



UNIVERSIDAD NACIONAL DE LA PLATA
FACULTAD DE CIENCIAS EXACTAS
DEPARTAMENTO DE MATEMÁTICA

Trabajo de Tesis Doctoral

**ESTIMACIÓN ROBUSTA EN MODELOS DE
SUPERVIVENCIA CON HAZARD ADITIVO**

Ferrario, Julieta

Director: Álvarez, Enrique E.

2015

Índice general

1. Introducción	3
2. Preliminares	5
2.1. Análisis de Supervivencia	5
2.2. Procesos de Conteo y Martingalas	13
3. Revisión de la literatura	21
3.1. Modelo de riesgo multiplicativo	22
3.2. Modelo de tiempo de falla acelerado	30
3.3. Modelo aditivo	37
4. Propuestas robustas para el Modelo con Hazard Aditivo	43
4.1. Introducción	44
4.2. Estimadores con función de influencia acotada (EFIA)	45
4.2.1. Función de Influencia	51
4.3. Estimadores con punto de ruptura no nulo (EPRNN)	54
4.3.1. En busca del Punto de Ruptura	59
5. Análisis de datos y simulación	61
5.1. Simulación para los EFIA	61
5.2. Simulación para los EPRNN	63
5.3. Análisis de datos	65
6. Conclusiones	69
7. Apéndice	73
8. Referencias	111

Capítulo 1

Introducción

Cuando uno modela una variable de supervivencia, esta es no negativa y representa el tiempo final en que una persona o evento experimenta un suceso terminal como la muerte, remisión completa de una enfermedad o retiro permanente del trabajo. Los modelos estadísticos para este tipo de datos son el objetivo principal de la rama de Estadística llamada *Análisis de Supervivencia*. Esta rama abarca una variedad de técnicas estadísticas para analizar variables aleatorias a valores positivos. Comúnmente el valor de esta variable es el tiempo de falla de un mecanismo físico (por ejemplo, el tiempo de falla de un componente electrónico o mecánico) o biológico (por ejemplo, la muerte de un animal, paciente o célula). Estos datos se presentan en diversos campos como la medicina, la biología, la salud pública, la epidemiología, la ingeniería, la economía y la demografía.

El enfoque más común en el Análisis de Supervivencia para analizar el tiempo de falla de algún evento se basa sobre la función de intensidad, tasa de riesgo o simplemente función hazard, que intuitivamente mide el riesgo instantáneo. Esta función es muy utilizada en los estudios clínicos y existen varios modelos para ella. Los modelos semiparamétricos más utilizados son: multiplicativo, acelerado y aditivo, donde se estima un parámetro de regresión y una función del tiempo de falla inicial, de forma funcional arbitraria, conocida como “función baseline”. En este trabajo solo nos focalizamos en la estimación del parámetro de regresión en el modelo aditivo sin la necesidad de estimar ni previamente ni conjuntamente la función baseline.

Los procedimientos clásicos para la estimación del parámetro se basan en el conocimiento del modelo que generan los datos. Pero estos procedimientos suelen ser muy sensibles a pequeñas desviaciones del modelo y/o ante observaciones atípicas (llama-

das outliers) que puede haber en los datos. Los procedimientos estadísticos robustos tienen como objetivo principal permitir inferencias válidas cuando el modelo no se cumple exactamente y al mismo tiempo ser altamente eficientes bajo el modelo. Por este motivo es de vital importancia la creación de procedimientos robustos que permitan no solo estimar el parámetro en el modelo sino también cuando el modelo se cumple aproximadamente y/o cuando hay outliers en los datos. En tal sentido, por falta de una estimación robusta para el parámetro de regresión del modelo de riesgo aditivo, aquí se proponen dos familias de estimadores robustos: EFIA (estimadores con función de influencia acotada) y EPRNN (estimadores con punto de ruptura no nulo).

En el siguiente capítulo se describirán las hipótesis del modelo aditivo que adoptamos para nuestras propuestas y definimos la Familia de Riesgo Aditivo. Además, introducimos las definiciones y resultados sobre Procesos de Conteo y Martingalas que se utilizarán aquí.

En el Capítulo 3, se encuentra una exhaustiva reseña bibliográfica (o revisión de la literatura) de métodos clásicos y robustos para la estimación del parámetro de regresión en los tres modelos de hazards en eventos simples y recurrentes, esto también se puede encontrar en Álvarez y Ferrario (2012). Esta búsqueda bibliográfica nos motivó a proponer las dos nuevas familias de estimadores robustos que se explican con detalles en el Capítulo 4 para el parámetro de regresión en eventos simples. Para ver el desempeño de nuestras propuestas robustas versus la versión clásica se han desarrollado, en el Capítulo 5, simulaciones y un análisis con datos reales.

En el Capítulo 6 se encuentran las conclusiones y discusiones de esta investigación. Finalmente, las demostraciones se encuentran en el Capítulo 7.

Capítulo 2

Preliminares

2.1. Análisis de Supervivencia

En el Análisis de Supervivencia se suele modelar y analizar datos referidos al tiempo final en el que ocurre un evento, este generalmente se suele referir como “tiempo de falla” del evento principal. Por ejemplo, en el desempeño de ciertos componentes mecánicos es de interés observar los tiempos de falla de estos o en un estudio clínico es de interés el tiempo de recuperación o de recaída de cierta enfermedad en estudio. O sea, la respuesta de interés, T^* , es una variable aleatoria (v.a.) no negativa y, en este trabajo, la suponemos continua. Además el evento de estudio que se observa puede ser recurrente (cuando el evento se repite en el tiempo, por ejemplo, en un estudio médico es de interés saber los tiempos en que ocurren los ataques epilépticos) o simple (cuando el evento no se repite en el tiempo, por ejemplo, la duración de las lámparas o la duración de la respuesta de un paciente al tratamiento de cierta enfermedad, o simplemente la muerte).

Una característica que distingue al Análisis de Supervivencia de otras ramas de la Estadística son las observaciones censuradas, que contienen sólo información parcial sobre la v.a. T^* . Esto comúnmente se debe a que los tiempos de falla de los eventos suelen ser muy grandes entonces en lugar de esperar hasta su falla se suelen censurar.

Hay varias clases de censura que se pueden clasificar en las dos siguientes clases:

1. **Censura fija:** sea τ un número fijo, llamado “tiempo fijo de censura”. En lugar de observar T^* se observa $T^* \wedge \tau$, el mínimo entre T^* y τ .

2. **Censura aleatoria:** sea C una v.a. independiente de T^* , llamada “tiempo de censura aleatorio asociado a T^* ”. En lugar de observar T^* se observa $T^* \wedge C$.

