \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T \leq t + \Delta t \mid T \geq t)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T \leq t + \Delta t)}{\Delta t P(T \geq t)} \\ &= \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t S(t)} \\ &= \frac{f(t)}{1 - F(t)} \\ &= \frac{f(t)}{S(t)}. \end{aligned} \quad (1.4)$$

Un primer resultado que se obtiene para la función de riesgo es el siguiente

**Teorema 1.5.** Si  $f(t) = -S'(t)$  entonces

$$i) \quad h(t) = -\frac{d \ln S(t)}{dt}, \quad (1.5)$$

$$ii) \quad \ln S(x)|_0^t = -\int_0^t h(x)dx, \quad (1.6)$$

para  $t > 0$ .

**Demostración.**

Haciendo uso de (1.4) se obtiene que la primera parte es

$$-\frac{d \ln S(x)}{dx} = -\left[ -\frac{S'(x)}{S(x)} \right] = \frac{f(x)}{S(x)} = h(x). \quad (1.7)$$

Continuemos con la segunda igualdad, para esto, partimos del lado izquierdo y de (1.7), con lo cual se obtiene lo siguiente

$$\begin{aligned} -\int_0^t h(x) dx &= -\int_0^t -\frac{d \ln S(x)}{dx} dx \\ &= \int_0^t \frac{d \ln S(x)}{dx} dx = \ln S(x)|_0^t. \end{aligned} \quad (1.8)$$

Por (1.7) y (1.8) se obtiene lo deseado. \*

Un resultado que determina una relación importante entre la función de supervivencia y la de riesgo es el siguiente

**Teorema 1.6.** Si  $S(0) = 1$  entonces

$$S(t) = \exp\left(-\int_0^t h(x)dx\right). \quad (1.9)$$

**Demostración.**

Partimos del lado izquierdo de la igualdad y utilizando la segunda relación del Teorema 1.5 se tiene que

$$\begin{aligned}
\exp\left(-\int_0^t h(x)dx\right) &= \exp\left(\ln S(x)\Big|_0^t\right) \\
&= \exp\left(\ln S(t) - \ln S(0)\right) \\
&= \exp\left(\ln S(t)\right) \\
&= S(t). \qquad \qquad \qquad *
\end{aligned}$$

**Definición 1.7.** La función de riesgo acumulativa correspondiente al tiempo de vida o de falla de  $T$  está definida por

$$H(t) = \int_0^t h(x)dx, \qquad t \geq 0. \qquad (1.10)$$

Una relación inmediata entre la función riesgo acumulativa y la de supervivencia es la siguiente.

**Corolario 1.8.** Sea  $T$  el tiempo de vida o de falla. Si  $S(t)$  y  $H(t)$  son sus funciones de supervivencia y riesgo acumulada respectivamente, entonces

$$S(t) = \exp(-H(t)), \quad t \geq 0. \qquad (1.11)$$

**Demostración.**

Para probar este resultado es conveniente hacer uso de (1.10) y de la primera igualdad del Teorema 1.5 de la siguiente manera

$$\begin{aligned}
\exp(-H(t)) &= \exp\left(-\int_0^t h(x)dx\right) \\
&= \exp\left(-\int_0^t -\frac{d \ln S(x)}{dx} dx\right) \\
&= \exp(\ln S(t)) \\
&= S(t),
\end{aligned}$$

que es lo que se deseaba probar. \*

El siguiente resultado nos indica que la función de densidad de  $T$  se puede determinar de manera única mediante la función de riesgo, la cual será de vital importancia durante el desarrollo del trabajo.

**Corolario 1.9.** Si  $f(t)$  y  $h(t)$  son las funciones de densidad y riesgo respectivamente de  $T$ , entonces

$$f(t) = h(t) \exp\left(-\int_0^t h(x)dx\right), \quad t \geq 0. \quad (1.12)$$

$$(1.13)$$

**Demostración.**

Para realizar la prueba se utilizan (1.4), (1.9) y el Teorema 1.5, entonces, haciendo uso de estos resultados se tiene que

$$\begin{aligned} h(t) \exp\left(-\int_0^t h(x)dx\right) &= -\frac{d \log S(t)}{dt} \exp\left(-\int_0^t h(x)dx\right) \\ &= -\frac{S'(t)}{S(t)} S(t) \\ &= f(t), \end{aligned}$$

que es lo deseado.

✱

### 1.1. Modelos discretos

En muchas ocasiones se tiene que los tiempos de vida  $T$ , de grupos o de medidas tienen un número de ciclos cortos, que generalmente son tratados como variables aleatorias discretas, para más detalles ver [8].

Supongamos que  $T$  puede tomar los valores  $t_1, t_2, \dots$ , con  $0 \leq t_1 \leq t_2, \dots$ , con función de probabilidad discreta dada por

$$f(t_j) = P(T = t_j), \quad j = 1, 2, \dots$$

**Definición 1.10.** Sea  $T$  una variable aleatoria discreta que determina el tiempo de vida o de falla de un individuo, la función de supervivencia para  $T$ , se define por

$$S(t) = P(T \geq t) = \sum_{j:t_j < t} f(t_j), \quad (1.14)$$

para  $j = 1, 2, 3, \dots$

Algunas propiedades de la función de supervivencia son las siguientes:

- \*  $S(t)$  es continua por la izquierda.
- \*  $S(t)$  es no decreciente.
- \*  $S(0) = 1$ .
- \*  $S(\infty) = 0$ .

**Definición 1.11.** Sea  $T$  el tiempo de vida o falla de un individuo, su función de riesgo está dada por,

$$h(t_j) = P(T = t_j | T \geq t_j) = \frac{f(t_j)}{S(t_j)}, \quad (1.15)$$

para cada  $j = 1, 2, \dots$

El siguiente resultado indica una relación existente entre la función de supervivencia y de riesgo.

**Teorema 1.12.** Si  $f(t_j) = S(t_j) - S(t_{j+1})$  entonces

$$h(t_j) = 1 - \frac{S(t_{j+1})}{S(t_j)} \quad (1.16)$$

y

$$S(t_j) = \prod_{j:t_j < t} [1 - h(t_j)]. \quad (1.17)$$

**Demostración.**

Primero se muestra (1.16), para lo cual se hace uso de la Definición 1.11 y se procede de la siguiente manera

$$\begin{aligned} h(t_j) &= \frac{f(t_j)}{S(t_j)} \\ &= \frac{S(t_j) - S(t_{j+1})}{S(t_j)} \\ &= 1 - \frac{S(t_{j+1})}{S(t_j)}, \end{aligned}$$

una vez obtenida la igualdad (1.16) se continua con la ecuación (1.17) partiendo del lado derecho se tiene que

$$\prod_{j:t_j < t} [1 - h(t_j)] = \prod_{j:t_j < t} \left[ 1 - \left( 1 - \frac{S(t_{j+1})}{S(t_j)} \right) \right] = \prod_{j:t_j < t} \frac{S(t_{j+1})}{S(t_j)} = S(t),$$

de esta manera queda demostrado el Teorema.

✱

En la sección anterior se obtuvo una relación entre la función de supervivencia y de riesgo, procediendo de manera análoga, se obtiene el siguiente resultado, teniendo en consideración que la variable aleatoria  $T$  ahora es discreta.

**Corolario 1.13.** *Sea  $T$  el tiempo de vida o de falla de un individuo. Si la función de supervivencia está dada por  $S(t_j) = \prod_{j:t_j < t} [1 - h(t_j)]$ , entonces la función de riesgo acumulativa es de la forma*

$$H(t) = -\log S(t).$$

La prueba del resultado se encuentra en [8].

## 1.2. Modelos paramétricos clásicos

Los modelos paramétricos son importante en el análisis de supervivencia, ya que si el conjunto de datos se ajusta a alguna distribución clásica se pue-

den obtener características particulares del modelo que aporten beneficios al análisis, entre las más empleadas se encuentran la distribución *Weibull*, *Exponencial* y *Gama*, en esta sección también se mencionan otras distribuciones que son usadas con menor recurrencia, puesto que en el estudio de estas funciones se complica encontrar los estimadores de sus parámetros, las funciones de riesgo, etc.

### 1.2.1. Distribución Normal

Hay muchos casos donde el tiempo de vida puede ajustarse a una distribución normal, por ejemplo el estudiar una falla debida a algunos efectos de desgaste. Esta distribución, se utiliza generalmente en la teoría de confiabilidad, para maás detalles ver [14].

Sea  $T$  el tiempo de vida o de falla de un individuo, su función de densidad de probabilidad de  $T$  está dada por

$$f(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[ \frac{-1}{2} \left( \frac{t - \mu}{\sigma} \right)^2 \right], \quad -\infty < t < \infty, \quad (1.18)$$

donde  $-\infty < \mu < \infty$ ,  $\sigma^2 > 0$ , son sus parámetros, entonces se dice que  $T$  tiene una función de densidad de probabilidad normal.

Para que el modelo sea aplicable se debe de cumplir que  $P(T < 0) = 0$ , se ha observado que la mayoría de las fallas se encuentran alrededor del tiempo promedio  $E(T) = \mu$ .

La función de supervivencia que corresponde a una distribución normal puede expresarse mediante su función de distribución acumulativa  $\Phi$  como sigue

$$\begin{aligned}
 S(t) &= P(T > t) = 1 - P(T \leq t) \\
 &= 1 - \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^t \exp\left[\frac{-1}{2} \left(\frac{x - \mu}{\sigma}\right)^2\right] dx \\
 &= 1 - \Phi\left(\frac{t - \mu}{\sigma}\right).
 \end{aligned}$$

Para tener una supervivencia alta, el tiempo de vida debe de ser considerablemente menor que la duración esperada  $\mu$ , ver [14].

### 1.2.2. Distribución Exponencial

Uno de los modelos más relevantes es aquel que analiza tiempos de vida los cuales siguen una distribución exponencial. Esta se puede definir de varias maneras pero la más simple es suponer que su función de riesgo es constante, es decir,  $h(t) = \alpha$ , donde  $\alpha$  es una constante positiva, (ver [8]).

Sea  $T$  el tiempo de vida o falla de un individuo con función de densidad de probabilidad dada por

$$f(t) = \alpha e^{-\alpha t}, \quad t > 0, \quad (1.19)$$

en este caso se dice que  $T$  tiene una distribución exponencial, con parámetro  $\alpha > 0$ .

**Teorema 1.14.** Si  $T$  tiene una función de supervivencia dada por

$$S(t) = 1 - F(t) = e^{-\alpha t} \quad (1.20)$$

entonces su función de riesgo es

$$h(t) = \frac{f(t)}{S(t)} = \alpha. \quad (1.21)$$

De esta manera se puede enunciar el siguiente resultado importante.

**Teorema 1.15.** *Sea  $T$  una variable aleatoria continua no negativa, que determina el tiempo de vida de un individuo. Entonces  $T$  tiene una distribución exponencial si y sólo si tiene una función de riesgo constante.*

**Demostración.**

$\Rightarrow$  ] Suponga que  $T$  tiene una distribución exponencial entonces

$$S(t) = 1 - F(t) = 1 - \int_0^t \alpha e^{-\alpha s} ds,$$

realizando un cambio de variable  $u = -\alpha s$  y  $du = -\alpha ds$  se tiene que

$$S(t) = 1 - \int_0^t -e^u du = 1 - (e^{-\alpha t} - 1)$$

por lo que

$$S(t) = e^{-\alpha t},$$

de (1.4) se obtiene que

$$h(t) = \frac{f(t)}{S(t)} = \frac{\alpha e^{-\alpha t}}{e^{-\alpha t}} = \alpha.$$

$\Leftarrow$  ] Se supone que  $T$  tiene una función de riesgo constante, entonces  $h(t) = \alpha$ , el objetivo es probar que  $T$  tiene una distribución exponencial, pero por el Corolario 1.9 se tiene que

$$f(t) = \alpha \exp \left[ - \int_0^t \alpha ds \right] = \alpha \exp ( - \alpha s \Big|_0^t ) = \alpha e^{-\alpha t},$$

por consiguiente  $T$  tiene una distribución exponencial. \*

Este resultado se puede interpretar como el hecho de que no hay efecto de desgaste cuando se estipula un modelo exponencial. Considere a  $\Delta t > 0$ , la probabilidad de falla o de muerte durante las siguientes  $\Delta t$  unidades de tiempo, dado que no falla en el instante  $t$ , está representada por

$$P(t < T \leq t + \Delta t | T > t).$$

Aplicando la definición de probabilidad condicional se tiene que

$$P(t \leq T \leq t + \Delta t | T > t) = \frac{e^{-\alpha t} - e^{-\alpha(t+\Delta t)}}{e^{-\alpha t}} = 1 - e^{-\alpha \Delta t}. \quad (1.22)$$

**Nota 1.16.** Esta probabilidad ya no depende de  $T$  y sólo depende de  $\Delta t$ , lo cual indica que en una ley exponencial de fallas, la posibilidad de fallar no depende de lo que ha ocurrido antes, es decir, mientras siga funcionando se considera tan bueno como al inicio.

Al desarrollar el segundo miembro de la expresión (1.22) en una serie de Maclaurin obtenemos que

$$\begin{aligned} P(t \leq T \leq t + \Delta t | T > t) &= 1 - \left( 1 - \alpha\Delta t + \frac{(\alpha\Delta t)^2}{2!} - \frac{(\alpha\Delta t)^3}{3!} + \dots \right) \\ &= \alpha\Delta t + h(\Delta t), \end{aligned}$$

donde  $h(\Delta t) = (\frac{(\alpha\Delta t)^2}{2!} - \frac{(\alpha\Delta t)^3}{3!} + \dots)$  y llega a ser despreciable para un  $\Delta t$  pequeño, por lo que la probabilidad anterior es directamente proporcional a  $\Delta t$ , ver [14].

### 1.2.3. Distribución Weibull

La distribución Weibull representa un modelo apropiado para los eventos en los cuales intervienen cierto número de componentes y cuando la falla se deba principalmente al defecto o accidente más grave, esta distribución es muy flexible para adaptarse a los datos puesto que está puede ser creciente, decreciente o constante, para más detalles ver [10].

Sea  $T$  el tiempo de vida o de falla de un individuo, se dice que tiene una distribución Weibull, si su función de densidad de probabilidad está dada por

$$f(t) = (\alpha\beta)t^{\beta-1}e^{-\alpha t^\beta}, \quad (1.23)$$

donde  $\lambda > 0$  y  $\beta > 0$  son constantes.

**Teorema 1.17.** Si  $T$  es el tiempo de vida o de falla de un individuo, con función de distribución Weibull entonces la función de supervivencia está dada por

$$S(t) = e^{-\alpha t^\beta}. \quad (1.24)$$

**Demostración.**

Observe que  $S(t) = 1 - F(t)$ , y haciendo el cambio de variable  $u = -\alpha x^\beta$ ,  $du = -\alpha\beta x^{\beta-1} dx$  se obtiene que

$$\begin{aligned} S(t) &= 1 - \left[ \int_0^t (\alpha\beta)x^{\beta-1}e^{-\alpha x^\beta} dx \right] \\ &= 1 - \left[ - \int_0^t e^u du \right] = 1 - \left[ -e^{-\alpha x^\beta} \Big|_0^t \right] \\ &= 1 - [1 - e^{-\alpha t^\beta}] \\ &= e^{-\alpha t^\beta}, \end{aligned}$$

lo que se deseaba mostrar. \*

**Teorema 1.18.** *Si  $T$  tiene una distribución Weibull entonces su función de riesgo es de la siguiente forma*

$$h(t) = (\alpha\beta)t^{\beta-1}. \quad (1.25)$$

Algunas observaciones de esta distribución son las siguientes:

\* La distribución exponencial es un caso particular de la distribución Weibull, basta tomar a  $\beta = 1$  en (1.23).

\*  $h(t)$  no es una constante pero es proporcional a las potencias de  $t$ , es decir;

para  $\beta = 1$  es constante,

para  $\beta > 1$  es creciente,

para  $0 < \beta < 1$  es decreciente.

**1.2.4. Distribución de Valores Extremos**

La distribución del valor extremo no predice el futuro, si no más bien da información de un caso extraordinario sobre los datos obtenidos, esta distribución es aplicable cuando después de un periodo de observación el paciente sufre un cambio notable, es decir, cruza el umbral o limite propuesto por el

responsable de la obsevación, (ver [10]).

Esta distribución está estrechamente relacionada con la distribución Weibull, es aquí donde surge nuestro interés, ya que, si  $T$  tiene un distribución Weibull entonces el  $\ln T$  tiene una distribución de valor extremo, ver [14].

