

**PONTIFICIA UNIVERSIDAD
CATÓLICA DEL PERÚ**

Escuela de Posgrado



**Estimación paramétrica en un modelo de eventos
recurrentes con riesgos competitivos: Aplicación para el
análisis de fallas de máquinas inyectoras**

Tesis para optar el grado académico de Maestro en Estadística
que presenta:

José Ignacio Vásquez Médico

Asesor:

Víctor Giancarlo Sal y Rosas Celi

Lima, 2025


Informe de Similitud

Yo, Víctor Giancarlo Sal y Rosas Celi, docente de la Escuela de Posgrado de la Pontificia Universidad Católica del Perú, asesor de la tesis de investigación titulado “Estimación paramétrica en un modelo de eventos recurrentes con riesgos competitivos: Aplicación para el análisis de fallas de máquinas inyectoras” del autor José Vásquez Medico deo constancia de lo siguiente:

- El mencionado documento tiene un índice de puntuación de similitud de 13%. Así lo consigna el reporte de similitud emitido por el software *Turnitin* el 11/09/2025.
- He revisado con detalle dicho reporte y la Tesis o Trabajo de Suficiencia Profesional, y no se advierte indicios de plagio.
- Las citas a otros autores y sus respectivas referencias cumplen con las pautas académicas.

Lugar y fecha:

Lima, 11 de Setiembre del 2025

Apellidos y nombres del asesor: <u>Sal y Rosas Celi Víctor Giancarlo</u>	
DNI: 40361284	
ORCID: 0000-0001-8636-7142	

Dedicatoria

A mis padres, hermanas y a mi novia, agradezco profundamente su amor incondicional y su constante apoyo. Su presencia y cariño han sido fundamentales, dándome fortaleza e inspiración para enfrentar desafíos. Les estoy eternamente agradecido por todo.



Agradecimientos

Agradezco a Dios por su constante guía y bendición, manteniéndome siempre en el buen camino.

A mis padres, hermanas y a mi novia, por su inquebrantable apoyo y comprensión en todo momento.

A los profesores de la maestría, por transmitirme valiosos conocimientos y enseñanzas a lo largo de mi formación académica.

A mi asesor Víctor Giancarlo Sal y Rosas Celi, por su interés para el desarrollo del presente trabajo.

A mis amigos y compañeros de la maestría, por compartir conmigo experiencias inolvidables y momentos de alegría que atesoraré siempre.

A todos ellos, mi más profundo agradecimiento, pues su contribución ha sido esencial en la culminación de este proyecto y en mi crecimiento personal y profesional.

Resumen

El análisis de supervivencia es una técnica estadística que tiene como objetivo modelar el tiempo hasta la ocurrencia de un determinado suceso. Este análisis es una de las disciplinas más antigua de la estadística, cuya aplicación generalmente se centra en la demografía y la medicina. Sin embargo, se ha demostrado que también es aplicable en otras áreas como en ingeniería y confiabilidad, específicamente en el análisis de fallas de equipos, donde la eficiencia es crucial debido a la alta competencia en todos los sectores industriales. Por lo tanto, como motivación a ello, se desarrolló y aplicó una estimación paramétrica de las funciones de distribución marginal de los tiempos entre fallas de máquinas inyectoras en una fábrica de plásticos en un modelo de riesgos competitivos recurrentes. De esta manera, el enfoque principal de este trabajo fue hacer inferencia sobre las propiedades distributivas de los tiempos entre fallas de estos equipos dentro este tipo de modelo, durante un tiempo de monitoreo aleatorio, considerando que luego de cada falla el equipo fue reparado bajo dos estrategias de reparación: perfecta y parcial.

Entre los tópicos estudiados en el presente trabajo se consideró la revisión de la literatura, la definición del modelo riesgos competitivos recurrentes, el proceso de estimación de parámetros e inferencia, un estudio de simulación para evaluar el buen desempeño de la estimación, y finalmente la aplicación del modelo a una base de datos de fallas de máquinas inyectoras.

Palabras-clave: Modelo, Supervivencia, Eventos Recurrentes, Riesgos Competitivos, Fallas de Equipos, Reparación Perfecta, Reparación Parcial.

Abstract

Survival analysis is a statistical technique aimed at modeling the time until the occurrence of a specific event. This analysis is one of the oldest disciplines in statistics, traditionally focused on demography and medicine. However, it has also been shown to be applicable in other fields such as engineering and reliability, specifically in the analysis of equipment failures, where efficiency is crucial due to high competition in all industrial sectors. As a motivation, a parametric estimation of the marginal distribution functions of the times between failures of injection machines in a plastics factory was developed and applied within a recurrent competing risks model. The main focus of this work was to make inferences about the distributional properties of the times between failures of these machines within this type of model during a random monitoring period, considering that after each failure, the equipment was repaired under two repair strategies: perfect and partial.

The topics studied in this work included a literature review, the definition of the recurrent competing risks model, the parameter estimation and inference process, a simulation study to evaluate the estimation's performance, and finally, the application of the model to a database of injection machine failures.

Keywords: Model, Survival, Recurrent Events, Competing Risks, Equipment Failures, Perfect Repair, Partial Repair.

Índice general

Lista de Abreviaturas	IX
Lista de Símbolos	X
Índice de figuras	XII
Índice de cuadros	XIII
1 Introducción	1
1.1 Antecedentes y Motivación	1
1.2 Objetivos	3
1.3 Organización del trabajo	3
2 Conceptos Preliminares	4
2.1 Estructura de datos	4
2.2 Función de Supervivencia	5
2.3 Modelos paramétricos univariados	6
2.4 Modelo de Riesgos Proporcionales de Cox	9
2.5 Modelo de Eventos Recurrentes	11
2.6 Modelo de Riesgos Competitivos	13
3 Modelo de Riesgos Competitivos Recurrentes	18
3.1 Modelo de Riesgos Competitivos de Evento Único	18
3.2 Estructura de los datos	23
3.2.1 Datos para el Modelo de Reparación Perfecta	23
3.2.2 Datos para el Modelo de Reparación Parcial	24
3.3 Planteamiento del problema	25
3.4 Construcción del Modelo	30
3.5 Estimación	32

4	Estudio de Simulación	34
4.1	Estructura de la simulación	34
4.2	Simulación	34
4.2.1	Simulación para 5 sistemas	34
4.2.2	Simulación para 10 sistemas	35
4.2.3	Simulación para 50 sistemas	37
5	Aplicación	38
5.1	Descripción de los datos	39
5.2	Modelo de Riesgos Competitivos Recurrentes	44
6	Conclusiones	46
7	Apéndice	48
7.1	Códigos R	48
7.1.1	Generación de datos para el ejemplo de Modelo de Riesgos Competi- tivos de Evento Único	48
7.1.2	Cálculo de las funciones de verosimilitud, supervivencia del sistema y subdistribución de los componentes 1 y 2 para el ejemplo de Modelo de Riesgos Competitivos de Evento Único	49
7.1.3	Simulador de data	50
7.1.4	Función de log-verosimilitud negativa	52
7.1.5	Maximización de verosimilitud de los datos generados	53
7.1.6	Simulación de 1000 réplicas	54
7.1.7	Aplicación del modelo	54
7.1.8	Comparaciones de modelos	57
7.1.9	Comparaciones de eficiencia	59
	Bibliografía	62

Lista de Abreviaturas

f.d. Función de densidad.

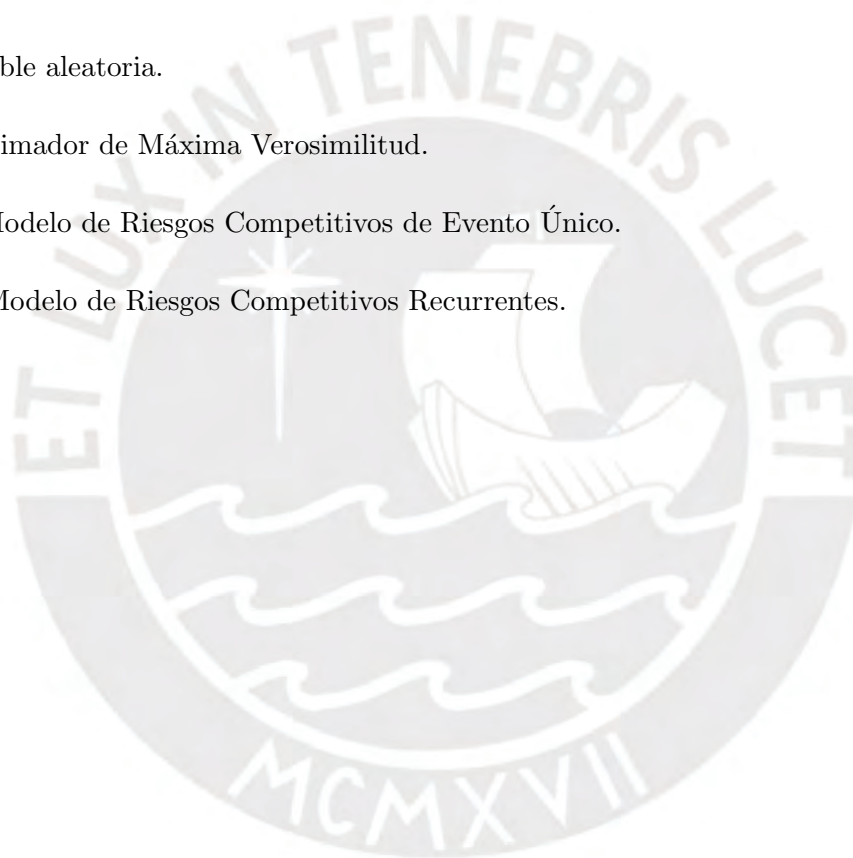
f.d.a. Función de distribución acumulada.

v.a. Variable aleatoria.

EMV Estimador de Máxima Verosimilitud.

SCRM Modelo de Riesgos Competitivos de Evento Único.

RCRM Modelo de Riesgos Competitivos Recurrentes.



Lista de Símbolos

C Causa del evento.

E Proceso de recurrencia hacia atrás.

f Función de densidad.

\check{f} Función de subdensidad.

F Función de distribución acumulada.

\check{F} Función de subdistribución acumulada.

\bar{F} Función de supervivencia.

$\bar{\bar{F}}$ Función de supervivencia específica.

g Función de densidad del tiempo de monitoreo.

G Función de distribución acumulada del tiempo de monitoreo.

\bar{G} Función de supervivencia del tiempo de monitoreo.

K Número de eventos observados.

ℓ Función de log-verosimilitud.

L Función de verosimilitud.

N Proceso de conteo.

R Conjunto de riesgo.

S Tiempo de vida útil del sistema.

T Tiempo hasta la ocurrencia del evento de interés.

V Tiempo hasta la censura.

Y Indicador de riesgo.

Z Tiempo observado.

δ Indicador de Censura.

λ Función de riesgo.

$\check{\lambda}$ Función de riesgo de causa específica.

λ_0 Función de riesgo base.

Λ Función de riesgo acumulado.

$\check{\Lambda}$ Función de riesgo acumulado de causa específica.

ϕ Función de normal estándar.

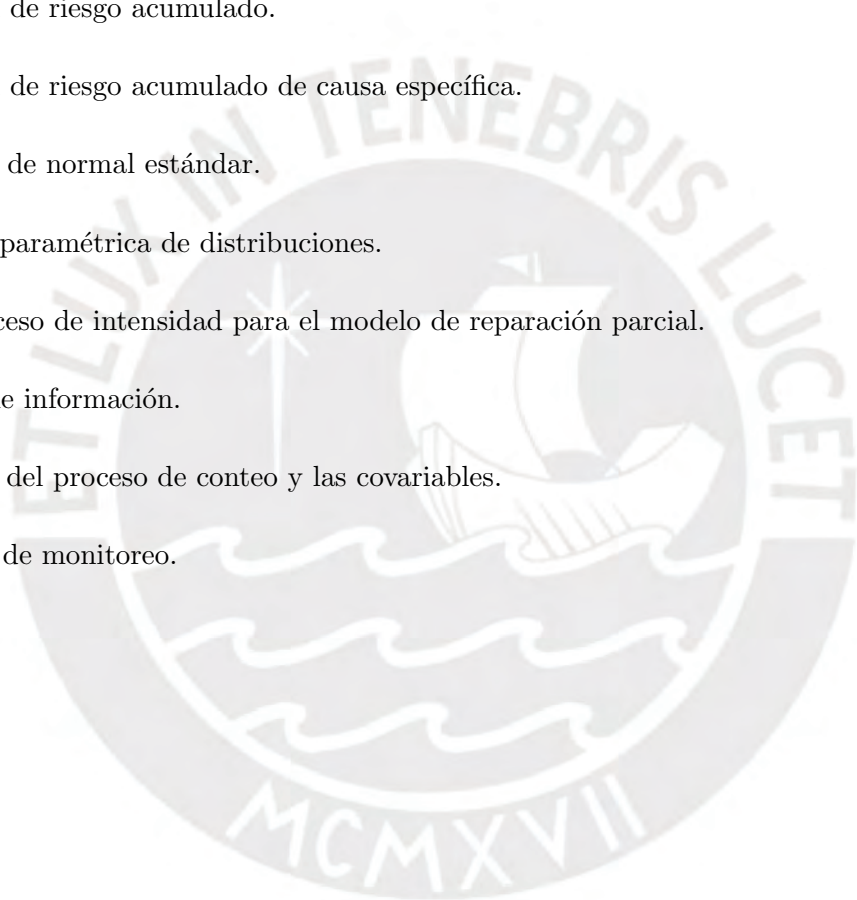
\mathcal{C} Familia paramétrica de distribuciones.

dN^A Proceso de intensidad para el modelo de reparación parcial.

\mathcal{F} Filtro de información.

\mathcal{H} Historia del proceso de conteo y las covariables.

τ Tiempo de monitoreo.



Índice de figuras

2.1	Tipos de censura.	4
3.1	Visualización gráfica de datos del Sistema de Riesgos Competitivos de Evento Único (SCRM) para 5 sistemas y 3 componentes.	21
3.2	Datos observados para un sistema con dos riesgos competitivos recurrentes bajo ambos modelos de reparación.	25
3.3	Planteamiento del problema bajo un modelo de reparación perfecta.	28
3.4	Planteamiento del problema bajo un modelo de reparación parcial.	29
4.1	Histogramas de las estimaciones de los parámetros para el caso de $n = 5$ sistemas y 1,000 réplicas utilizando $\alpha_1 = 2, \beta_1 = 2, \alpha_2 = 3, \beta_2 = 3, \alpha_3 = 4, \beta_3 = 4$	35
4.2	Histogramas de las estimaciones de los parámetros para el caso de $n = 10$ sistemas y 1,000 réplicas utilizando $\alpha_1 = 2, \beta_1 = 2, \alpha_2 = 3, \beta_2 = 3, \alpha_3 = 4, \beta_3 = 4$	36
4.3	Histogramas de las estimaciones de los parámetros para el caso de $n = 50$ sistemas y 1,000 réplicas utilizando $\alpha_1 = 2, \beta_1 = 2, \alpha_2 = 2,1, \beta_2 = 2,1, \alpha_3 = 2,2, \beta_3 = 2,2$	37
5.1	Fallas de los componente en cada máquina.	39
5.2	Distribuciones estimadas de los tiempos entre fallas de los 5 componentes y de la máquina.	45

Índice de cuadros

3.1	Datos del Sistema de Riesgos Competitivos de Evento Único (SCRM) para 3 sistemas y 2 componentes.	20
4.1	Resultados de las estimaciones de los parámetros para el caso de $n = 5$ sistemas y 1,000 réplicas utilizando $\alpha_1 = 2, \beta_1 = 2, \alpha_2 = 3, \beta_2 = 3, \alpha_3 = 4, \beta_3 = 4$	35
4.2	Resultados de las estimaciones de los parámetros para el caso de $n = 10$ sistemas y 1,000 réplicas utilizando $\alpha_1 = 2, \beta_1 = 2, \alpha_2 = 3, \beta_2 = 3, \alpha_3 = 4, \beta_3 = 4$	36
4.3	Resultados de las estimaciones de los parámetros para el caso de $n = 50$ sistemas y 1,000 réplicas utilizando $\alpha_1 = 2, \beta_1 = 2, \alpha_2 = 2,1, \beta_2 = 2,1, \alpha_3 = 2,2, \beta_3 = 2,2$	37
5.1	Estructura de los datos disponibles.	39
5.2	Fallas de máquina por componente - Parte 1.	40
5.3	Fallas de máquina por componente - Parte 2.	41
5.4	Fallas de máquina por componente - Parte 3.	42
5.5	Fallas de máquina por componente - Parte 4.	43
5.6	Fallas totales por componente.	43
5.7	Estimación de parámetros de los tiempos entre fallas de cada componente asumiendo que siguen distribuciones Weibull.	44
5.8	Comparación de modelos de riesgos competitivos recurrentes en donde los tiempos entre fallas siguen distribuciones Weibull, Log-Normal, Gamma y Log-Logístico.	44

Capítulo 1

Introducción

1.1. Antecedentes y Motivación

De acuerdo con el Fondo Monetario Internacional, el crecimiento del Perú se reducirá al 2.4 % en 2023, debido a las condiciones externas, que cada vez son más restrictivas, principalmente derivadas de la guerra entre Rusia y Ucrania, una brusca desaceleración mundial, y a la incertidumbre política, que sigue siendo elevada, principalmente a causa de los acontecimientos políticos y los desastres naturales, que podrían frenar la actividad económica y poner en riesgo la consolidación fiscal en el mediano plazo (Instituto Peruano de Economía, 2023).

Asimismo, la Sociedad Nacional de Industrias del Perú, según informó en su última Memoria Anual, en el 2022 la industria manufacturera registró un crecimiento del 1 %, debido al incremento del subsector no primario, el cual fue de 2.3 %, y la contracción productiva del subsector primario, el cual fue de -3.0 %. No obstante, a nivel de las divisiones industriales, varias presentaron decrecimiento, en la que destaca la elaboración de productos de caucho y plástico con -4.6 % (Salazar Nishi, 2023).

Así también, el Instituto Nacional de Estadística e Informática, reportó que la producción de la industria del plástico registró una reducción de -5.5 % en el 2022 (Bonett, 2022). Esto último debido a la baja demanda que hay actualmente en el mercado, por lo que las empresas del sector deben ser más eficientes en la fabricación de sus productos, y así evitar mayores costos.

Respecto a esto último, es común que se consulte sobre los parámetros o condiciones de proceso que se deberían controlar para saber si el proceso es estable y repetitivo, y en definitiva si está funcionando según lo esperado sin tener más adelante problemas en los

productos terminados (Echeverri y Jaramillo, 1980). Por lo tanto, un punto de partida para realizar un mejor control del proceso, y así lograr esta mayor eficiencia necesaria hoy en día, es analizando las fallas de los equipos productivos.

De esta manera, el análisis de supervivencia es un método estadístico utilizado para estudiar tanto la ocurrencia como el momento en que se producen ciertos eventos (Hosmer et al., 2008), el cual tiene aplicación en diversos campos, como el análisis de historia de eventos, análisis de confiabilidad, análisis de tiempos de falla, análisis de duración y análisis de transición (Allison, 2010).

La principal característica de los datos de supervivencia es que la variable de respuesta es el tiempo desde un origen bien definido hasta un evento bien definido. Además, una segunda característica del análisis de supervivencia es la censura, la cual surge cuando no se observan con precisión los eventos iniciales o finales (Prentice y Zhao, 2019).

El tipo más común de censura en este análisis es la censura por la derecha, la cual ocurre cuando solo se sabe que el tiempo hasta el evento excede un valor particular (Kleinbaum y Klein, 2012). En este contexto, si T representa el tiempo hasta el evento y V representa el tiempo hasta la censura, entonces el tiempo observado está definido por $Z = \min(T, V)$. Asimismo, se utiliza el indicador de censura $\delta = I[T < V]$, cuyo valor es 0 si T es un tiempo censurado, y 1 si T es un tiempo de fallo observado (Singer y Willett, 2003).