La clase de censura fija surge en aplicaciones de Ingeniería. Por ejemplo, se testea la duración de un lote de transistores. Algunos de ellos pueden tener una duración muy larga y no deseamos esperar para terminar el experimento, puesto que éste podría ser demasiado largo. Por lo tanto, podemos parar el experimento en el tiempo τ preestablecido.

Las censuras aleatorias surgen en aplicaciones médicas, en tratamientos clínicos o estudios de animales o células. Por ejemplo, en un tratamiento clínico, los pacientes pueden entrar al estudio en diferentes tiempos; luego cada uno es tratado con una terapia. Pero aquí las censuras pueden ocurrir por:

- Pérdida de seguimiento: el paciente puede mudarse y nunca lo vemos de nuevo.
- Abandono: el paciente se rehusa a continuar con la terapia o esta no tiene los efectos deseados y es necesario interrumpir el tratamiento.
- Final de estudio.

Por eso es crucial aquí suponer la independencia entre T^* y C cuando solo se observan los tiempos, o la independencia condicional ante posibles variables independientes X , cuando se disponen de ellas. Sin esta suposición se pueden conseguir pocos resultados. Esto parece justificado en el caso de valores aleatorios y pérdidas de seguimiento que se presentan aleatoriamente. Pero, si la razón del abandono esta relacionado con el curso de la terapia, puede haber dependencia entre T^* y C .

En particular, la función de distribución continua de T^* , F_* , puede ser especificada por tres funciones útiles en aplicaciones de Supervivencia:

1. Función de Supervivencia:

$$S_*(t) := \text{Prob}(T^* > t) = 1 - F_*(t) \quad (2.1)$$

2. Función de densidad de probabilidad:

$$f_*(t) := \frac{d}{dt}F_*(t)$$

3. Función de riesgo o hazard:

$$\lambda_*(t) := \lim_{\Delta t \rightarrow 0^+} \frac{\text{Prob}(T^* < t + \Delta t | T^* \geq t)}{\Delta t} = \frac{f_*(t)}{1 - F_*(t)} = \frac{f_*(t)}{S_*(t)} \quad (2.2)$$

Notar que estas tres funciones se relacionan entre sí mediante las siguientes igualdades:

$$\lambda_*(t) = -\frac{d}{dt} \log(S_*(t)),$$

o bien

$$S_*(t) = \exp\left(-\int_0^t \lambda_*(u) du\right) = \exp(-\Lambda_*(t)), \quad (2.3)$$

donde $\Lambda_*(t) = \int_0^t \lambda_*(u) du$ es la función de riesgo acumulado,

$$f_*(t) = -\frac{d}{dt} S_*(t), \quad (2.4)$$

$$f_*(t) = \lambda_*(t) \exp\left(-\int_0^t \lambda_*(u) du\right) \quad (2.5)$$

y, por último,

$$f_*(t) = \lambda_*(t) S_*(t). \quad (2.6)$$

Entonces $f_*(t)$, $S_*(t)$ y $\lambda_*(t)$ están todas en correspondencia uno a uno y por lo tanto son equivalentes para determinar $F_*(t)$.

El enfoque en que se ha realizado esta investigación fue através de la función hazard $\lambda_*(t)$ que es fundamental en Análisis de Supervivencia. Esta función es también conocida como la tasa de falla condicional en Confiabilidad, la influencia de mortalidad en Demografía, la función de intensidad en Procesos Estocásticos, la tasa de falla por edad en Epidemiología, la inversa de la tasa de Mill en Economía o simplemente la tasa de riesgo. Esta función representa el riesgo instantáneo en el tiempo t , ya que, de la definición en (2.2), $\lambda_*(t)\Delta t$ es, para Δt chico, la probabilidad aproximada de que ocurra un evento en $(t, t + \Delta t]$, dado que $T^* > t$. Esta función es particularmente útil para determinar la distribución de los tiempos de falla que utilizan información cualitativa sobre el mecanismo de falla y para describir la forma en que el evento cambia con el tiempo. Hay muchas formas generales de la tasa de

riesgo. La única restricción a la función $\lambda_*(t)$ es que sea no negativa y con integral igual a infinito.

Algunas de las formas generales de la tasa de riesgo para distribuciones paramétricas son:

1. Exponencial: $\lambda_*(t) = \alpha$, por la relación (2.5) ya que $f_*(t) = \alpha \exp(-\alpha t)$, donde $\alpha > 0$.
2. Weibull: $\lambda_*(t) = \alpha \beta t^{\alpha-1}$, por la relación (2.5) ya que $f_*(t) = \alpha \beta t^{\alpha-1} \exp(-\beta t^\alpha)$, donde $\alpha, \beta > 0$.
3. Log-logística: $\lambda_*(t) = \frac{\alpha \beta t^{\alpha-1}}{1 + \beta t^\alpha}$, por la relación (2.5) ya que $f_*(t) = \frac{\alpha \beta t^{\alpha-1}}{(1 + \beta t^\alpha)^2}$, donde $\alpha, \beta > 0$.

Observar que en el caso de la Weibull, cuando $\alpha = 1$ la función de tasa de riesgo es constante en el tiempo y corresponde a una función de riesgo exponencial. Además, para $\alpha < 1$ la función de riesgo disminuye con el tiempo (puede servir por ejemplo para modelar la aparición de fallas en el software, que no envejece, salvo Windows claro), y para $\alpha > 1$ la función de riesgo aumenta con el tiempo (por ejemplo para modelar la supervivencia del hardware, que envejece con el tiempo). Sin embargo en la vida real a veces la función de riesgo puede variar su sentido. Por ejemplo, después de una intervención quirúrgica es posible que inicialmente el riesgo sea alto y posteriormente, pasada una fase crítica, el riesgo disminuya. En estos casos no nos sirven los modelos de Weibull, pero existen otras alternativas. La más empleada es el modelo Log-logístico. Todo esto se ve en los gráficos en los que se hicieron algunos ejemplos, ver la Figura 2.1.