La función de densidad de probabilidad de la distribución de valor extremo es la siguiente

$$f(t) = b^{-1} \exp \left[ \frac{t-u}{b} - \exp \left( \frac{t-u}{b} \right) \right], \quad -\infty < t < \infty. \quad (1.26)$$

**Teorema 1.19.** Si  $T$  tiene una distribución Weibull con parámetros  $\alpha$  y  $\beta$ ,  $y$   $Y = \ln T$  entonces  $Y$  tiene una distribución de valor extremo.

**Demostración.**

Empleando el Teorema de cambio de variable (ver [1]), se obtiene que

$$\begin{aligned} g(y) &= f(t) \left| \frac{dt}{dy} \right| = (\alpha\beta)(\alpha t)^{\beta-1} \exp [ - (\alpha t)^\beta ] e^y \\ &= \alpha^\beta \beta \exp [y(\beta - 1)] \exp [y - \alpha^\beta \exp (y\beta)] \\ &= \alpha^\beta \beta \exp [y\beta - \alpha^\beta \exp (y\beta)] \\ &= \alpha^\beta \beta \exp [y\beta - \alpha^\beta \exp (y\beta)], \end{aligned}$$

tomando a  $\beta = 1/b$ ,  $u = -\log \lambda$  entonces se tiene que  $e^{-u} = \lambda$  por lo que, la expresión anterior se puede expresar de la siguiente forma

$$\begin{aligned} \alpha^\beta \beta \exp (y\beta - \alpha^\beta \exp y\beta) &= \lambda^{\frac{1}{b}} \frac{1}{b} \exp \left[ \frac{y}{b} \right] - (e^{-u})^{\frac{1}{b}} e^{\frac{y}{b}} \\ &= \frac{1}{b} \exp \left[ - \frac{y}{b} \right] \exp \left[ \frac{y}{b} - \exp \left( - \frac{y-u}{b} \right) \right] \\ &= \frac{1}{b} \exp \left[ \frac{y-u}{b} \right] - \exp \left[ \frac{y-u}{b} \right], \end{aligned}$$

por lo cual se concluye que la distribución de  $Y$  es de valor extremo. \*

La función de supervivencia correspondiente a  $Y$  está dada por

$$S(y) = \exp \left[ - \exp \left( \frac{y-u}{b} \right) \right], \quad -\infty < y < \infty. \quad (1.27)$$

Por lo que la función de riesgo para  $Y$  está expresada por

$$h(y) = b^{-1} \exp\left(\frac{y-u}{b}\right), \quad -\infty < y < \infty. \quad (1.28)$$

### 1.2.5. Distribución Log-logística

Sea  $T$  el tiempo de vida o de falla de un individuo, entonces se puede tratar como un evento binario por lo cual es natural considerar el modelo lineal *log-logístico*, (ver [?]). Se dice que  $T$  tiene una distribución *log-logística* si su función de densidad de probabilidad está dada por

$$f(t) = \frac{\left(\frac{\beta}{\alpha}\right) \left(\frac{t}{\alpha}\right)^{\beta-1}}{\left[1 + \left(\frac{t}{\alpha}\right)^\beta\right]^2}, \quad t > 0, \quad (1.29)$$

donde  $\alpha > 0$  y  $\beta > 0$  son sus parámetros.

La función de supervivencia asociada a  $T$  es

$$S(t) = \left[1 + \left(\frac{t}{\alpha}\right)^\beta\right]^{-1}.$$

Haciendo uso de (1.4) se puede obtener la función de riesgo para  $T$ , la cual esta expresada de la siguiente manera

$$h(t) = \frac{f(t)}{s(t)} = \frac{\frac{\left(\frac{\beta}{\alpha}\right) \left(\frac{t}{\alpha}\right)^{\beta-1}}{\left[1 + \left(\frac{t}{\alpha}\right)^\beta\right]^2}}{\left[1 + \left(\frac{t}{\alpha}\right)^\beta\right]^{-1}} = \frac{\left(\frac{\beta}{\alpha}\right) \left(\frac{t}{\alpha}\right)^{\beta-1}}{\left[1 + \left(\frac{t}{\alpha}\right)^\beta\right]}.$$

**Nota 1.20.** La función *log-logística* obtiene su nombre, del hecho de que  $Y = \ln T$ , donde  $T$  tiene una distribución *logística* con función de densidad de probabilidad dada por

$$f(y) = \frac{b^{-1} \exp\left(\frac{y-u}{b}\right)}{\left[1 + \exp\left(\frac{y-u}{b}\right)\right]^2}, \quad -\infty < y < \infty, \quad (1.30)$$

donde  $u = \ln \alpha$ ,  $b = \beta^{-1}$ ,  $-\infty < u < \infty$  y  $b > 0$ .

**Demostración.**

Empleando el Teorema de cambio de variable se obtiene que

$$\begin{aligned} g(t) &= f(y) \left| \frac{dy}{dt} \right| = \frac{b^{-1} \exp\left(\frac{y-u}{b}\right)}{\left[1 + \exp\left(\frac{y-u}{b}\right)\right]^2} \frac{1}{t} \\ &= \frac{\beta \exp\left(\frac{y-\ln \alpha}{\beta^{-1}}\right)}{\left[1 + \exp\left(\frac{y-\ln \alpha}{\beta^{-1}}\right)\right]^2} \frac{1}{t} \\ &= \frac{\beta \left(\frac{e^y}{\alpha}\right)^\beta}{\left[1 + \left(\frac{e^y}{\alpha}\right)^\beta\right]^2} \frac{1}{t} \\ &= \frac{\beta \left(\frac{t}{\alpha}\right)^\beta}{\left[1 + \left(\frac{t}{\alpha}\right)^\beta\right]^2} \frac{1}{t} \\ &= \frac{\frac{\beta}{\alpha} \left(\frac{t}{\alpha}\right)^{\beta-1}}{\left[1 + \left(\frac{t}{\alpha}\right)^\beta\right]^2}, \end{aligned}$$

lo que nos indica que  $Y$  tiene una distribución *log-logística*, con parámetros  $\alpha$  y  $\beta$ , positivos. \*

**1.2.6. Distribución Gama**

En esta distribución se mide el tiempo hasta que se produce  $p$  veces un determinado evento, (ver [17]).

Sea  $T$  el tiempo de vida o de falla de un individuo, se dice que,  $T$  tiene una distribución *gama* si su función de densidad de probabilidad es de la siguiente forma

$$f(t) = \frac{\lambda(\lambda t)^{k-1} \exp(-\lambda t)}{\Gamma(k)}, \quad t > 0, \quad (1.31)$$

donde  $k$ ,  $\lambda$  son sus parámetros,  $\Gamma(k)$  es la función gama, ver [1],  $\lambda$  es conocido como el parámetro de escala y  $k$  es algunas veces llamado el parámetro de forma.

Esta distribución al igual que la distribución *Weibull* incluye como un caso especial a la distribución exponencial, cuando se toma  $k = 1$ .

La función de riesgo y la de supervivencia están dadas mediante la función *Gama incompleta* la cual tiene la siguiente forma

$$I(k, x) = \frac{1}{\Gamma(k)} \int_0^x u^{k-1} \exp^{-u} du.$$

La función de supervivencia está dada por

$$S(t) = 1 - I(k, \lambda t).$$

La función de riesgo para  $T$  tiene la forma siguiente

$$h(t) = \frac{\lambda(\lambda t)^{k-1} \exp(-\lambda t)}{\Gamma(k)(1 - I(k, \lambda t))}.$$

### 1.2.7. Distribución Gama generalizada

Para generalizar (1.31) considere la siguiente transformación, si  $Y = \log T$  donde  $T$  tiene una distribución gama entonces  $Y = \log T \Leftrightarrow Y = \alpha + W$  donde  $\alpha = \ln \frac{1}{\lambda}$ , (ver [17]) y  $W$  tiene una función de densidad de probabilidad dada por

$$f(w) = \frac{\exp(kw - \exp(w))}{\Gamma(k)}, \quad w > 0.$$

Se toma  $T = \alpha + \sigma w$  con  $\sigma = p^{-1} > 0$  y  $\alpha = \log(1/\lambda)$  entonces se sigue que  $Y$  tiene una densidad dada por

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

donde  $p, k$  son los parámetros de forma y  $\lambda > 0$  es un parámetro de escala.

De esta manera  $T$  tiene una distribución *Gama Generalizada*, ver [17].

La función de supervivencia y de riesgo están dadas en términos de la función gama incompleta.

La función de supervivencia para  $T$  se expresa mediante

$$S(t) = 1 - \left( \frac{I(k, (\lambda t)^p)}{\Gamma(k)} \right).$$

La función de riesgo está dada por

$$h(t) = 1 - \left( \frac{\lambda p (\lambda t)^{pk-1} \exp(-(\lambda t)^p)}{\Gamma(k)} - I(k, (\lambda t)^p) \right).$$

### 1.3. Censura

La censura es de gran ayuda para poder analizar el evento de interés, en particular en el análisis de supervivencia. Los datos del tipo evento-tiempo tienen diferentes variantes durante el periodo de observación.

**Definición 1.21.** *Los datos en los cuales solo se conocen algunas características durante ciertos intervalos de tiempo, son los que serán llamados datos censurados.*

En el estudio del análisis de supervivencia los datos se examinarán mediante las parejas  $(t_i, \delta_i)$ , donde  $\delta_i$  es conocida como la función indicadora de censura, es decir, es uno si los datos no son censurados y vale cero en otro caso, y  $t_i$  es el mínimo entre el tiempo de vida y de censura, ver [6], [5].

**Nota 1.22.** *Teniendo en consideración la definición anterior podemos concluir que si conocemos en su totalidad a los tiempos de vida durante el evento de interés, estos no serán descartados o censurados.*

Existen diferentes tipos de censura, censura a la derecha, censura a la izquierda y censura por intervalos, en esta sección sólo se mencionan algunos tipos de censura por la derecha, como lo son censura tipo I, censura tipo II y censura aleatoria independiente.

### 1.3.1. Censura por la derecha

La censura por la derecha se da en los límites inferiores de los tiempos de vida y sólo son permitidos para algunos individuos, esto puede ocurrir por varias razones, cuando el investigador establece criterios bajo los cuales descartara al individuo, al tomar una decisión de prueba de vida antes de que muera, en algún caso no planificado, etc.

Cuando una persona es estudiada bajo cierto propósito, y esta queda censurada o descartada por motivos externos al estudio, se dice que se tiene una *pérdida de seguimiento*, esto es, que se encuentran dentro de los criterios de evaluación del análisis.

Para realizar el análisis, es necesario introducir notación especial, suponga que los tiempos de vida de  $n$  individuos, son denotados por  $T_1, \dots, T_n$  y los tiempos de censura, por  $C_1, \dots, C_n$ , entonces se tiene lo siguiente.

Se define a

$$\delta_i = \begin{cases} 1, & \text{si } T_i \leq C_i, \\ 0, & \text{si } T_i > C_i. \end{cases}$$

donde  $\delta_i$  es la función indicadora de censura.

Los datos obtenidos se analizan mediante las parejas  $(T_i, C_i)$  desde  $i = 1, \dots, n$ .

### 1.3.2. Censura tipo I

La censura tipo I es empleada cuando se llevan acabo estudios sobre un periodo de tiempo específico.

La censura tipo I es aplicable cuando cada individuo tiene un tiempo  $t_i$  y un potencial fijo de censura  $\delta_i > 0$ , el cual se denota por

$$t_i = \min(T_i, C_i), \quad \delta_i = I(T_i \leq C_i), \quad (1.32)$$

donde  $T_i$  y  $C_i$  representan el tiempo de vida y de censura respectivamente.

La función de verosimilitud para la muestra con censura tipo I está basada en la función de probabilidad de  $(t_i, \delta_i)$ , para  $i = 1, \dots, n$ , la cual está dada por

$$L = f(t_i)^{\delta_i} P(T_i > C_i)^{1-\delta_i}. \quad (1.33)$$

**Observación 1.23.** Si  $C_i$  es una constante fija y  $T_i$  toma valores menores o iguales a  $C_i$  se tiene que

$$\begin{aligned} P(t_i = C_i, \delta_i = 0) &= P(T_i > C_i), \\ P(t_i, \delta_i = 1) &= f(t_i), \quad t_i \leq C_i, \end{aligned}$$

donde la segunda expresión denota la función de densidad de probabilidad para  $T_i$ .

Asumimos que los tiempos de vida  $T_1, \dots, T_n$  son estadísticamente independientes, así obtenemos que la función de verosimilitud toma la forma

$$\begin{aligned} L &= \prod_{i=1}^n f(t_i)^{\delta_i} P(T_i > C_i)^{1-\delta_i} \\ &= \prod_{i=1}^n f(t_i)^{\delta_i} S(t_{i+})^{1-\delta_i}, \end{aligned} \quad (1.34)$$

donde  $S(t_{i+}) = P(T_i > t_i)$ . En general, si  $S(t)$  es continua entonces  $S(t_{i+}) = S(t_i)$ .

### 1.3.3. Censura aleatoria independiente

El proceso de censura aleatoria es muy simple ya que con frecuencia se realiza para cada individuo con un tiempo de vida  $T$  y un tiempo de censura  $C$ , que son variables aleatorias continuas e independientes, con funciones de supervivencia y de distribución  $S(t)$  y  $G(t)$ , respectivamente. Para todos los tiempos de vida y de censura se asume independencia, por lo que  $G(t)$  no depende de ninguno de los parámetros de  $S(t)$ .

Esto nos indica que es un caso de censura tipo I

Una vez más es necesario considerar a  $(t_i, \delta_i)$  los cuales están representados mediante la expresión (1.32).

Los datos de las observaciones de individuos consisten de  $n$  parejas  $(t_i, \delta_i)$  con  $i = 1, \dots, n$ .

El mismo resultado es obtenido si  $C_i$  es adecuado para todo  $i = 1, \dots, n$ , y la función de densidad de probabilidad es obtenida para  $(t_i, \delta_i)$ .

Si  $f(t)$  y  $g(t)$  son las funciones de densidad de probabilidad de  $T_i$  y  $C_i$  entonces,

$$\begin{aligned} P(t_i = t, \delta_i = 0) &= P(C_i = t, T_i > C_i) = g(t)s(t) \\ P(t_i = t, \delta_i = 1) &= P(T_i = t, T_i \leq C_i) = f(t)G(t) \end{aligned}$$

Lo anterior se puede expresar de la siguiente manera

$$P(t_i = t, \delta_i) = [f(t)G(t)]^{\delta_i} [g(t)S(t)]^{1-\delta_i},$$

y la verosimilitud de  $(t_i, \delta_i)$ , para  $i = 1, \dots, n$ , es la siguiente

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

Dado que  $G(t)$  y  $g(t)$  no involucran parámetros de  $f(t)$ , la función de verosimilitud está dada por

$$L = \prod_{i=1}^n f(t_i)^{\delta_i} S(t_i)^{1-\delta_i}, \quad (1.35)$$

que es análoga a (1.34).

El modelo de censura aleatoria independiente es de gran utilidad en los procesos de tiempos de falla o de muerte, aunque la independencia del modelo de censura aleatoria es a menudo razonable, ver [14].

#### 1.3.4. Censura tipo II

La censura tipo II, sólo considera los  $r$  primeros tiempos de falla más pequeños  $t_{(1)} \leq \dots \leq t_{(r)}$  de una muestra aleatoria de tamaño  $n$ , donde  $r$  es un

entero entre 1 y  $n$ .

Este tipo de censura surge cuando  $n$  individuos inician un estudio en algún tiempo específico  $t$ , el estudio es terminado cuando se han obtenido los primeros  $r$  tiempos de fallas. El análisis de tiempos de vida que son planteados con censura tipo II, tiene una desventaja en la práctica, ya que en este tipo de censura se considera que el tiempo  $t_{(r)}$  es dado de manera aleatoria, lo cual no permite establecer un tiempo de inicio.

En la censura tipo II el valor de  $r$  es elegido antes de hacer la recolección de los datos, y se eligen los primeros  $r$  tiempos de falla más pequeños de la muestra aleatoria.

Para las distribuciones continuas podemos denotar al tiempo de falla más pequeño como  $T_{(1)} \leq \dots \leq T_{(r)}$  y dar el siguiente resultado.

**Teorema 1.24.** Si  $T_i$  tiene una función de densidad de probabilidad  $f(t)$  y función de supervivencia  $S(t)$ , entonces su función de densidad de probabilidad conjunta de  $T_{(1)} \leq \dots \leq T_{(r)}$  es

$$f_{t_1, \dots, t_r}(t_1, \dots, t_r) = \frac{n!}{(n-r)!} \left[ \prod_{i=1}^r f(t_i) \right] S(t_{(r)})^{n-r}. \quad (1.36)$$

donde  $(t_1, \dots, t_r)$  son los  $r$  tiempos de vida más pequeños.