Sin embargo, el análisis de supervivencia se vuelve complejo cuando se tiene recurrencia en los eventos, es decir que el estudio de una determinada entidad no termina hasta que le dé el evento único, como es usualmente, ya que éste se puede repetir; y se puede complicar aún más cuando estos eventos además compiten entre sí, lo cual suele suceder en las máquinas industriales (Kalbfleisch y Prentice, 2002).

Por lo tanto, en este trabajo se propuso realizar una estimación paramétrica de los tiempos entre fallas de máquinas inyectoras en una empresa que fabrica productos plásticos en el Perú. El estudio se enmarcó en un contexto de fallas (eventos) recurrentes con riesgos competitivos (causas de fallas), con el objetivo de obtener estimadores de las funciones de distribución marginal para cada causa de falla en competencia. A partir de estos estimadores, también se buscó determinar la función de distribución de vida del equipo, considerando que, tras cada falla, el equipo puede ser reparado de forma parcial, es decir, reparando únicamente la falla ocurrida (Taylor y Peña, 2013).

1.2. Objetivos

El objetivo general de la presente tesis fue obtener los estimadores de las funciones de distribución marginal de los tiempos entre fallas por cada causa de falla en competencia y, en consecuencia, de la función de distribución de vida de los equipos industriales en un modelo de riesgos competitivos recurrentes, considerando que cada falla es reparada parcialmente inmediatamente después de ocurrida. De manera específica:

- Revisar la literatura sobre el modelo base de supervivencia y sus variaciones cuando se presentan eventos recurrentes y riesgos competitivos.
- Proponer, estudiar las propiedades e implementar la estimación paramétrica de las funciones de distribución marginal de los tiempos entre fallas de equipos industriales que presentan riesgos competitivos recurrentes.
- Realizar estudios de simulación para analizar el comportamiento de las propiedades de muestra pequeña de los estimadores, utilizando computación intensiva sobre diferentes escenarios.
- Aplicar el modelo a un conjunto de datos de fallas de máquinas inyectoras.
- Conclusiones y recomendaciones.

1.3. Organización del trabajo

Entre los tópicos estudiados en el presente trabajo se consideró en primer lugar la revisión de la literatura (Capítulo 2). Luego, se procedió con la definición matemática del modelo de riesgos competitivos recurrentes así como su proceso de estimación de parámetros e inferencia (Capítulo 3). Posteriormente, se realizó un estudio de simulación para evaluar el desempeño de la estimación (Capítulo 4). A continuación, se aplicó el modelo a una base de datos de fallas de máquinas inyectoras del año 2023 en una fábrica de plásticos del Perú (Capítulo 5). Finalmente, se discuten las conclusiones y recomendaciones del presente trabajo de tesis (Capítulo 6).

Capítulo 2

Conceptos Preliminares

En este capítulo, se presentan los conceptos y las consideraciones que serán utilizados en el desarrollo del presente trabajo de tesis.

2.1. Estructura de datos

En general, la estructura de los datos para el análisis de supervivencia está conformada por el tiempo transcurrido desde un origen bien definido hasta un evento bien definido, y la información sobre si dicho evento fue observado o censurado (Kleinbaum y Klein, 2012).

Respecto a la censura, ésta se define como la situación en la que no se dispone de la información exacta sobre el tiempo de ocurrencia del evento. Existen principalmente tres tipos de censura, los cuales se pueden observar en la figura 2.1.

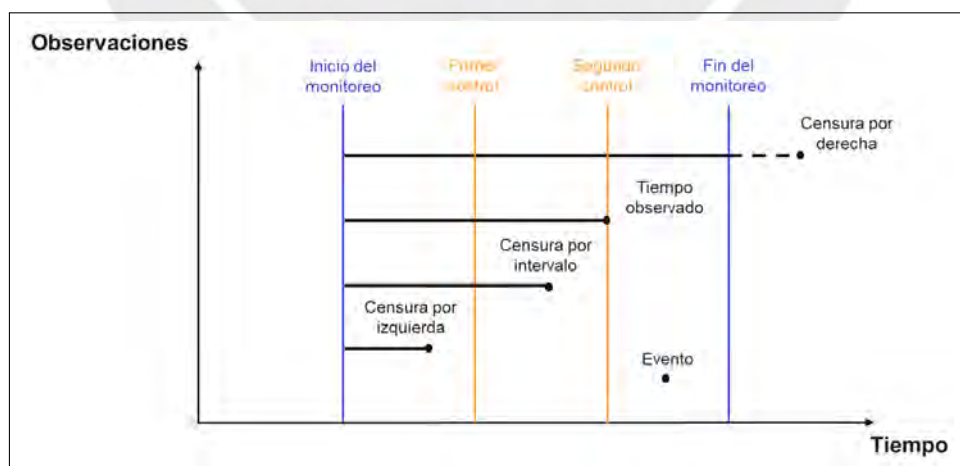


Figura 2.1: Tipos de censura.

La censura por derecha ocurre cuando el evento definido no se observa dentro del tiempo de monitoreo o seguimiento, pero que se sabe que ocurrirá en algún momento posterior. Por

otro lado, la censura por izquierda sucede cuando el evento no se observa en el momento exacto en que ocurrió, pero se sabe que ocurrió en algún momento antes de ese tiempo. Finalmente, la censura por intervalo se da cuando el evento tampoco se observa en el tiempo exacto en que sucedió, pero que se conoce que ocurrió dentro de un intervalo de tiempo específico antes (Moore, 2016).

El tipo de censura más común en el análisis de supervivencia es la censura por derecha. Formalizando a partir de esto, sean T el tiempo hasta el evento, V el tiempo hasta la censura, y δ una variable indicadora que especifica si el evento definido fue observado o no. De esta forma, $Z = \min(T, V)$ es el tiempo observado y $\delta = I[T < V]$, donde δ es 0 si T es un tiempo censurado ($T > V$), ó 1 si T es un tiempo observado ($T < V$).

2.2. Función de Supervivencia

Sea T una variable aleatoria no negativa que representa el tiempo hasta la ocurrencia del evento de interés. La función de supervivencia, denotada por $\bar{F}(t)$, se define como la probabilidad de que el evento no ocurra hasta el tiempo t (Klein et al., 2014).

$$\bar{F}(t) = P(T > t), \quad 0 < t < \infty \quad (2.1)$$

Teóricamente, esta función es una curva no creciente que comienza en 1 cuando $t = 0$ (la probabilidad de que el evento aún no haya ocurrido) y tiende a 0 conforme t se aproxima a infinito (Kleinbaum y Klein, 2012).

Por lo tanto, la función de distribución acumulada (f.d.a.), denotada por $F(t)$, se define como la probabilidad de que el evento ocurra antes del tiempo t . Está definida por:

$$F(t) = P(T \leq t), \quad 0 < t < \infty \quad (2.2)$$

Si la función de distribución acumulada es diferenciable, la función de densidad (f.d.), denotada por $f(t)$, se obtiene derivando la f.d.a.

$$f(t) = \frac{d}{dt}F(t) = -\frac{d}{dt}\bar{F}(t) \quad (2.3)$$

Esto representa la tasa de cambio de la función de distribución acumulada, o el negativo de la tasa de cambio de la función de supervivencia.

Asimismo, la función de riesgo, denotada por $\lambda(t)$, es la tasa instantánea de falla de que

el evento ocurra en un pequeño de tiempo Δt , dado que no ha ocurrido hasta el tiempo t .

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T < t + \Delta t \mid T > t)}{\Delta t} \quad (2.4)$$

Esta función se relaciona con la función de densidad y la función de supervivencia de la siguiente manera:

$$\lambda(t) = \frac{f(t)}{\bar{F}(t)} \quad (2.5)$$

Además, la función de riesgo acumulada, denotada por $\Lambda(t)$, se define como el área bajo la curva de la función de riesgo hasta el tiempo t .

$$\Lambda(t) = \int_0^t \lambda(u) du \quad (2.6)$$

Finalmente, la función de supervivencia puede definirse en términos de la función de riesgo y de la función de riesgo acumulada como:

$$\bar{F}(t) = \exp\left(-\int_0^t \lambda(u) du\right) = \exp(-\Lambda(t)) \quad (2.7)$$

2.3. Modelos paramétricos univariados

A continuación, se presentan los modelos paramétricos de supervivencia más comunes, en los cuales se asume que el tiempo hasta la ocurrencia del evento sigue una distribución conocida.

Distribución Exponencial

Sea T una variable aleatoria continua que sigue una distribución exponencial con parámetro β . Esta distribución se denota como $T \sim \text{Exp}(\beta)$ y su función de densidad se define así:

$$f(t) = \begin{cases} \beta e^{-\beta t}, & \text{si } t \geq 0 \\ 0, & \text{si } t < 0 \end{cases}, \beta > 0 \quad (2.8)$$

Por consiguiente, su media y varianza son, respectivamente:

$$E[T] = \frac{1}{\beta} \text{ y } V[T] = \frac{1}{\beta^2} \quad (2.9)$$

Asimismo, su función de distribución acumulada en el intervalo $[0, +\infty[$ está dada por:

$$F(t) = P(T \leq t) = \int_0^t \beta e^{-\beta u} du = 1 - e^{-\beta t}, \quad 0 \leq t < \infty \quad (2.10)$$

De este modo, la función de supervivencia es:

$$\bar{F}(t) = P[T > t] = 1 - F(t) = e^{-\beta t}, \quad 0 \leq t < \infty \quad (2.11)$$

Distribución Weibull

La variable aleatoria continua T se modela mediante una distribución weibull con parámetros α y β . Por lo tanto, $T \sim W(\alpha, \beta)$, y su función de densidad se expresa como:

$$f(t) = \begin{cases} \alpha \beta t^{\alpha-1} e^{-\beta t^\alpha}, & \text{si } t \geq 0, \\ 0, & \text{si } t < 0 \end{cases}, \quad \alpha > 0, \beta > 0 \quad (2.12)$$

A partir de esto, su media y varianza son, respectivamente:

$$E[T] = \frac{\Gamma(1 + \frac{1}{\alpha})}{\beta^{\frac{1}{\alpha}}} \quad y \quad V[T] = \frac{\Gamma(1 + \frac{2}{\alpha}) - [\Gamma(1 + \frac{1}{\alpha})]^2}{\beta^{\frac{2}{\alpha}}} \quad (2.13)$$

Asimismo, su función de distribución acumulada en el intervalo $[0, +\infty[$ está dada por:

$$F(t) = P[T \leq t] = \int_0^t \alpha \beta u^{\alpha-1} e^{-\beta u^\alpha} du = 1 - e^{-\beta t^\alpha}, \quad 0 \leq t < \infty \quad (2.14)$$

Para el intervalo $]-\infty, 0[$, esta función es cero.

Por consiguiente, su función de supervivencia se presenta como:

$$\bar{F}(t) = P[T > t] = 1 - F(t) = e^{-\beta t^\alpha}, \quad 0 \leq t < \infty \quad (2.15)$$

Distribución Log-Normal

Se define que una variable aleatoria continua T posee distribución log-normal cuando está caracterizada por los parámetros no negativos μ y σ^2 . Esta distribución se representa como $T \sim \text{Ln}(\mu, \sigma^2)$, y su función de densidad se especifica de la siguiente manera:

$$f(t) = \frac{1}{\sqrt{2\pi\sigma}} t^{-1} e^{-\frac{(\ln t - \mu)^2}{2\sigma^2}}, \quad t > 0 \quad (2.16)$$

Por tanto, su media y varianza son respectivamente:

$$E[T] = e^{(\mu + \sigma^2)/2} \quad y \quad V[T] = \mu^2(e^{\sigma^2} - 1) \quad (2.17)$$

Además, su función de distribución acumulada está dada por:

$$F(t) = P[T \leq t] = P[\ln(T) \leq \ln(t)] = P[Z \leq (\ln(t) - \mu)/\sigma] = \Phi((\ln(t) - \mu)/\sigma), \quad (2.18)$$

donde $\Phi(\cdot)$ es la función de distribución acumulada de la distribución Normal estándar.

De este modo, la función de supervivencia se formula como:

$$\bar{F}(t) = P[T > t] = 1 - F(t) = 1 - \Phi((\ln(t) - \mu)/\sigma) \quad (2.19)$$

Distribución Log-Logística

Sea T una variable aleatoria continua que sigue una distribución log-logística, cuyos parámetros no negativos son λ y k . Esta distribución se representa como $T \sim \log(\lambda, k)$, y su función de densidad se especifica de la siguiente forma:

$$f(t) = \frac{\lambda k (\lambda t)^{k-1}}{(1 + (\lambda t)^k)^2}, \quad t > 0 \quad (2.20)$$

Por tanto, su media y varianza son, respectivamente:

$$E[T] = \frac{\pi}{k\lambda \left(\sin\left(\frac{\pi}{k}\right)\right)} \quad y \quad V[T] = \frac{\pi \left(2k \left(1 - \cos\left(\frac{\pi}{k}\right)^2\right) + \pi \sin\left(\frac{\pi(k+2)}{k}\right)\right)}{\left(\sin\left(\frac{\pi(k+2)}{k}\right)\right) \left(\cos\left(\frac{\pi}{k}\right)^2 - 1\right) (\lambda k)^2} \quad (2.21)$$

Además, su función de distribución acumulada está dada por:

$$F(t) = P[T \leq t] = \frac{(\lambda t)^k}{1 + (\lambda t)^k}, \quad t > 0 \quad (2.22)$$

De este modo, la función de supervivencia se formula como:

$$\bar{F}(t) = P[T > t] = 1 - F(t) = 1 - \frac{(\lambda t)^k}{1 + (\lambda t)^k} = \frac{1}{1 + (\lambda t)^k}, \quad t > 0 \quad (2.23)$$

Distribución Gamma

Una variable aleatoria continua T se considera que sigue una distribución gamma con parámetros α y β . Esta función se denota como $T \sim G(\alpha, \beta)$, y su función de densidad está definida de la siguiente manera:

$$f(t) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t}, & \text{si } t \geq 0 \\ 0, & \text{si } t < 0 \end{cases}, \quad \alpha > 0, \beta > 0 \quad (2.24)$$

En consecuencia, su media y varianza son, respectivamente:

$$E[T] = \frac{\alpha}{\beta} \quad y \quad V[T] = \frac{\alpha}{\beta^2} \quad (2.25)$$

Además, su función de distribución acumulada se presenta de la siguiente manera:

$$F(t) = P(T \leq t) = \frac{1}{\Gamma(\alpha)} \int_0^{\beta t} u^{\alpha-1} e^{-u} du \quad (2.26)$$

Por lo tanto, su función de supervivencia es:

$$\bar{F}(t) = P[T > t] = 1 - F(t) = 1 - \frac{1}{\Gamma(\alpha)} \int_0^{\beta t} u^{\alpha-1} e^{-u} du, \quad t > 0 \quad (2.27)$$

2.4. Modelo de Riesgos Proporcionales de Cox

Este modelo proporciona una forma de calcular la función de riesgo en el tiempo t para un sistema dado un conjunto de covariables (Cox, 1972). Se expresa de la siguiente manera:

$$\lambda(t | X) = \lambda_0(t) \exp(X^\top \beta), \quad (2.28)$$

donde $\lambda(t | X)$ es la función de riesgo, $\lambda_0(t)$ es la función de riesgo base, $X = (X_1, \dots, X_k)^\top$ es el vector de covariables, y $\beta = (\beta_1, \dots, \beta_k)$ es el vector de coeficientes que miden el impacto de cada covariable. De esta manera, se tiene un modelo semiparamétrico ya que asume una forma paramétrica (lineal) para el efecto de las covariables a través de $X^\top \beta$ pero no asume ninguna forma específica para la función de riesgo base $\lambda_0(t)$, la cual es tratada de manera no paramétrica.

Si se toma el logaritmo en (2.28), $\lambda(t | X)$ es una función lineal de las covariables, por lo que se puede expresar como una combinación lineal de estas.

$$\log[\lambda(t | X)] = \log[\lambda_0(t)] + X^\top \beta \quad (2.29)$$

Descomponiendo aún más (2.29), se puede escribir de manera explícita para cada covariable X_j y su coeficiente correspondiente β_j como:

$$\log [\lambda(t | X)] = \log [\lambda_0(t)] + X_1\beta_1 + \dots + X_k\beta_k \quad (2.30)$$

Aquí $\lambda_0(t)$ es independiente de las covariables, es decir que $\lambda_0(t)$ describe la tasa de riesgo en función del tiempo para un sistema con valores de covariables iguales a cero. Mientras que β_j es el cambio en el logaritmo del riesgo de ocurrencia del evento cuando X_j aumenta en una unidad, considerando que las demás covariables se mantienen constantes.

Por otro lado, sean $Z_i = \min(T_i, V_i)$ el tiempo observado (falla o censura) para el sistema i , $\delta_i = 1$ si el evento ocurrió y $\delta_i = 0$ si está censurado, $f(z_i | X_i)$ la función de densidad dado el vector de covariables X_i , y $\bar{F}(z_i | X_i)$ la función de supervivencia dado el vector de covariables X_i . De esta manera, para estimar λ_0 y β , se parte de la función de verosimilitud total para n sistemas asumiendo que la censura no es informativa (Klein y Moeschberger, 2003) la cual está definida por:

$$L_n(\beta, \lambda_0) = \prod_{i=1}^n \left[f(z_i | X_i)^{\delta_i} \bar{F}(z_i | X_i)^{1-\delta_i} \right] \quad (2.31)$$

Por otro lado, la función de densidad $f(z_i | X_i)$ se puede descomponer como el producto de la función de riesgo $\lambda(z_i | X_i)$ y la función de supervivencia $\bar{F}(z_i | X_i)$:

$$f(z_i | X_i) = \lambda(z_i | X_i) \bar{F}(z_i | X_i) \quad (2.32)$$

Sustituyendo (2.32) en (2.31) se obtiene:

$$L_n(\beta, \lambda_0) = \prod_{i=1}^n \left[\lambda(z_i | X_i)^{\delta_i} \bar{F}(z_i | X_i) \right] \quad (2.33)$$

Reemplazando (2.28) en (2.33), se tiene:

$$L_n(\beta, \lambda_0) = \prod_{i=1}^n \left[\left(\lambda_0(z_i) \exp(X_i^\top \beta) \right)^{\delta_i} \exp \left(- \int_0^{z_i} \lambda_0(u) \exp(X_i^\top \beta) du \right) \right] \quad (2.34)$$

Para estimar los coeficientes β a partir de (2.34), se puede construir la función de verosimilitud parcial, la cual se enfoca en los términos que dependen únicamente de β . Esto se logra al factorizar los términos de (2.34), de tal manera que los componentes que dependen de $\lambda_0(t)$ se ignoran al maximizar con respecto a β . La función de verosimilitud parcial queda expresada como:

$$L(\beta) = \prod_{i=1}^D \frac{\exp(X_{(i)}^\top \beta)}{\sum_{j \in R(t_i)} \exp(X_j^\top \beta)}, \quad (2.35)$$

donde D es el número de eventos observados (es decir, el número de sistemas que fallaron) y $R(t_i)$ es el conjunto de riesgo en t_i , que incluye a todos los sistemas que estaban en riesgo justo antes de t_i .