Además, asociado a cada tiempo de falla T^* ocasionalmente se recoge un vector de covariables $\mathbf{Z} \in \mathbb{R}^p$ que puede incluir variables cuantitativas, cualitativas, que dependan o no del tiempo e incluso pueden ser variables externas (independientes del proceso del evento recurrente) y/o internas (si no es externa). Una generalización de los modelos para la función de intensidad del tiempo de falla T^* asociada con el vector \mathbf{Z} incluye variables regresoras. Esta generalización puede ser formada de varias maneras y las tres más comunes son, donde $\lambda_*(t, \mathbf{Z})$ denotamos a la función de intensidad del tiempo de falla dado \mathbf{Z} :

1. Modelo de riesgo proporcional o riesgo multiplicativo de Cox:

$$\lambda_*(t, \mathbf{Z}) = \lambda_0(t) \exp(\beta'_0 \mathbf{Z}(t))$$

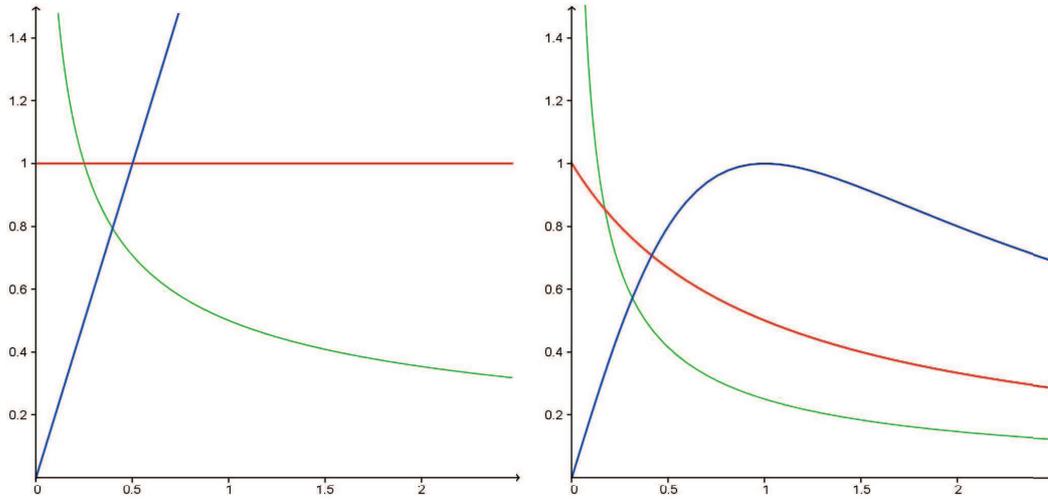


Figura 2.1: Función de riesgo Weibull (derecha) y Log-logística (izquierda) para $\beta = 1$ y para los distintos valores de α : 1 (—), 0.5 (—) y 2 (—).

2. Modelo de tiempo de falla acelerado:

$$\lambda_*(t, \mathbf{Z}) = \lambda_0(t \exp(\beta'_0 \mathbf{Z}(t))) \exp(\beta'_0 \mathbf{Z}(t))$$

3. Modelo de riesgo aditivo:

$$\lambda_*(t, \mathbf{Z}) = \lambda_0(t) + \beta'_0 \mathbf{Z}(t)$$

donde $\beta_0 \in \mathbb{R}^{p+}$ es el vector de parámetros de regresión y $\lambda_0(t) = \lambda_*(t, \mathbf{0})$ es la función baseline, desconocida, arbitraria y no negativa en función del tiempo. En los tres casos se desea estimar el vector de parámetros de regresión, β_0 , y la función baseline, $\lambda_0(t)$.

Una descripción y revisión más amplia de estos tres modelos se realizará en el Capítulo 3. En esta investigación se han desarrollado dos familias de estimadores robustos para el parámetro de regresión del modelo de riesgo aditivo pero en el caso en que la covariable es fija en el tiempo, es decir, cuando el valor de la covariable $\{\mathbf{Z} = \mathbf{z}\}$, la función de riesgo bajo el modelo aditivo es

$$\lambda_*(t, \mathbf{z}) = \lambda_0(t) + \beta'_0 \mathbf{z}. \quad (2.7)$$

La descripción de estas dos familias de estimadores robustos se encuentra en el Capítulo 4.

En vía de las propuestas robustas que aquí presentamos necesitamos caracterizar una nueva familia de distribuciones que nos permitirá contaminar de distintas maneras. Para esto primero necesitamos las expresiones de las siguientes funciones, bajo el modelo (2.7):

1. Tasa de riesgo acumulado:

$$\Lambda_*(t, \mathbf{z}) := \int_0^t \lambda_*(u, \mathbf{z}) du = \Lambda_0(t) + \beta'_0 \mathbf{z}t, \quad (2.8)$$

donde $\Lambda_0(t) := \int_0^t \lambda_0(u) du$ es la función de riesgo acumulado baseline.

2. Función de supervivencia:

$$S_*(t, \mathbf{z}) \stackrel{(2.3)}{=} \exp(-\Lambda_*(t, \mathbf{z})) \stackrel{(2.8)}{=} S_0(t) \exp(-\beta'_0 \mathbf{z}t), \quad (2.9)$$

donde $S_0(t) := \exp(-\Lambda_0(t))$.

3. Función de distribución:

$$F_*(t, \mathbf{z}) \stackrel{(2.1)}{=} 1 - S_*(t, \mathbf{z}) \stackrel{(2.9)}{=} 1 - S_0(t) \exp(-\beta'_0 \mathbf{z}t).$$

4. Función de densidad:

$$f_*(t, \mathbf{z}) \stackrel{(2.6)}{=} \lambda_*(t, \mathbf{z}) S_*(t, \mathbf{z}) \stackrel{(2.9)}{=} S_0(t) \exp(-\beta'_0 \mathbf{z}t) (\lambda_0(t) + \beta'_0 \mathbf{z}).$$

Por otro lado, las covariables \mathbf{Z} son independientes, no negativas y fijas en el tiempo, con función de distribución F_Z y densidad f_Z . Además, hay una v.a. latente de censura C con distribución F_C (y densidad f_C) que puede depender de \mathbf{Z} , pero no se supone que C es independiente de T^* . Adoptamos un supuesto más débil, que la censura es *no informativa* en el sentido de que, dado el vector de covariable que toma el valor $\mathbf{Z} = \mathbf{z}$, la censura y el tiempo del evento son independientes, es decir, que para todo \mathbf{z}

$$T^* \perp C \mid \mathbf{Z} = \mathbf{z} \quad (2.10)$$

Esto implica que

$$\begin{aligned} F_{[T^*, C] | Z} &= F_{T^* | Z} F_{C | Z} \\ F_{T^*, C, Z} &= F_{T^* | Z} F_{C | Z} F_Z. \end{aligned}$$