**Demostración.**

Haciendo uso de la probabilidad conjunta se tiene que

$$\begin{aligned} f_{t_1, \dots, t_r}(t_1, \dots, t_r) &= n! f(t_1) \cdot f(t_2) \cdot \dots \cdot f(t_{(r)}) \frac{1}{(n-r)!} [1 - F(t_{(r)})]^{n-r} \\ &= \frac{n!}{(n-r)!} \prod_{i=1}^r f(t_i) [1 - F(t_{(r)})]^{n-r} \\ &= \frac{n!}{(n-r)!} \prod_{i=1}^r f(t_i) [S(t_{(r)})]^{n-r}, \end{aligned}$$

lo que se quería probar.

✱

## 1.4. Análisis secuencial

Los primeros métodos estadísticos secuenciales tuvieron la intención de efectuar muestreos más eficientes que los conocidos con tamaños de muestra fijo, este tipo de análisis es parte de la estadística desde hace unos 60 años.

La idea del análisis secuencial es que en lugar de tomar un número determinado de observaciones en una sola toma, se forman a los elementos de la muestra secuencialmente, en función de las observaciones mismas, terminando en el momento que sea conveniente o que se cumpla determinada condición, ver [13].

La esencia de la detección de los puntos de cambio consiste en observar un proceso aleatorio  $X_t$ , del cual se toma una muestra del tal manera que sea una sucesión con alguna característica de interés, el observador debe de tomar una decisión lo antes posible sobre el momento adecuado para detenerse, ver [7].

Se observa una sucesión  $X_1, \dots, X_n$  de variables aleatorias independientes. Para cada prueba de hipótesis  $H_i$  con  $1 \leq i \leq n$  las variables aleatorias  $X_1, \dots, X_{i-1}$  tienen alguna función de densidad  $f_0$  y las  $X_i, \dots, X_n$  tienen una función  $f_1$ .

Se denota por  $H_0$  la hipótesis nula. Entonces la prueba de razón de verosimilitud para  $H_0$  es

$$S_k = \sum_{j=1}^k \ln \frac{f_0(x_j)}{f_1(x_j)}, \quad k = 1, \dots, n, \quad (1.37)$$

la cual ayuda a indicar si se acepta o se rechaza la hipótesis nula.

## 1.5. Prueba chi-cuadrada

La prueba de razón de verosimilitud será una herramienta necesaria para poder encontrar los estimadores y definir el modelo final que mejor se ajuste a los tiempos de vida o falla.

Si  $X_1, \dots, X_n$  es una muestra aleatoria con función de densidad  $f_0$  o  $f_1$ , entonces la siguiente prueba de hipótesis

$$H_0 : X_i \sim f_0 \quad \text{Vs} \quad H_1 : X_i \sim f_1,$$

es definida por una prueba de razón de verosimilitud simple, dada por

$$\begin{aligned} \lambda &= \lambda(x_1, \dots, x_n) \\ &= \frac{\prod f_0(x_i)}{\prod f_1(x_i)} \\ &= \frac{L_0(x_1, \dots, x_n)}{L_1(x_1, \dots, x_n)} \\ &= \frac{L_0}{L_1}. \end{aligned} \tag{1.38}$$

Observe que  $L_j$  es la verosimilitud de la función de densidad  $f_{ji}$ .

La regla de decisión para la prueba de hipótesis está dada por

$$\text{Rechazamos } H_0 \text{ si } \lambda < k,$$

donde  $k$  es una constante positiva.

Si  $X_1, \dots, X_n$  es una muestra aleatoria con función de densidad conjunta  $f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta)$ , donde  $\theta \in \Theta$ , que satisfacen las condiciones de regularidad.

**Nota 1.25.** Las condiciones de regularidad son las siguientes

- \* Las funciones de densidad de probabilidad son distintas, es decir,  $\theta \neq \theta'$  entonces  $f(x_i, \theta) \neq f(x_i, \theta')$ .
- \* Las funciones de densidad de probabilidad tienen soporte común para todo  $\theta$ .

En la prueba de hipótesis

$$H_0 : \theta \in \Theta_0 \quad \text{Vs} \quad H_{1i} : \theta \in \Theta_1 = \Theta - \Theta_0,$$

se tiene que  $-2 \ln \Lambda_n$ , tiene aproximadamente una distribución  $\chi_r^2$  con  $r$  grados de libertad, ver [15], donde  $\Lambda_n$  esta dada por

$$\Lambda_n = \frac{\sup_{\theta \in \Theta_0} L(\theta; x_1, \dots, x_n)}{\sup_{\theta \in \Theta} L(\theta; x_1, \dots, x_n)}. \quad (1.39)$$

Los grados de libertad de la distribución asintótica *Chi-cuadrada* se puede considerar de dos maneras:

- \* La primera, tiene el número de parámetros especificados por  $H_0$ .
- \* La segunda, es la diferencia de las dimensiones de  $\Theta$  y  $\Theta_0$ .

El principio de la generalización de la prueba de razón de verosimilitud indica que

$H_0$  es rechazada para  $\Lambda_n$  pequeña.

Sin embargo,  $-2 \log \Lambda_n$  aumenta a medida que  $\Lambda_n$  decrece, por esta razón en este trabajo se realiza una prueba equivalente a la de razón de verosimilitud (1.39) y rechazamos  $H_0$  para  $-2 \log \Lambda_n$  grande.

Para un  $\alpha$  dado, una prueba está dada por

$$\text{Rechazamos } H_0 \text{ si y sólo si } -2 \log \Lambda_n > \chi_{1-\alpha}^2(r).$$

Donde  $\chi_{1-\alpha}^2(r)$  es el  $(1 - \alpha)$  cuántil de la distribución *Chi-cuadrada* con  $r$  grados de libertad.

## Capítulo 2

# Modelo constante y lineal por tramos con múltiples puntos de cambio

En este capítulo se considera el modelo constante y lineal por tramos con múltiples puntos de cambio, sin embargo, antes de iniciar con el desarrollo de este trabajo, es conveniente analizar los conceptos de supervivencia de las distribuciones continuas, discretas y mixtas en un mismo sistema, a la par, se considera la prueba de tipo Wald y algunos métodos de optimización que serán importantes para la metodología propuesta.

### 2.1. Formulación de supervivencia

Las distribuciones continuas, discretas y mixtas las podemos tratar en un mismo sistema, para esto, vamos a introducir dos tipos de integrales que son, la *integral de Riemann-Stieltjes* y la *integral producto*, las cuales serán de gran utilidad para definir las funciones de riesgo y de supervivencia.

Sea  $G(u)$  una función no decreciente, continua por la derecha y con límite a la izquierda con un número finito de discontinuidades en un conjunto de medida de *Lebesgue* cero. Asumimos que  $g(u) = G'(u)$  existe excepto en

los puntos de discontinuidad de  $G$ , dichos puntos tienen la siguiente forma  $G(a_j) - G(a_j-) = g_j$  donde  $G(a) = \lim_{\Delta a \rightarrow 0} G(a - \Delta a)$ .

**Definición 2.1.** La integral de Riemann-Stieltjes de  $dG$  sobre el intervalo  $(a, b]$  está definida por

$$\int_{(a,b]} dG(u) = \int_a^b g(u)du + \sum_{j:a \leq a_j \leq b} g_j. \quad (2.1)$$

En general, una función de distribución  $F(t) = P(T \leq t)$  es continua por la derecha, no decreciente, con saltos en los puntos  $a_j$  para los cuales  $P(T = a_j) = f_j$  para  $f_j > 0$  y la función de densidad de probabilidad es  $f(u) = F'(u)$  en los puntos donde  $F(u)$  es continua, de esta forma

$$\begin{aligned} P(a < T \leq b) &= F(b) - F(a) = \int_{(a,b]} dF(u) \\ &= \int_a^b f(u)du + \sum_{j:a \leq a_j \leq b} f_j. \end{aligned} \quad (2.2)$$

**Observación 2.2.** Si  $F(t)$  es continua entonces no tiene discontinuidades, y si  $T$  tiene una distribución discreta entonces  $F(u)$  es una función discreta con  $f(u) = 0$  en los puntos donde es continua. Además las discontinuidades son numerables y por lo tanto, tiene medida de Lebesgue cero.

Para un tratamiento general de la función de riesgo se introduce la integral producto.

**Definición 2.3.** Sea  $a = u_0 < u_1 < \dots < u_m = b$  una partición del intervalo  $(a, b]$  con  $\Delta u_i = u_i - u_{i-1}$  y el  $\max(\Delta u_i) \rightarrow 0$  cuando  $m \rightarrow \infty$ . La función de riesgo está definida mediante la integral producto de  $dG(u)$  dada por

$$\prod_{(a,b]} [1 + dG(u)] = \lim_{m \rightarrow \infty} \prod_{i=1}^m [1 + G(u_i) - G(u_{i-1})]. \quad (2.3)$$

**Teorema 2.4.** Si  $G(u)$  es continua para todo  $u$  en  $(a, b]$  y  $dG(u) = g(u)du$  entonces

$$\begin{aligned} \prod_{(a,b]} [1 + g(u)du] &= \lim_{m \rightarrow \infty} \prod_{i=1}^m [1 + g(u_i)\Delta u_i + o(\Delta u_i)] \\ &= \lim_{m \rightarrow \infty} \prod_{i=1}^m [1 + g(u_i)\Delta u_i], \end{aligned}$$

donde  $o(x)$  es una función de medida de  $w(x)$  tal que  $\frac{w(x)}{x} \rightarrow 0$  cuando  $x \rightarrow 0$ , para más detalles de la demostración ver [8].

**Demostración.**

Haciendo uso de la Definición 2.3 se tiene que

$$\begin{aligned} \prod_{i=1}^m [1 + g(u)du] &= \prod_{i=1}^m [1 + dG(u)du] \\ &= \lim_{m \rightarrow \infty} \prod_{i=1}^m [1 + G(u_i) - G(u_{i-1})] \\ &= \lim_{m \rightarrow \infty} \prod_{i=1}^m [1 + P(T \leq u_i) - P(T \leq u_{i-1})] \\ &= \lim_{m \rightarrow \infty} \prod_{i=1}^m [1 + P(u_{i-1} \leq T \leq u_i)] \\ &= \lim_{m \rightarrow \infty} \prod_{i=1}^m [1 + g(u_i)\Delta u_i + o(\Delta u_i)] \\ &= \lim_{m \rightarrow \infty} \prod_{i=1}^m [1 + g(u_i)\Delta u_i]. \end{aligned}$$

que era lo que se deseaba probar. \*

**Corolario 2.5.** Si  $\ln [1 + g(u_i)\Delta u_i] = g(u_i)\Delta u_i + o(\Delta u_i)$  para un  $\Delta u_i$  pequeño entonces para el caso continuo se tiene que

$$\prod_{(a,b]} [1 + g(u)du] = \exp \left[ \int_a^b g(u)du \right]. \tag{2.4}$$

**Demostración.**

Tomando el logaritmo en (2.4) y aplicando el Teorema 2.4 se tiene que

$$\begin{aligned}
 \ln \prod_{(a,b]} [1 + g(u)du] &= \ln \lim_{m \rightarrow \infty} \prod [1 + g(u_i)\Delta u_i] \\
 &= \ln \lim_{m \rightarrow \infty} \sum_{i=1}^m [1 + g(u_i)\Delta u_i] \\
 &= \lim_{m \rightarrow \infty} \sum_{i=1}^m [g(u_i)\Delta u_i] \\
 &= \int_a^b g(u)du,
 \end{aligned}$$

entonces

$$\ln \prod_{(a,b]} [1 + g(u)du] = \int_a^b g(u)du,$$

de esta manera

$$\prod_{(a,b]} [1 + g(u)du] = \exp \left[ \int_a^b g(u)du \right],$$

obteniendo así el resultado. \*

**Corolario 2.6.** Si  $G(u)$  es una función que tiene saltos en los puntos  $a_j$  con  $j = 1, 2, \dots$ , de tamaño  $g_j$  entonces

$$\prod_{(a,b]} [1 + dG(u)] = \prod_{(a,b]} [1 + g(u)du] \prod_{j:a < a_j \leq b} (1 + g_j). \quad (2.5)$$

**Demostración.**

Se puede notar que

$$(1 + g(u)du)(1 + g_j) = 1 + g(u)du + g_j = 1 + dG(u).$$

Por lo que se da la igualdad siguiente

$$\prod_{(a,b]} [1 + dG(u)] = \prod_{(a,b]} [1 + g(u)du] \prod_{j:a < a_j \leq b} (1 + g_j),$$

de esta forma se tiene lo deseado. \*

Recordemos que

$$h(u) = \frac{f(u)}{S(u)},$$

es la función de riesgo para  $T$  en los puntos donde  $F(u)$  y  $S(u)$  son continuas y  $h_j = P(T = a_j | t \geq a_j)$  en los valores de tiempo discreto con saltos en los puntos  $a_j$ .

**Definición 2.7.** La función de riesgo acumulativa se define mediante la integral de Riemman-Stieltjes como

$$H(t) = \int_0^t dH(u) = \int_0^t h(u)du + \sum_{j:a_j \leq t} h_j. \quad (2.6)$$

Dada la función de riesgo acumulativa, podemos obtener la función de supervivencia, para esto es fundamental tomar una partición  $0 = u_0 < u_1 < \dots < u_m = t$  y definir

$$P(T \geq t) = \prod_{i=1}^m P(T \geq u_i | T \geq u_{i-1}). \quad (2.7)$$

Ahora,  $\Delta u_i = u_i - u_{i-1}$  y para un intervalo  $[u_{i-1}, u_i)$  suficientemente pequeño en donde puede haber a lo más un salto, se tiene que

$$\begin{aligned} P(T \geq u_i | T \geq u_{i-1}) &= 1 - \frac{P(u_{i-1} \leq T \leq u_i)}{P[T \geq u_{i-1}]} \\ &= 1 - [H(u_i) - H(u_{i-1}-)] + o(\Delta u_i). \end{aligned}$$

Por lo tanto, para  $u \in (0, t)$ , se tiene que

$$P(T \geq t) = \prod_{(0,t)} [1 - dH(u)], \quad (2.8)$$

que es la integral producto sobre el intervalo abierto  $(0, t)$ .

**Definición 2.8.** Sea  $S(t) = P(T \geq t)$ , continua por la izquierda y  $H(u)$  es continua por la derecha, entonces definimos a

$$S(t) = P(T \geq t) = \exp \left[ - \int_0^t h(u)du \right] \prod (1 - h_j),$$

como la función de supervivencia para  $T$ .

**Observación 2.9.** Para  $a < b$  tenemos que la función de supervivencia queda expresada por

$$P(T \geq b | T \geq a) = \prod_{[a,b)} [1 - dH(u)].$$

**Teorema 2.10.** Sea  $a_0 < a_1 < \dots < a_m$  valores específicos con  $a_0 = 0$  y  $a_m = \infty$  y con función de riesgo de la forma

$$h(t) = \lambda_j, \quad (2.9)$$

donde  $a_{j-1} \leq t < a_j$  y  $\lambda_j > 0$  para  $j = 1, 2, \dots, m$ , entonces  $T$  tienen una función de riesgo constante por tramos.

Los detalles de la demostración se encuentran en [8].

Este modelo puede parecer inverosímil dado que  $h(t)$  es discontinua en los puntos de corte  $a_1, \dots, a_{m-1}$  pero con un valor apropiado de  $m$  y con una selección de puntos de cambio se puede encontrar la función de riesgo y de supervivencia.

La función de riesgo constante por tramos es  $h(t) = \lambda_j$ , de (1.9) se obtiene que la función de supervivencia está dada por,

$$S(t) = \exp \left[ - \sum_{j=1}^{m-1} \lambda_j ((a_j - a_{j-1}) - \lambda_{m(t)}(t - a_{m(t)-1})) \right]. \quad (2.10)$$

Para cada intervalo  $(a_{j-1}, a_j)$  la función de riesgo toma un valor  $\lambda_j$ , por lo cual la función de densidad de probabilidad es

$$f(t) = \lambda_{m(t)} S(t), \quad (2.11)$$

que es una exponencial por tramos, ver Teorema 1.15.

Otro factor de interés es la longitud del  $j$ -ésimo intervalo, la cual se obtendrá mediante la siguiente expresión

$$\Delta_j(t) = \int_{a_{j-1}}^{a_j} I(u \leq t) du, \quad (2.12)$$

para  $j = 1, \dots, m$ , y  $I(u \leq t)$  es una función indicadora de no censura.