Tomando el logaritmo en (2.35), la función log-verosimilitud parcial es:

$$\log[L(\beta)] = \sum_{i=1}^D (X_i^\top \beta) - \sum_{i=1}^D \log \left(\sum_{j \in R(t_i)} \exp(X_j^\top \beta) \right) \quad (2.36)$$

Por otra parte, la estimación de la función de riesgo base, $\lambda_0(t)$, se obtiene también a partir de (2.34). De esta forma, Breslow (1972) propuso un método práctico para realizar esta estimación, la cual se expresa de la siguiente forma:

$$\hat{\lambda}_0(t_i) = \frac{1}{\sum_{j \in R(t_i)} \exp(X_j^\top \hat{\beta})} \quad (2.37)$$

Finalmente, el modelo de Cox a menudo se denomina modelo de riesgos proporcionales porque si se observan dos sistemas con covariables X y X^* , respectivamente, la relación entre sus funciones de riesgo es:

$$\frac{\lambda(t | X)}{\lambda(t | X^*)} = \frac{\lambda_0(t) \exp \left[\sum_{p=1}^k \beta_p X_p \right]}{\lambda_0(t) \exp \left[\sum_{p=1}^k \beta_p X_p^* \right]} = \exp \left[\sum_{p=1}^k \beta_p (X_p - X_p^*) \right] \quad (2.38)$$

2.5. Modelo de Eventos Recurrentes

El análisis de eventos recurrentes es una extensión del análisis de supervivencia que se utiliza para estudiar situaciones en las que un sistema puede experimentar múltiples eventos de interés durante el tiempo de monitoreo (Klein y Moeschberger, 2003). Este modelo permite estudiar el tiempo hasta múltiples eventos recurrentes para cada sistema. El principal enfoque utilizado para analizar este tipo de datos es el de Procesos de Conteo.

Sea T_{ij} el tiempo del j -ésimo evento para el i -ésimo sistema y sea $N_i(t)$ el proceso de conteo para el i -ésimo sistema, que cuenta el número de eventos observados en el sistema i -ésimo hasta el tiempo t . La historia del proceso de conteo y las covariables hasta el tiempo t se denota por $\mathcal{J}_i(t)$.

La función de riesgo para el i -ésimo sistema se define como:

$$\lambda_i(t|\mathcal{J}_i(t)) = \lambda_0(t) \exp(\beta^\top X_i(t)), \quad (2.39)$$

donde $\lambda_0(t)$ es la función de riesgo base, $X_i(t)$ son las covariables observadas en el tiempo t para el i -ésimo sistema, β es el vector de coeficientes de regresión y $\mathcal{J}_i(t)$ representa la historia del proceso de conteo y las covariables $X_i(t)$ hasta el tiempo t .

Asimismo, sea $R(T_{ij})$ el conjunto de riesgo en el tiempo T_{ij} , el cual incluye a todos los sistemas que están en riesgo de experimentar el j -ésimo evento en el tiempo T_{ij} , es decir, aquellos sistemas que no han experimentado el j -ésimo evento antes del tiempo T_{ij} .

Así, para construir la función de verosimilitud, se parte de la probabilidad condicional de que el evento ocurra en el tiempo T_{ij} dado que el sistema estaba en riesgo justo antes de T_{ij} . Por ello, para cada evento observado T_{ij} , esta probabilidad se puede expresar como:

$$L = \frac{\lambda_i(T_{ij}|\mathcal{J}_i(T_{ij}))}{\sum_{l \in R(T_{ij})} \lambda_l(T_{ij}|\mathcal{J}_l(T_{ij}))} \quad (2.40)$$

Sustituyendo (2.39) en (2.40) y simplificando $\lambda_0(T_{ij})$ se obtiene:

$$L = \frac{\lambda_0(T_{ij}) \exp(\beta^\top X_i(T_{ij}))}{\sum_{l \in R(T_{ij})} \lambda_0(T_{ij}) \exp(\beta^\top X_l(T_{ij}))} = \frac{\exp(\beta^\top X_i(T_{ij}))}{\sum_{l \in R(T_{ij})} \exp(\beta^\top X_l(T_{ij}))} \quad (2.41)$$

De esta manera, la función de verosimilitud total $L(\beta)$, que depende implícitamente del tiempo t , se obtiene multiplicando las contribuciones de cada evento observado en todos los sistemas de la siguiente forma:

$$L(\beta) = \prod_{i=1}^n \prod_{j=1}^{N_i(t)} \frac{\exp(\beta^\top X_i(T_{ij}))}{\sum_{l \in R(T_{ij})} \exp(\beta^\top X_l(T_{ij}))}, \quad (2.42)$$

donde n es el número de sistemas y $N_i(t)$ es el número total de eventos recurrentes observados para el i -ésimo sistema hasta el tiempo t . Cabe resaltar que $N_i(t)$ representa el número máximo de eventos observados, donde el tiempo t es diferente para cada sistema.

Finalmente, para simplificar el proceso de estimación de los coeficientes β , se suele trabajar con el logaritmo de la función de verosimilitud (2.42), convirtiéndose en función de log-verosimilitud:

$$\log[L(\beta)] = \sum_{i=1}^n \sum_{j=1}^{N_i(t)} \left[\beta^\top X_i(T_{ij}) - \log \left(\sum_{l \in R(T_{ij})} \exp(\beta^\top X_l(T_{ij})) \right) \right] \quad (2.43)$$

2.6. Modelo de Riesgos Competitivos

A diferencia del modelo clásico de supervivencia, en este modelo los datos de supervivencia presentan como variable de respuesta el tiempo desde un origen bien definido hasta un evento bien definido en presencia de otros eventos, denominados riesgos competidores (Porta et al., 2007). En estos casos, las técnicas clásicas de análisis de supervivencia pueden no ser apropiadas, por lo que se utilizan métodos específicos para el tratamiento de estos datos. Los riesgos competidores, o también denominados fallas, pueden clasificarse en Q ($Q > 1$) causas mutuamente exclusivas (Crowder, 2012).

Sea T el tiempo hasta la falla y C la causa de la falla que experimenta un sistema. En este modelo, se asume que T es una variable aleatoria continua y positiva, y C es una variable aleatoria discreta y positiva que toma valores en el conjunto finito $\{1, \dots, Q\}$. Además, se supone que el sistema falla por una y sólo una causa. Por lo tanto, la distribución conjunta de (T, C) se puede expresar completamente a través de las funciones de subdistribución de la causa q , $\check{F}_q(t)$, o por medio de las funciones de riesgo específica de la causa q , $\check{\lambda}_q(t)$.

La función de subdistribución de la causa q , $\check{F}_q(t)$, que corresponde a la probabilidad de que el sistema falle por la causa q en presencia de los demás riesgos competidores, es:

$$\check{F}_q(t) = P(T \leq t, C = q), \quad q = 1, \dots, Q \quad (2.44)$$

La función de riesgo específica de la causa q , que representa la tasa de ocurrencia solo de la causa q dado que no se ha presentado la falla hasta el tiempo t , se define como:

$$\check{\lambda}_q = \lim_{\Delta t \rightarrow 0} \frac{P(T < t + \Delta t, C = q \mid T \geq t)}{\Delta t}, \quad q = 1, \dots, Q \quad (2.45)$$

Por ende, la función de riesgo acumulada específica de la causa q , $\check{\Lambda}_q$, se define como:

$$\check{\Lambda}_q(t) = \int_0^t \check{\lambda}_q(u) du, \quad q = 1, \dots, Q \quad (2.46)$$

Partiendo de (2.45) y (2.46), la función de riesgo general, $\lambda(t)$, la función riesgo acumulado general, $\Lambda(t)$, y la función de supervivencia general, $\bar{F}(t)$, se definen, respectivamente, como:

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(T < t + \Delta t \mid T \geq t)}{\Delta t} = \sum_{q=1}^Q \check{\lambda}_q(t) \quad (2.47)$$

$$\Lambda(t) = \int_0^t \lambda(u) du = \sum_{q=1}^Q \check{\Lambda}_q(t) \quad (2.48)$$

$$\bar{F}(t) = P(T > t) = e^{-\Lambda(t)} \quad (2.49)$$

Por otro lado, sea $\bar{\bar{F}}_q(t)$ la función de supervivencia específica de la causa q , que corresponde a la probabilidad de que el sistema no falle por la causa q en presencia de los demás riesgos competidores. Se expresa como:

$$\bar{\bar{F}}_q(t) = e^{-\check{\Lambda}_q(t)} \quad (2.50)$$

De esta manera, reemplazando (2.48) en (2.49), la función de supervivencia general, $\bar{F}(t)$, se puede factorizar en las siguientes funciones de supervivencia específica de la causa q , $\bar{\bar{F}}_q(t)$, expresándose de la siguiente forma:

$$\bar{F}(t) = e^{-\sum_{q=1}^Q \check{\Lambda}_q(t)} = \prod_{q=1}^Q e^{-\check{\Lambda}_q(t)} = \prod_{q=1}^Q \bar{\bar{F}}_q(t) \quad (2.51)$$

No obstante, las funciones de supervivencia específica, $\bar{\bar{F}}_q(t)$, no representan ninguna variable aleatoria observable, por lo que $\bar{\bar{F}}_q(t) \neq 1 - \check{F}_q(t)$. Esto se debe a que $\bar{\bar{F}}_q(t)$ reflejaría la probabilidad de no fallar debido a la causa q sin que haya otro evento competidor, lo cual no es el caso aquí.

Finalmente, la función de subdensidad de la causa q , $\check{f}_q(t)$, la función de distribución acumulada de T , $F(t)$, y la función de distribución marginal de C , F_q , se definen como:

$$\check{f}_q(t) = \frac{d}{dt} \check{F}_q(t) = \check{\lambda}_q(t) \bar{F}(t) \quad (2.52)$$

$$F(t) = P(T \leq t) = \sum_{q=1}^Q \check{F}_q(t) \quad (2.53)$$

$$F_q(t) = P(C = q) = \lim_{t \rightarrow \infty} \check{F}_q(t), \quad q = 1, \dots, Q \quad (2.54)$$

Por lo tanto, calculando la integral en (2.52), (2.44) puede obtenerse de la función de riesgo específica de la causa q , $\check{\lambda}_q$, y de la función de supervivencia general, $\bar{F}(t)$, como:

$$\check{F}_q(t) = \int_0^t \check{\lambda}_q(u) \bar{F}(u) du, \quad q = 1, \dots, Q \quad (2.55)$$

Sea T_i el tiempo hasta la falla y C_i la causa de la falla para el sistema i , por lo que para n sistemas se tienen $(T_1, C_1), \dots, (T_n, C_n)$. Asimismo, para cada uno existe un tiempo de censura a la derecha no negativo V_i , por lo que se define a $\delta_i = I(T_i \leq V_i)$ como el indicador de censura, y a $\tilde{C}_i = \delta_i C_i$, como la causa de la falla para los sistemas que fallan (si estos no fallan, entonces \tilde{C}_i sería 0). Así, los datos observados para toda la muestra son:

$$\{Z_i = \min(T_i, V_i), \delta_i, \tilde{C}_i, \quad i = 1, \dots, n\}$$

Se asume lo siguiente: V_i es independiente de (T_i, C_i) ; si $Z_i = T_i$ (es decir, T_i no está censurado), entonces C_i es observado (excluyendo los casos en los que el tiempo de falla es observado sin información sobre la causa de la falla); y los soportes de T y V son disjuntos, es decir que cada tiempo observado corresponde a un solo tipo de evento (falla o censura).

Para hallar la función de verosimilitud, se parte primero de la probabilidad de que en un tiempo z , un sistema falle ($\delta = 1$) por la causa q . Dado que T y V son continuas, entonces $\check{f}_q(t)$ es la función de subdensidad de la causa q de T , y $\bar{G}(v)$ es la función de supervivencia para V . Por lo tanto, para cada falla no observada, esta probabilidad se puede expresar como:

$$\begin{aligned} P(Z = z, \delta = 1, \tilde{C} = q) &= P(T = z, T \leq V, \tilde{C} = q) = P(T = z, V \geq z, \tilde{C} = q) \\ &= P(T = z, \tilde{C} = q) P(V \geq z) = \check{f}_q(z) \bar{G}(z) \end{aligned} \quad (2.56)$$

En segundo lugar, se tiene la probabilidad de que el sistema no falle ($\delta = 0$ y $\tilde{C} = 0$) en el tiempo z . Si $\bar{F}(t)$ es la función de supervivencia de T y $g(v)$ es la función de densidad de V , entonces para cada falla observada esta probabilidad se puede expresar como:

$$\begin{aligned} P(Z = z, \delta = 0, \tilde{C} = 0) &= P(Z = z, \delta = 0) = P(V = z, T > V) \\ &= P(V = z, T > z) = P(T > z) P(V = z) \\ &= \bar{F}(z) g(z) \end{aligned} \quad (2.57)$$

Por lo tanto, de (2.56) y (2.57), la función de verosimilitud para un sistema tiene la siguiente forma:

$$\mathcal{L} = (\check{f}_c(z) \bar{G}(z))^\delta (\bar{F}(z) g(z))^{1-\delta} \quad (2.58)$$

Reordenando (2.58), y extendiendo la función de verosimilitud \mathcal{L} para más de un sistema,

se multiplican las contribuciones de todos los sistemas de la siguiente forma:

$$\mathcal{L} = \prod_{i=1}^n \check{f}_{c_i}(z_i)^{\delta_i} \bar{F}(z_i)^{1-\delta_i} \prod_{i=1}^n \bar{G}(z_i)^{\delta_i} g(z_i)^{1-\delta_i} \quad (2.59)$$

Dado que el tiempo de censura V es independiente del tiempo de falla T , y sus soportes son disjuntos, entonces las censuras no proporcionan información sobre las causas de falla a la función de verosimilitud. Por lo tanto, la función de verosimilitud es proporcional a:

$$\mathcal{L} \propto L = \prod_{i=1}^n \check{f}_{c_i}(z_i)^{\delta_i} \bar{F}(z_i)^{1-\delta_i}. \quad (2.60)$$

Así, si el sistema i falla en s_i por la causa de falla $c_i = q$, entonces su contribución a la verosimilitud será dada por la función de subdensidad de la causa q , $\check{f}_q(z_i)$. Sin embargo, si el sistema i está censurado en el tiempo z_i , el sistema aún está en riesgo de fallar por cualquier causa, por lo que su contribución a la verosimilitud está dada por la función de supervivencia general $\bar{F}(z_i)$.

Finalmente, sea $\delta_{iq} = I(C_i = q)$ el indicador de la causa de falla q en el sistema i , donde $\delta_i = \sum_{q=1}^Q \delta_{iq}$. Reemplazando (2.51) en (2.60) y definiendo $\psi_q(t) = -\bar{F}'_q(t) = \check{\lambda}_q(t) \bar{F}_q(t)$, la función de verosimilitud total puede reescribirse como el producto de Q causas separadas de la siguiente forma:

$$\begin{aligned} L &= \prod_{i=1}^n \left(\prod_{q=1}^Q \check{f}_q(z_i)^{\delta_{iq}} \right) \bar{F}(z_i)^{1-\delta_i} = \prod_{i=1}^n \left(\prod_{q=1}^Q (\check{\lambda}_q(z_i) \bar{F}(z_i))^{\delta_{iq}} \right) \bar{F}(z_i)^{1-\delta_i} \\ &= \prod_{i=1}^n \left\{ \left(\prod_{q=1}^Q (\check{\lambda}_q(z_i) \bar{F}(z_i))^{\delta_{iq}} \right) \left(\prod_{q=1}^Q \bar{F}_q(z_i)^{1-\delta_i} \right) \right\} \\ &= \prod_{i=1}^n \prod_{q=1}^Q \left(\check{\lambda}_q(z_i)^{\delta_{iq}} \left[\prod_{\ell=1}^Q \bar{F}_\ell(z_i)^{\delta_{i\ell}} \right] \bar{F}_q(z_i)^{1-\delta_i} \right) \\ &= \prod_{i=1}^n \prod_{q=1}^Q \check{\lambda}_q(z_i)^{\delta_{iq}} \bar{F}_q(z_i)^{\delta_{iq}} \bar{F}_q(z_i)^{\delta_{iq}} \bar{F}_q(z_i)^{\sum_{\ell \neq q} \delta_{i\ell} + 1 - \delta_i} \\ &= \prod_{i=1}^n \prod_{q=1}^Q \psi_q(z_i)^{\delta_{iq}} \bar{F}_q(z_i)^{1-\delta_{iq}} = \prod_{q=1}^Q \left(\prod_{i=1}^n \psi_q(z_i)^{\delta_{iq}} \bar{F}_q(z_i)^{1-\delta_{iq}} \right) = \prod_{q=1}^Q L_q \end{aligned} \quad (2.61)$$

Esta expresión proporciona una relación de la función de verosimilitud total L con las funciones de verosimilitud específicas por causa q , L_q . Si los tiempos de falla de otras causas de falla se consideran como tiempos de censura, entonces cada L_q corresponde a la función de verosimilitud obtenida de la muestra $\{(Z_i, \delta_{iq}), i = 1, \dots, n\}$, donde las funciones de riesgo, densidad y supervivencia específicas son, respectivamente, $\check{\lambda}_q(t)$, $\psi_q(t)$ y $\bar{F}_q(t)$. Esta

factorización muestra cómo $\check{\lambda}_q(t)$ y $\check{\Lambda}_q(t)$ son directamente estimables a partir de los datos $\{(Z_i, \delta_{iq})\}$, tratando las fallas de otras causas en Z_i como observaciones censuradas.



Capítulo 3

Modelo de Riesgos Competitivos Recurrentes

En el presente capítulo, se presenta una propuesta de estimación paramétrica en un modelo de riesgos competitivos recurrentes en un contexto de ingeniería y confiabilidad (Taylor y Peña, 2013). Se comienza presentando el modelo de riesgos competitivos de evento único; luego, se detalla la estructura de los datos para el modelo de riesgos competitivos recurrentes; posteriormente, se realiza el planteamiento del problema; a continuación, se construye el modelo; y, finalmente, se describe el procedimiento de estimación de los estimadores de máxima verosimilitud de los parámetros del modelo propuesto.

3.1. Modelo de Riesgos Competitivos de Evento Único

Sean T_1, T_2, \dots, T_Q los tiempos (posiblemente latentes y continuos) hasta la falla de Q componentes que pertenecen a un sistema. Asumiendo que estos tiempos son independientes entre sí, la probabilidad de que el componente Q falle antes del tiempo T_Q es la función de distribución marginal $F_q(\cdot)$.

Asimismo, sea S el tiempo de vida útil del sistema, el cual se puede obtener del mínimo de los T_q , donde $q \in \{1, \dots, Q\}$. De esta manera, la probabilidad de que el sistema falle antes del tiempo t es la función de distribución F_S , la cual se puede expresar como el producto de probabilidades de que los q componentes del sistema fallen antes del tiempo t .

$$F_S(t) = P(S \leq t) = 1 - P(S > t) = 1 - \prod_{q=1}^Q [1 - F_q(t)] \quad (3.1)$$

Por otra parte, sea λ_q la función de riesgo marginal del componente q definida por:

$$\lambda_q(t) = \frac{f_q(t)}{1 - F_q(t)} = \frac{f_q(t)}{\bar{F}_q(t)}, \quad (3.2)$$

donde $f_q(t)$ es la función de densidad marginal de $F_q(t)$.

De esta forma, la función de riesgo acumulada marginal del componente q , Λ_q , es:

$$\Lambda_q(t) = -\log(1 - F_q(t)), \quad q = 1, 2, \dots, Q \quad (3.3)$$

Sea $\check{F}_q(t)$ la función de subdistribución del componente q , que es la probabilidad de que el sistema falle antes del tiempo t debido a la falla del componente q considerando la competencia de las fallas de los demás componentes. Si \bar{F}_S es la función de supervivencia del sistema, $\check{F}_q(t)$ se puede expresar en función de \bar{F}_S y λ_q como:

$$\check{F}_q(t) = P\{S \leq t, S = T_q\} = \int_0^t \bar{F}_S(w) \lambda_q(w) dw, \quad q = 1, 2, \dots, Q \quad (3.4)$$

Por otro lado, la función de riesgo específica del componente q , que considera la competencia de las fallas de los demás componentes dado que el sistema ha sobrevivido hasta el tiempo t , $\check{\lambda}_q(t)$, se define como:

$$\check{\lambda}_q(t) = \lim_{\Delta t \rightarrow 0^+} \frac{1}{\Delta t} P\{S \in [t, t + \Delta t), S = T_q | S \geq t\}, \quad q = 1, 2, \dots, Q \quad (3.5)$$

De esta forma, la función de riesgo acumulada específica del componente q , que considera la competencia de las fallas de los demás componentes, $\check{\Lambda}_q(t)$, se define mediante:

$$\check{\Lambda}_q(t) = \int_0^t \check{\lambda}_q(w) dw \quad (3.6)$$

Se puede deducir que $\lambda_q(\cdot) = \check{\lambda}_q(\cdot)$ y $\check{\Lambda}_q(t) = \Lambda_q(t)$ para $q = 1, 2, \dots, Q$ debido a la independencia entre las T_q . Sin embargo, cuando hay múltiples componentes en competencia ($Q > 1$), $\check{\Lambda}_q \neq -\log(1 - \check{F}_q)$ debido a que hay interacción entre los distintos componentes. Aquí \check{F}_q no solo depende del riesgo acumulado debido a la causa q , sino que también está afectada por los riesgos acumulados debido a los demás componentes q' ($q' \neq q$).