Debido a esta censura aleatoria, los tiempos observados son $T := T^* \wedge C$. Condicionado sobre $\{\mathbf{Z} = \mathbf{z}\}$, su función de supervivencia es

$$\text{Prob}(T > t | \mathbf{Z} = \mathbf{z}) \stackrel{(2.10)}{=} \text{Prob}(T^* > t | \mathbf{Z} = \mathbf{z}) \text{Prob}(C > t | \mathbf{Z} = \mathbf{z}) = S_*(t, \mathbf{z}) S_C(t) \quad (2.11)$$

$$\stackrel{(2.9)}{=} S_0(t) S_C(t) \exp(-\boldsymbol{\beta}'_0 \mathbf{z} t) = \tilde{S}(t) \exp(-\boldsymbol{\beta}'_0 \mathbf{z} t), \quad (2.12)$$

donde $\tilde{S}(t) := S_0(t) S_C(t)$. Esto implica las siguientes funciones de distribución condicional y de densidad, respectivamente:

$$\begin{aligned} F_{T | Z}(t, \mathbf{z}) &= \text{Prob}(T \leq t | \mathbf{Z} = \mathbf{z}) = 1 - \text{Prob}(T > t | \mathbf{Z} = \mathbf{z}) \\ &\stackrel{(2.11)}{=} 1 - S_*(t, \mathbf{z}) S_C(t) \stackrel{(2.12)}{=} 1 - \tilde{S}(t) \exp(-\boldsymbol{\beta}'_0 \mathbf{z} t), \\ f_{T | Z}(t, \mathbf{z}) &= \frac{\partial}{\partial t} F_{T | Z}(t, \mathbf{z}) = -\frac{d}{dt} \tilde{S}(t) \exp(-\boldsymbol{\beta}'_0 \mathbf{z} t) + \tilde{S}(t) \boldsymbol{\beta}'_0 \mathbf{z} \exp(-\boldsymbol{\beta}'_0 \mathbf{z} t) \\ &\stackrel{(2.4)}{=} (f_0(t) S_C(t) + S_0(t) f_C(t)) \exp(-\boldsymbol{\beta}'_0 \mathbf{z} t) + \tilde{S}(t) \boldsymbol{\beta}'_0 \mathbf{z} \exp(-\boldsymbol{\beta}'_0 \mathbf{z} t) \\ &= \left(\frac{f_0(t)}{S_0(t)} \tilde{S}(t) + \frac{f_C(t)}{S_C(t)} \tilde{S}(t) \right) \exp(-\boldsymbol{\beta}'_0 \mathbf{z} t) + \tilde{S}(t) \boldsymbol{\beta}'_0 \mathbf{z} \exp(-\boldsymbol{\beta}'_0 \mathbf{z} t) \\ &= \tilde{S}(t) \exp(-\boldsymbol{\beta}'_0 \mathbf{z} t) [\lambda_0(t) + \lambda_C(t) + \boldsymbol{\beta}'_0 \mathbf{z}] \\ &= \tilde{S}(t) \exp(-\boldsymbol{\beta}'_0 \mathbf{z} t) [\tilde{\lambda}(t) + \boldsymbol{\beta}'_0 \mathbf{z}], \end{aligned} \quad (2.13)$$

donde $\tilde{\lambda}(t) := \lambda_0(t) + \lambda_C(t)$. Por lo tanto, la densidad conjunta no condicionada de T y \mathbf{Z} es

$$\begin{aligned} f_{T, Z}(t, \mathbf{z}) &= f_{T | Z}(t, \mathbf{z}) f_Z(\mathbf{z}) \\ &= \tilde{S}(t) \exp(-\boldsymbol{\beta}'_0 \mathbf{z} t) [\tilde{\lambda}(t) + \boldsymbol{\beta}'_0 \mathbf{z}] f_Z(\mathbf{z}). \end{aligned} \quad (2.14)$$

Llamamos $H(t, \mathbf{z})$ a la función de distribución acumulada bivarada con densidad $f_{T, Z}(t, \mathbf{z})$ dada por la ecuación (2.14).

Por otro lado, sea $\Delta = I(T^* \leq C)$ el indicador de una observación no censurada.

Considerar

$$\begin{aligned}
\text{Prob}(T \leq t; \Delta = 0 | \mathbf{Z} = \mathbf{z}) &= \text{Prob}(T^* \wedge C \leq t; T^* > C | \mathbf{Z} = \mathbf{z}) \\
&= \text{Prob}(C \leq t; T^* > C | \mathbf{Z} = \mathbf{z}) \\
&= \text{Prob}(C \leq t; T^* > C; T^* \leq t | \mathbf{Z} = \mathbf{z}) + \\
&\quad \text{Prob}(C \leq t; T^* > C; T^* > t | \mathbf{Z} = \mathbf{z}) \\
&= \text{Prob}(C < T^* \leq t | \mathbf{Z} = \mathbf{z}) + \\
&\quad \text{Prob}(C \leq t < T^* | \mathbf{Z} = \mathbf{z}) \\
&= \mathbb{E} \left[\text{Prob}(C < T^* \leq t | \mathbf{Z} = \mathbf{z}) | T^* \right] + \\
&\quad \mathbb{E} \left[\text{Prob}(C \leq t < T^* | \mathbf{Z} = \mathbf{z}) | T^* \right] \\
&= \mathbb{E} \left[F_C(T^*) I(T^* \leq t) \right] + \mathbb{E} \left[F_C(t) I(T^* > t) \right] \\
&= \int_0^t F_C(u) f_*(u, \mathbf{z}) du + F_C(t) \int_t^\infty f_*(u, \mathbf{z}) du \\
&= \int_0^t F_C(u) f_*(u, \mathbf{z}) du + F_C(t) S_*(t, \mathbf{z}),
\end{aligned}$$

y

$$\begin{aligned}
F_{[T, \Delta] | Z}(t, \Delta = 1) &= \text{Prob}(T \leq t; \Delta = 1 | \mathbf{Z} = \mathbf{z}) \\
&= \text{Prob}(T \leq t | \mathbf{Z} = \mathbf{z}) - \text{Prob}(T \leq t; \Delta = 0 | \mathbf{Z} = \mathbf{z}) \\
&= F_{T|Z}(t, \mathbf{z}) - \int_0^t F_C(u) f_*(u, \mathbf{z}) du - F_C(t) S_*(t, \mathbf{z}).
\end{aligned}$$