Haciendo uso de (2.11) y (2.12) podemos definir a la función de supervivencia como

$$S(t) = \exp \left[ - \sum_{j=1}^m \lambda_j \Delta_j(t) \right], \quad (2.13)$$

que es similar a la dada en (2.10).

La función de riesgo constante por tramos y la función de densidad de probabilidad, suelen presentar algunas dificultades al momento de analizarlas, para que estas sean más flexibles se hace uso de un *Spline*, este último consiste de polinomios constantes por tramos con cortes en los puntos de cambio  $a_1, \dots, a_{m-1}$ .

Un *Spline cúbico*  $g(t)$  definido en el intervalo  $(a_0, a_m)$ , consiste de un polinomio cúbico por tramos, asignado a  $g(t)$  donde las primeras dos derivadas son continuas en todas partes de  $(a_0, a_m)$ . En particular, son continuas en todo punto  $a_1, \dots, a_{m-1}$  que hacen referencia a los nudos en la terminología de *Spline*.

También se puede definir el *Spline* en otros ordenes (cuadráticos o lineales), pero el *Spline cúbico* es favorable ya que sólo  $m + 3$  parámetros son necesarios para especificar un *Spline* con  $m - 1$  nudos  $a_1, \dots, a_{m-1}$ . Una expresión de este *Spline* es la siguiente

$$g(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \alpha_3 t^3 + \sum_{j=1}^{m-1} \beta_j (t - a_j)_+^3, \quad (2.14)$$

donde  $x_+ = \max(x, 0)$ .

Observe que (2.14) no es una buena expresión de un *Spline* cuando en este, se desean estimar los parámetros  $\alpha_1, \dots, \alpha_3$  y  $\beta_1, \dots, \beta_{m-1}$ , por lo que una expresión más flexible está dada de la siguiente manera

$$g(t) = \sum_{i=1}^{m+3} \theta_j B_j(t),$$

donde  $B_j(t)$  son funciones base específicas, cúbicas por pedazos. Un enfoque común es considerar a  $B$  como una función de *Spline básico*.

Los *Spline* utilizados en los modelos de riesgo, consideran a  $h(t)$  o el  $\ln h(t)$ , y los que se aplican en los estudios de tiempos de vida generalmente son los *Spline cúbicos*.

Cuando los modelos manejan a  $h(t)$  o  $f(t)$  como un *Spline*, se tiene que  $H(t)$  y  $F(t)$  son fáciles de obtener, pero los parámetros de  $h(t)$  y  $f(t)$  se tienen que mantener positivos, para detalles adicionales ver [8].

## 2.2. Generalización de censura por la derecha

La idea para la aproximación general es considerar los procesos de falla y de censura sobre un tiempo determinado, para un grupo de individuos.

Se supone que  $n$  individuos tienen un tiempo  $t$ , de falla o de censura, los tiempos de vida y los de censura son discretos, por conveniencia y sin pérdida de generalidad admitimos que los valores para  $t$  están dados por  $t = 0, 1, 2, \dots$ . Se introduce la notación para el análisis de los procesos de falla y de censura en el tiempo  $t$ , con el producto usual.

- $Y_i(t) = I(T_i \geq t)$ , el individuo  $i$  no tiene censura después del tiempo  $t$  ).
- $dN_i(t) = Y_i(t)I(T_i(t) = t)$ .
- $dC_i(t) = Y_i(t)I(\text{el individuo } i \text{ es censurado en un tiempo } t )$ .

La variable  $Y_i(t)$  es llamada el indicador de riesgo.

**Observación 2.11.** *El indicador de riesgo es 1 si y sólo si el individuo  $i$  está vivo y no es censurado justo después del tiempo  $t$ .*

Las variables  $dN_i(t)$  y  $dC_i(t)$  registran las fallas observadas y los tiempos de censura respectivamente.

Definamos los vectores

- $dN(t) = (dN_1(t), \dots, dN_n(t))$
- $dC(t) = (dC_1(t), \dots, dC_n(t))$
- $H(t) = \{(dN(s), dC(s)), s = 0, 1, \dots, t - 1\}$ ,

donde  $H(t)$  representa las fallas y las censuras hasta un instante antes del tiempo  $t$ .

**Observación 2.12.** *Todos los datos observados pueden ser representados mediante*

$$Datos = ((dN(t), dC(t)); t = 0, 1, 2, \dots).$$

Hasta este momento no se han hecho suposiciones de los mecanismos de censura, en este momento se supone que la censura será tomada de la manera estándar, es decir. censura tipo I, en el análisis de tiempos de vida y se requiere que

$$P(dN(t)|H(t)) = \prod_{i=1}^n h_i(t)^{dN_i(t)} [1 - h_i(t)]^{Y_i(t)[1-dN_i(t)]}, \quad (2.15)$$

donde los mecanismos de falla para cada individuo en un tiempo  $t$  son independientes para  $t = 0, 1, 2, \dots$  es decir,

$$P(dN_i = 1|H(t)) = Y_i(t)h_i(t). \quad (2.16)$$

Se denota convencionalmente que  $0^0 = 1$  y lo usamos en (2.15), asumiendo que  $Y_i(t) = 0$ , esto no da información del individuo en el tiempo  $t$ , y el término se muestra en la verosimilitud, dado que el valor es igual a uno.

**Observación 2.13.** *El valor de  $Y_i(t)$  está determinado por  $H(t)$ .*

La ecuación (2.16) representa una probabilidad condicional, haciendo notar la independencia de los tiempos de falla y los tiempos de censura. Bajo (2.16), la probabilidad de que un individuo siga vivo o sea censurado justo en un tiempo  $t$  es  $h_i(t)$ .

\* Si la falla ocurre en algún tiempo  $t$  está será denotada por

$$dN_i(t) = I(s \leq t).$$

\* Si se censura en el tiempo  $t$  esta dado por

$$dN_i(t) = 0, \quad Y_i(s) = I(s \leq t).$$

Usando (1.11) y (1.17) podemos observar que

$$S_i(t) = \prod_{s=0}^{t-1} (1 - h_i(s)), \quad f_i(t) = h_i(t)S_i(t). \quad (2.17)$$

Por consiguiente, empleando la notación  $(t_i, \delta_i)$ , se tiene que la verosimilitud está dada por

$$L = \prod_{i=1}^n f_i(t_i)^{\delta_i} S_i(t_{i+1})^{1-\delta_i}. \quad (2.18)$$

Dado que  $S_i(t_{i+1}) = S_i(t+)$ , se tiene que la verosimilitud es exactamente igual a la dada en la forma (1.34), es decir, coincide con la verosimilitud de la censura tipo I y los otros tipos de censura, esto es, la verosimilitud se expresa siempre en términos de la función de densidad y de supervivencia.

## 2.3. Optimización

Para llevar acabo la estimación de los parámetros en el modelos constante por tramos con múltiples puntos de cambio se hará uso de algunos métodos de optimización, principalmente en el método de Nelder Mead.

### 2.3.1. Multiplicadores de Lagrange

Sean  $f : U \subset \mathbf{R}^n \rightarrow \mathbf{R}$  y  $g : U \subset \mathbf{R}^n \rightarrow \mathbf{R}$ , funciones suaves con valores en los reales. Sean  $\mathbf{x}_0 \in U$ ,  $g(\mathbf{x}_0) = c$ , y  $S_n$  el conjunto de nivel para la función  $g$  con valor  $c$ . Se supone que  $\nabla g(\mathbf{x}_0) \neq 0$ .

Si  $f$  es restringida a  $S_n$ , está tiene un máximo o un mínimo sobre  $S_n$ , en  $\mathbf{x}_0$ , entonces existe un número  $\lambda$  tal que

$$\nabla f(\mathbf{x}_0) = \lambda \nabla g(\mathbf{x}_0). \quad (2.19)$$

Este método consiste en buscar un punto  $\mathbf{x}_0$  y una constante  $\lambda$ , distinta de cero, llamada *Multiplicador de Lagrange*, tal que  $\nabla f(\mathbf{x}_0) = \lambda \nabla g(\mathbf{x}_0)$ .

En la ecuación (2.19) se dice que las derivadas parciales de  $f$  son proporcionales a las de  $g$ . Para hallar los puntos  $\mathbf{x}_0$  en los que esto ocurre, es necesario resolver el siguiente sistema de ecuaciones, ver [9];

$$\begin{aligned} \frac{\partial f}{\partial x_1}(x_1, \dots, x_n) &= \lambda \frac{\partial g}{\partial x_1}(x_1, \dots, x_n), \\ \frac{\partial f}{\partial x_2}(x_1, \dots, x_n) &= \lambda \frac{\partial g}{\partial x_2}(x_1, \dots, x_n), \\ &\vdots \\ \frac{\partial f}{\partial x_n}(x_1, \dots, x_n) &= \lambda \frac{\partial g}{\partial x_n}(x_1, \dots, x_n), \\ g(x_1, \dots, x_n) &= c, \end{aligned}$$

para  $x_1, \dots, x_n$  y  $\lambda$ .

### 2.3.2. Método Simplex

Cualquier segmento de recta que una dos puntos esquinas se llama *Arista* del conjunto de restricciones o *Simplex*.

El procedimiento para un simplex es el siguiente

Si  $f$  no toma su valor máximo en el punto esquina  $A$  entonces existe una arista que parte de  $A$  a lo largo de la cual aumenta  $f$ , el objetivo es encontrar

el vector  $x = (x_1, \dots, x_n)$  con  $n$  componentes que maximice a la función objetivo

$$f = c_1x_1 + \dots + c_nx_n,$$

donde  $c_i$  son constantes y además que satisfagan las  $m + n$  desigualdades lineales,

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &\leq b_1, \\ \vdots & \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &\leq b_m, \end{aligned} \tag{2.20}$$

donde  $x_1 \geq 0, x_2 \geq 0, \dots, x_n \geq 0$  y  $b_1 \geq 0, b_2 \geq 0, \dots, b_m \geq 0$ .

Para esto, se introducen las variables  $s_1, \dots, s_m$ , de esta forma el problema se transforma en

$$f = c_1x_1 + \dots + c_nx_n + 0s_1 + \dots + 0s_m,$$

sujetas a

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n + s_1 &= b_1 \\ \vdots & \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n + \dots + s_m &= b_m \end{aligned} \tag{2.21}$$

donde  $x_1 \geq 0, x_2 \geq 0, \dots, x_n \geq 0, s_1 \geq 0, \dots, s_m \geq 0$ ,  $s_1, \dots, s_m$  son las variables iniciales de nuestro simplex, ver [3].

Despejamos a  $s_1, \dots, s_m$  en términos de  $x_1, \dots, x_n$  y ahora las variables  $s_1, \dots, s_m$  serán llamadas variables básicas, y las variables no básicas resultan ser  $x_1, \dots, x_n$ , considerando que

$$b_1 \geq 0, b_2 \geq 0, \dots, b_m \geq 0$$

e igualando a cero las variables no básicas se obtiene el punto esquina inicial  $(0, 0, \dots, 0)$ .

### ALGORITMO SIMPLEX

1.- Se escoge a cualquier columna del sistema (2.21) de tal manera que cada elemento de esta columna sea positivo.

- 2.- Suponga que se escogió la  $j$ -ésima columna, se define al pivote como  $\frac{b_i}{a_{ij}}$ , siendo este cociente el mínimo de todos los elementos de la columna.
- 3.- Se utilizan métodos de álgebra para poder hacer cero al resto de la columna.
- 4.- Este procedimiento se realiza hasta que se encuentran los  $c_i$  que minimizan a la función.

### 2.3.3. Método de Nelder-Mead

El método simplex de *Nelder-Mead* está basado en la búsqueda directa para la minimización sin restricciones de funciones multidimensionales.

Desde su publicación en 1965, ver [16] este método se ha convertido en uno de los métodos más usados para la optimización no lineal sin restricciones. El método de *Nelder-Mead* trata de minimizar una función escalar no lineal de  $n$  variables usando solo valores de la función, sin tener ninguna información acerca de sus derivadas.

Este método parte de un espacio de dimensión  $n$ , donde se encuentran  $n$  puntos distintos, por lo que se busca construir su envolvente convexa en el espacio  $n - dimensional$ , es decir, el poliedro más simple no degenerado, formado por dichos vértices, en cada paso se va modificando el poliedro, de tal manera que los nuevos vértices de la función objetivo mejoren, ver [16].

#### ALGORITMO DE NELDER-MEAD

La deformación y variación del poliedro se hace mediante 4 operaciones básicas; reflexión, expansión, contracción y encogimiento, estos pasos llevan asociados respectivamente 4 parámetros denotados por  $\rho$ ,  $\chi$ ,  $\gamma$  y  $\sigma$ , los cuales deben satisfacer lo siguiente

$$* \rho > 0, \chi > 1, 0 < \gamma < 1,$$

$$* 0 < \sigma < 1, \chi > \rho.$$

En las siguientes líneas se ilustra como se realiza la  $k$ -ésima iteración.

Partiremos de la  $dim + 1$  que es el número de variables más uno, estos serán los puntos que formarán un simplex no degenerado de  $R^{dim}$ , si  $k = 1$  tendremos los  $dim + 1$  vértices iniciales. El objetivo de la iteración es obtener un nuevo simplex variando uno o más de los vértices que lo forman, ver [16].

\* Paso 1: ORDENACIÓN

Primero se ordenan los puntos de tal manera que los valores de la función objetivo en dichos puntos estén ordenados de forma creciente, es decir,  $x_1, \dots, x_{dim}, x_{dim+1}$  tal que  $f(x_1) \leq \dots \leq f(x_{dim}) \leq f(x_{dim+1})$ . Puesto que el problema consiste en minimizar a  $f$  entonces, se debe entender que el primer punto es el mejor, y que el último es el peor.

\* Paso 2: REFLEXIÓN

Encontrar el punto de reflexión que será el punto de simetría, el cual esta ponderado por  $\rho$ , y está dado por

$$\rho = \frac{1}{dim} \sum_{i=1}^{dim} x_i,$$

entonces tenemos que nuestro nuevo punto de reflexión está determinado por

$$x_{refle} = \bar{x} + \rho(\bar{x} - x_{dim+1}),$$

después se evalúa la función objetivo en  $f(x_{refle})$ .

CASO 1: Si el punto de reflexión supera al mejor

$$f(x_{refle}) < f(x_1).$$

\* Paso 3: EXPANSIÓN

Se calcula el punto de expansión con respecto al centro, esto se hace para ver si aún se puede mejorar más,

$$x_{expan} = \bar{x} + \chi(x_{refle} - \bar{x}).$$

Una vez más, se evalúa en la función objetivo  $f(x_{expan})$ . Si se ha mejorado, i.e., si  $f(x_{expan}) < f(x_{refle})$  entonces se acepta el punto de expansión y eliminamos al peor punto para formar el nuevo simplex.

Si no hemos mejorado, i.e.,  $f(x_{expan}) \geq f(x_{refle})$  aceptamos al punto reflejado y eliminamos al peor punto para formar el nuevo simplex. Con esto se termina la iteración.

CASO II: El punto de reflexión aún no es preferible al mejor pero si ha sido superior a alguno de los  $n$  mejores, es decir

$$f(x_1) \leq f(x_{refle}) < f(x_{dim}).$$

Se acepta el punto reflejado y eliminamos al peor punto para formar el nuevo simplex. Concluyendo de esta forma la iteración.

CASO III: El punto de reflexión no ha excedido a ninguno de los  $n$  mejores

$$f(x_{refle}) \geq f(x_{dim}).$$

\* Paso 4: CONTRACCIÓN

La contracción se hace hacia el punto reflejado o hacia el peor punto dependiendo de cual sea el mejor de los dos.

**Contracción hacia afuera**

Si el punto reflejado es mejor que el peor, i.e.,  $f(x_{refle}) < f(x_{dim+1})$ , entonces se calcula el punto

$$x_{con_f} = \bar{x} + \gamma(x_{refle} - \bar{x}).$$

Enseguida se evalúa a la función en el punto de contracción  $f(x_{con_f})$ .

Si hemos mejorado, i.e.,  $f(x_{con_f}) < f(x_{refle})$  se acepta el punto de contracción y se elimina al peor punto para formar el nuevo simplex. Con esto termina la iteración.

Si no hemos mejorado, i.e.,  $f(x_{con_f}) \geq f(x_{refle})$  se continua al paso 5.

**Contracción hacia adentro**

Si el punto reflejado iguala o empeora al peor, i.e.,  $f(x_{refle}) \geq f(x_{dim+1})$  y se obtiene el punto

$$x_{con_d} = \bar{x} - \gamma(x_{refle} - \bar{x}).$$

Se evalúa a la función en el punto de contracción  $f(x_{con_d})$ .