Por lo tanto, para hacer inferencias sobre F_S a partir de las F_q o las \check{F}_q , se monitorean n sistemas independientes durante un tiempo de monitoreo aleatorio τ_i para cada uno. Así, el sistema i -ésimo se monitorea durante este tiempo $[0, \tau_i]$, en el cual se asume que $\tau_1, \tau_2, \dots, \tau_n$ son independientes y siguen una distribución común G , la cual no proporciona información

sobre las F_q , por lo que se deduce que los τ_i son independientes de los tiempos entre fallas.

Asimismo, el sistema i -ésimo se monitorea hasta que falle en S_i o hasta τ_i . Además, se asume que cuando el sistema falla, se puede identificar el componente que falló. Así, el tiempo observado se define con $\{(Z_i, \delta_i), i = 1, 2, \dots, n\}$, donde $Z_i = \min(S_i, \tau_i)$ y δ_i es el indicador del componente que falló; por lo que $\delta_i = 0$ si $Z_i = \tau_i$, indicando que el tiempo vida del sistema S_i está censurado por la derecha por τ_i ; y $\delta_i = q$ si $Z_i = T_{iq}$, indicando que el tiempo de vida del sistema S_i es el tiempo hasta la falla del componente q , T_{iq} .

Dadas estas observaciones aleatorias, se pueden hacer inferencias sobre las F_q , de las cuales se inferirá el tiempo de vida del sistema a través de (3.1). Por ello, si el interés está en las funciones de subdistribución del componente q , \check{F}_q , entonces sus estimadores pueden obtenerse de los estimadores de \bar{F}_S y λ_q mediante (3.4) (Taylor y Peña, 2013).

Ejemplo: Tres sistemas y dos componentes

Se tienen 3 sistemas, cada uno con 2 componentes en competencia. Los tiempos de falla de los componentes siguen distribuciones exponenciales con tasas de falla $\beta_1 = 0,095$ y $\beta_2 = 0,07$ para los componentes 1 y 2, respectivamente, y el tiempo de monitoreo τ_i sigue una distribución exponencial con tasa $\beta_3 = 0,085$.

En la cuadro 3.1, se muestra la estructura de datos del enunciado, la cual se generó aleatoriamente:

i	T_1	T_2	$S = \min(T_1, T_2)$	τ	$Z = \min(S, \tau)$	δ
1	3.05	15.7	3.05	13.86	3.05	1
2	10.2	6.5	6.5	12.0	6.5	2
3	18.5	20.0	18.5	10.0	10.0	0

Cuadro 3.1: Datos del Sistema de Riesgos Competitivos de Evento Único (SCRM) para 3 sistemas y 2 componentes.

Asimismo, en la Figura 3.1 se puede apreciar de forma gráfica el contexto de los datos del ejemplo.

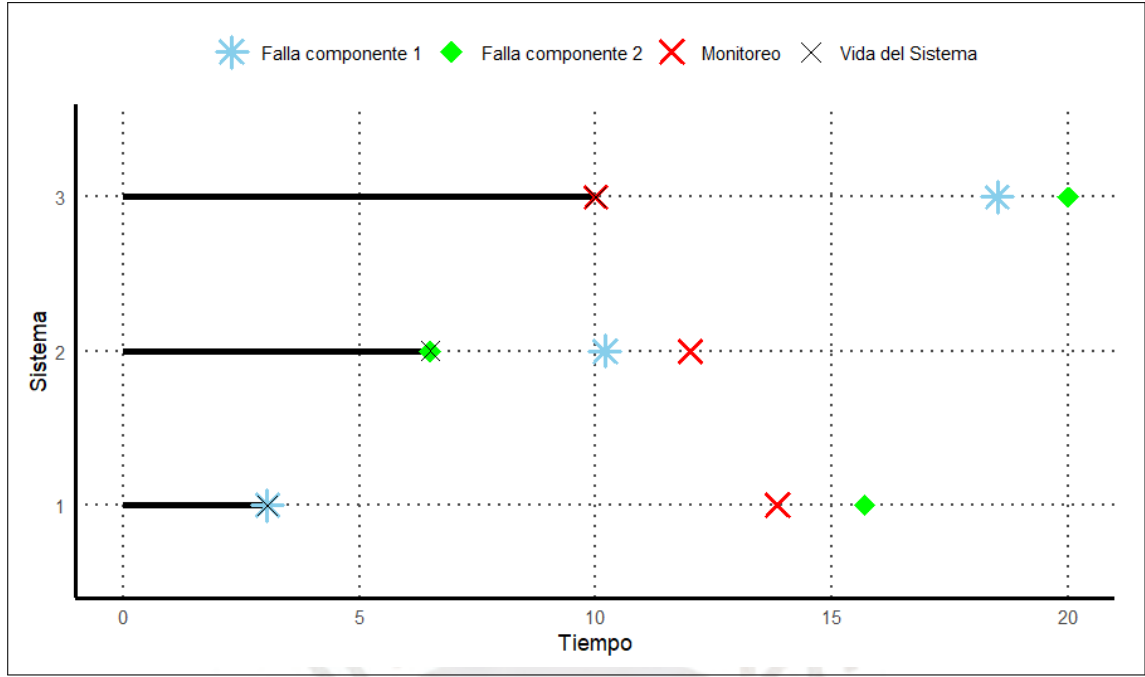


Figura 3.1: Visualización gráfica de datos del Sistema de Riesgos Competitivos de Evento Único (SCRM) para 5 sistemas y 3 componentes.

Dadas estas observaciones aleatorias, se buscará expresar las funciones de subdistribución de los componentes 1 y 2, \check{F}_1 y \check{F}_2 , respectivamente.

En primer lugar, se definen las funciones de supervivencia del sistema y de subdensidad de los componentes 1 y 2.

Sea T_1 el tiempo hasta la falla del componente 1. Se denota como $T_1 \sim \text{Exp}(\beta_1)$, donde β_1 es la tasa de falla. La función de distribución marginal se define como:

$$F_1(t) = 1 - e^{-\Lambda_1(t)} = 1 - e^{-\beta_1 t}, \quad t \geq 0$$

Sea T_2 el tiempo hasta la falla del componente 2. Se denota como $T_2 \sim \text{Exp}(\beta_2)$, donde β_2 es la tasa de falla. La función de distribución marginal se define como:

$$F_2(t) = 1 - e^{-\Lambda_2(t)} = 1 - e^{-\beta_2 t}, \quad t \geq 0$$

Por lo tanto, la función de supervivencia del sistema se define como:

$$\bar{F}_S(t) = [1 - F_1(t)] \times [1 - F_2(t)] = [1 - (1 - e^{-\beta_1 t})] \times [1 - (1 - e^{-\beta_2 t})] = e^{-(\beta_1 + \beta_2)t}$$

De esta manera, las funciones de subdensidad de los componentes 1 y 2 se expresan:

$$\begin{aligned}\check{f}_1(t) &= \check{\lambda}_1(t)\bar{F}_S(t) = \lambda_1(t)\bar{F}_S(t) = \beta_1 e^{-(\beta_1+\beta_2)t}, \quad t \geq 0, \\ \check{f}_2(t) &= \check{\lambda}_2(t)\bar{F}_S(t) = \lambda_2(t)\bar{F}_S(t) = \beta_2 e^{-(\beta_1+\beta_2)t}, \quad t \geq 0,\end{aligned}$$

En segundo lugar, se calculan las funciones de subdensidad y de supervivencia del sistema en los tiempos observados para cada sistema Z_1 , Z_2 y Z_3 .

Para el sistema 1, la función de subdensidad en el punto $Z_1 = 3,05$, debido al componente 1, se calcula de la siguiente forma:

$$\check{f}_1(Z_1 | \delta_1 = 1) = \beta_1 e^{-(\beta_1+\beta_2)Z_1} = \beta_1 e^{-(\beta_1+\beta_2) \times 3,05}$$

De manera similar, para el sistema 2, la función subdensidad en el punto $Z_2 = 6,5$, debido al componente 2, se calcula como:

$$\check{f}_2(Z_2 | \delta_2 = 2) = \beta_2 e^{-(\beta_1+\beta_2)Z_2} = \beta_2 e^{-(\beta_1+\beta_2) \times 6,5}$$

Por último, para el sistema 3, que no ha fallado hasta su tiempo de monitoreo, la función de supervivencia del sistema hasta $Z_3 = 10,0$ se expresa como:

$$\bar{F}_S(Z_3 | \delta_3 = 0) = e^{-(\beta_1+\beta_2)\tau_3} = e^{-(\beta_1+\beta_2) \times 10,0}$$

En tercer lugar, se obtiene la función de verosimilitud considerando los tres sistemas.

La función de verosimilitud total se expresa como:

$$L(\beta_1, \beta_2) = \check{f}_1(Z_1 | \delta_1 = 1) \times \check{f}_2(Z_2 | \delta_2 = 2) \times \bar{F}_S(Z_3 | \delta_3 = 0)$$

Sustituyendo las expresiones obtenidas en el paso 2, la función de verosimilitud se expresa:

$$L(\beta_1, \beta_2) = \left(\beta_1 e^{-(\beta_1+\beta_2) \times 3,05} \right) \times \left(\beta_2 e^{-(\beta_1+\beta_2) \times 6,5} \right) \times \left(e^{-(\beta_1+\beta_2) \times 10,0} \right)$$

Si luego de maximizar la función de verosimilitud se obtienen los valores aproximados de los parámetros $\hat{\beta}_1 \approx 0,1$ y $\hat{\beta}_2 \approx 0,08$, con estos valores se pueden obtener la función de supervivencia del sistema y las funciones de subdistribución de los componentes 1 y 2.

En cuarto lugar, se procede a calcular la función de supervivencia estimada del sistema:

$$\bar{F}_S(t) = e^{-(\hat{\beta}_1+\hat{\beta}_2)t} = e^{-0,18t}$$

Finalmente, se expresan las funciones de subdistribución para los componentes 1 y 2.

La función de subdistribución para el componente 1 se calcula como:

$$\check{F}_1(t) = \int_0^t \bar{F}_S(w) \lambda_q(w) dw = \int_0^t e^{-0,18w} \times 0,1 dw = 0,1 \left(\frac{1 - e^{-0,18t}}{0,18} \right) \approx 0,556(1 - e^{-0,18t})$$

La función de subdistribución para el componente 2 se calcula como:

$$\check{F}_2(t) = \int_0^t \bar{F}_S(w) \lambda_q(w) dw = \int_0^t e^{-0,18w} \times 0,08 dw = 0,08 \left(\frac{1 - e^{-0,18t}}{0,18} \right) \approx 0,444(1 - e^{-0,18t})$$

De esta manera, a partir de los datos observados y después de maximizar la función de verosimilitud, se obtienen las estimaciones de los parámetros β_1 y β_2 que permiten calcular las funciones de distribución del tiempo de vida del sistema, F_S , y de subdistribución para los componentes 1 y 2, \check{F}_1 y \check{F}_2 , respectivamente.

3.2. Estructura de los datos

En el contexto de confiabilidad e ingeniería, cuando un sistema falla durante el tiempo de monitoreo $[0, \tau_i]$ éste se repara para que continúe funcionando. Como resultado, las fallas de los componentes se repiten, lo que hace que el modelo de riesgos competitivos de evento único ya no sea válido en este contexto.

Para plantear correctamente el modelo de riesgos competitivos recurrentes en el contexto mencionado, resultó necesario considerar que, después de cada falla, el sistema se repara inmediatamente (Taylor y Peña, 2013). A partir de esto, existen dos tipos de modelos que incluyen reparaciones: el modelo de reparación perfecta, en el cual la reparación consiste en el reemplazo completo del sistema por uno nuevo luego de cada falla (esquema similar al de un modelo de eventos recurrentes sin riesgos competitivos); y el modelo de reparación parcial, en el cual la reparación consiste en el reemplazo solo del componente del sistema que falló por uno nuevo (esquema para el modelo de riesgos competitivos recurrentes).

3.2.1. Datos para el Modelo de Reparación Perfecta

Sean S_{ij} el tiempo de la j -ésima falla en el sistema i -ésimo, T_{ij} el tiempo de la j -ésima falla en el sistema i -ésimo desde el tiempo de la j -ésima falla anterior ($S_{i,j-1}$), y τ_i el tiempo de monitoreo para el sistema i -ésimo. De este modo, los tiempos entre fallos en el sistema i -ésimo dentro de $[0, \tau_i]$ se pueden obtener de $T_{ij} = S_{ij} - S_{i,j-1}$, teniendo en cuenta que $S_{i0} = 0$ debido a que no existen fallas al inicio del monitoreo.

Asimismo, sea K_i el número total de fallas en el sistema i -ésimo dentro de $[0, \tau_i]$ causadas por cualquier componente. De esta forma, K_i se puede obtener como $K_i = \max \{k : S_{ik} \leq \tau_i\}$, donde S_{ik} es el tiempo en el que ocurre la k -ésima falla en el sistema i .

Finalmente, $\tau_i - S_{i,K_i}$ será la variable de censura por derecha para los tiempos entre fallas posteriores a la última ocurrida (T_{i,K_i}).

3.2.2. Datos para el Modelo de Reparación Parcial

Por otro lado, sean S_{ijq} el tiempo de la j -ésima falla en el sistema i -ésimo debido a la falla del componente q , T_{ijq} el tiempo de la j -ésima falla en el sistema i -ésimo desde el tiempo de la j -ésima falla anterior ($S_{i,j-1,q}$) debido a la falla del componente q . De este modo, los tiempos entre fallas en el sistema i -ésimo dentro de $[0, \tau_i]$ debido a la falla del componente q , se pueden obtener de $T_{ijq} = S_{ijq} - S_{i,j-1,q}$, teniendo en cuenta que $S_{i0q} = 0$ debido a que no existen fallas debido al componente q al inicio del monitoreo.

Así también, sean K_{iq} el número de fallas en el sistema i -ésimo dentro de $[0, \tau_i]$ causadas por las fallas del componente q , y δ_{ij} el indicador del componente que falló en la j -ésima falla. De esta forma, K_{iq} se puede obtener de $K_{iq} = \sum_{j=1}^{K_i} I \{\delta_{ij} = q\}$.

Finalmente, $\tau_i - S_{i,K_{iq},q}$ será la variable de censura por derecha para los tiempos entre fallas posteriores a la última ocurrida debido al componente q , ($T_{i,K_{iq},q}$).

La notación general de los datos para el sistema i -ésimo será:

$$D_i = (K_i, \tau_i, S_{i1}, S_{i2}, \dots, S_{i,K_i}, \tau_i - S_{i,K_i}, \delta_{i1}, \delta_{i2}, \dots, \delta_{i,K_i})$$

Ejemplo

En la Figura 3.2, se muestran de forma gráfica los datos de un sistema con dos riesgos competitivos recurrentes bajo ambos modelos de reparación. Las fallas de los componentes 1 y 2 se muestran en estrellas azules y diamantes verdes, respectivamente; y la censura se muestra con una X roja. Bajo el modelo de reparación perfecta, los tiempos de falla (S_{ij}) y los tiempos entre fallas (T_{ij}) se ven en color negro. Bajo el modelo de reparación parcial, los tiempos de falla (S_{ijq}) y los tiempos entre fallas (T_{ijq}) por componente se ven en sus colores respectivos.

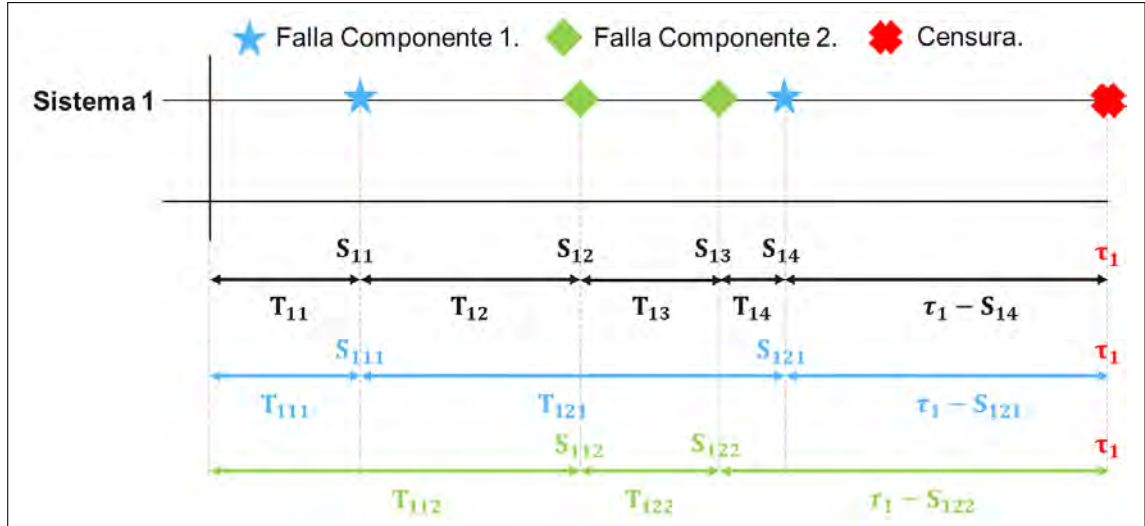


Figura 3.2: Datos observados para un sistema con dos riesgos competitivos recurrentes bajo ambos modelos de reparación.

En el ejemplo, bajo el modelo de reparación perfecta, el número total de fallos en el sistema K_1 será 4 ($K_1 = \max \{k : S_{14} \leq \tau_1\}$); mientras que bajo el modelo de reparación parcial, el número de fallos en el sistema debido al componente 1 será 2 ($K_{11} = \sum_{j=1}^{K_1} I \{\delta_{1j} = 1\}$) y debido al componente 2 será 2 ($K_{12} = \sum_{j=1}^{K_1} I \{\delta_{1j} = 2\}$).

3.3. Planteamiento del problema

Al igual que el modelo de riesgos competitivos de evento único, se asume que los tiempos T_{ijq} son independientes entre sí. Por ende, la probabilidad de que el componente q falle antes del tiempo t es la función de distribución marginal desconocida $F_q(\cdot)$ (Taylor y Peña, 2013).

Asimismo, dado que F_q es desconocida, se asume que ésta pertenece a una familia paramétrica de distribuciones $\mathcal{C}_q = \{F_q(\cdot; \theta_q) : \theta_q \in \Theta_q\}$, en el cual Θ_q es un subconjunto abierto de \mathfrak{R}^{p_q} , donde p_q es el número de parámetros relacionados con F_q .

Sea $f_q(t, \theta_q)$ la función densidad marginal correspondiente a $F_q(t, \theta_q) = P(T_{ijq} \leq t)$, por lo que su función de supervivencia marginal es $\bar{F}_q(t, \theta_q) = 1 - F_q(t, \theta_q)$.