Esto implica que

$$\begin{aligned}
\frac{\partial}{\partial t} F_{[T, \Delta] | Z}(t, \Delta = 1) &= f_{T|Z}(t, \mathbf{z}) - F_C(t) f_*(t, \mathbf{z}) - f_C(t) S_*(t, \mathbf{z}) + F_C(t) f_*(t, \mathbf{z}) \\
&\stackrel{(2.13)}{=} \tilde{S}(t) \exp(-\boldsymbol{\beta}'_0 \mathbf{z} t) [\tilde{\lambda}(t) + \boldsymbol{\beta}'_0 \mathbf{z}] - \lambda_C(t) S_C(t) S_0(t) \exp(-\boldsymbol{\beta}'_0 \mathbf{z} t) \\
&= \tilde{S}(t) \exp(-\boldsymbol{\beta}'_0 \mathbf{z} t) [\tilde{\lambda}(t) + \boldsymbol{\beta}'_0 \mathbf{z} - \lambda_C(t)] \\
&= \tilde{S}(t) \exp(-\boldsymbol{\beta}'_0 \mathbf{z} t) [\lambda_0(t) + \boldsymbol{\beta}'_0 \mathbf{z}],
\end{aligned}$$

del cual obtenemos la *subdensidad* (con esta función nos referimos a la densidad conjunta evaluada en un valor particular de la v.a. Δ)

$$\begin{aligned}
f_{T, Z, \Delta}(t, \mathbf{z}, \Delta = 1) &= f_{[T, \Delta] | Z}(t, \Delta = 1) f_Z(\mathbf{z}) \\
&= \tilde{S}(t) (\lambda_0(t) + \boldsymbol{\beta}'_0 \mathbf{z}) \exp(-\boldsymbol{\beta}'_0 \mathbf{z} t) f_Z(\mathbf{z}). \quad (2.15)
\end{aligned}$$

Llamamos $H(t, \mathbf{z}, \Delta = 1)$ a la función de distribución acumulada conjunta bivariada con densidad $f_{T,Z,\Delta}(t, \mathbf{z}, \Delta = 1)$ dada por la ecuación (2.15).

Una función de distribución acumulada conjunta $H(t, \mathbf{z}, \Delta)$ para la terna (T, \mathbf{Z}, Δ) es un miembro de la **Familia de Riesgo Aditivo** con parámetro $\beta_0 \in \mathbb{R}^{p^+}$ y lo denotamos $H \in \mathcal{A}_{\beta_0}$, si posee:

1. una función de distribución conjunta bivariada para (t, \mathbf{z}) dada por $H(t, \mathbf{z})$, con función de densidad

$$f_{T,Z}(t, \mathbf{z}) = \tilde{S}(t) \exp(-\beta'_0 \mathbf{z}t) [\tilde{\lambda}(t) + \beta'_0 \mathbf{z}] f_Z(\mathbf{z}),$$

y

2. una función de subdistribución $H(t, \mathbf{z}, \Delta = 1)$ con función de densidad

$$f_{T,Z,\Delta}(t, \mathbf{z}, \Delta = 1) = \tilde{S}(t) \exp(-\beta'_0 \mathbf{z}t) [\lambda_0(t) + \beta'_0 \mathbf{z}] f_Z(\mathbf{z}).$$

Es de destacar que la familia aditiva \mathcal{A}_{β_0} es semiparamétrica, ya que no solo está indexada por β_0 , sino que también por las funciones de supervivencia no negativas arbitrarias $S_0(\cdot)$, $S_C(\cdot)$ y $S_Z(\cdot)$. Fuera de \mathcal{A}_{β_0} , los riesgos pueden ser no aditivos, o puede haber dependencia de T^* y C , aún condicionando sobre $\{\mathbf{Z} = \mathbf{z}\}$. Esta caracterización también muestra diferentes maneras de contaminar algún miembro $H \in \mathcal{A}_{\beta_0}$, esto se verá en la Sección 4.3.

2.2. Procesos de Conteo y Martingalas

Los conjuntos de datos que observamos o que recolectamos de un experimento sobre un período de tiempo son usualmente modelados a través de los procesos de conteo, N . Antes de dar la definición formal de este proceso y conceptos básicos, veamos la idea intuitiva de su relación con martingalas.

Al proceso de conteo se lo puede descomponer como $N = A + M$ (conocida como descomposición de Doob-Meyer), donde A es el compensador de N y M es una martingala. Luego la representación de los datos $N(t) = A(t) + M(t)$, puede ser pensada como: observación = modelo + error. El Teorema Central del Límite para Martingalas facilita la descripción asintótica del comportamiento de estadísticos específicos. En particular, los estadísticos de la forma $\sum_i \int H_i dM_i$, donde H_i es un proceso y

M_i es la martingala de un proceso de conteo, que bajo ciertas propiedades de acotación, integrabilidad y predictibilidad, resulta ser también una martingala. Como nuestro estadístico de estimación es de esa forma, a continuación se desarrollará las definiciones y resultados necesarios en vía de dar el comportamiento asintótico del estimador propuesto aquí. Estos conceptos se extrajeron de Fleming y Harrington (1991), y Martinussen y Scheicke (2006).

Para realizar la formulación rigurosa de estos conceptos, sea (Ω, \mathcal{F}, P) un espacio de probabilidad, donde Ω es un espacio de las respuestas del experimento aleatorio, \mathcal{F} es una σ -álgebra de los eventos sobre Ω y P es una medida de probabilidad sobre Ω .

Definición: Una función Z de Ω a \mathbb{R} es llamada *variable aleatoria* (v.a) o *medible* (relativa a \mathcal{F}) si $\{Z \leq x\} = \{\omega \in \Omega : Z(\omega) \leq x\} \in \mathcal{F}$, para todo x . Esto es, Z es una función medible de (Ω, \mathcal{F}, P) a \mathbb{R} equipada con una σ -álgebra Borel, \mathcal{B} . Para abreviar esto, diremos simplemente Z es \mathcal{F} -medible.

Definición: Un *proceso estocástico* es una familia de v.a. $X = \{X(t) : t \in \Gamma\}$ indexada por un conjunto Γ , definidas sobre el mismo espacio de probabilidad (Ω, \mathcal{F}, P) . El conjunto Γ casi siempre indica el tiempo, y es usual que sea $\{0, 1, 2, \dots\}$ (procesos de tiempos discretos) o \mathbb{R}^+ (procesos de tiempos continuos). Estos últimos los denotaremos como $\{X(t) : t \geq 0\}$.

Por ejemplo, $X(t)$ puede representar el número de personas que se recuperan de cierta enfermedad al instante t , o la posición de una partícula en el tiempo t , o el estado de un componente mecánico en el instante t .

Definición: Para un proceso estocástico X , las funciones aleatorias $X(\cdot, \omega) : \mathbb{R}^+ \rightarrow \mathbb{R}$, con $\omega \in \Omega$, son llamadas *trayectorias* de X .

Definición: Un proceso será llamado continuo a derecha o izquierda, de variación acotada, creciente, o con límite a derecha o izquierda si el conjunto de trayectorias con la correspondiente propiedad tiene probabilidad 1.