Si hemos mejorado, i.e.,  $f(x_{con_d}) < f(x_{dim+1})$  se acepta el punto de contracción y se elimina al peor punto para formar el nuevo simplex. Concluyendo de esta forma la iteración.

Si no se a mejorado, i.e.,  $f(x_{con_d}) \geq f(x_{dim+1})$  se va al paso 5.

\* Paso 5:

Se toman  $n$  nuevos puntos de la forma

$$v_i = x_1 + \sigma(x_i - x_1),$$

para  $i = 2, \dots, dim + 1$ .

El nuevo simplex tiene como vértices  $x_1, v_1, \dots, v_{dim+1}$ . Terminando la iteración.

Este es un algoritmo que permite minimizar a una función objetivo, y que no requiere condiciones adicionales, ver [16].

## 2.4. Modelo constante por tramos con múltiples puntos de cambio

El modelo paramétrico con un punto de cambio más simple y el que ocurre con más frecuencia en la literatura, es aquel que supone que, en una muestra aleatoria de tamaño  $n$ ,  $k$  sujetos que tienen una función de riesgo común  $\alpha_1$  y los  $(n - k)$  restantes, tienen otra función de riesgo común  $\alpha_2$ , con  $\alpha_1 \neq \alpha_2$ . Este modelo se puede expresar de la forma siguiente

$$\lambda(t) = \begin{cases} \alpha_1, & 0 \leq t < \tau, \\ \alpha_2, & t \geq \tau. \end{cases} \quad (2.22)$$

El interés del trabajo se centra en el modelo con múltiples puntos de cambio, donde todos los parámetros son desconocidos y se desean estimar, este modelo se describe a continuación.

Sean  $X_1, \dots, X_n$  que denotan los tiempos de vida o de supervivencia los cuales son independientes e idénticamente distribuidos y sean  $C_1, \dots, C_n$  los tiempos de censura, estos son independientes de los tiempos de supervivencia.

**Definición 2.14.** Sea  $T$  el tiempo de vida o de falla. Se dice que  $T$  tiene múltiples puntos de cambio si su función de riesgo  $\lambda(t)$  está dada por

$$\lambda(t) = \begin{cases} \alpha_1, & 0 \leq t < \tau_1, \\ \alpha_2, & \tau_1 \leq t < \tau_2, \\ \vdots & \vdots \\ \alpha_{k+1}, & t \geq \tau_k, \end{cases} \quad (2.23)$$

donde  $0 < \tau_1 < \dots < \tau_k$ ,  $k$  es el número de puntos de cambio en el modelo, y  $\alpha_j$  es el valor que toma la función de riesgo entre  $\tau_{j-1}$  y  $\tau_j$ , donde los puntos de cambio y el valor que toma la función de riesgo en cada tramo son desconocidos.

**Nota 2.15.** Los  $\tau_j$  pueden ser considerados como los estadísticos de orden para los puntos de cambio en la función de riesgo. Además al modelo (2.23) se le llama modelo constante por tramos con múltiples puntos de cambio.

Una vez definida la función de riesgo por tramos con múltiples puntos de cambio, se obtendrá la correspondiente función de densidad de probabilidad de  $T$ .

Para poder obtener la función de densidad de  $T$  se hará uso de un resultado previo el cual indica que si se tiene la función de riesgo se puede determinar de manera única a la función de densidad, ver Corolario 1.9.

Utilizando dicho resultado se tiene que la función de densidad de  $T$  está dada por

$$f(t) = \begin{cases} \alpha_1 \exp[-\alpha_1 t], & 0 \leq t < \tau_1, \\ \alpha_2 \exp[-\alpha_1 \tau_1 - \alpha_2(t - \tau_1)], & \tau_1 \leq t < \tau_2, \\ \vdots & \vdots \\ \alpha_{k+1} \exp[-\alpha_1 \tau_1 - \alpha_2(\tau_2 - \tau_1) - \dots - \alpha_{k+1}(t - \tau_k)], & t \geq \tau_k, \end{cases} \quad (2.24)$$

que es una función de densidad exponencial, con  $K$  puntos de cambio.

A continuación se presenta un bosquejo de la construcción de esta función de densidad, solo se hará para los dos primeros casos y para el resto se obtienen de manera análoga

\* Para  $0 \leq t < \tau_1$

$$\begin{aligned} f_1(t) &= \lambda(t) \exp \left[ - \int_0^t \lambda(u) du \right] = \alpha_1 \exp \left[ - \int_0^t \alpha_1 du \right] \\ &= \alpha_1 \exp [-\alpha_1 t]. \end{aligned}$$

\* Para  $\tau_1 \leq t < \tau_2$

$$\begin{aligned} f_2(t) &= \lambda(t) \exp \left[ - \int_0^t \lambda(u) du \right] = \alpha_2 \exp \left[ - \int_0^{\tau_1} \alpha_1 du - \int_{\tau_1}^t \alpha_2 du \right] \\ &= \alpha_2 \exp [-\alpha_1 \tau_1 - \alpha_2 (t - \tau_1)]. \end{aligned}$$

De esta manera se va construyendo la función de densidad  $f(t)$  para cada punto de cambio.

El número de fallas observadas hasta el instante de tiempo  $t$  estará dado por

$$X(t) = \sum_{i=1}^k I(T_i < t) \delta_i,$$

donde  $\delta_i$  es la función indicadora de censura.

**Observación 2.16.** Si  $X(t)$  denota el número de fallas hasta el tiempo  $t$  entonces se tiene que

$$X(\tau_j) = \sum_{i=1}^k I(T_i < \tau_j) \delta_i,$$

denotará el número de fallas observadas hasta el punto de cambio  $\tau_j$ .

Ahora, el problema es similar a un problema de análisis secuencial, de esta forma se realizará una prueba de hipótesis, si esta es rechazada, se continua

con la siguiente prueba de hipótesis. Si no se rechaza, el proceso se detiene y se concluye que se ha encontrado el modelo final.

La prueba de hipótesis es la siguiente

$$H_0 : \alpha_{k-1} = \alpha_k \quad V_s \quad H_1 : \alpha_{k-1} \neq \alpha_k.$$

Que es equivalente a probar la hipótesis de que  $\tau_{k-1} = 0$  para  $k = 2, \dots, K$ .

Extendiendo la metodología existente, se desarrolla una prueba de razón de verosimilitud, la cual será de ayuda para poder realizar la prueba de hipótesis.

El primer paso será encontrar a la función de densidad  $f_0$ , y así poder aplicar (2.26).

La función de densidad para el número de fallas hasta un determinado tiempo  $\tau_j$ , esta dada de la siguiente manera

$$f_1(X(\tau_j)) = \begin{cases} \alpha_1 \exp [X(\tau_1)], & 0 \leq X(\tau_j) < \tau_1, \\ \alpha_2, \exp [X(\tau_2) - X(\tau_1)], & \tau_1 \leq X(\tau_j) < \tau_2, \\ \vdots & \vdots \\ \alpha_{k+1} \exp [n_u - X(\tau_{k-1})], & X(\tau_j) \geq \tau_k. \end{cases} \quad (2.25)$$

Donde  $n_u$  es el número de datos no censurados.

El tipo de censura que se va a utilizar en esta situación es la censura por la derecha, por lo cual definiremos a  $f(t_1) = (T_i \wedge \tau_1)$  y  $s_1(t_i) = I(T_i > \tau_1)$ , puesto que estos son los tiempos de vida y de censura hasta un determinado tiempo  $\tau_1$ , de esta forma se tiene que la función de densidad está dada por

$$f_0(\tau_j) = \exp \left( - \alpha_i \sum_{i=1}^n f(t_1) s_1(t_i) \right).$$

Una vez que se ha establecido a  $f_0(\tau_j)$  y  $f_1(X(\tau_j))$ , se procede a calcular la *log-verosimilitud*

$$L_k = \sum_{j=1}^k \ln \frac{f_0(\tau_j)}{f_1(X(\tau_j))}. \quad (2.26)$$

Dado que la función de densidad de  $T$  está definida por tramos, se lleva a cabo el proceso de máxima verosimilitud para cada tramo.

$$\begin{aligned}
 l_1 &= \frac{e^{-\alpha_1 \sum_{i=1}^n (T_i \wedge \tau_1)}}{\alpha_1 e^{-X(\tau_1)}}, \\
 l_2 &= \frac{e^{-\alpha_2 \sum_{i=1}^n (T_i \wedge \tau_2 - \tau_1) I(T_i > \tau_1)}}{\alpha_1 e^{-(X(\tau_2) - X(\tau_1))}}, \\
 &\vdots \\
 l_k &= \frac{e^{-\alpha_k \sum_{i=1}^n (T_i \wedge \tau_{k-1}) I(T_i > \tau_{k-1})}}{\alpha_k e^{-(n_u - X(\tau_{k-1}))}},
 \end{aligned}$$

donde  $n_u$  denota el número de datos no censurados.

Por otro lado, se tiene que

$$L_k = \sum_{j=1}^k \ln \frac{f_0(x_j)}{f_1(x_j)} = \ln l_1 + \ln l_2 + \dots + \ln l_k,$$

luego se procede a calcular,  $\ln l_1 + \ln l_2 + \dots + \ln l_k$ , los cuales resultan ser

$$\begin{aligned}
 \ln l_1 &= (-\alpha_1 \sum_{i=1}^n (T_i \wedge \tau_1)) - (-X(\tau_1) \ln \alpha_1) \\
 &= X(\tau_1) \ln \alpha_1 - \alpha_1 \sum_{i=1}^n (T_i \wedge \tau_1), \\
 \ln l_2 &= (-\alpha_2 \sum_{i=1}^n (T_i \wedge \tau_2 - \tau_1) I(T_i > \tau_1)) - (-(X(\tau_2) - X(\tau_1)) \ln \alpha_2) \\
 &= (X(\tau_2) - X(\tau_1)) \ln \alpha_2 - \alpha_2 \sum_{i=1}^n (T_i \wedge \tau_2 - \tau_1) I(T_i > \tau_1), \\
 &\vdots \\
 \ln l_k &= [-\alpha_k \sum_{i=1}^n (T_i \wedge \tau_{k-1}) I(T_i > \tau_{k-1})] - [-(n_u - X(\tau_{k-1})) \ln \alpha_{k-1}] \\
 &= (n_u - X(\tau_{k-1})) \ln \alpha_{k-1} - \alpha_k \sum_{i=1}^n (T_i \wedge \tau_{k-1}) I(T_i > \tau_{k-1}).
 \end{aligned}$$

Por lo que,

$$\begin{aligned}
 L_k &= \sum_{j=1}^k \ln \frac{f_1}{f_0} = X(\tau_1) \ln \alpha_1 + [X(\tau_2) - X(\tau_1)] \ln \alpha_2 \\
 &+ \dots + [n_u - X(\tau_{k-1})] \ln \alpha_{k-1} - \alpha_1 \sum_{i=1}^n (T_i \wedge \tau_1) - \\
 &\alpha_2 \sum_{i=1}^n (T_i \wedge \tau_2 - \tau_1) I(T_i > \tau_1) - \dots - \alpha_k \sum_{i=1}^n (T_i \wedge \tau_{k-1}) I(T_i > \tau_{k-1}).
 \end{aligned} \tag{2.27}$$

Una vez que es obtenida la función *log-verosimilitud*, se procede a calcular los estimadores de máxima verosimilitud, los cuales se pueden encontrar al derivar e igualar a cero a la función antes descrita, dando como resultado

$$\begin{aligned}
 \hat{\alpha}_1 &= \frac{X(\tau_1)}{\sum_{i=1}^n (T_i \wedge \tau_1)}, \\
 \hat{\alpha}_2 &= \frac{X(\tau_2) - X(\tau_1)}{\sum_{i=1}^n (T_i \wedge \tau_2 - \tau_1) I(T_i > \tau_1)}, \\
 &\vdots \\
 \hat{\alpha}_k &= \frac{n_u - X(\tau_{k-1})}{\sum_{i=1}^n (T_i \wedge \tau_{k-1}) I(T_i > \tau_{k-1})}.
 \end{aligned} \tag{2.28}$$

Ahora se procede a verificar que la función de verosimilitud tiene una distribución *Chi-cuadrada*, para esto, vamos a sustituir los estimadores de máxima verosimilitud de los parámetros de la siguiente manera:

Cuando se tienen  $k$  puntos de cambio la función de verosimilitud resultaría ser la siguiente

$$\begin{aligned}
 L_1(\tau_1, \dots, \tau_{k-1}) &= X(\tau_1) \ln \frac{X(\tau_1)}{\sum_{i=1}^n (T_i \wedge \tau_1)} + \\
 &(X(\tau_2) - X(\tau_1)) \ln \frac{X(\tau_2) - X(\tau_1)}{\sum_{i=1}^n (T_i \wedge \tau_2 - \tau_1) I(T_i > \tau_1)} + \\
 &\vdots \\
 &(n_u - X(\tau_{k-1})) \ln \frac{n_u - X(\tau_{k-1})}{\sum_{i=1}^n (T_i \wedge \tau_{k-1}) I(T_i > \tau_{k-1})} - n_u.
 \end{aligned} \tag{2.29}$$

Considerando a  $\alpha_{k-1} = \alpha_k$  y los  $k - 1$  puntos de cambio, se tiene entonces que la función se puede expresar como

$$\begin{aligned} L_2(\tau_1, \dots, \tau_{k-1}) &= X(\tau_1) \ln \frac{X(\tau_1)}{\sum_{i=1}^n (T_i \wedge \tau_1)} + \\ &\quad (X(\tau_2) - X(\tau_1)) \ln \frac{X(\tau_2) - X(\tau_1)}{\sum_{i=1}^n (T_i \wedge \tau_2 - \tau_1) I(T_i > \tau_1)} + \\ &\quad \vdots \\ &\quad (n_u - X(\tau_{k-2})) \ln \frac{n_u - X(\tau_{k-2})}{\sum_{i=1}^n (T_i \wedge \tau_{k-2}) I(T_i > \tau_{k-2})} - n_u. \end{aligned}$$

Por lo tanto, la razón de verosimilitud queda expresada mediante

$$\begin{aligned} L_1(\tau_1, \dots, \tau_{k-1}) - L_2(\tau_1, \dots, \tau_{k-1}) &= [X(\tau_{k-1}) - X(\tau_{k-2})] \ln \hat{\alpha}_{k-1} + \\ &\quad [n_u - X(\tau_{k-1})] \ln \hat{\alpha}_k - [n_u - X(\tau_{k-2})] \ln \hat{\alpha}_{k-1}. \end{aligned}$$

Dado que el modelo reducido tiene dos parámetros menos que el modelo completo, la prueba de razón de verosimilitud asintótica nos lleva a concluir que la función de verosimilitud tiene una distribución  $\chi_2^2$ , ver [7].

### 2.4.1. Prueba de tipo Wald para el modelo constante por tramos

El interés esta basado en las pruebas de los puntos de cambio para el siguiente modelo

$$\lambda(t) = \begin{cases} \alpha_1, & 0 \leq t < \tau_1, \\ \alpha_2, & \tau_1 \leq t < \tau_2, \\ \vdots & \vdots \\ \alpha_{k+1}, & t \geq \tau_k, \end{cases}$$

y una prueba equivalente seria analizar si la falla se encuentra antes o después de estos puntos de cambio.

Para probar la hipótesis nula de que no hay puntos de cambio contra la alternativa de que hay un punto de cambio, se tiene que probar que

$$H_0 : \alpha_1 = \alpha_2 \quad \text{o} \quad H_0 : \alpha_1 - \alpha_2 = 0,$$

para lo cual se propone el uso de una prueba de tipo Wald.

Sea  $\theta' = [\alpha_1, \alpha_2, \dots, \alpha_{K+1}, \tau_1, \tau_2, \dots, \tau_K]$ , el vector de parámetros, para probar la hipótesis nula  $H_0 : C'\theta = M$ , donde  $C'$  es una matriz de  $s \times p$ ,  $s \leq p$  y  $M$  es un vector solución de  $s \times 1$ , ver [7].

La prueba de tipo Wald de la hipótesis  $H_0$  es

$$\chi_w^2 = (C'\hat{\theta} - M)' [C' \sum_{\hat{\theta}} C]^{-1} (C'\hat{\theta} - M) \sim \chi_s^2.$$

Aunque el método propuesto es un procedimiento para múltiples puntos de cambio, este proceso se realiza paso a paso, por lo que se probará una hipótesis a la vez,

$$H_0 : \alpha_1 - \alpha_2 = 0 \quad Vs \quad H_1 : \alpha_1 - \alpha_2 \neq 0.$$

Para esto se utilizó el estadístico de tipo Wald de la forma

$$X_W = \frac{(\alpha_{k-1} - \hat{\alpha}_k)^2}{Var((\alpha_{k-1} - \hat{\alpha}_k))} \sim \chi_1^2. \quad (2.30)$$

Esta prueba determinara si hay puntos de cambio en cada intervalo  $(\tau_{j-1}, \tau_j)$  para  $j = 1, \dots, k$ .