Sea $\lambda_q(t; \theta_q)$, la función de riesgo marginal del componente q , donde θ_q es un vector de parámetros distintos de dimensión $p. \times 1$, donde $p. = \sum_{q=1}^Q p_q$, definido como $\theta = (\theta_1^\top, \theta_2^\top, \dots, \theta_Q^\top)^\top$, los cuales están asociados a λ_q . De esta forma, se tiene la siguiente ecuación (similar a 3.2):

$$\lambda_q(t, \theta_q) = \frac{f_q(t, \theta_q)}{\bar{F}_q(t, \theta_q)}, \quad q = 1, 2, \dots, Q \quad (3.7)$$

Sea $Y_i^\dagger(t) = I \{\tau_i \geq t\}$ la variable indicadora de si el sistema i está en riesgo de fallar

dentro de $[0, \tau_i]$. Bajo el modelo de reparación perfecta, sea $N_i^E = \{N_i^E(t) : t \geq 0\}$ el proceso de conteo que cuenta el número de fallas en el sistema i dentro de $[0, t]$ causadas por cualquier componente. Por ende:

$$N_i^E(t) = \sum_{j=1}^{K_i} I\{S_{ij} \leq t \wedge \tau_i\} \quad (3.8)$$

Bajo el modelo de reparación parcial, sea $N_{iq}^A = \{N_{iq}^A(t) : t \geq 0\}$ el proceso de conteo que cuenta el número de fallas en el sistema i dentro de $[0, t]$ causadas por el componente q .

$$N_{iq}^A(t) = \sum_{j=1}^{K_{iq}} I\{S_{ijq} \leq t \wedge \tau_i\} \quad (3.9)$$

Además, bajo el modelo de reparación perfecta, sea $E_i^E = \{E_i^E(t) : t \geq 0\}$ el proceso de recurrencia hacia atrás que mide el tiempo desde la última falla en el sistema i por cualquier componente hasta t . De esta forma se tiene:

$$E_i^E(t) = t - S_{i, N_i^E(t-)} \quad (3.10)$$

Aquí $t-$ es el instante de tiempo anterior a t , por lo que $N_i^E(t-)$ representa el número de fallas que han ocurrido en el sistema i por cualquier componente hasta justo antes de t . Así, $S_{i, N_i^E(t-)}$ es el tiempo calendario en el que ocurrió la última falla en el sistema i por cualquier componente antes de t . De esta forma, (3.10) define exactamente el tiempo que ha transcurrido desde la última falla en el sistema i por cualquier componente hasta t .

Por otro lado, bajo el modelo de reparación parcial, sea $E_{iq}^A = \{E_{iq}^A(t) : t \geq 0\}$ el proceso de recurrencia hacia atrás que mide el tiempo desde la última falla en el sistema i causada por el componente q hasta t . De esta manera se tiene:

$$E_{iq}^A(t) = t - S_{i, N_{iq}^A(t-), q} \quad (3.11)$$

Similar al modelo anterior, si $t-$ es el instante de tiempo anterior a t , entonces $N_{iq}^A(t-)$ es el número de fallas que han ocurrido en el sistema i por el componente q hasta justo antes de t . Así, $S_{i, N_{iq}^A(t-), q}$ es el tiempo calendario en el que ocurrió la última falla en el sistema i por el componente q antes de t . De esta forma, (3.11) define exactamente el tiempo que ha transcurrido desde la última falla en el sistema i por el componente q hasta t .

Por otra parte, bajo el modelo de reparación parcial, sea $dN_{iq}^A(t)$ el proceso de intensidad que cuenta el número de fallas en el sistema i por el componente q en $[t, t + dt)$, donde $dt > 0$

es un infinitesimal. Esto se puede expresar como:

$$dN_{iq}^A(t) = N_{iq}^A((t + dt)-) - N_{iq}^A(t-) \quad (3.12)$$

Aquí $(t + dt)-$ es el instante de tiempo anterior a $(t + dt)$, por lo que $N_{iq}^A((t + dt)-)$ representa el número de fallas que han ocurrido en el sistema i por el componente q hasta justo antes de $(t + dt)$. De esta manera, dado que $[t, t + dt)$ es un intervalo de tiempo muy pequeño, se asume que el sistema i o no falla o falla solo una vez, por lo que $dN_{iq}^A(t)$ toma solo los valores 0 ó 1, respectivamente.

Finalmente, sea $\mathcal{F} = \{\mathfrak{J}_t : t \geq 0\}$ un filtro que añade información al espacio de probabilidad básico $(\Omega, \mathfrak{J}, P)$. De esta forma, \mathfrak{J}_t es el campo σ que contiene toda la información disponible hasta t . Esto se puede definir como:

$$\mathfrak{J}_t = \mathfrak{J}_0 \vee \left\{ \bigvee_{i=1}^n \sigma \left\{ (N_{iq}^A(v), Y_i^A(v+)) : q = 1, 2, \dots, Q; 0 \leq v \leq t \right\} \right\}$$

Aquí \mathfrak{J}_0 es el campo σ que contiene toda la información disponible en $t = 0$, $N_{iq}^A(v)$ es el proceso de conteo de fallas en el sistema i por el componente q hasta el tiempo v , y $Y_i^A(v+)$ es la variable indicadora que indica si el sistema i está en riesgo de fallar inmediatamente después del tiempo v . Por lo tanto, $\sigma \left\{ (N_{iq}^A(v), Y_i^A(v+)) : q = 1, 2, \dots, Q; 0 \leq v \leq t \right\}$ es el campo σ generado por $N_{iq}^A(v)$ y $Y_i^A(v+)$ desde $v = 0$ hasta $v = t$, el cual contiene toda la información de ambos para el sistema i entre 0 y t . Por último, $\bigvee_{i=1}^n$ es la unión de todos los campos σ generados por todos los sistemas desde $v = 0$ hasta $v = t$.

En la Figura 3.3, se muestra de forma gráfica el planteamiento general del problema bajo un modelo de reparación perfecta a partir de la figura 3.2, donde se tiene un sistema con dos componentes que presentan 2 fallas cada uno. Asimismo, para realizar el gráfico, se utilizaron los parámetros β_1 y β_2 para generar los tiempos de falla de los componentes 1 y 2, respectivamente, y el parámetro β_3 para el tiempo de monitoreo.

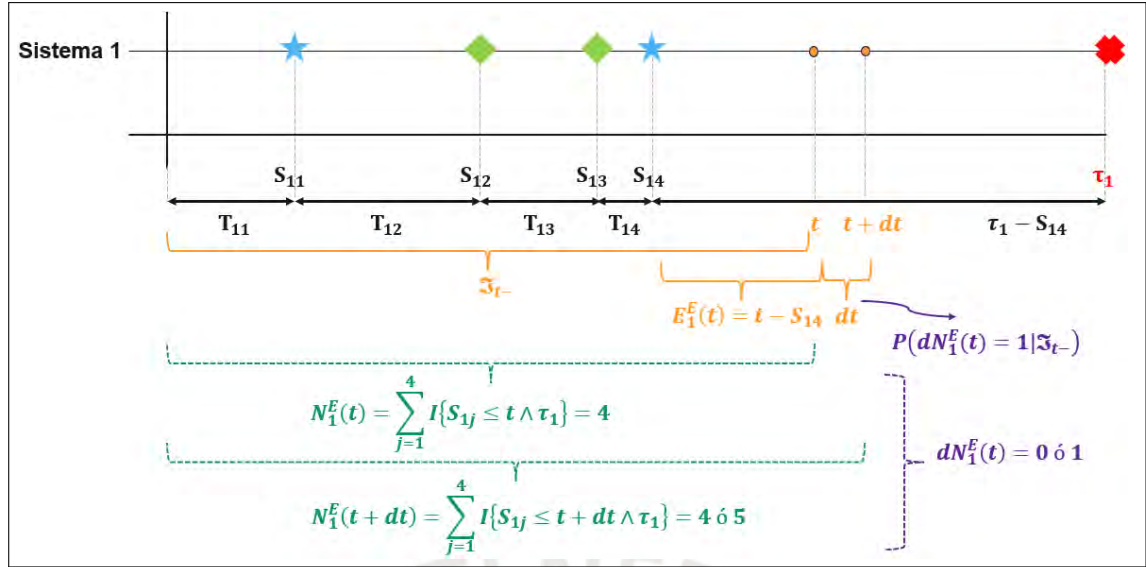


Figura 3.3: Planteamiento del problema bajo un modelo de reparación perfecta.

De la figura, se puede apreciar que para un tiempo t que está dentro de $[0, \tau_1]$, se cuentan las fallas ocurridas hasta ese tiempo mediante el proceso de conteo $N_1^E(t)$, el cual en este caso es 4. Luego, dado que los T_{ij} son independientes, para mantener este supuesto, se calcula el tiempo transcurrido desde la última falla ocurrida, en este caso 4, mediante el proceso de recurrencia hacia atrás $E_1^E(t)$, que es $t - S_{i, N_1^E(t-)} = t - S_{14}$. Posteriormente, para el tiempo $t + dt$, se puede ver que el número de fallas ocurridas hasta ese tiempo, $N_1^E(t + dt)$, puede ser 4 ó 5, dependiendo de si en el instante dt ocurre una falla o no; por lo tanto, $dN_1^E(t)$ puede ser 0 ó 1. Finalmente, \mathfrak{Z}_{t-} es el histórico de fallas (información) hasta justo antes de t .

Por otro lado, en la Figura 3.4, se muestra el planteamiento general del problema bajo un modelo de reparación parcial a partir de la figura 3.3, donde se tiene un sistema con dos componentes que presentan 2 fallas cada uno. Las fallas de los componentes 1 y 2 se muestran en estrellas azules y diamantes verdes, respectivamente; y la censura se muestra con una X roja.

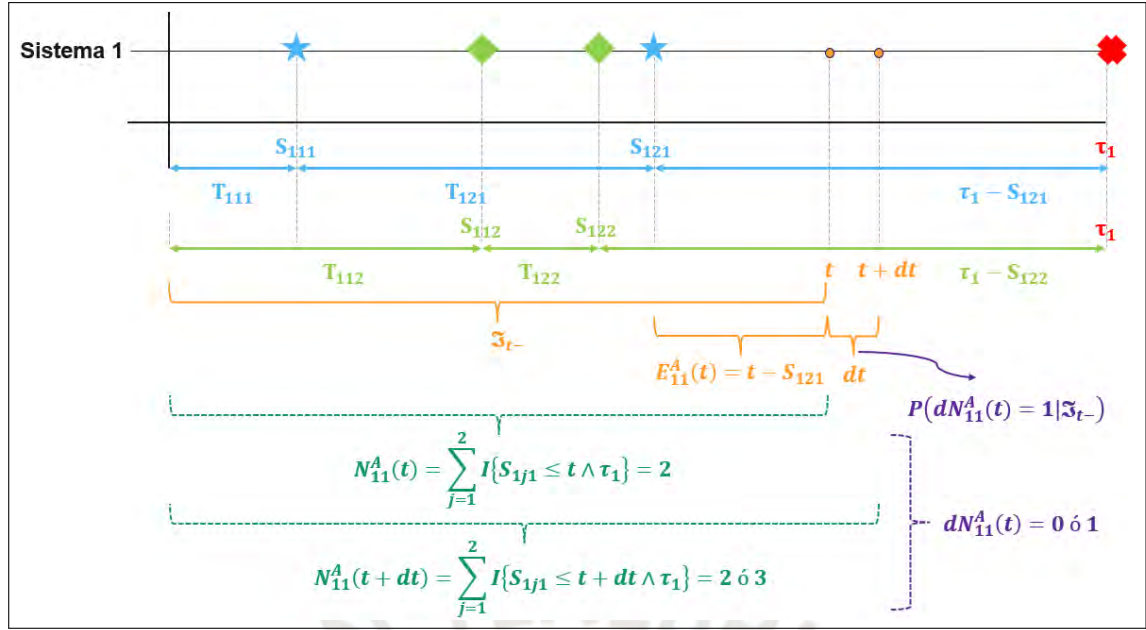


Figura 3.4: Planteamiento del problema bajo un modelo de reparación parcial.

De la figura, se puede apreciar que para un tiempo t que está dentro de $[0, \tau_1]$, se cuentan las fallas ocurridas debido al componente q , en este caso el componente 1, hasta ese tiempo mediante el proceso de conteo $N_{11}^A(t)$, el cual en este caso es 2. Luego, dado que los T_{ijq} son independientes entre sí, para mantener este supuesto, se calcula el tiempo transcurrido desde la última falla ocurrida debido al componente 1, en este caso 2, mediante el proceso de recurrencia hacia atrás $E_{11}^A(t)$, que es $t - S_{i, N_{11}^A(t-), q} = t - S_{121}$. Posteriormente, para el tiempo $t + dt$, se puede ver que el número de fallas ocurridas debido al componente 1 hasta ese tiempo, $N_{11}^A(t + dt)$, puede ser 2 ó 3, dependiendo de si en el instante dt ocurre una falla o no debido al componente 1; por lo tanto, $dN_{11}^A(t)$ puede ser 0 ó 1. Finalmente, \mathfrak{I}_{t-} sigue siendo el histórico de fallas (información) hasta el tiempo t .

De esta manera, dado que deseamos estimar los parámetros de las funciones de distribución marginal F_q , asumiendo que estas pertenecen a una familia paramétrica de distribuciones conocida (Por ejemplo: Weibull, Log-Normal, entre otros), y con la presencia de riesgos competitivos recurrentes (modelo bajo reparación parcial), se partió de la estimación de los parámetros de las funciones de riesgo marginal λ_q en el intervalo de tiempo $[t, t + dt)$, donde $t \in [0, \tau_i]$. Por lo tanto, si los T_{ijq} son independientes entre sí, entonces se utilizó el S_{ijq} como punto de partida y el proceso de recurrencia hacia atrás $E_{iq}^A(t)$ como el tiempo desde la partida hasta el tiempo t de cada función de riesgo marginal, $\lambda_q(E_{iq}^A(t), \theta)$.

3.4. Construcción del Modelo

Para construir el modelo de riesgos competitivos recurrentes, se partió de la probabilidad de que ocurra una falla en el sistema i debido al componente q en el intervalo de tiempo $[t, t + dt)$, $dN_{iq}^A(t)$, dado el historial de fallas hasta justo antes de t , \mathfrak{J}_{t-} . A continuación, se muestra la notación matemática de esta probabilidad:

$$\begin{aligned}
 P(dN_{iq}^A(t) = 1 \mid \mathfrak{J}_{t-}) &= P \{ dN_{iq}^A(t) = 1; \cap_{q' \neq q} [dN_{iq'}^A(t) = 0] \mid \mathfrak{J}_{t-} \} \\
 &= (Y_i^A(t) \lambda_q(E_{iq}^A(t), \theta_q) dt + o(dt)) \\
 &\quad \times \prod_{v=t}^{t+dt} \left\{ 1 - \sum_{q' \neq q} \lambda_{q'}(E_{iq'}^A(v), \theta_{q'}) dv \right\} \\
 &= (Y_i^A(t) \lambda_q(E_{iq}^A(t), \theta_q) dt + o(dt)) \times \epsilon
 \end{aligned} \tag{3.13}$$

Aquí, esta probabilidad condicional se descompone en la probabilidad de que ocurra una falla debido al componente q en el intervalo de tiempo $[t, t + dt)$, $dN_{iq}^A(t)$, y la probabilidad de que no ocurran fallas debido a los otros componentes $q' (q' \neq q)$ dado el historial de fallas hasta justo antes de t , \mathfrak{J}_{t-} , en el mismo intervalo de tiempo $[t, t + dt)$. La primera probabilidad se obtiene multiplicando la variable indicadora $Y_i^A(t)$ de si el sistema i está en riesgo de fallar por la función de riesgo marginal del componente q , que depende del tiempo transcurrido desde la última falla y de los parámetros del modelo θ_q , y por dt , más $o(dt)$ que es un término que denota otras contribuciones al resultado que se vuelven insignificantes cuando dt tiende a cero. Mientras que la segunda, dado que dt es infinitesimal, la probabilidad de que no ocurra una falla por cualquier otro componente q' se aproxima a 1; no obstante, se simplifica a ϵ para representar los efectos combinados de los términos adicionales que se vuelven insignificantes en dt , pero que igual se incluyen por formalidad.

Si al resultado de (3.13) se divide por dt y luego se toma el límite $dt \rightarrow 0$:

$$\begin{aligned}
 P \{ dN_{iq}^A(t) = 1 \mid \mathfrak{J}_{t-} \} &= \lim_{0 < dt \rightarrow 0} \frac{1}{dt} (Y_i^A(t) \lambda_q(E_{iq}^A(t), \theta_q) dt + o(dt)) \times \epsilon \\
 &= Y_i^A(t) \lambda_q(E_{iq}^A(t), \theta_q)
 \end{aligned} \tag{3.14}$$

De esta forma, esta expresión vendría a ser la función de riesgo marginal del componente q , mientras el sistema i se encuentre en riesgo de fallar, denominado también proceso de intensidad de $N_{iq}^\dagger(\cdot)$ con respecto a \mathcal{F} . En consecuencia, la probabilidad de que no ocurra

una falla en el sistema i debido al componente q , $dN_{iq}^A(t)$, en el intervalo de tiempo $[t, t + dt)$, dado el historial de fallas hasta justo antes de t , \mathfrak{J}_{t-} , es:

$$P \{dN_{iq}^A(t) = 0 \mid \mathfrak{J}_{t-}\} = 1 - Y_i^A(t)\lambda_q(E_{iq}^A(t), \theta_q) \quad (3.15)$$

Para construir la función de verosimilitud del modelo, se partió de las probabilidades especificadas en (3.14) y (3.15) para un sistema i . Asimismo, se incluyó a dN_{iq}^A y a $1 - dN_{iq}^A$ como exponentes para indicar si hubo o no una falla en el sistema i , por lo que para un sistema i y una falla debido al componente q , se tuvo:

$$L_{iq}(\theta, t) = [Y_i^A(t)\lambda_q(E_{iq}^A(t), \theta_q)]^{dN_{iq}^A(t)} \times [1 - Y_i^A(t)\lambda_q(E_{iq}^A(t), \theta_q)]^{1-dN_{iq}^A(t)} \quad (3.16)$$

De esta manera, para considerar todos los componentes de un sistema en la función de verosimilitud, estos se pueden expresar como el productorio de las contribuciones de cada componente q en (3.16) como:

$$L_{iq}(\theta, t) = \prod_{q=1}^Q \left\{ [Y_i^A(t)\lambda_q(E_{iq}^A(t), \theta_q)]^{dN_{iq}^A(t)} \times [1 - Y_i^A(t)\lambda_q(E_{iq}^A(t), \theta_q)]^{1-dN_{iq}^A(t)} \right\} \quad (3.17)$$

Asimismo, dado que se necesita la expresión de la función de verosimilitud para toda una muestra de n sistemas, ésta se puede expresar como el productorio de las contribuciones de cada sistema i en (3.17).

$$L_{iq}(\theta, t) = \prod_{i=1}^n \prod_{q=1}^Q \left\{ [Y_i^A(t)\lambda_q(E_{iq}^A(t), \theta_q)]^{dN_{iq}^A(t)} \times [1 - Y_i^A(t)\lambda_q(E_{iq}^A(t), \theta_q)]^{1-dN_{iq}^A(t)} \right\} \quad (3.18)$$

Finalmente, para considerar las recurrencias de las fallas, estas se pueden expresar como el productorio de las contribuciones a lo largo del intervalo de tiempo desde $v = 0$ hasta $v = t$ en (3.18).

$$L(\theta, t) = \prod_{v=0}^t \prod_{i=1}^n \prod_{q=1}^Q \left\{ [Y_i^A(v)\lambda_q(E_{iq}^A(v); \theta_q)]^{dN_{iq}^A(v)} \times [1 - Y_i^A(v)\lambda_q(E_{iq}^A(v); \theta_q)]^{1-dN_{iq}^A(v)} \right\}, \quad (3.19)$$

donde $\prod_{i=1}^n$ es de n valores, en el que cada n corresponde a un sistema.