Definición: Un proceso X es una función de $\mathbb{R}^+ \times \Omega$ a \mathbb{R} y $X(t, \omega)$ es el valor de la v.a. $X(t)$ para la respuesta $\omega \in \Omega$. Si $X(\cdot, \cdot) : \mathbb{R}^+ \times \Omega \rightarrow \mathbb{R}$ es una función medible cuando \mathbb{R}^+ está equipada con la σ -álgebra Borel \mathcal{B} , entonces el proceso X es llamado *medible*.

Las siguientes tres definiciones son propiedades importantes de integrabilidad y de acotación para los procesos estocásticos. Un proceso estocástico X es:

1. *Integrable*: si

$$\sup_{0 \leq t < \infty} \mathbb{E}(|X(t)|) < \infty.$$

2. *Con Cuadrado Integrable (CCI)*: si

$$\sup_{0 \leq t < \infty} \mathbb{E}(X(t)^2) < \infty.$$

3. *Acotado*: si existe una constante finita Γ tal que

$$\text{Prob} \left(\sup_{0 \leq t < \infty} |X(t)| < \Gamma \right) = 1.$$

Las siguientes definiciones nos permitirá la formulación rigurosa del concepto de la información acumulada en el tiempo.

Definición: Una familia de sub σ -álgebra $\{\mathcal{F}_t : t \geq 0\}$ de una σ -álgebra \mathcal{F} es llamada *creciente* si $s \leq t$ implica $\mathcal{F}_s \subset \mathcal{F}_t$ (es decir, si para $s \leq t$, $A \in \mathcal{F}_s$ implica que $A \in \mathcal{F}_t$). Una familia creciente de sub σ -álgebra es llamada una *filtración*. Cuando $\{\mathcal{F}_t : t \geq 0\}$ es una filtración, la σ -álgebra $\bigcap_{h \geq 0} \mathcal{F}_{t+h}$ es usualmente denotado por \mathcal{F}_{t+} .

El correspondiente límite a izquierda, \mathcal{F}_{t-} , es la σ -álgebra más chica que contiene a todos los conjuntos en $\bigcup_{h > 0} \mathcal{F}_{t-h}$ y se escribe $\sigma \left\{ \bigcup_{h > 0} \mathcal{F}_{t-h} \right\}$.

Definición: Una filtración $\{\mathcal{F}_t : t \geq 0\}$ es *continua a derecha* si, para cualquier t , $\mathcal{F}_{t+} = \mathcal{F}_t$.

Definición: Una *base estocástica* es un espacio de probabilidad (Ω, \mathcal{F}, P) equipado con una filtración continua a derecha $\{\mathcal{F}_t : t \geq 0\}$, y es denotada por $(\Omega, \mathcal{F}, \{\mathcal{F}_t : t \geq 0\}, P)$.

Definición: Las filtraciones más naturales son las *historias* de los procesos estocásticos, o familias con $\mathcal{F}_t = \sigma\{X(s) : 0 \leq s \leq t\}$, la σ -álgebra más chica con respecto a la cual cada variable $X(s)$, $0 \leq s \leq t$, es medible. En este caso, \mathcal{F}_t “contiene la información” generada por el proceso X en $[0, t]$.

Definición: Un proceso estocástico $\{X(t) : t \geq 0\}$ es *adaptado* a una filtración si, $\forall t \geq 0$, $X(t)$ es \mathcal{F}_t -medible.

Claramente, cualquier proceso es adaptado a su historia.

Con todas estas definiciones ahora definimos formalmente un proceso de conteo.

Definición: Un *proceso de conteo* es un proceso estocástico $\{N(t) : t \geq 0\}$ adaptado a una filtración $\{\mathcal{F}_t : t \geq 0\}$ con $N(0) = 0$ y $N(t) < \infty$ casi seguro (c.s.), y cuyas trayectorias son, con probabilidad uno, continuas a derecha, constantes a trozo, y tienen solo discontinuidades por salto de tamaño 1.

Para un proceso de conteo, definimos $dN(t)$ tal que $\int_0^t dN(u)$ representa la suma de los saltos de N en el intervalo $[0, t]$.

Ejemplo 2.1 (datos con censura aleatoria): Sean T^* (tiempos de falla del evento en estudio), C (censura aleatoria) dos v.a. no negativas independientes entre sí, $T = T^* \wedge C$ (tiempo de falla observado) y $\Delta = I(T^* \leq C)$ (indicador de censura). El proceso $\{N(t) : t \geq 0\}$ definido como $N(t) = I(T \leq t, \Delta = 1)$ es un proceso de conteo con historias dadas por

$$\mathcal{F}_t = \sigma \left\{ I(T \leq s, \Delta = 0), I(T \leq s, \Delta = 1) : s \leq t \right\}.$$

Observemos que, haciendo unos simples cálculos se llega a

$$\mathbb{E}(N(t)) = \mathbb{E} \left[\int_0^t Y(u) \lambda_*(u) du \right] =: \mathbb{E}[A(t)],$$

donde $Y(u) := I(T \geq u)$, conocido como *proceso de riesgo* (ya que indica con uno si para el tiempo u no se ha producido ni el tiempo de falla ni el tiempo de censura, es decir, que todavía está en riesgo) y $\lambda_*(u)$ es la función de riesgo de T^* . El proceso $Y(u)\lambda_*(u)$ es llamado *proceso de intensidad del proceso de conteo* $N(t)$.

□

Los procesos $N(t)$ y $A(t)$ juegan un rol importante en el marco de martingalas para datos censurados como veremos a continuación.

Definición: Sea $X = \{X(t) : t \geq 0\}$ un proceso estocástico continuo a derecha y con límite a izquierda y $\{\mathcal{F}_t : t \geq 0\}$ una filtración, definidos sobre un espacio de probabilidad común. X es llamada una *martingala* con respecto a $\{\mathcal{F}_t : t \geq 0\}$ si:

1. X es adaptado a $\{\mathcal{F}_t : t \geq 0\}$,
2. $\mathbb{E}(|X(t)|) < \infty, \forall t < \infty$,
3. $\mathbb{E}(X(t+s)|\mathcal{F}_t) = X(t)$ c.s. $\forall s, t \geq 0$.

Para abreviar, diremos simplemente X es una \mathcal{F}_t -martingala.

Observemos que el punto 2 hace referencia a que X es integrable.

Ejemplo 2.2: Sea $X(t) = \sum_{i=1}^n Y_i(t)$, donde las v.a. Y_i son independientes y con media cero. Entonces $X(t)$ es una martingala.

□

Ejemplo 2.3 (continuación del ejemplo 2.1): El proceso $M(t) := N(t) - A(t)$ es una \mathcal{F}_t -martingala.