### 2.4.2. Estimación para el modelo constante por tramos

Para estimar los parámetro en el modelo constante por tramos se utilizara optimización basada en el algoritmo Simplex de *Nelder-Mead*, el cual consiste en minimizar la función negativa de *log-verosimilitud* evaluada en los estimadores de máxima verosimilitud de los  $\alpha_k$  que permite encontrar el valor de  $\tau_j$  que minimiza a la función.

## 2.5. Modelo lineal por tramos con múltiples puntos de cambio

El modelo lineal por tramos es ligeramente más exhaustivo que el modelo constante por tramos, ya que este modelo toma en consideración factores externos que influyen en el análisis de supervivencia que se está llevando

acabo, por lo cual este modelo es uno de los que se puede encontrar con más frecuencia en la práctica.

Este modelo es *log-lineal* y tiene una función de riesgo continua que varía después de cada punto de cambio, por lo que se permite una adición de co-variables en el modelo, ya que estas influyen en el tiempo de vida.

Los datos observados se pueden representar mediante las parejas  $(T_i, \delta_i)$ , donde  $T_i$  y  $\delta_i$  son como en 1.32, y con función de riesgo  $\lambda(t_i)$ . Se hará una transformación de esta función para facilitar su análisis, sea

$$\eta = \ln \lambda,$$

donde  $\eta$  es un *Spline lineal* por tramos con nudos en los puntos de cambio  $\tau_1, \dots, \tau_k$ , definido como

$$\begin{aligned} \eta(t) &= \eta(t; \alpha_0, \alpha_1, \dots, \alpha_{k+1}, \tau_1, \dots, \tau_k, \beta) \\ &= \alpha_0 + \alpha_1 t + \sum_{i=1}^k \alpha_{k+1} (t - \tau_k)_+ + \beta' \mathbf{Z}. \end{aligned}$$

Para un  $k$  fijo, sea  $x_+ = \max(0, x)$ , donde  $\mathbf{Z}$  es el vector de covariables y  $\beta$  es el vector de los coeficientes de regresión de los parámetros para el efecto de los factores externos, es decir, de las covariables.

Definamos el modelo lineal por tramos con  $K$  puntos de cambio el cual está dado por  $\eta$ , esto es,

$$\eta(t) = \begin{cases} \alpha_0 + \alpha_1 t + \beta' \mathbf{Z}, & 0 \leq t < \tau_1, \\ \alpha_0 + \alpha_1 t + \alpha_2 (t - \tau_1)_+ + \beta' \mathbf{Z}, & \tau_1 \leq t < \tau_2, \\ \vdots & \\ \alpha_0 + \alpha_1 t + \alpha_2 (t - \tau_1)_+ + \dots + \alpha_{k+1} (t - \tau_k)_+ + \beta' \mathbf{Z}, & t \geq \tau_k. \end{cases}$$

Expresando a

$$\eta(t) = \alpha_{0i} + \alpha_{1i} t + \beta' \mathbf{Z}, \quad \tau_i \leq t < \tau_{i+1},$$

donde  $\alpha_{0i} = \alpha_0 - \sum_{j=1}^{i+1} \alpha_j \tau_{j-1}$  y  $\alpha_{1i} = \sum_{j=1}^{i+1} \alpha_j$ ,  $i = 0, 1, \dots, k$  y  $\tau_0 = 0$ .

Utilizando (1.9) se obtiene que la *log-verosimilitud* para  $\eta(t)$  está dada por

$$\begin{aligned} \ln L &= \ln L(\alpha_0, \alpha_1, \dots, \alpha_{k+1}, \tau_1, \dots, \tau_k, \beta) \\ &= \sum_{i=1}^n \left[ \delta_i \eta(T_i) - \int_0^{T_i} e^{\eta(u)} du \right], \end{aligned}$$

donde  $\delta_i$  es el indicador de censura, y  $T_i = \min(X_i, C_i)$ .

La existencia de los puntos de cambio  $\tau_{k-1}$  se puede mostrar, probando la hipótesis de que  $\tau_{k-1} = 0$ , para  $k = 2, \dots, K$ . Para lo cual se hace uso de la prueba estadística de razón de verosimilitud, la cual está dada por

$$\ln L(\alpha_0, \alpha_1, \dots, \alpha_k, \tau_1, \dots, \tau_{k-1}, \beta) - \ln L(\alpha_0, \alpha_1, \dots, \alpha_{k-1}, \tau_1, \dots, \tau_{k-2}, \beta). \quad (2.31)$$

En el modelo lineal por tramos, donde  $\alpha_{k-1} \neq \alpha_k$ , la función *log-verosimilitud* tiene la siguiente forma,

$$\begin{aligned} &\ln L(\alpha_0, \alpha_1, \dots, \alpha_k, \tau_1, \dots, \tau_{k-1}, \beta) \quad (2.32) \\ &= \sum_{i=1}^n \left\{ \delta_i (\alpha_0 + \alpha_1 T_i + \alpha_2 (T_i - \tau_1)_+ + \dots + \alpha_{k+1} (T_i - \tau_k)_+ + \beta' \mathbf{Z}) \right. \\ &\quad \left. - \int_0^{T_i} \exp(\alpha_0 + \alpha_1 u + \alpha_2 (u - \tau_1)_+ + \dots + \alpha_{k+1} (u - \tau_k)_+ + \beta' \mathbf{Z}) du \right\}. \end{aligned}$$

A continuación se da un bosquejo de como obtener la función *log-verosimilitud*.

Por el momento solo se desarrolla la parte de la integral, la cual se realizará para cada punto de cambio, ya que es lineal por tramos.

Para  $0 \leq T_i < \tau_1$  se tiene que evaluar la siguiente integral

$$\int_0^{T_i} e^{\alpha_0 + \alpha_1 u + \beta' \mathbf{Z}} du, \quad (2.33)$$

para esto se realiza un cambio de variable tomando a  $v = \alpha_0 + \alpha_1 u$  y luego tenemos que (2.33) resulta ser

$$\begin{aligned} \int_0^{T_i} e^{\alpha_0 + \alpha_1 u + \beta' \mathbf{Z}} du &= \frac{1}{\alpha_1} \int_0^{T_i} e^{v + \beta' \mathbf{Z}} dv \\ &= \frac{1}{\alpha_1} e^{\alpha_0 + \beta' \mathbf{Z}} [e^{\alpha_1 T_i + \beta' \mathbf{Z}} - 1]. \end{aligned} \quad (2.34)$$

Para  $\tau_1 \leq T_i < \tau_2$  se tiene que evaluar la siguiente integral,

$$\int_{\tau_1}^{T_i} e^{\alpha_0 + \alpha_1 u + \alpha_2 (u - \tau_1)_+ + \beta' \mathbf{Z}} du, \quad (2.35)$$

para lo cual se hace de nuevo un cambio de variable tomando a  $v = \alpha_0 + \alpha_1 u + \alpha_2 (u - \tau_1)_+$  y se tiene que (2.35) resulta ser:

$$\begin{aligned} \int_{\tau_1}^{T_i} e^{\alpha_0 + \alpha_1 u + \alpha_2 (u - \tau_1)_+ + \beta' \mathbf{Z}} du &= \frac{1}{\alpha_1 + \alpha_2} \int_{\tau_1}^{T_i} e^{v + \beta' \mathbf{Z}} dv \\ &= \frac{e^{\alpha_0 + \beta' \mathbf{Z}}}{\alpha_1 + \alpha_2} [e^{\alpha_1 T_i + \alpha_2 (T_i - \tau_1)} - e^{\alpha_1 \tau_1}]. \end{aligned}$$

Este mismo procedimiento se realiza para los  $k$  puntos de cambio donde  $\tau_{k-1} \leq t_i < \tau_k$

$$\int_{\tau_{k-1}}^{T_i} e^{\alpha_0 + \alpha_1 u + \alpha_2 (u - \tau_1)_+ + \dots + \alpha_k (t - \tau_{k-1})_+ + \beta' \mathbf{Z}} du, \quad (2.36)$$

haciendo un cambio de variable, se toma a  $v = \alpha_0 + \alpha_1 u + \alpha_2 (u - \tau_1)_+ + \dots + \alpha_k (t - \tau_{k-1})_+$  y se tiene que (2.36) se puede expresar ahora como

$$\begin{aligned} \int_{\tau_{k-1}}^{T_i} e^{\alpha_0 + \alpha_1 u + \alpha_2 (u - \tau_1)_+ + \dots + \alpha_k (t - \tau_{k-1})_+ + \beta' \mathbf{Z}} du & \quad (2.37) \\ &= \frac{1}{\alpha_1 + \dots + \alpha_k} \int_{\tau_{k-1}}^{T_i} e^{v + \beta' \mathbf{Z}} dv \\ &= \frac{e^{\alpha_0 + \beta' \mathbf{Z}}}{\alpha_1 + \dots + \alpha_k} \left[ e^{\alpha_1 T_i + \alpha_2 (T_i - \tau_1)_+ + \dots + \alpha_k (t - \tau_{k-1})_+} \right. \\ & \quad \left. - e^{\alpha_1 \tau_{k-1} + \alpha_2 (\tau_{k-1} - \tau_1)_+ + \dots + \alpha_{k-1} (\tau_{k-2} - \tau_{k-1})_+} \right]. \end{aligned}$$

De donde, se ha expresado la parte de la integral en (2.32) como

$$\begin{aligned}
 & \int_0^{T_i} e^{\alpha_0 + \alpha_1 u + \alpha_2 (u - \tau_1)_+ + \dots + \alpha_{k+1} (u - \tau_k)_+ + \beta' \mathbf{Z}} du \\
 &= \frac{1}{\alpha_1} e^{\alpha_0 + \beta' \mathbf{Z}} [e^{\alpha_1 T_i + \beta' \mathbf{Z}} - 1] \\
 &+ \frac{e^{\alpha_0 + \beta' \mathbf{Z}}}{\alpha_1 + \alpha_2} [e^{\alpha_1 T_i + \alpha_2 (T_i - \tau_1)} - e^{\alpha_1 \tau_1}] \\
 &+ \dots + \\
 &\frac{e^{\alpha_0 + \beta' \mathbf{Z}}}{\alpha_1 + \dots + \alpha_k} \left[ e^{\alpha_1 T_i + \alpha_2 (T_i - \tau_1)_+ + \dots + \alpha_k (t - \tau_{k-1})_+} \right. \\
 &\left. - e^{\alpha_1 \tau_{k-1} + \alpha_2 (\tau_{k-1} - \tau_1)_+ + \dots + \alpha_{k-1} (\tau_{k-2} - \tau_{k-1})_+} \right].
 \end{aligned}$$

Ahora bien, se analizará la siguiente expresión

$$e^{\alpha_1 T_i + \alpha_2 (T_i - \tau_1)_+}.$$

Observe que  $(T_i - \tau_1)_+ = \max(0, T_i - \tau_1)$ , así tenemos que  $T_i = \min(X_i, C_i)$  que indica el tiempo de vida o de censura, y otra manera de poder representar este hecho es mediante la siguiente expresión  $(T_i \wedge \tau_1)$ .

Por lo cual se tiene que

$$\begin{aligned}
 & \int_0^{T_i} e^{\alpha_0 + \alpha_1 u + \alpha_2 (u - \tau_1)_+ + \dots + \alpha_{k+1} (u - \tau_k)_+ + \beta' \mathbf{Z}} du \quad (2.38a) \\
 &= \frac{e^{\alpha_0 + \beta' \mathbf{Z}}}{\alpha_1} \left( e^{\alpha_1 (T_i \wedge \tau_1)} - 1 \right) \\
 &+ \frac{e^{\alpha_0 + \beta' \mathbf{Z}}}{\alpha_1 + \alpha_2} \left( e^{\alpha_1 (T_i \wedge \tau_2) + \alpha_2 (T_i \wedge \tau_2 - \tau_1) - e^{\alpha_1 \tau_1}} \right) I(T_i > \tau_1) + \dots \\
 &+ \frac{e^{\alpha_0 + \beta' \mathbf{Z}}}{\alpha_1 + \alpha_2 + \dots + \alpha_k} \left( e^{\alpha_1 T_i + \alpha_2 (T_i - \tau_1)_+ + \dots + \alpha_k (t - \tau_{k-1})_+} \right. \\
 &\left. \exp^{\alpha_1 \tau_{k-1} + \alpha_2 (\tau_{k-1} - \tau_1)_+ + \dots + \alpha_{k-1} (\tau_{k-2} - \tau_{k-1})_+} \right) I(T_i > \tau_{k-1}).
 \end{aligned}$$

Por lo tanto,

$$\begin{aligned}
& \ln L(\alpha_0, \alpha_1, \dots, \alpha_k, \tau_1, \dots, \tau_{k-1}, \beta) \tag{2.39a} \\
&= \sum_{i=1}^n \delta_i (\alpha_0 + \alpha_1 T_i + \alpha_2 (T_i - \tau_1)_+ + \dots + \alpha_{k+1} (T_i - \tau_k)_+ + \beta' \mathbf{Z}) \\
&\quad - \frac{e^{\alpha_0 + \beta' \mathbf{Z}}}{\alpha_1} \left( e^{\alpha_1 (T_i \wedge \tau_1)} - 1 \right) \\
&\quad - \frac{e^{\alpha_0 + \beta' \mathbf{Z}}}{\alpha_1 + \alpha_2} \left( e^{\alpha_1 (T_i \wedge \tau_2) + \alpha_2 (T_i \wedge \tau_2 - \tau_1) - e^{\alpha_1 \tau_1}} \right) I(T_i > \tau_1) - \dots \\
&\quad - \frac{e^{\alpha_0 + \beta' \mathbf{Z}}}{\alpha_1 + \alpha_2 + \dots + \alpha_k} \left( e^{\alpha_1 T_i + \alpha_2 (T_i - \tau_1) + \dots + \alpha_k (T_i - \tau_{k-1}) +} \right. \\
&\quad \left. - e^{\alpha_1 \tau_{k-1} + \alpha_2 (\tau_{k-1} - \tau_1) + \dots + \alpha_{k-1} (\tau_{k-1} - \tau_{k-2})} \right) I(T_i > \tau_{k-1}).
\end{aligned}$$

Para la función *log-verosimilitud* donde  $\alpha_{k-1} = \alpha_k$  se procede de manera análoga como en el caso cuando  $\alpha_{k-1} \neq \alpha_k$ , por consiguiente la función *log-verosimilitud* tiene la siguiente forma

$$\begin{aligned}
& \ln L(\alpha_0, \alpha_1, \dots, \alpha_{k-1}, \tau_1, \dots, \tau_{k-2}, \beta) \tag{2.40a} \\
&= \sum_{i=1}^n \delta_i (\alpha_0 + \alpha_1 T_i + \alpha_2 (T_i - \tau_1)_+ + \dots + \alpha_{k+1} (T_i - \tau_{k-2})_+ + \beta' \mathbf{Z}) \\
&\quad - \int_0^{T_i} e^{\alpha_0 + \alpha_1 u + \alpha_2 (u - \tau_1)_+ + \dots + \alpha_{k+1} (u - \tau_{k-2})_+ + \beta' \mathbf{Z}} du \\
&= \sum_{i=1}^n \delta_i (\alpha_0 + \alpha_1 T_i + \alpha_2 (T_i - \tau_1)_+ + \dots + \alpha_{k+1} (T_i - \tau_{k-2})_+ + \beta' \mathbf{Z}) \\
&\quad - \frac{e^{\alpha_0 + \beta' \mathbf{Z}}}{\alpha_1} \left( e^{\alpha_1 (T_i \wedge \tau_1)} - 1 \right) \\
&\quad - \frac{e^{\alpha_0 + \beta' \mathbf{Z}}}{\alpha_1 + \alpha_2} \left( e^{\alpha_1 (T_i \wedge \tau_2) + \alpha_2 (T_i \wedge \tau_2 - \tau_1) - e^{\alpha_1 \tau_1}} \right) I(T_i > \tau_1) - \dots \\
&\quad - \frac{e^{\alpha_0 + \beta' \mathbf{Z}}}{\alpha_1 + \alpha_2 + \dots + \alpha_k} \left( e^{\alpha_1 T_i + \alpha_2 (T_i - \tau_1) + \dots + \alpha_{k-1} (T_i - \tau_{k-2}) +} \right. \\
&\quad \left. - e^{\alpha_1 \tau_{k-2} + \alpha_2 (\tau_{k-2} - \tau_1) + \dots + \alpha_{k-2} (\tau_{k-2} - \tau_{k-3})} \right) I(T_i > \tau_{k-2}).
\end{aligned}$$

De esta manera después de haber calculado  $\ln L(\alpha_0, \alpha_1, \dots, \alpha_k, \tau_1, \dots, \tau_{k-1}, \beta)$  y  $\ln L(\alpha_0, \alpha_1, \dots, \alpha_{k-1}, \tau_1, \dots, \tau_{k-2}, \beta)$  obtenemos una prueba de razón de verosimilitud.