Factorizando (3.19), tenemos:

$$L(\theta, t) = \prod_{i=1}^n \prod_{q=1}^Q \left\{ \prod_{v=0}^t [Y_i^A(v) \lambda_q(E_{iq}^A(v); \theta_q)]^{dN_{iq}^A(v)} \right. \\ \left. \times \prod_{v=0}^t [1 - Y_i^A(v) \lambda_q(E_{iq}^A(v); \theta_q)]^{1-dN_{iq}^A(v)} \right\} \quad (3.20)$$

Finalmente, la función de verosimilitud se puede expresar como:

$$L(\theta, t) = \prod_{i=1}^n \prod_{q=1}^Q \left\{ \left(\prod_{v=0}^t [Y_i^A(v) \lambda_q(E_{iq}^A(v); \theta_q)]^{dN_{iq}^A(v)} \right) \right. \\ \left. \times \exp \left\{ - \int_0^t Y_i^A(v) \lambda_q(E_{iq}^A(v); \theta_q) dv \right\} \right\} \quad (3.21)$$

Por lo tanto, la función de log-verosimilitud $\{\ell(\theta, t) : t \geq 0\}$ queda definida como:

$$\ell(\theta, t) = \sum_{i=1}^n \left\{ \int_0^t \sum_{q=1}^Q \log[Y_i^A(v) \lambda_q(E_{iq}^A(v); \theta_q)] dN_{iq}^A(v) \right. \\ \left. - \int_0^t Y_i^A(v) \sum_{q=1}^Q \lambda_q(E_{iq}^A(v); \theta_q) dv \right\}. \quad (3.22)$$

3.5. Estimación

Para obtener los estimadores de máxima verosimilitud y sus propiedades, se busca el proceso $p. \times 1$ de vector de puntuación $U = \{U(\theta, t) : t \geq 0\}$ y el proceso $p. \times p.$ de matriz de información observada $I(\theta) = \{I(\theta, t) : t \geq 0\}$. Para ello, se define el operador de gradiente $\nabla_a \equiv \frac{\partial}{\partial a}$ para un vector a . Además, para $q = 1, 2, \dots, Q$, se define:

$$\rho_q(t; \theta_q) = \nabla_{\theta_q} \log[\lambda_q(t; \theta_q)] \text{ y } V_q(t; \theta_q) = \nabla_{\theta_q^\top} \nabla_{\theta_q} \log[\lambda_q(t; \theta_q)]. \quad (3.23)$$

Por lo tanto, $\rho_q(t; \theta_q)$ es un vector de funciones $p_q \times 1$, mientras que $V_q(t; \theta_q)$ es una matriz de funciones $p_q \times p_q$. Dado que el vector del proceso de puntuación se obtiene mediante $U(\theta, t) = \nabla_{\theta} \ell(\theta, t)$, se puede expresar como $U(\theta, t) = (U_1^\top(\theta_1, t), U_2^\top(\theta_2, t), \dots, U_Q^\top(\theta_Q, t))^\top$, donde $U_q(\theta_q, t)$ es el vector $p_q \times 1$ definido por:

$$U_q(\theta_q, t) = \sum_{i=1}^n \int_0^t \rho_q(E_{iq}^A(v); \theta_q) dM_{iq}^A(v; \theta_q). \quad (3.24)$$

Por otro lado, el proceso de la matriz de información de Fisher observada se define como $I(\theta, t) = -\nabla_{\theta^\top} U(\theta, t)$. Se puede verificar que $I(\theta, t)$ es una matriz diagonal de bloques, siendo

la matriz de bloques (q, q) ésima la matriz $p_q \times p_q$, donde:

$$\begin{aligned}
I_{qq}(\theta_q, t) &= \nabla_{\theta_q^\top} U_q(\theta_q, t) \\
&= \sum_{i=1}^n \int_0^t Y_i^A(v) \rho_q(E_{iq}^A(v); \theta_q)^{\otimes 2} \lambda_q(E_{iq}^A(v); \theta_q) dv \\
&\quad - \sum_{i=1}^n \int_0^t V_q(E_{iq}^A(v); \theta_q) dM_{iq}^A(v; \theta_q)
\end{aligned} \tag{3.25}$$

donde para un vector a , tenemos $a^{\otimes 2} = aa^\top$. El estimador de máxima verosimilitud de θ_q dada la ejecución de los procesos hasta el tiempo calendario t^* , denotado como $\hat{\theta}_q(t^*)$, es:

$$U_q(\theta_q, t^*) = 0 \tag{3.26}$$

Generalmente se requieren métodos numéricos, como el algoritmo de Newton-Raphson, para calcular la estimación $\hat{\theta}_q(t^*)$ de θ_q basada en la ecuación anterior.

$$\hat{\theta}_q^{new} \leftarrow \hat{\theta}_q^{old} + I_{qq}(\hat{\theta}_q^{old}, t^*)^{-1} U_q(\hat{\theta}_q^{old}, t^*). \tag{3.27}$$

Capítulo 4

Estudio de Simulación

En este capítulo, se presenta el estudio de simulación en el que se generaron datos con el objetivo de validar la precisión en la estimación de los parámetros inicialmente establecidos.

4.1. Estructura de la simulación

Para demostrar las propiedades de los estimadores de muestra pequeña, se realizaron estudios de simulación. El objetivo de cada estudio fue demostrar las propiedades de los procedimientos de estimación para muestras pequeñas. Para cada simulación, se consideraron n sistemas con tres componentes, los cuales fallan y reciben reparación parcial inmediatamente después de cada falla. Los tiempos entre fallas siguen distribuciones Weibull con distintos parámetros para cada componente, representados por (α_1, β_1) , (α_2, β_2) y (α_3, β_3) . Asimismo, para realizar esta estimación, se generaron 1000 bases de datos.

4.2. Simulación

A continuación, se presentan 3 escenarios en los que se realizaron los estudios de simulación, los cuales fueron llevados a cabo para muestras de 5, 10 y 50 sistemas.

4.2.1. Simulación para 5 sistemas

En esta primera simulación, se consideraron $n = 5$ sistemas, en donde los tiempos entre fallas de sus 3 componentes se simularon bajo distribuciones Weibull con $\theta = (\alpha_1 = 2, \beta_1 = 2, \alpha_2 = 3, \beta_2 = 3, \alpha_3 = 4, \beta_3 = 4)$, respectivamente.

Asimismo, los tiempos de monitoreo de cada sistema se generaron aleatoriamente bajo distribuciones exponenciales con media 4. En el cuadro 4.1. se pueden ver los resultados.

Parámetro	Número de Fallas	Real	Estimado	Sesgo	Error Estándar
Componente 1					
α_1	50.272	2	2.104	0.104	0.304
β_1	50.272	2	1.994	-0.006	0.188
Componente 2					
α_2	73.877	3	3.106	0.106	0.383
β_2	73.877	3	3.014	0.014	0.179
Componente 3					
α_3	91.193	4	4.099	0.099	0.407
β_3	91.193	4	4.010	0.010	0.138

Cuadro 4.1: Resultados de las estimaciones de los parámetros para el caso de $n = 5$ sistemas y 1,000 réplicas utilizando $\alpha_1 = 2$, $\beta_1 = 2$, $\alpha_2 = 3$, $\beta_2 = 3$, $\alpha_3 = 4$, $\beta_3 = 4$.

Además, en la figura 4.1 se pueden ver los histogramas de las estimaciones de los parámetros para el caso de $n = 5$ sistemas.

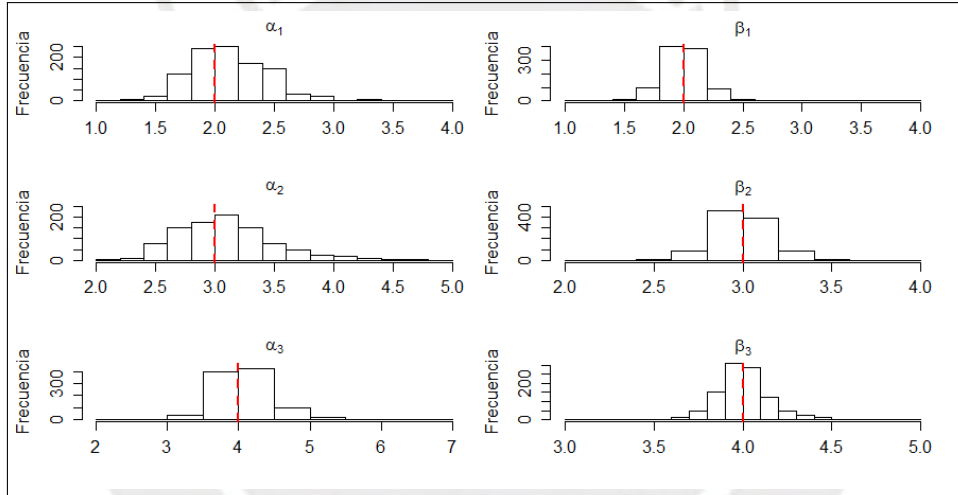


Figura 4.1: Histogramas de las estimaciones de los parámetros para el caso de $n = 5$ sistemas y 1,000 réplicas utilizando $\alpha_1 = 2$, $\beta_1 = 2$, $\alpha_2 = 3$, $\beta_2 = 3$, $\alpha_3 = 4$, $\beta_3 = 4$.

De esta forma, se pudo ver, para el caso de $n = 5$, en promedio 50, 73 y 91 observaciones de los 3 componentes, respectivamente. A pesar del número relativamente grande de observaciones, se pudo apreciar que las distribuciones de muestreo estimadas de los estimadores para cada uno de los parámetros muestran cierto sesgo hacia la derecha para los parámetros de forma $\alpha_1 = 2$, $\alpha_2 = 3$ y $\alpha_3 = 4$, mientras que en el caso de los parámetros de escala $\beta_1 = 2$, $\beta_2 = 3$ y $\beta_3 = 4$ se ve una normalidad aproximada.

4.2.2. Simulación para 10 sistemas

En la segunda simulación, se consideraron $n = 10$ sistemas, en donde los tiempos entre fallas de sus 3 componentes se simularon bajo distribuciones Weibull con $\theta = (\alpha_1 = 2, \beta_1 =$

2, $\alpha_2 = 3$, $\beta_2 = 3$, $\alpha_3 = 4$, $\beta_3 = 4$), respectivamente.

Asimismo, los tiempos de monitoreo de cada sistema se generaron también aleatoriamente a partir de una distribución exponencial con media 4 (ver resultados en cuadro 4.2).

Parámetro	Número de Fallas	Real	Estimado	Sesgo	Error Estándar
Componente 1					
α_1	96.385	2	2.068	0.068	0.218
β_1	96.385	2	2.011	0.011	0.167
Componente 2					
α_2	154.950	3	3.078	0.078	0.266
β_2	154.950	3	3.017	0.017	0.114
Componente 3					
α_3	197.757	4	4.087	0.087	0.293
β_3	197.757	4	4.007	0.007	0.119

Cuadro 4.2: Resultados de las estimaciones de los parámetros para el caso de $n = 10$ sistemas y 1,000 réplicas utilizando $\alpha_1 = 2$, $\beta_1 = 2$, $\alpha_2 = 3$, $\beta_2 = 3$, $\alpha_3 = 4$, $\beta_3 = 4$.

Además, en la figura 4.2 se pueden ver los histogramas de las estimaciones de los parámetros para el caso de $n = 10$ sistemas.

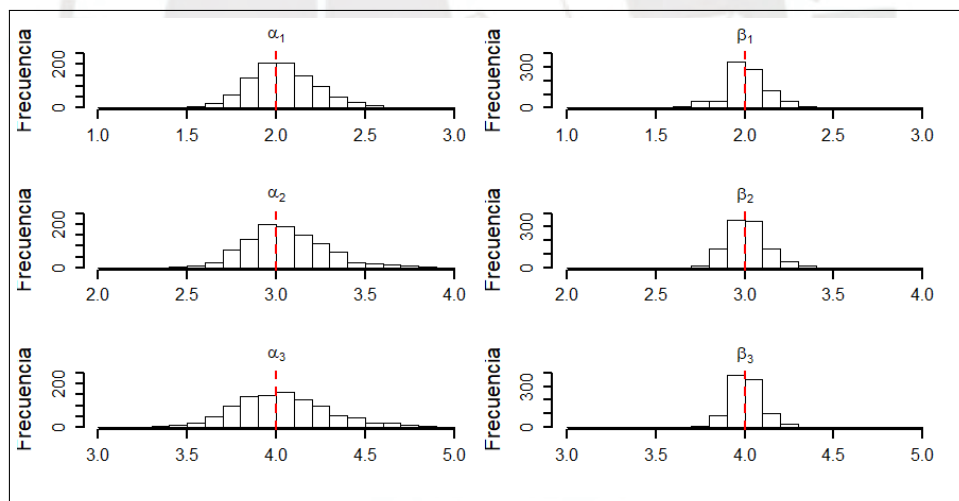


Figura 4.2: Histogramas de las estimaciones de los parámetros para el caso de $n = 10$ sistemas y 1,000 réplicas utilizando $\alpha_1 = 2$, $\beta_1 = 2$, $\alpha_2 = 3$, $\beta_2 = 3$, $\alpha_3 = 4$, $\beta_3 = 4$.

Aquí se pudo ver, para el caso de $n = 10$, que el número de observaciones aumentó en promedio a 96, 154 y 197 de cada una de las tres fallas de Weibull, respectivamente. En consecuencia, se pudo apreciar que las distribuciones de muestreo estimadas para cada uno de los parámetros, tanto de forma como de escala, son aproximadamente normales.

4.2.3. Simulación para 50 sistemas

Finalmente, en la tercera simulación, se consideraron $n = 50$ sistemas, en donde los tiempos entre fallas de sus 3 componentes se simularon bajo distribuciones Weibull con $\theta = (\alpha_1 = 2, \beta_1 = 2, \alpha_2 = 2,1, \beta_2 = 2,1, \alpha_3 = 2,2, \beta_3 = 2,2)$, respectivamente.

Los tiempos de monitoreo de cada sistema se generaron igual que en los escenarios anteriores pero con media 6 (ver resultados en cuadro 4.3).

Parámetro	Número de Fallas	Real	Estimado	Sesgo	Error Estándar
Componente 1					
α_1	479.385	2	2.009	0.009	0.084
β_1	479.385	2	2.005	0.005	0.056
Componente 2					
α_2	494.337	2.1	2.114	0.014	0.081
β_2	494.337	2.1	2.106	0.006	0.058
Componente 3					
α_3	507.104	2.2	2.209	0.009	0.092
β_3	507.104	2.2	2.207	0.007	0.061

Cuadro 4.3: Resultados de las estimaciones de los parámetros para el caso de $n = 50$ sistemas y 1,000 réplicas utilizando $\alpha_1 = 2, \beta_1 = 2, \alpha_2 = 2,1, \beta_2 = 2,1, \alpha_3 = 2,2, \beta_3 = 2,2$.

Asimismo, en la figura 4.3 se pueden ver los histogramas de las estimaciones de los parámetros para el caso de $n = 50$ sistemas.

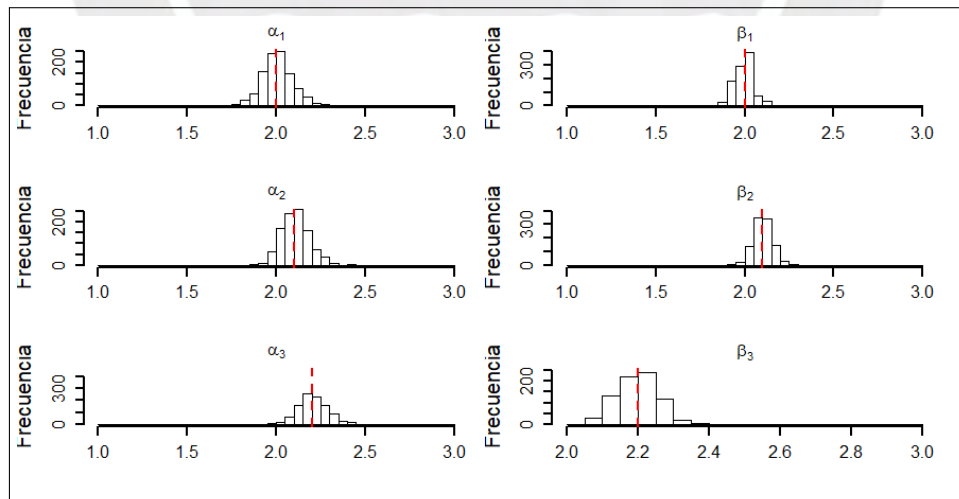


Figura 4.3: Histogramas de las estimaciones de los parámetros para el caso de $n = 50$ sistemas y 1,000 réplicas utilizando $\alpha_1 = 2, \beta_1 = 2, \alpha_2 = 2,1, \beta_2 = 2,1, \alpha_3 = 2,2, \beta_3 = 2,2$.

En este último caso, se demuestra que las distribuciones de muestreo estimadas para cada uno de los parámetros, tanto de forma como de escala, mantienen la normalidad aproximada.

Capítulo 5

Aplicación

En el presente capítulo, se presenta la aplicación del modelo de riesgos competitivos recurrentes a una base de datos de fallas de máquinas inyectoras en una de las fábricas de plástico más importantes del país.

Para entender el contexto de la situación actual, la compañía cuenta con 20 máquinas, las cuales están disponibles todo el año, cuyo funcionamiento se ve interrumpido frecuentemente por las fallas de sus respectivos componentes. Cada máquina tiene aproximadamente 90 componentes; no obstante, para simplificar los cálculos, se tuvieron que agrupar estos en 5. Por lo tanto, se llevó a cabo la estimación paramétrica de las funciones de distribución marginal de los tiempos entre fallas de estos 5 componentes en un contexto en el que sus fallas ocurren recurrentemente y, a su vez, compiten con las fallas de los demás componentes.

Para realizar ello, se utilizó el registro de fallas de todas las máquinas de la compañía del último año, es decir desde el 1 de enero del 2023 hasta el 31 de diciembre del 2023. Se asumió que todas las máquinas entran en funcionamiento por primera vez desde el inicio del presente periodo, lo cual sería equivalente a asumir que las máquinas o son nuevas o recibieron mantenimiento total inmediatamente antes del inicio del tiempo de monitoreo.

Finalmente, las circunstancias en las que fueron recolectados estos datos no fueron completamente conocidas, salvo que fueron registradas manualmente por los maquinistas, por lo que es posible que estos sufran algún sesgo de selección, es decir que solo se tuvieron disponibles los datos de las fallas registradas por los maquinistas (es posible que hayan habido más pero que no fueron registradas).

5.1. Descripción de los datos

La estructura de los datos para aplicar el modelo propuesto se muestra en el cuadro 5.1, en donde “Máquina” representa al sistema en el que ocurrió la falla, “Componente” es el componente del sistema que causó la falla, “Tiempo Inicial” es el tiempo en el que sucedió la falla anterior, y “Tiempo Final” es el tiempo en el que sucedió la falla.

Máquina	Tiempo Acumulado	Componente	Tiempo Inicial	Tiempo Final
INY1	350	Componente 2	0	350
INY1	510	Componente 3	350	510
⋮	⋮	⋮	⋮	⋮
INY20	200,000	Componente q	296,000	300,000

Cuadro 5.1: Estructura de los datos disponibles.

Con el objetivo de entender el cuadro presentado, se tiene en la primera fila, por ejemplo, que la INY1 falló por el componente 2 en el tiempo 350.

En la Figura 5.1, se aprecian las fallas de los componentes en cada máquina, en donde con círculos rojos se tienen las fallas del componente 1, con triángulos azules las fallas del componente 2, con cuadrados verdes las fallas del componente 3, con rombos naranjas las fallas del componente 4 y con estrellas amarillas las fallas del componente 5.