□

Definición: X es llamada una *submartingala* si el punto 3 es reemplazado por $\mathbb{E}(X(t+s)|\mathcal{F}_t) \geq X(t)$ c.s. $\forall s, t \geq 0$, y simplemente diremos X es una \mathcal{F}_t -submartingala.

Para una submartingala X , a menudo es posible encontrar un proceso creciente A tal que $X - A$ es una martingala. Utilizando algunas restricciones adicionales sobre X y A , especificadas en el Teorema 2.1, A es única (conocida como *compensador* de N) y la única descomposición de X como $M + A$ es llamada su *descomposición de Doob-Meyer*. Necesitamos entonces introducir las nociones de predictibilidad, para los cuales esta descomposición se verifica.

Definición: Sea (Ω, \mathcal{F}, P) un espacio de probabilidad con una filtración $\{\mathcal{F}_t : t \geq 0\}$. La σ -álgebra sobre $\mathbb{R}^+ \times \Omega$ generada por todos los conjuntos de la forma $[0] \times A$, $A \in \mathcal{F}_0$ y $(a, b] \times A$, $0 \leq a < b < \infty$, $A \in \mathcal{F}_a$, es llamada la σ -álgebra *predecible* para la filtración $\{\mathcal{F}_t : t \geq 0\}$.

Definición: Un proceso X es llamado *predecible* con respecto a la filtración si, pensado como una función de $\mathbb{R}^+ \times \Omega$ a \mathbb{R} , es medible con respecto a la σ -álgebra predecible generada por esa filtración. Llamaremos a X un proceso \mathcal{F}_t -predecible.

Definición: Una colección de v.a. $\{X_t : t \in T\}$, donde T es un conjunto arbitrario de índices, es *uniformemente integrable* si

$$\lim_{n \rightarrow \infty} \sup_{t \in T} \mathbb{E}(|X_t| I(|X_t| > n)) = 0.$$

La siguiente descomposición de Doob-Meyer no es una versión general ya que solo se establece para el caso de submartingalas, no negativas y continuas a derecha, pero permite desprenderse de ella un concepto nuevo y útil para el compensador del

cuadrado de una martingala. Una versión extendida de este teorema será dada más adelante.

Teorema 2.1. (Teorema de Descomposición de Doob-Meyer). *Sea X una submartingala no negativa y continua a derecha con respecto a una base estocástica $(\Omega, \mathcal{F}, \{\mathcal{F}_t : t \geq 0\}, P)$. Entonces existe una martingala M continua a derecha y un proceso A creciente, predecible, continuo a derecha tal que $\mathbb{E}(A(t)) < \infty$ y $X(t) = M(t) + A(t)$ c.s. para cualquier t . Si $A(0) = 0$ c.s., y si $X(t) = M'(t) + A'(t)$ es otra descomposición con $A'(0) = 0$, entonces para cualquier $t \geq 0$, $\text{Prob}[M'(t) \neq M(t)] = 0 = \text{Prob}[A'(t) \neq A(t)]$. Si además X es acotado, entonces M es uniformemente integrable y A es integrable.*

De este teorema se desprende el concepto de proceso de variación predecible: Si $M(t)$ es una martingala CCI entonces su cuadrado $M^2(t)$ es una submartingala, a su único compensador denotado por $\langle M \rangle(t)$ se lo conoce como el *proceso de variación predecible de $M(t)$* . Este nombre proviene del hecho que $\text{Var}(dM(t)|\mathcal{F}_{t-}) = d\langle M \rangle(t)$. Si además, M es de variación finita, es decir, si $\forall t > 0$, $\int_0^t |dM(s)| < \infty$, el *proceso de variación cuadrática de $M(t)$* se lo define como $[M](t) = \sum_{s \leq t} (\Delta M(s))^2$, donde $\Delta M(s) := M(s) - \lim_{u \nearrow s} M(u)$. Estos dos nuevos proceso son de utilidad en el Teorema Central del Límite para Martingalas enunciado al final de esta sección. Este teorema requiere ciertas propiedades de regularidad en los procesos estocásticos que se verifican localmente. Para ello necesitamos definir:

Definición: Sea $\{\mathcal{F}_t : t \geq 0\}$ una filtración sobre un espacio de probabilidad. Una v.a. no negativa τ es un *tiempo de parada (stopping time)* con respecto a $\{\mathcal{F}_t\}$ si $\{\tau \geq 0\} \in \mathcal{F}_t, \forall t \geq 0$.

Definición: Una secuencia creciente de tiempos aleatorios $\tau_n, n = 1, 2, \dots$, es llamada *secuencia localizada con respecto a una filtración* si:

1. cada τ_n es un tiempo de parada con respecto a una filtración, y
2. $\lim_{n \rightarrow \infty} \tau_n = \infty$ c.s.

Definición: Un proceso estocástico $M = \{M(t) : t \geq 0\}$ es una *martingala (submartingala) local con respecto a la filtración $\{\mathcal{F}_t : t \geq 0\}$* si existe una secuencia

localizada $\{\tau_n\}$ tal que, para cada n , $M_n = \{M(t \wedge \tau_n) : t \geq 0\}$ es una \mathcal{F}_t -martingala (submartingala).

Definición: Si al proceso M_n definido arriba es una \mathcal{F}_t -martingala y es un proceso CCI, es llamado una *martingala con cuadrado integrable*, y a M es llamada una *martingala con cuadrado integrable localmente*. Diremos que M es una \mathcal{F}_t -MCCIL.

Definición: Un proceso adaptado $X = \{X(t) : t \geq 0\}$ es llamado *localmente acotado* si, para una adecuada secuencia localizada $\{\tau_n\}$, $X_n = \{X(t \wedge \tau_n) : t \geq 0\}$ es un proceso acotado para cada n .

A continuación se enunciará la extensión del Teorema de Descomposición de Doob-Meyer. Esta versión implica la existencia de un compensador A para cualquier proceso de conteo N tal que $N - A$ es una martingala local, y esto implica que cualquier MCCIL tenga un único proceso de variación predecible $\langle M \rangle$ tal que $M^2 - \langle M \rangle$ es una martingala local y $\langle M \rangle(0) = 0$.

Teorema 2.2. (Teorema extendido de la Descomposición de Doob-Meyer). Sea $X = \{X(t) : t \geq 0\}$ una \mathcal{F}_t -submartingala local, no negativa y continua a derecha con secuencia localizada $\{\tau_n\}$, donde $\{\mathcal{F}_t : t \geq 0\}$ es una filtración continua a derecha. Entonces existe un único proceso A creciente, predecible, continuo a derecha tal que $A(0) = 0$ c.s., $\text{Prob}(A(t) < \infty) = 1 \forall t$, y $X - A$ es una martingala local continua a derecha. Para cada t , $A(t)$ puede tomarse como $\lim_{n \rightarrow \infty} A_n(t)$, donde A_n es el compensador de la submartingala detenida $X(\cdot \wedge \tau_n)$.