### 2.5.1. Prueba de tipo Wald para el modelo lineal por tramos

Esta prueba esta basada en el estimador de máxima verosimilitud para el vector  $\theta$ , el cual se define como  $\theta' = [\alpha_0, \alpha_1, \dots, \alpha_{k+1}, \tau_1, \tau_2, \dots, \tau_k, \beta]$ , por consiguiente la prueba de hipótesis para el  $j$ -ésimo elemento de  $\theta$  es

$$X_W = \frac{(\hat{\theta} - \theta_0)^2}{\widehat{V}(\hat{\theta}_j)} \sim X_1^2, \quad (2.41)$$

donde  $\widehat{V}(\hat{\theta}_j) = [I(\hat{\theta})^{-1}]_{jj}$ .

En el modelo lineal por tramos una prueba de tipo Wald se usa para la hipótesis nula, no hay puntos de cambio, contra la alternativa de que hay por lo menos un punto de cambio, esto es equivalente a probar la hipótesis  $\alpha_2 = 0$ .

Una prueba para la existencia de los  $\tau_k$  puntos de cambio se lleva a cabo probando

$$H_0 : \alpha_{k+1} = 0 \quad V_s \quad H_1 : \alpha_{k+1} \neq 0, \quad (2.42)$$

para esto se utiliza una prueba de tipo Wald de la forma

$$X_W = \frac{\alpha_{k+1}^2}{\widehat{Var}(\hat{\alpha}_{k+1})} \sim X_1^2. \quad (2.43)$$

Dado que la prueba de tipo Wald está basada en la varianza del estimador entonces el nivel de significancia de la prueba de tipo Wald puede ser aceptada a un 95 % de confianza, para más detalles ver [7].

### 2.5.2. Estimación para el modelo lineal por tramos

La estimación en el modelo lineal por tramos se hará de manera similar a la del modelo constante por tramos, minimizando la función negativa de *log-verosimilitud* y encontrando los valores de  $\alpha_j$ ,  $\tau_j$  y  $\beta_j$  que optimicen la función de verosimilitud, ver [16].

## 2.6. Preservando el error tipo I

El procedimiento implementado consiste en realizar múltiples comparaciones, por lo que se tiene que hacer algunas correcciones con la finalidad de preservar el error de Tipo I.

Para mantener el nivel  $\alpha$  de la prueba, se considera la metodología del análisis secuencial. Se propone una técnica para el nivel de significancia necesario para rechazar la hipótesis nula, cada prueba tiene que ser menor que  $\alpha$ , y esté aumenta a medida que el estudio progresa.

Se propone el uso de

$$\alpha^*(k) = \alpha s^*(k), \quad (2.44)$$

donde  $\alpha^*(k)$  es el nivel de significancia para la  $k$ -ésima prueba de hipótesis y

$$s^*(k) = \frac{k}{K}$$

es la función que se va a emplear, donde  $\alpha^*(1) < \alpha^*(2) < \dots < \alpha^*(K)$ . El interés se centra en una función de  $\alpha$ , donde  $\alpha^*(1) > \alpha^*(2) > \dots > \alpha^*(K)$ , es una función decreciente en  $\alpha$ , la prueba para cada punto de cambio adicional se lleva a cabo con un estricto nivel de significancia  $\alpha$ , sea

$$\alpha^*(k) = \frac{\alpha}{2^{k-1}}, \quad (2.45)$$

donde  $\alpha^*(k)$  es el nivel de significancia para la  $k$ -ésima prueba de hipótesis.

Una ventaja de encontrar a este  $\alpha$  es que no depende del número de pruebas de hipótesis, por lo cual, no se necesita saber el número total de pruebas para llegar al modelo final.

Se observa que el error de tipo I, se puede considerar como la elección incorrecta de un modelo que tiene más puntos de cambio que el modelo verdadero.

Se tiene que el error tipo I no excede a  $\alpha$ , esto es la probabilidad de elegir un modelo con al menos uno punto de cambio dado que el modelo verdadero

no tiene puntos de cambio; la probabilidad de elegir un modelo con  $k$  puntos de cambio dado que el modelo verdadero no tiene puntos de cambios, y la probabilidad de elegir un modelo con más de  $k$  puntos de cambios dado que el verdadero modelo tiene  $k$  puntos de cambio, todas estas probabilidades son menores que  $\alpha$ .

Sea  $M_i$  es el evento: el modelo tiene  $i$  puntos de cambio. Entonces, aplicando la definición de probabilidad condicional se tiene que

$$P(M_{i \geq 1} | M_0) = 1 - P(M_0 | M_0) = 1 - (1 - \alpha) = \alpha$$

$$P(M_k | M_0) = \left(1 - \frac{\alpha}{2^k}\right) \prod_{i=1}^k \frac{\alpha}{2^{i-1}} < \frac{\alpha^k}{\prod_{i=1}^k 2^{i-1}} < \alpha$$

y para  $k \geq 1$

$$P(M_{i > k} | M_k) = \sum_{j=1}^{\infty} \left[ \left(1 - \frac{\alpha}{2^{k+j}}\right) \prod_{i=1}^{k+j} \frac{\alpha}{2^{i-1}} \right] < \alpha.$$

Esto muestra que la metodología propuesta, nos proporciona un error tipo I, pequeño o al menos menor que nuestro nivel de significancia.



## Capítulo 3

# Simulación

El análisis de supervivencia tiene una gran aplicación en los tiempos de vida de individuos sometidos a algún tipo de estudio, sin embargo, en la práctica conseguir datos reales de tiempos de vida es complicado ya que las bases de datos son considerados como confidenciales.

Para verificar la metodología propuesta, se simularon tiempos de vida para el modelo constante por tramos.

El análisis de este método se realiza de la siguiente manera:

### 1. Simulación de los tiempos vida

Se realizó una simulación de 500 tiempos de vida aleatorios a los cuales se les aplicó la función inversa de distribución, la cual está dada por

$$y = -\frac{\ln(1-x)}{\alpha}, \quad (3.1)$$

donde  $0 < x < 1$ , y  $\alpha$  es el valor medio de los tiempos de vida aleatorios.

Por lo que los tiempos de vida resultantes se ajustan a una distribución exponencial.

### 2. Simulación de los tiempos de censura

Los tiempos de censura fueron obtenidos mediante la distribución uniforme (0,28) y a cada tiempo de vida se le asigna un tiempo de censura, esta asignación es de manera arbitraria.

### 3. Función negativa de *log-verosimilitud*

Se obtiene la función negativa *log-verosimilitud* para el modelo constante por tramos con dos puntos de cambio, haciendo uso de (2.27), donde  $k = 2$ , y la cual resulta ser

$$\begin{aligned}
 L_2 = & - \left[ X(\tau_1) \ln \alpha_1 + [X(\tau_2) - X(\tau_1)] \ln \alpha_2 [n_u - X(\tau_1)] \ln \alpha_1 \right. \\
 & - \alpha_1 \sum_{i=1}^n (T_i \wedge \tau_1) - \alpha_2 \sum_{i=1}^n (T_i \wedge \tau_2 - \tau_1) I(T_i > \tau_1) \\
 & \left. - \alpha_3 \sum_{i=1}^n (T_i \wedge \tau_1) I(T_i > \tau_1) \right]. \tag{3.2}
 \end{aligned}$$

### 4. Estimadores de máxima verosimilitud

Los estimadores de máxima verosimilitud de la función negativa *log-verosimilitud* están dados por

$$\begin{aligned}
 \hat{\alpha}_1 &= \frac{X(\tau_1)}{\sum_{i=1}^n (T_i \wedge \tau_1)}, \\
 \hat{\alpha}_2 &= \frac{X(\tau_2) - X(\tau_1)}{\sum_{i=1}^n (T_i \wedge \tau_2 - \tau_1) I(T_i > \tau_1)}, \\
 \hat{\alpha}_3 &= \frac{n_u - X(\tau_1)}{\sum_{i=1}^n (T_i \wedge \tau_1) I(T_i > \tau_1)}. \tag{3.3}
 \end{aligned}$$

### 5. Verosimilitud de los puntos de cambio

Para poder estimar los puntos de cambio que minimicen a la función negativa *log-verosimilitud*, se sustituyen los estimadores (3.3), en la función negativa *log-verosimilitud* (3.2), obteniendo

$$\begin{aligned}
 L_2(\tau_1, \tau_2) = & X(\tau_1) \ln \frac{X(\tau_1)}{\sum_{i=1}^n (T_i \wedge \tau_1)} + \\
 & (X(\tau_2) - X(\tau_1)) \ln \frac{X(\tau_2) - X(\tau_1)}{\sum_{i=1}^n (T_i \wedge \tau_2 - \tau_1) I(T_i > \tau_1)} + \\
 & (n_u - X(\tau_2)) \ln \frac{n_u - X(\tau_2)}{\sum_{i=1}^n (T_i \wedge \tau_2) I(T_i > \tau_2)} - n_u. \tag{3.4}
 \end{aligned}$$

### 6. Estimación

Una vez encontrada la función (3.4), se hace uso del método de optimización de *Nelder-Mead* para encontrar los valores estimados de nuestros

puntos de cambio.

La estimación se realiza en  $R$ , (ver apéndice A), de acuerdo a los datos y al modelo seleccionado se obtuvo que  $\tau_1 = 3.145$  y  $\tau_2 = 4.147$ .

Se toman los valores estimados de  $\tau_1$  y  $\tau_2$ , los cuales son sustituidos en (3.3), obteniendo de esta manera el valor de la función de riesgo en cada tramo, obteniendo un error aproximado de 0.007 y la función esta dada por

$$\lambda(t) = \begin{cases} 0.13, & 0 \leq t < 3.145, \\ 0.385, & 3.145 \leq t < 4.147, \\ 0.351, & t \geq 4.147. \end{cases} \quad (3.5)$$

### 7. Pruebas de hipótesis

Se realizan las pruebas de hipótesis, para obtener el modelo final.

\* La prueba de hipótesis es

$$H_0 : \alpha_1 - \alpha_2 = 0 \quad Vs \quad H_1 : \alpha_1 - \alpha_2 \neq 0, \quad (3.6)$$

para realizar esta prueba de hipótesis es necesario obtener el nivel de significancia dado en (2.45), que con las estimaciones obtenidas resulta ser

$$\hat{\alpha}^*(1) = \frac{\alpha_1}{2^{1-1}} = \alpha_1 = 0.13,$$

se utiliza la prueba estadística de razón de verosimilitud (1.5) por lo que

$$\text{Se rechaza } H_0 \text{ si y sólo si } -2 \log \Lambda_n > \chi_{(1-\alpha^*, 2)}^2.$$

donde

$$-2 \log \Lambda_1 = 0.51$$

y

$$\chi_{(1-0.13, 2)}^2 = \chi_{(0.8652, 2)}^2 = 0.21,$$

con estos resultados se rechaza la hipótesis nula, siguiendo la metodología propuesta, se realiza la siguiente prueba de hipótesis dada por

$$H_0 : \alpha_2 - \alpha_3 = 0 \quad Vs \quad H_0 : \alpha_2 - \alpha_3 \neq 0, \quad (3.7)$$

El nivel de significancia para la segunda prueba de hipótesis es

$$\alpha^*(2) = \frac{\alpha_2}{2^{2-1}} = 0.192,$$

se utiliza el mismo estadístico de prueba y se obtiene que

$$\text{Se rechaza } H_0 \text{ si y sólo si } -2 \log \Lambda_n > \chi_{(1-\alpha^*, 2)}^2. \quad (3.8)$$

donde

$$-2 \log \Lambda_1 = -0.068$$

y

$$\chi_{(1-0.192, 2)}^2 = \chi_{(0.808, 2)}^2 = 0.21.$$

En este caso se acepta la hipótesis nula y esto indica que hay puntos de cambio, y se tiene que el modelo final es

$$f(t) = \begin{cases} 0.13 \exp(-0.13t), & 0 \leq t < 3.145, \\ .385 \exp(0.694 - 0.351t), & 3.145 \leq t < 4.147, \\ .351 \exp(0.6625 - 0.351t), & t \geq 4.147, \end{cases} \quad (3.9)$$

por lo que la función de supervivencia esta dada por

$$S(t) = \begin{cases} \exp(-0.13t), & 0 \leq t < 3.145, \\ \exp(0.694 - 0.351t), & 3.145 \leq t < 4.147, \\ \exp(0.6625 - 0.351t), & t \geq 4.147. \end{cases} \quad (3.10)$$

El modelo que se obtuvo fue usando el estadístico de prueba de razón de verosimilitud, ahora se procede a realizar el mismo juego de hipótesis y se obtiene un modelo final haciendo uso del estadístico de tipo Wald.

\* La prueba de hipótesis es

$$H_0 : \alpha_1 - \alpha_2 = 0 \quad Vs \quad H_0 : \alpha_1 - \alpha_2 \neq 0, \quad (3.11)$$

Para este caso tenemos que el nivel de significancia es

$$\hat{\alpha}^*(1) = \frac{\alpha_1}{2^{1-1}} = \alpha_1 = 0.13,$$

se utiliza la prueba estadística de tipo Wald dada en (2.30) por lo que

$$\text{Se rechaza } H_0 \text{ si y sólo si } X_W > \chi_{(1-\alpha^*,2)}^2.$$

donde

$$X_W = 2$$

y

$$\chi_{(1-0.13,2)}^2 = \chi_{(0.8652,2)}^2 = 0.21,$$

con estos resultados se tiene que la hipótesis nula es rechazada, y se prosigue a realizar la siguiente prueba de hipótesis dada por

$$H_0 : \alpha_2 - \alpha_3 = 0 \quad Vs \quad H_0 : \alpha_2 - \alpha_3 \neq 0,$$

El nivel de significancia para la segunda prueba de hipótesis es

$$\alpha^*(2) = \frac{\alpha_2}{2^{2-1}} = 0.192,$$

y utilizando el estadístico de prueba se tiene que

$$\text{Se rechaza } H_0 \text{ si y sólo si } X_w > \chi_{(1-\alpha^*,2)}^2.$$

donde

$$X_W = 0.0355$$

y

$$\chi_{(1-0.192,2)}^2 = \chi_{(0.808,2)}^2 = 0.21,$$

En este caso no se rechaza la hipótesis nula, esto indica que hay puntos de cambio.

La función de densidad obtenida fue la misma que (3.12) y la función de supervivencia esta dada de igual manera que en (3.10).

## 8. Gráficas

El comportamiento de la función de supervivencia, se puede observar en la Figura 1

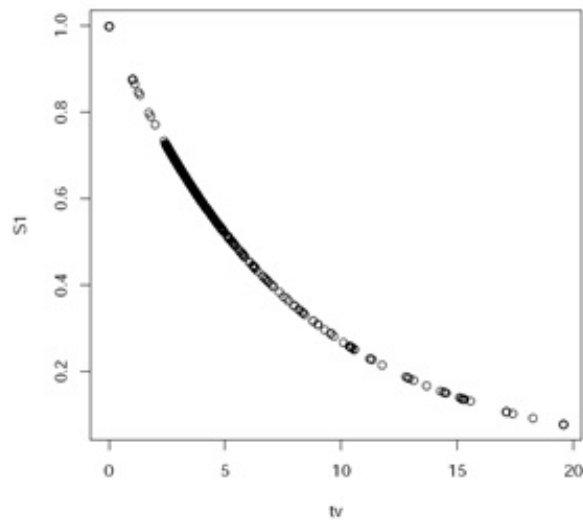


Figura 1: Función de supervivencia

Debido a que se tienen observaciones exponenciales, se tiene este tipo de forma.

El desarrollo de la función de riesgo durante un determinado tiempo  $t$ , se observa en el Figura 2, donde  $TV$  representan los tiempos de vida y el eje  $Y$  es el valor de la función de riesgo, podemos ver que existen dos puntos de cambio durante los primeros 5 años.