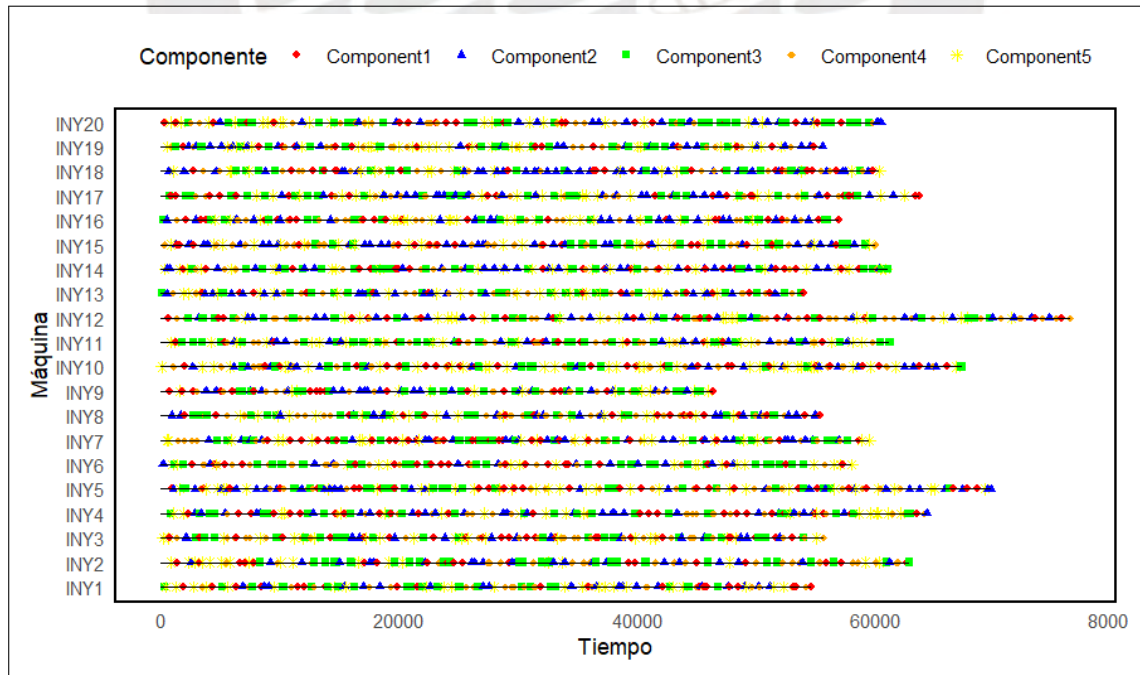


Figura 5.1: Fallas de los componente en cada máquina.

De esta forma, en los cuadros 5.2, 5.3, 5.4 y 5.5, se muestran el número total de fallas,

el tiempo promedio entre fallas y la desviación estándar del tiempo entre fallas de cada componente en cada máquina.

Componente	Número de Fallas	Tiempo Promedio entre Fallas	Desviación Estándar Tiempo entre Fallas
INY1			
Component1	20	2729.95	2449.20
Component2	20	2557.00	2050.68
Component3	21	2227.71	2060.58
Component4	22	2465.5	1800.30
Component5	24	2180.13	1872.12
Total INY1	107		
INY2			
Component1	19	3295.16	2878.32
Component2	22	2782.00	1970.00
Component3	26	2417.88	2044.86
Component4	24	2578.79	2406.60
Component5	18	3349.78	4314.09
Total INY2	109		
INY3			
Component1	26	2070.46	1273.68
Component2	14	3714.21	3182.83
Component3	24	2247.92	1807.63
Component4	28	1987.61	1757.39
Component5	21	2620.33	2223.66
Total INY3	113		
INY4			
Component1	25	2540.56	2123.91
Component2	28	2295.79	1821.70
Component3	27	2330.96	1716.67
Component4	22	2915.64	2373.01
Component5	20	3103.85	3799.50
Total INY4	122		
INY5			
Component1	27	2539.52	1869.70
Component2	30	2324.43	2399.50
Component3	25	2649.20	2911.02
Component4	22	3018.41	2596.47
Component5	31	2232.29	1556.22
Total INY5	135		
INY6			
Component1	21	2724.48	2434.85
Component2	12	4009.25	2961.41
Component3	26	2075.35	1953.77
Component4	19	2934.00	2479.82
Component5	19	3053.42	3421.73
Total INY6	97		

Cuadro 5.2: Fallas de máquina por componente - Parte 1.

Componente	Número de Fallas	Tiempo Promedio entre Fallas	Desviación Estándar Tiempo entre Fallas
INY7			
Component1	29	2020.07	2053.28
Component2	27	2109.26	1897.80
Component3	24	2401.83	2052.94
Component4	15	3804.80	4440.67
Component5	23	2587.61	2365.20
Total INY7	118		
INY8			
Component1	18	3075.61	2614.63
Component2	20	2748.95	2615.04
Component3	22	2341.00	1965.09
Component4	26	2016.77	1590.98
Component5	19	2528.95	1549.10
Total INY8	105		
INY9			
Component1	18	2574.11	2456.34
Component2	23	1919.13	1338.93
Component3	23	1969.83	1777.15
Component4	16	2662.69	2238.17
Component5	12	3831.83	2979.97
Total INY9	92		
INY10			
Component1	26	18648.38	9237.66
Component2	22	2957.73	2310.08
Component3	17	3955.29	3140.12
Component4	20	3245.65	1915.50
Component5	35	1906.26	1642.20
Total INY10	120		
INY11			
Component1	15	3705.40	2922.02
Component2	18	3202.44	2671.02
Component3	32	1912.88	1358.89
Component4	25	2390.20	1633.12
Component5	28	2103.25	1937.42
Total INY11	118		
INY12			
Component1	22	3440.00	3386.59
Component2	27	20752.81	11596.49
Component3	24	2898.92	2210.80
Component4	35	2182.49	2010.81
Component5	22	3246.45	2988.00
Total INY12	130		

Cuadro 5.3: Fallas de máquina por componente - Parte 2.

Componente	Número de Fallas	Tiempo Promedio entre Fallas	Desviación Estándar Tiempo entre Fallas
INY13			
Component1	13	4149.77	2873.94
Component2	22	2345.95	2543.12
Component3	22	2444.14	1549.44
Component4	22	2409.36	1945.49
Component5	23	2031.65	1448.60
Total INY13	102		
INY14			
Component1	22	2665.00	2469.98
Component2	28	2155.21	1600.84
Component3	24	2544.71	2213.09
Component4	14	4293.79	3537.02
Component5	26	2341.00	1800.09
Total INY14	114		
INY15			
Component1	17	3338.88	2875.85
Component2	32	1820.16	1696.59
Component3	16	3698.69	4790.75
Component4	31	1935.32	1433.34
Component5	20	2976.40	2160.21
Total INY15	116		
INY16			
Component1	20	2846.25	2539.73
Component2	21	2590.33	2327.29
Component3	20	2801.75	2094.30
Component4	22	2445.41	2295.64
Component5	23	2177.00	2261.20
Total INY16	106		
INY17			
Component1	21	3033.81	3541.40
Component2	26	2363.35	2459.09
Component3	25	2341.48	1721.82
Component4	23	2475.39	2346.27
Component5	21	2971.29	2576.16
Total INY17	116		
INY18			
Component1	20	2989.60	2637.97
Component2	34	1746.97	1768.72
Component3	19	3092.21	1810.73
Component4	19	3104.16	2911.61
Component5	18	3353.39	3283.19
Total INY18	110		

Cuadro 5.4: Fallas de máquina por componente - Parte 3.

Componente	Número de Fallas	Tiempo Promedio entre Fallas	Desviación Estándar Tiempo entre Fallas
INY19			
Component1	16	3428.13	1959.72
Component2	23	2414.96	2407.21
Component3	19	2508.37	1819.08
Component4	21	2520.67	2094.98
Component5	28	1860.61	2094.19
Total INY19	107		
INY20			
Component1	19	3007.16	2572.20
Component2	21	2879.33	2315.77
Component3	28	2123.14	2210.60
Component4	19	3148.26	3111.50
Component5	21	2613.00	2414.14
Total INY20	108		

Cuadro 5.5: Fallas de máquina por componente - Parte 4.

Asimismo, en el cuadro 5.6 se puede apreciar el resumen del número total de fallas de cada componente, así como también los tiempos promedio entre fallas y las desviaciones estándar de los tiempos entre fallas de cada componente.

Componente	Número Fallas	Tiempo Promedio entre Fallas	Desviación Estándar Tiempo entre Fallas
Component1	414	3861.07	5124.30
Component2	470	3504.32	5536.19
Component3	464	2482.32	2243.11
Component4	445	2622.07	2390.19
Component5	452	2554.07	2502.40
Total	2245		

Cuadro 5.6: Fallas totales por componente.

En total se tuvieron 2245 fallas entre todas las máquinas en el año 2023, de las cuales 414 fueron causadas por el componente 1, 470 por el componente 2, 464 por el componente 3, 445 por el componente 4 y 452 por el componente 5. Así también, los promedios y las desviaciones estándar de los tiempos entre fallas de los componentes fueron: 3861.07 (5124.30) del componente 1, 3504.32 (5536.19) del componente 2, 2482.32 (2243.11) del componente 3, 2622.07 (2390.19) del componente 4, y 2544.07 (2502.40) del componente 5.

5.2. Modelo de Riesgos Competitivos Recurrentes

Los datos se modelaron bajo el esquema de reparación parcial (Modelo de Riesgos Competitivos Recurrentes) debido a que cuando sucede una falla en la fábrica, los técnicos de mantenimiento usualmente solo reparan el componente que causó la falla.

Por otra parte, se consideró que el tiempo de monitoreo de cada máquina terminó con la ocurrencia de la última falla de cada una.

De esta forma, se llevó a cabo la aplicación del modelo a los datos de fallas de máquinas inyectoras, considerando que los tiempos entre fallas de los componentes siguen distribuciones Weibull. Las estimaciones de los parámetros de los tiempos entre falla de cada componente se ven en el cuadro 5.7.

Componente	Parámetros	Real	Desviación Estándar
Component1	$\hat{\alpha}_1$	0.490	0.039
Component1	$\hat{\beta}_1$	6357.028	598.17
Component2	$\hat{\alpha}_2$	0.552	0.048
Component2	$\hat{\beta}_2$	7742.440	703.23
Component3	$\hat{\alpha}_3$	0.553	0.045
Component3	$\hat{\beta}_3$	7242.904	687.44
Component4	$\hat{\alpha}_4$	0.499	0.041
Component4	$\hat{\beta}_4$	6669.854	651.39
Component5	$\hat{\alpha}_5$	0.522	0.046
Component5	$\hat{\beta}_5$	6871.834	621.56

Cuadro 5.7: Estimación de parámetros de los tiempos entre fallas de cada componente asumiendo que siguen distribuciones Weibull.

De la misma manera, se realizó la aplicación del modelo considerando que los tiempos entre fallas siguen distribuciones Log-normal, Gamma y Log logística (ver apéndice 7.3).

Posteriormente, se realizó la comparación del modelo de riesgos competitivos recurrentes considerando que los tiempos entre fallas siguen las distribuciones mencionadas (ver cuadro 5.8).

Modelo	AIC	BIC
Weibull	6940.095	6878.056
Log-Normal	7452.893	7490.853
Gamma	522331.744	522369.704
Log-Logístico	7934.336	7972.296

Cuadro 5.8: Comparación de modelos de riesgos competitivos recurrentes en donde los tiempos entre fallas siguen distribuciones Weibull, Log-Normal, Gamma y Log-Logístico.

Se pudo apreciar que el modelo Weibull presenta los valores más bajos de AIC y BIC, lo que sugiere que este modelo paramétrico se ajusta mejor que los demás modelos comparados. En segundo lugar, se encuentra el modelo Log-Normal; en el tercero, se encuentra el modelo Log-logístico; y en el último, se encuentra el modelo Gamma.

Finalmente, en la Figura 5.2 se pueden ver las curvas de distribución marginal de los tiempos entre fallas de cada componente y la curva de distribución de los tiempos entre fallas de la máquina.

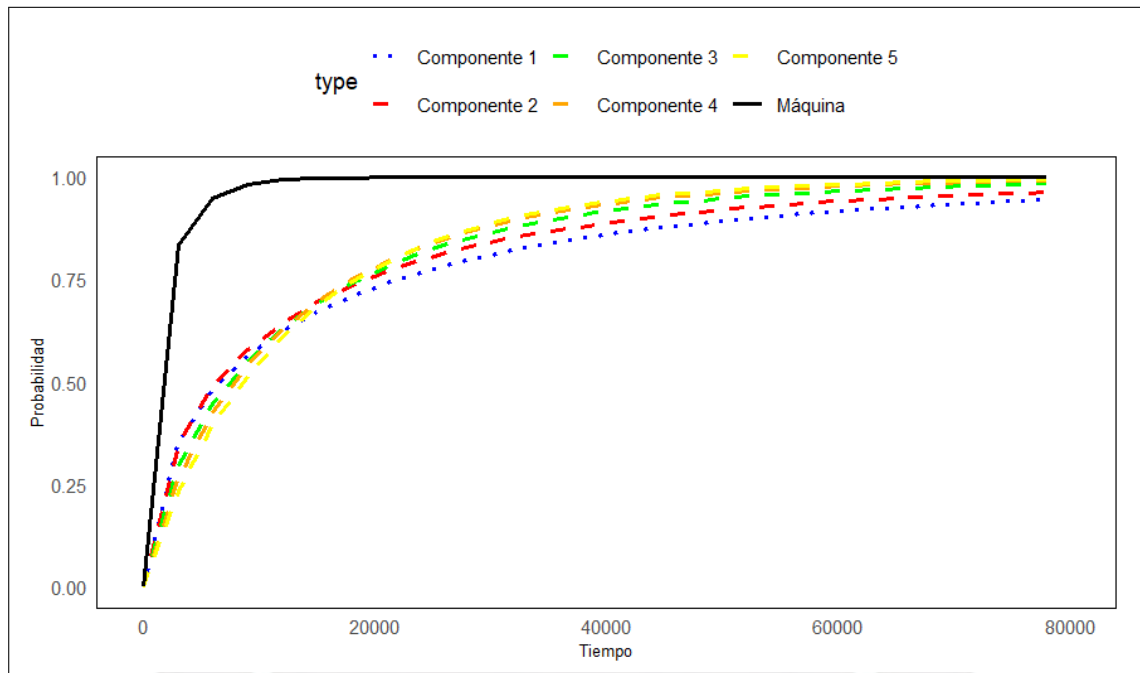


Figura 5.2: Distribuciones estimadas de los tiempos entre fallas de los 5 componentes y de la máquina.

De la figura presentada, se puede ver que a lo largo del tiempo la máquina tiene más probabilidades de fallar debido principalmente al deterioro de sus componentes, los cuales también a lo largo del tiempo tienen más probabilidades de fallar.

Capítulo 6

Conclusiones

Esta tesis complementa el estudio realizado por Taylor y Peña (2013) sobre los modelos de riesgos competitivos recurrentes, considerando modelos paramétricos distintos para los tiempos entre fallas, y comparándolos entre sí, y realizando la aplicación a una base de datos más amplia y diferente como la de fallas de máquinas inyectoras en una empresa real.

En particular, los riesgos competitivos recurrentes permiten un mejor uso del tiempo de monitoreo mediante estrategias de reparación (perfecta y parcial), por lo que maximizan el uso de los recursos al continuar el monitoreo después de una falla. A partir de ello, se demuestra que este enfoque conduce a estimaciones más precisas de las funciones de distribución marginal de la vida útil de los componentes y, en consecuencia, de la distribución de vida del sistema, en comparación con el modelo de riesgos competitivos de evento único, pues en el contexto de una fábrica los riesgos competitivos recurrentes son una realidad. Por lo tanto, se muestra que la distribución muestral de los estimadores es aproximadamente normal para pequeños tamaños de muestra. Así, los estudios de simulación también demostraron e ilustraron las propiedades de los estimadores en muestras pequeñas.

Asimismo, se aplicó el modelo a los datos de fallas de máquinas inyectoras para estimar las funciones distribución marginal de los tiempos entre fallas para cada componente. En este análisis, se asumió que las fallas siguieron distribuciones marginales de Weibull, y que ocurrieron bajo una estrategia de reparación parcial, además de suponer la independencia entre las variables de tiempo latente hasta la falla.

Por otra parte, un enfoque más robusto para analizar conjuntos de datos de riesgos competitivos recurrentes podría ser el uso de métodos no paramétricos, que no imponen supuestos paramétricos específicos. Asimismo, una limitación importante es que los datos de riesgos en competencia suelen ser insuficientes para verificar empíricamente la suposición de

independencia entre las variables de tiempo de falla latente. No obstante, esta limitación no resulta tan crítica en aplicaciones de ingeniería y confiabilidad, como es este caso, aunque de todas formas es relevante desarrollar métodos que permitan considerar posibles dependencias entre los diferentes riesgos, lo cual también es una realidad en una fábrica.

Finalmente, cuando los sistemas pueden ser monitoreados por períodos que se extienden más allá de la primera falla, es ventajoso permitir que los sistemas sean reparados y observar las recurrencias de las fallas durante el tiempo de monitoreo. De esta manera, en este estudio, el conjunto de datos utilizado se refiere a un escenario de riesgos competitivos recurrentes con reparación parcial, observando las fallas debido a múltiples causas.



Capítulo 7

Apéndice

7.1. Códigos R

7.1.1. Generación de datos para el ejemplo de Modelo de Riesgos Competitivos de Evento Único

```
# Parámetros de las distribuciones exponenciales

beta_1 <- 0.095 # Tasa de falla del componente 1
beta_2 <- 0.07  # Tasa de falla del componente 2
beta_3 <- 0.085 # Tasa de monitoreo

# Generar tiempos de falla y tiempos de monitoreo

T_1 <- rexp(3, rate = beta_1)
T_2 <- rexp(3, rate = beta_2)
tau <- rexp(3, rate = beta_3)

# Calcular el tiempo de vida del sistema i, el tiempo observado Z y el
  indicador del componente que falló delta

S <- pmin(T_1, T_2)
Z <- pmin(S, tau)
delta <- ifelse(Z == T_1, 1, ifelse(Z == T_2, 2, 0))

# Datos del Sistema de Riesgos Competitivos de Evento Único (SCRM) para 3
  sistemas y 2 componentes.

datos <- data.frame(
  i = 1:3,
```

```

T_1 = T_1,
T_2 = T_2,
S = S,
tau = tau,
Z = Z,
delta = delta
)
print(datos)

```

	i	T_1	T_2	S	tau	Z	delta
1	1	3.050929	15.7436765	3.0509291	13.4524218	3.0509291	1
2	2	10.246297	6.5380032	6.5380032	12.0390315	6.5380032	2
3	3	18.468157	20.0033105	18.4681576	10.0048508	10.0048508	0

7.1.2. Cálculo de las funciones de verosimilitud, supervivencia del sistema y subdistribución de los componentes 1 y 2 para el ejemplo de Modelo de Riesgos Competitivos de Evento Único

```

# Funci\on de verosimilitud

L <- function(beta_1, beta_2, datos) {
  likelihood <- 1
  for (i in 1:nrow(datos)) {
    if (datos$delta[i] == 1) {
      likelihood <- likelihood * (beta_1 * exp(-beta_1 * datos$Z[i]))
    } else if (datos$delta[i] == 2) {
      likelihood <- likelihood * (beta_2 * exp(-beta_2 * datos$Z[i]))
    } else {
      likelihood <- likelihood * exp(-(beta_1 + beta_2) * datos$Z[i])
    }
  }
  return(likelihood)
}

# C\alculo de la funci\on de verosimilitud para los valores dados

verosimilitud <- L(beta_1, beta_2, datos)
print(verosimilitud)

# Funci\on de supervivencia del sistema

```

```

F_S <- function(t, beta_1, beta_2) {
  return(exp(-(beta_1 + beta_2) * t))
}

# Funci\`on de subdistribuci\`on para el componente 1

F_1 <- function(t, beta_1, beta_2) {
  return(beta_1 * (1 - exp(-(beta_1 + beta_2) * t)) / (beta_1 + beta_2))
}

# Funci\`on de subdistribuci\`on para el componente 2

F_2 <- function(t, beta_1, beta_2) {
  return(beta_2 * (1 - exp(-(beta_1 + beta_2) * t)) / (beta_1 + beta_2))
}

```

7.1.3. Simulador de data

```

library(survival)
library(tidyverse)

# Funci\`on para generar tiempos entre eventos de Weibull

generate_weibull <- function(n, shape, scale) {
  rweibull(n, shape = shape, scale = scale)
}

# Funci\`on para generar tiempos de monitoreo

generate_censor_times <- function(n, rate) {
  rexp(n, rate = 1 / rate)
}

# Funci\`on para calcular la funci\`on de riesgo de Weibull

riesgo_weibull <- function(E_iq_A, shape, scale) {
  (shape / scale) * (E_iq_A / scale)^(shape - 1)
}

```

```

# Funci\on para simular datos de sistemas bajo riesgos competitivos
recurrentes

simulate_rcrm_data <- function(n_units, n_events, shapes, scales, censor_mean
) {
  data <- list()
  for (i in 1:n_units) {
    times_1 <- generate_weibull(n_events, shapes[1], scales[1])
    times_2 <- generate_weibull(n_events, shapes[2], scales[2])
    times_3 <- generate_weibull(n_events, shapes[3], scales[3])

    censor_time <- generate_censor_times(1, censor_mean)

    event_times <- c(times_1, times_2, times_3)
    event_labels <- c(rep(1, n_events), rep(2, n_events), rep(3, n_events))

    df <- data.frame(
      unit = i,
      time = event_times,
      event = event_labels
    )

    df <- df %>% arrange(time)

    df <- df %>% group_by(event) %>% mutate(
      E_iq_A = ifelse(row_number() == 1, 0, time - lag(time))
    )

    df$censor_time <- censor_time
    data[[i]] <- df
  }
  do.call(rbind, data)
}

# Par\ametros de simulaci\on

set.seed(123)
shapes <- c(2, 3, 4)
scales <- c(2, 3, 4)
censor_mean <- 4

# Simulaci\on para n = 5 unidades

```

```

data_n5 <- simulate_rcrm_data(n_units = 5, n_events = 10, shapes = shapes,
  scales = scales, censor_mean = censor_mean)
data_n10 <- simulate_rcrm_data(n_units = 10, n_events = 10, shapes = shapes,
  scales = scales, censor_mean = censor_mean)

# M\'aximo tiempo observado

max_time <- max(data_n5$time)

```

7.1.4. Función de log-verosimilitud negativa

```

neg_log_likelihood <- function(params, data) {
  shape1 <- params[1]
  scale1 <- params[2]
  shape2 <- params[3]
  scale2 <- params[4]
  shape3 <- params[5]
  scale3 <- params[6]

  log_lik <- 0
  for (unit_data in split(data, data$unit)) {
    censor_time <- unique(unit_data$censor_time)
    integral_term <- 0

    for (v in seq(0, max(unit_data$time), by = 0.01)) {

      if (v <= censor_time) {

        events1 <- unit_data %>% filter(event == 1 & time == v)
        events2 <- unit_data %>% filter(event == 2 & time == v)
        events3 <- unit_data %>% filter(event == 3 & time == v)

        if (nrow(events1) > 0) {
          log_lik <- log_lik + log(riesgo_weibull(events1$E_iq_A, shape =
            shape1, scale = scale1))
        }
        if (nrow(events2) > 0) {
          log_lik <- log_lik + log(riesgo_weibull(events2$E_iq_A, shape =
            shape2, scale = scale2))
        }
      }
    }
  }
}

```

```

}
if (nrow(events3) > 0) {
  log_lik <- log_lik + log(riesgo_weibull(events3$E_iq_A, shape =
    shape3, scale = scale3))
}

lambda1 <- ifelse(nrow(events1) > 0, riesgo_weibull(events1$E_iq_A,
  shape = shape1, scale = scale1), 0)
lambda2 <- ifelse(nrow(events2) > 0, riesgo_weibull(events2$E_iq_A,
  shape = shape2, scale = scale2), 0)
lambda3 <- ifelse(nrow(events3) > 0, riesgo_weibull(events3$E_iq_A,
  shape = shape3, scale = scale3), 0)

sum_lambda <- lambda1 + lambda2 + lambda3
integral_term <- integral_term + sum_lambda * 0.01
}
}

log_lik <- log_lik - integral_term
}

return(-log_lik)
}

```

7.1.5. Maximización de verosimilitud de los datos generados

```

# Estimación de parámetros usando constrOptim

initial_params <- c(2, 2, 3, 3, 4, 4)
lower_bounds <- rep(0.1, 6)

fit_n5 <- constrOptim(initial_params, neg_log_likelihood, NULL, ui = diag(6),
  ci = lower_bounds, data = data_n5)
fit_n10 <- constrOptim(initial_params, neg_log_likelihood, NULL, ui = diag(6)
  , ci = lower_bounds, data = data_n10)

fit_n5$par
fit_n10$par

```

7.1.6. Simulación de 1000 réplicas

```
# Función para realizar simulaciones y almacenar resultados

run_simulations <- function(n_sims, n_units, n_events, shapes, scales, censor_
  _mean) {
  results <- matrix(NA, n_sims, 6)
  colnames(results) <- c("shape1", "scale1", "shape2", "scale2", "shape3", "
    scale3")

  for (i in 1:n_sims) {
    sim_data <- simulate_rcrm_data(n_units, n_events, shapes, scales, censor_
      mean)
    fit <- constrOptim(initial_params, neg_log_likelihood, NULL, ui = diag(6)
      , ci = lower_bounds, data = sim_data)
    results[i, ] <- fit$par
  }

  results
}

# Realizar simulaciones

n_sims <- 1000
results_n5 <- run_simulations(n_sims, n_units = 5, n_events = 10, shapes,
  scales, censor_mean)
results_n10 <- run_simulations(n_sims, n_units = 10, n_events = 10, shapes,
  scales, censor_mean)
```

7.1.7. Aplicación del modelo

```
library(survival)
library(tidyverse)
library(openxlsx)

# Registro de datos de fallas de m'quina

file_path <- "Paradas_INY_2023.xlsx"
data_excel <- read.xlsx(file_path)
```

```

# Transformaci\on de los datos para adaptarlos al modelo

data_transformed <- data_excel %>%
  rename(unit = Maquina, time = Produccion_Acum, event = Tipo_Falla) %>%
  mutate(
    event = as.numeric(factor(event, levels = c("Component1", "Component2", "
      Component3", "Component4", "Component5"))),
    censor_time = Produccion_Final
  ) %>%
  select(unit, time, event, censor_time)

# Funci\on de log-verosimilitud negativa para Weibull con 5 riesgos

# Funci\on de verosimilitud negativa para Weibull
neg_log_likelihoood <- function(params, data) {
  shape1 <- params[1]
  scale1 <- params[2]
  shape2 <- params[3]
  scale2 <- params[4]
  shape3 <- params[5]
  scale3 <- params[6]
  shape4 <- params[7]
  scale4 <- params[8]
  shape5 <- params[9]
  scale5 <- params[10]

  log_lik <- 0
  for (unit_data in split(data, data$unit)) {
    censor_time <- unique(unit_data$censor_time)
    integral_term <- 0

    for (v in seq(0, max(unit_data$time), by = 0.01)) {

      if (v <= censor_time) {

        events1 <- unit_data %>% filter(event == 1 & time == v)
        events2 <- unit_data %>% filter(event == 2 & time == v)
        events3 <- unit_data %>% filter(event == 3 & time == v)
        events4 <- unit_data %>% filter(event == 4 & time == v)
        events5 <- unit_data %>% filter(event == 5 & time == v)

        if (nrow(events1) > 0) {

```

```

    log_lik <- log_lik + log(riesgo_weibull(events1$E_iq_A, shape =
      shape1, scale = scale1))
  }
  if (nrow(events2) > 0) {
    log_lik <- log_lik + log(riesgo_weibull(events2$E_iq_A, shape =
      shape2, scale = scale2))
  }
  if (nrow(events3) > 0) {
    log_lik <- log_lik + log(riesgo_weibull(events3$E_iq_A, shape =
      shape3, scale = scale3))
  }
  if (nrow(events4) > 0) {
    log_lik <- log_lik + log(riesgo_weibull(events4$E_iq_A, shape =
      shape4, scale = scale4))
  }
  if (nrow(events5) > 0) {
    log_lik <- log_lik + log(riesgo_weibull(events5$E_iq_A, shape =
      shape5, scale = scale5))
  }

  lambda1 <- ifelse(nrow(events1) > 0, riesgo_weibull(events1$E_iq_A,
    shape = shape1, scale = scale1), 0)
  lambda2 <- ifelse(nrow(events2) > 0, riesgo_weibull(events2$E_iq_A,
    shape = shape2, scale = scale2), 0)
  lambda3 <- ifelse(nrow(events3) > 0, riesgo_weibull(events3$E_iq_A,
    shape = shape3, scale = scale3), 0)
  lambda4 <- ifelse(nrow(events4) > 0, riesgo_weibull(events4$E_iq_A,
    shape = shape4, scale = scale4), 0)
  lambda5 <- ifelse(nrow(events5) > 0, riesgo_weibull(events5$E_iq_A,
    shape = shape5, scale = scale5), 0)

  sum_lambda <- lambda1 + lambda2 + lambda3 + lambda4 + lambda5
  integral_term <- integral_term + sum_lambda * 0.01
}
}

log_lik <- log_lik - integral_term
}

return(-log_lik)
}

```

```

# Parámetros iniciales para optimización

initial_params <- c(2, 2, 3, 3, 4, 4, 5, 5, 6, 6)
lower_bounds <- rep(0.1, 10)

# Dividir los datos en conjuntos de entrenamiento y prueba (80-20)

set.seed(123)
train_index <- sample(seq_len(nrow(data_transformed)), size = 0.8 * nrow(data_
_transformed))
train_data <- data_transformed[train_index, ]
test_data <- data_transformed[-train_index, ]

# Ajuste del modelo con los datos de entrenamiento

fit_train <- constrOptim(initial_params, neg_log_likelihood, NULL, ui = diag
(10), ci = lower_bounds, data = train_data)

# Evaluar el modelo con los datos de prueba

test_log_likelihood <- neg_log_likelihood(fit_train$par, test_data)
print(test_log_likelihood)

```

7.1.8. Comparaciones de modelos

```

# Ajuste de cada modelo

fit_weibull <- constrOptim(initial_params_weibull, neg_log_likelihood_weibull
, NULL, ui = diag(10), ci = lower_bounds, data = data_transformed)

fit_lognormal <- constrOptim(initial_params_lognormal, neg_log_likelihood_
lognormal, NULL, ui = diag(10), ci = lower_bounds, data = data_transformed
)

fit_gamma <- constrOptim(initial_params_gamma, neg_log_likelihood_gamma, NULL
, ui = diag(10), ci = lower_bounds, data = data_transformed)

fit_loglogistic <- constrOptim(initial_params_loglogistic, neg_log_likelihood
_loglogistic, NULL, ui = diag(10), ci = lower_bounds, data = data_
transformed)

```

```

# Cálculo del AIC, BIC y la log-verosimilitud para cada modelo

calculate_aic_bic <- function(log_likelihood, num_params, num_obs) {
  aic <- -2 * log_likelihood + 2 * num_params
  bic <- -2 * log_likelihood + log(num_obs) * num_params
  return(c(AIC = aic, BIC = bic))
}

num_obs <- nrow(data_transformed)

log_likelihood_weibull <- -fit_weibull$value
log_likelihood_lognormal <- -fit_lognormal$value
log_likelihood_gamma <- -fit_gamma$value
log_likelihood_loglogistic <- -fit_loglogistic$value

aic_bic_weibull <- calculate_aic_bic(log_likelihood_weibull, length(initial_
  params_weibull), num_obs)
aic_bic_lognormal <- calculate_aic_bic(log_likelihood_lognormal, length(
  initial_params_lognormal), num_obs)
aic_bic_gamma <- calculate_aic_bic(log_likelihood_gamma, length(initial_
  params_gamma), num_obs)
aic_bic_loglogistic <- calculate_aic_bic(log_likelihood_loglogistic, length(
  initial_params_loglogistic), num_obs)

# Comparación de estadísticas

model_comparison <- data.frame(
  Model = c("Weibull", "Log-Normal", "Gamma", "Log-Logística"),
  AIC = c(aic_bic_weibull["AIC"], aic_bic_lognormal["AIC"], aic_bic_gamma["
    AIC"], aic_bic_loglogistic["AIC"]),
  BIC = c(aic_bic_weibull["BIC"], aic_bic_lognormal["BIC"], aic_bic_gamma["
    BIC"], aic_bic_loglogistic["BIC"]),
  LogLikelihood = c(log_likelihood_weibull, log_likelihood_lognormal, log_
    likelihood_gamma, log_likelihood_loglogistic)
)

print(model_comparison)

```

7.1.9. Comparaciones de eficiencia

```
library(survival)
library(ggplot2)
library(dplyr)
library(tidyr)
library(tidyverse)

set.seed(123)

n <- 30
N <- 1000

# Parámetros de las distribuciones Weibull para dos riesgos competitivos

alpha1 <- 2
beta1 <- 1/2
alpha2 <- 3
beta2 <- 1/3

# Tasa de la distribución exponencial para el tiempo de monitoreo

gamma_vals <- c(2, 1, 1/2, 1/5, 1/10)

# Función para generar tiempos de falla y censura

generate_data <- function(n, alpha1, beta1, alpha2, beta2, gamma) {
  T1 <- rweibull(n, shape = alpha1, scale = beta1)
  T2 <- rweibull(n, shape = alpha2, scale = beta2)
  C <- rexp(n, rate = gamma)
  time <- pmin(T1, T2, C)
  event <- ifelse(time == T1, 1, ifelse(time == T2, 2, 0))
  data.frame(time, event)
}

# Función para calcular MSE

calculate_mse <- function(estimates, true_values) {
  mean((estimates - true_values)^2)
}

# Simulación
```

```

results <- list()
for (gamma in gamma_vals) {
  mse_values <- data.frame(matrix(ncol = 4, nrow = 0))
  colnames(mse_values) <- c("Percentile", "Estimate", "Cause", "Gamma")

  for (i in 1:N) {
    data <- generate_data(n, alpha1, beta1, alpha2, beta2, gamma)

    for (cause in 1:2) {
      cause_data <- data[data$event == cause | data$event == 0,]
      fit <- survfit(Surv(time, event == cause) ~ 1, data = cause_data)
      percentiles <- quantile(fit$surv, probs = seq(0.01, 0.99, by = 0.01))
      mse_cause <- data.frame(
        Percentile = seq(0.01, 0.99, by = 0.01),
        Estimate = percentiles,
        Cause = cause,
        Gamma = gamma
      )

      mse_values <- rbind(mse_values, mse_cause)
    }
  }

  results[[as.character(gamma)]] <- mse_values
}

all_results <- bind_rows(lapply(names(results), function(gamma) {
  res <- results[[gamma]]
  res
}))

grouped_data <- all_results %>%
  group_by(Gamma, Percentile, Cause) %>%
  summarise(MeanEstimate = mean(Estimate, na.rm = TRUE), .groups = 'drop')

pivoted_data <- grouped_data %>%
  pivot_wider(names_from = Cause, values_from = MeanEstimate)

# Eficiencia relativa para cada Gamma y Percentile

relative_efficiency <- pivoted_data %>%

```

```

mutate(RelativeEfficiency = pmin('2' / '1', 1)) %>%
select(Gamma, Percentile, RelativeEfficiency)

# Gráfico de la eficiencia relativa para la Causa 1

ggplot(relative_efficiency, aes(x = Percentile, y = RelativeEfficiency,
  linetype = as.factor(Gamma))) +
  geom_smooth(size = 1, method = "loess", se = FALSE) +
  labs(title = "Eficiencia relativa simulada para la causa 1",
    x = "Percentil",
    y = "Eficiencia Relativa") +
  theme_minimal()

# Eficiencia relativa para el Sistema

relative_efficiency_system <- pivoted_data %>%
  mutate(SystemRelativeEfficiency = pmin(('1' + '2') / 2, 1)) %>%
  select(Gamma, Percentile, SystemRelativeEfficiency)

```



Bibliografía

Allison, P. D. (2010). *Survival Analysis Using SAS: A Practical Guide*, 2nd edn, SAS Institute Inc., Cary, NC.

URL: <http://support.sas.com/publishing>

Bonett, D. C. (2022). Industria del plástico en el Perú, Presentación de PowerPoint.

URL: <https://www.inei.gob.pe/>

Breslow, N. E. (1972). Discussion of the paper by d. r. cox, *Journal of the Royal Statistical Society: Series B (Methodological)* **34**: 216–217.

Cox, D. R. (1972). Regression models and life-tables, *Journal of the Royal Statistical Society: Series B (Methodological)* **34**(2): 187–220.

URL: <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1972.tb00899.x>

Crowder, M. J. (2012). *Multivariate Survival Analysis and Competing Risks*, CRC Press, Boca Raton, FL.

Echeverri, G. y Jaramillo (1980). La industria de productos de plástico. Unidad de Estudios Industriales, Departamento Nacional de Planeación, Colombia. Documento elaborado con la colaboración de Ricardo Obregón.

URL: <https://www.dnp.gov.co/>

Hosmer, D. W., Lemeshow, S. y May, S. (2008). *Applied Survival Analysis: Regression Modeling of Time-to-Event Data*, Wiley Series in Probability and Statistics, 2nd edn, John Wiley & Sons, Inc., Hoboken, NJ.

Instituto Peruano de Economía (2023). Crecimiento del pbi será de apenas 0.8 % en el 2023. Recuperado de El Comercio.

URL: <https://www.elcomercio.pe/economia/peru/crecimiento-del-pbi-sera-de-apenas-08-en-el-2023-informe-ipe-noticia/>

- Kalbfleisch, J. D. y Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*, Wiley Series in Probability and Statistics, 2nd edn, John Wiley & Sons, Hoboken, NJ.
- Klein, J. P. y Moeschberger, M. L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data*, Statistics for Biology and Health, 2nd edn, Springer-Verlag, New York, NY.
- Klein, J. P., van Houwelingen, H. C., Ibrahim, J. G. y Scheike, T. H. (eds) (2014). *Handbook of Survival Analysis*, Handbooks of Modern Statistical Methods, Chapman and Hall/CRC, Boca Raton, FL.
- Kleinbaum, D. G. y Klein, M. (2012). *Survival Analysis: A Self-Learning Text*, Statistics for Biology and Health, 3rd edn, Springer Science+Business Media, LLC, New York, NY.
- Moore, D. F. (2016). *Applied Survival Analysis Using R, Use R!*, 1st edn, Springer International Publishing, Cham, Switzerland.
- Porta, N., Gómez, G., Calle, M. L. y Malats, N. (2007). Competing risks methods, *Technical Report DR 2007/14*, Universitat Politècnica de Catalunya, Barcelona, Spain. Partially supported by Grant 050831 from La Marato de TV3 Foundation and by grant MTM2005-0886 from the Ministerio de Ciencia y Tecnología. Núria Porta is a recipient of a research fellowship from DURSI.
URL: <http://www-eio.upc.es/~nporta/>
- Prentice, R. L. y Zhao, S. (2019). *The Statistical Analysis of Multivariate Failure Time Data: A Marginal Modeling Approach*, 1st edn, CRC Press, Boca Raton, FL.
- Salazar Nishi, J. (2023). Memoria anual 2022-2023. Recuperado de la Sociedad Nacional de Industrias.
URL: <https://www.sni.org.pe/>
- Singer, J. D. y Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*, Oxford University Press, New York, NY.
- Taylor, L. L. y Pena, E. A. (2013). Parametric estimation in a recurrent competing risks model, *Journal of the Iranian Statistical Society (JIRSS)* **12**(1): 153–181. Published in final edited form as: JIRSS, 2013, 12(1): 153–181.