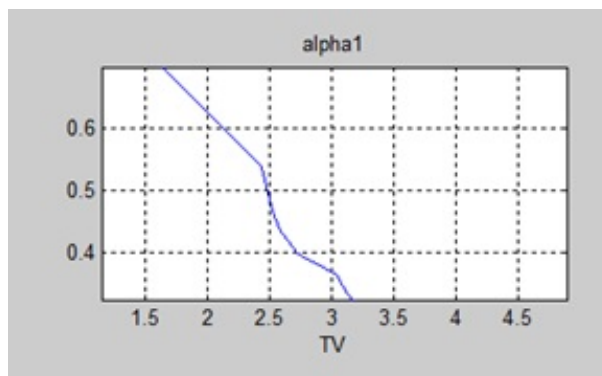


Figura 2: Función de riesgo

Además se encontraron los valores estimados de la función de riesgo empleando *regresión lineal simple*. Esta técnica proporciona estimaciones para los parámetros  $\alpha_1, \alpha_2, \alpha_3$ .

■ **Función de supervivencia**

La función de supervivencia descrita para  $\tau_1$  y  $\tau_2$  está dada por

$$S(t) = \begin{cases} \exp[-\alpha_1 t], & 0 \leq t < 3.145 \\ \exp[-\alpha_1 3.145 - \alpha_2(t - 3.145)], & 3.145 \leq t < 4.147 \\ \exp[-\alpha_1 3.145 - \alpha_2 1.002 - \alpha_3(t - 4.147)], & t \geq 4.147. \end{cases} \quad (3.12)$$

■ **Transformación logarítmica**

Se transforma a  $S(t)$  en una función lineal, empleando una transformación logarítmica, para esto se analizará a  $S(t)$  en cada tramo.

\* $0 \leq t < 3.145$  se tiene que

$$\ln S(t) = -\alpha_1 t; \quad (3.13)$$

\*\*  $3.145 \leq t < 4.147$  la función esta dada por

$$\ln S(t) = -\alpha_1 3.145 - \alpha_2(t - 3.145); \quad (3.14)$$

\*\*\*  $t \geq 3.38151$  se tiene que

$$\ln S(t) = -\alpha_1 3.145 - \alpha_2(1.001) - \alpha_3(t - 4.147). \quad (3.15)$$

■ **Estimación de los parámetros**

Para estimar de manera gráfica a los parámetros, se realiza lo siguiente

\*Se ordenan los 500 tiempos de vida o de falla de manera ascendente y se le asigna un número de orden  $i$ , de 1 a 500.

\*\*Se calcula la proporción  $F_i$  con  $i = 1 \dots 500$ , de la siguiente forma

$$F_i = \frac{i - 0.5}{500}$$

$F_i$  representa el porcentaje de fallas que han tenido lugar hasta antes del  $i$ -ésimo tiempo de falla correspondiente.

\*\*\* Se aplica regresión lineal a los puntos  $x_i, y_i$ .

- $0 \leq t_i < 3.145$ , donde  $i = 1, \dots, 500$

$$x_i = t_i, \quad y_i = \ln \left[ \frac{1}{1 - F(i)} \right].$$

- $3.145 \leq t_i < 4.147$ , donde  $i = 1, \dots, 500$

$$x_i = (t - 3.145), \quad y_i = \ln \left[ \frac{1}{1 - F(i)} \right].$$

- $t_i \geq 4.147$ , donde  $i = 1, \dots, 500$  se tiene que,

$$x_i = (t - 4.147), \quad y_i = \ln \left[ \frac{1}{1 - F(i)} \right].$$

Los valores estimados de  $\alpha_1, \alpha_2, \alpha_3$ , se obtienen mediante los resultados de la regresión realizada en el paquete estadístico  $R$ .

Por lo que se tiene que  $\widehat{\alpha}_1 = 0.2805, \widehat{\alpha}_2 = 0.1001, \widehat{\alpha}_3 = 0.1608$

#### ■ Prueba de hipótesis

Se realizan las pruebas de hipótesis, para obtener el modelo final.

\* La prueba de hipótesis es

$$H_0 : \alpha_1 - \alpha_2 = 0 \quad Vs \quad H_0 : \alpha_1 - \alpha_2 \neq 0, \quad (3.16)$$

para realizar la prueba de hipótesis es necesario obtener el nivel significancia dado en (2.45), el cual está dado por

$$\alpha^*(1) = \frac{\alpha_1}{2^{1-1}} = \alpha_1 = 0.2805,$$

se utiliza la prueba estadística de razón de verosimilitud de (1.5) entonces se tiene que;

$$\text{Se rechaza } H_0 \text{ si y sólo si } -2 \log \Lambda_n > \chi_{(1-\alpha^*, 2)}^2.$$

donde

$$-2 \log \Lambda_1 = -0.361$$

y

$$\chi_{(1-0.2805, 2)}^2 = \chi_{(0.7195, 2)}^2 = 1.39,$$

esto nos indica que la hipótesis nula no se rechaza, entonces, se tiene que el modelo final es

$$f(t) = 0.2805 \exp[-0.2805t], \quad 0 \leq t < 3.145, \quad (3.17)$$

así la función de supervivencia está dada por

$$S(t) = \exp[-0.2805t], \quad 0 \leq t < 3.145. \quad (3.18)$$

■ **Gráficas**

El comportamiento de la función de supervivencia, con respecto a los tiempos de vida se observa en el Figura 4

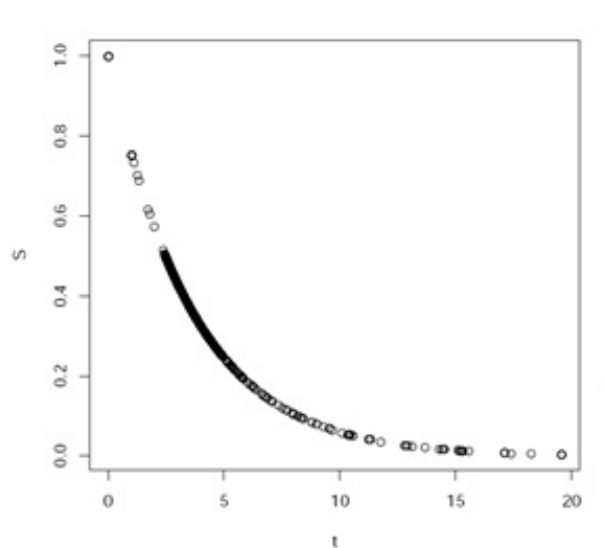


Figura 4: Función de supervivencia

Es importante analizar el comportamiento de la función de riesgo para cada individuo, en el Figura 5, podemos observar que la función de riesgo consta de dos puntos de cambio, los cuales se localizan en 2.45 y 3.5.

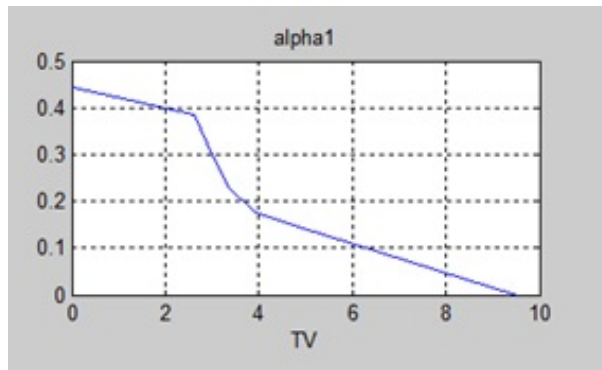


Figura 5: Función de riesgo

Como se puede observar en el Figura 5, la función de riesgo consta de dos puntos de cambio, pero la prueba estadística nos indica que en ese tramo no hay puntos de cambio. Por lo cual si se desea modelar usando regresión lineal simple es importante realizar otra prueba estadística para detectar los puntos de cambio, ya que de esta manera no se obtiene un buen ajuste de los datos.

# Conclusiones

El análisis de supervivencia es muy importante para el estudio de tiempos de vida o de falla, ya que proporciona información sobre el riesgo de falla existente en la población que se está estudiando. Por lo que es conveniente obtener información del riesgo en determinados intervalos de tiempo.

En este trabajo se presentó una metodología para encontrar y estimar los puntos de cambio, la prueba está basada en la función *log-verosimilitud* y en la prueba de tipo Wald, estas dos herramientas fueron de vital importancia para encontrar el modelo que mejor describe el comportamiento del tiempo de vida de un individuo.

En el capítulo 4, se desarrolló la metodología propuesta, para la cual se utilizaron dos estadísticos de prueba, se demostró que el modelo final obtenido fue el mismo en ambos casos, lo cual indica que la estimación de los parámetros fue realizada de manera correcta. Empleando otra metodología para obtener la estimación de los parámetros el modelo que resultó no es adecuado, para el comportamiento de los datos.

La prueba estadística de tipo Wald y la de razón de verosimilitud, son de las más empleadas en el análisis de supervivencia, y como se logra ver en la simulación nos proporcionan los mismos resultados, la elección de alguna de estas pruebas depende del tipo de datos que se tengan ya que en ocasiones se presentan dificultades, para poder calcular estadísticos, (como podría ser la varianza o la razón de verosimilitud, por mencionar algunas).

Concluimos que el modelo de falla con múltiples puntos de cambio proporcionan más y mejor información al emplearse en el análisis de tiempos de vida (esto siempre y cuando la estimación de los parámetros se realice de forma correcta).

La estimación correcta de este modelo en el análisis de datos reales, repercutirá positivamente y permitirá mejorar el tiempo de vida de los individuos bajo estudio, puesto que al identificar el tiempo donde exista mayor posibilidad de fallar, él o los investigadores (médicos), tomaran las medidas necesarias para reducir el riesgo de falla.

Es indudable que el efecto de covariables juega un papel importante en los tiempos de vida o de falla, así que, un trabajo futuro podría ser analizar un modelo con múltiples puntos de cambio que involucran covariables.

# Apéndice A

## Algoritmo

Se muestra el algoritmo de Simplex de *Nelder-Mead*, el cual fue utilizado para encontrar el modelo final.

```
clearall;clc
sigma_plot = 10;
load('datosF.mat')
tau1 = 1 : 0.1 : 10;
tau2 = 11 : 0.1 : 18;
[TU1,TU2] = meshgrid(tau1,tau2);
Z = zeros(size(TU1))';
Alp1 = Z;
Alp2 = Z;
Alp3 = Z;
fori = 1 : numel(tau1)
forj = 1 : numel(tau2)
[Z(i,j), Alp1(i,j), Alp2(i,j), Alp3(i,j)] = fcn2(TVdat, Cdat, tau1(i), tau2(j));
end
end
subplot(3,3,1); mesh(tau1, tau2, Z'); colorbar
xlabel('tau1')
ylabel('tau2')
title('Funciondecosto')
```

**70**

```
subplot(3,3,2); mesh(tau1, tau2, Alp1'); colorbar
xlabel('tau1')
ylabel('tau2')
title('Alp1')
subplot(3,3,3); mesh(tau1, tau2, Alp2'); colorbar
xlabel('tau1')
ylabel('tau2')title('Alp2')
subplot(3,3,4); mesh(tau1, tau2, Alp3'); colorbar
xlabel('tau1')
ylabel('tau2')
title('Alp3')
break
taus = [010;
001];
fork = 1 : 20
k
[Z, pIdx] = sort([fcn2(TVdat, Cdat, taus(1,1), taus(2,1))...
fcn2(TVdat, Cdat, taus(1,2), taus(2,2))...
fcn2(TVdat, Cdat, taus(1,3), taus(2,3))]);
X = taus(:, pIdx);
plot(X(1,1), X(2,1), 'or', ...
X(1,2), X(2,2), 'og', ...
X(1,3), X(2,3), 'ob')
axis([-3 * sigma_plot3 * sigma_plot...
-3 * sigma_plot3 * sigma_plot])
gridon
pause(1)
M = sum(X(:, 1 : end - 1), 2) / (size(X, 2) - 1)
plot(X(1,1), X(2,1), 'or', ...
X(1,2), X(2,2), 'og', ...
X(1,3), X(2,3), 'ob', ...
M(1,1), M(2,1), 'ok')
```

```

axis([-3 * sigma_plot3 * sigma_plot...
-3 * sigma_plot3 * sigma_plot])
gridon
pause(1)
rho = 1;
R = M + rho * (M - X(:, end));
plot(X(1, 1), X(2, 1), 'or', ...
X(1, 2), X(2, 2), 'og', ...
X(1, 3), X(2, 3), 'ob', ...
M(1, 1), M(2, 1), 'ok', ...
R(1, 1), R(2, 1), 'om')
axis([-3 * sigma_plot3 * sigma_plot...
-3 * sigma_plot3 * sigma_plot])
gridon
pause(1)
Fr = fnc(R);
gamma = 0.5;
if Z(1) <= FryFr <= Z(end - 1)
X0 = [X(:, 1 : end - 1)R];
elseif Fr <= Z(1)
chi = 2;
Xe = M + chi * (M - X(:, end));
Fxe = fnc(Xe);
if Fxe > Fr
X0 = [X(:, 1 : end - 1)R];
else
X0 = [X(:, 1 : end - 1)Xe];
end
elseif Z(end - 1) <= FryFr <= Z(end)
Xoc = M + gamma * (M - X(:, end));
Fxoc = fnc(Xoc);
if Fxoc <= Fr

```

**72**

```
X0 = [X(:, 1 : end - 1)Xoc];
else
end
elseif Fr >= Z(end)
Xic = M - gamma * (M - X(:, end));
Fxic = fnc(Xic);
if Fxic <= Z(end)
X0 = [X(:, 1 : end - 1)Xic];
else
end
else
C0x = [(Mx - X(3));
(Rx - Mx)]/2;
C0y = [(My - Y(3));
(Ry - My)]/2;
[ZC, pcIdx] = sort(fnc(C0x, C0y));
Cx = C0x(pcIdx(1));
Cy = C0y(pcIdx(1));
plot([X(1 : 2)Cx], [Y(1 : 2)Cy])
axis([-3 * sigma_plot3 * sigma_plot...
-3 * sigma_plot3 * sigma_plot])
grid on
pause(1)
end
end
```

# Bibliografía

- [1] Ash R., *Basic Probability Theory*, Mineola, New York, (1970).
- [2] Antoniadis A., Gijbels I., Macgibbon B., *Non-Parametric Estimation for the Location of a Change-Point in an Otherwise Smooth Hazard Function under Random Censoring*, Scandinavian Journal of Statistics, Vol. 27, pp. 501-519, (2000).
- [3] Grossman S., *Aplicaciones de álgebra lineal*. Grupo editorial Iberoamericana, (1987).
- [4] Hogg R., McKean J., Craig A., *Introduction to mathematical Statistics*, Pearson, (2005).
- [5] Kalbfleisch J., Prentice R., *The Statistical Analysis of Failure Time Data*, John Wiley and Sons, (1980).
- [6] Klein J., Moeschberger M., *Survival Analysis Techniques for Censored and Truncated Data*, Springer, (2005).
- [7] Lanchin J., *Biostatistical Methods The Assessment of Relative Risks*, John Wiley - Sons, Inc., (2000).
- [8] Lawless J., *Statistical Models and Methods for Lifetime Data*, Wiley Series in Probability and Statistics, (2003).
- [9] Marsden J., Tromba A., *Cálculo vectorial*, Addison-Wesley Iberoamericana, (1991).

- [10] Mazucheli J., Coelho E., Alberto Jorge., *Inferences for the Change-Point of the Exponentiated Weibull Hazard Function*, Revista Statical Journal, Volumen 10, (2012).
- [11] Melody S., *Survival Analysis whith Cahnge Point Hazard Functions*, Revista Harvard University Biostatistics, Paper 40, (2006).
- [12] Melody S., *Detecting Multiple Change Points in Piecewise Constant Hazard Functions*, Revista Harvard University Biostatistics, (2010).
- [13] Mendoza M., *Importancia del análisis estadística secuencial*, Comunicación interna, Departamento de Matemáticas-facultad de Ciencias, UNAM (1978).
- [14] Meyer R., Prado C., Ardila G., Octavio S., *Probabilidad y aplicaciones estadísticas*, Addison-Wesley Iberoamericana, (1970).
- [15] Mood A., Graybill., Boes D., *Introduction to theory of statistics*, McGraw, (1974).
- [16] Moreno Z., *Análisis de algunas metaheurísticas creadas a partir de Optimización Gravitatoria*, VII Encuentro internacional anuales de ASEPUMA, n<sup>o</sup> 19.
- [17] Palmeros O., *Algunos modelos en el análisis de supervivencia y una aplicación del modelo Weibull*, Tesis de maestría (2007).
- [18] Zhao X., Wu X., Zhou X., *A change-point model for survival data with long-term survivors*, Statistica Sinica, pp 377-390, (2009).