Los siguientes dos resultados garantizan la existencia de la martingala $M = N - A$ del Teorema 2.5.

Teorema 2.3. Sea N un proceso de conteo arbitrario.

1. Entonces existe un único proceso A creciente, predecible, continuo a derecha tal que $A(0) = 0$ c.s., $A(t) < \infty$ c.s. para cualquier t , y el proceso $M = N - A$ es una martingala local.
2. Si el proceso A de arriba es localmente acotado, M es una MCCIL.

Teorema 2.4. *Sea N un proceso de conteo y sea A su único compensador por el Teorema 2.2. Entonces A es un proceso localmente acotado y $\Delta A(t) \leq 1$ c.s. $\forall t \geq 0$.*

Como mencionamos anteriormente, nuestro estadístico de estimación es de la forma $\sum_i \int H_i dM_i$, donde H_i es un proceso y M_i es la martingala de un proceso de conteo. A continuación se enunciará el teorema en donde se expresa bajo que condiciones este estadístico resulta ser una MCCIL.

Teorema 2.5. *Sea $(\Omega, \mathcal{F}, \{\mathcal{F}_t : t \geq 0\}, P)$ una base estocástica con filtración continua a derecha $\{\mathcal{F}_t : t \geq 0\}$, sea H un proceso estocástico \mathcal{F}_t -predecible y localmente acotado, y sea N un proceso de conteo. Sea además $M = N - A$ una \mathcal{F}_t -MCCIL. Entonces $\int H dM$ es una MCCIL.*

Por último, daremos el Teorema Central del Límite para Martingalas que facilita la descripción asintótica del estadístico de estimación.

Teorema 2.6. *Sea $\{W^{(n)}(t) : t \in [0, \infty)\}$ una secuencia de MCCIL de \mathbb{R}^p . Supongamos que*

$$\langle W^{(n)} \rangle (t) \xrightarrow{Prob} \Sigma(t), \quad \text{cuando } n \rightarrow \infty, \forall t \in [0, \infty) \quad (2.16)$$

y

$$\langle W_{\epsilon l}^{(n)} \rangle (t) \xrightarrow{Prob} 0, \quad \text{cuando } n \rightarrow \infty, \forall t \in [0, \infty), l = 1, \dots, p \text{ y } \epsilon > 0 \quad (2.17)$$

donde $W_{\epsilon l}^{(n)}(t) = \sum_{s \leq t} \Delta W_l^{(n)}(s) I(|\Delta W_l^{(n)}(s)| > \epsilon)$. Entonces

$$W^{(n)} \xrightarrow{\mathcal{D}} W, \quad \text{cuando } n \rightarrow \infty,$$

donde W es una martingala normal con función de covarianza Σ . Más aún, $\langle W^{(n)} \rangle$ y $[W^{(n)}]$ convergen uniformemente sobre subconjuntos compactos de $[0, \infty)$, en probabilidad, a Σ .

Capítulo 3

Revisión de la literatura

Como hemos mencionado en el capítulo anterior, una generalización de los modelos para la función de intensidad, λ , del tiempo de falla, T , asociada con el vector de covariables, $\mathbf{Z} \in \mathbb{R}^p$, incluye variables regresoras. Las tres generalizaciones más comunes y a las cuales nos referiremos aquí son: el modelo de riesgo proporcional, el modelo de tiempo de falla acelerado y el modelo de riesgo aditivo, todos anteriormente definidos. En los tres casos se desea estimar el vector de parámetros de regresión, β_0 , y la función baseline, $\lambda_0(t)$. Si bien esta función puede ser considerada en forma paramétrica (por ejemplo, los modelos exponencial, Weibull y log-logística, descritos en el capítulo anterior), aquí solo veremos modelos en los que se la considera en forma no paramétrica. Es decir, en esta revisión solo veremos modelos semiparamétricos, ya que se supone un modelo paramétrico solamente para el efecto de la variable independiente \mathbf{Z} . Observemos que la función de intensidad es modificada por una proporción (en el modelo de Cox), una reescalación (en el modelo de falla acelerado) y una traslación (en el modelo aditivo).

Haremos, para cada uno de los modelos semiparamétricos anteriores, una revisión y análisis de algunos trabajos focalizándonos solamente sobre los resultados estadísticos de la estimación del parámetro de regresión, motivados por la búsqueda de estimadores menos sensibles que los métodos clásicos ante observaciones extremas y de alta palanca y/o a perturbaciones en la distribución de F , que se han estudiado hasta el momento. Si bien es más difícil estimar en modelos semiparamétricos (ya que en ellos hay al menos un parámetro de dimensión infinita, como es el caso de $\lambda_0(\cdot)$) que en modelos paramétricos, cabe destacar que los procedimientos de estimación descritos aquí estiman al parámetro de regresión β_0 sin la necesidad de

estimar conjuntamente ni previamente a la función λ_0 . Algunos procedimientos robustos basados en modelos paramétricos se pueden encontrar por ejemplo en Huber (1981), Beran (1981) y Maronna, Martin y Yohai (2006), entre otros.

En la siguiente sección desarrollamos las propuestas de estimación del parámetro para el modelo de riesgo proporcional en el que exponemos, para el caso en que el evento sea simple, el conocido estimador de verosimilitud parcial (o estimador de Cox (1972 y 1975)) y dos propuestas robustas, la de Sasieni (1993a y 1993b) y la de Bednarski (1993). También desarrollamos la extensión del estimador de Cox, que realizan Huang y Chen (2003), a eventos recurrentes. En la Sección 3.2 desarrollamos las propuestas robustas (para eventos simples) de Salibian-Barrera y Yohai (2008), y (para eventos recurrentes) de Lin, Wei y Ying (1998) y Strawderman (2005) para el modelo de falla acelerado. Por último, en la Sección 3.3 para la estimación en el modelo aditivo daremos las ideas de Lin y Ying (1994), para eventos simples y Sun, Park y Sun (2006) para eventos recurrentes.

3.1. Modelo de riesgo multiplicativo

En el modelo clásico de riesgo proporcional de Cox (1972), los datos provienen de un evento simple y la variable explicativa es fija en el tiempo. Se observan las n ternas independientes $(t_i, \mathbf{Z}_i, \Delta_i)$, $i = 1, \dots, n$, donde para cada individuo i , t_i es el tiempo de falla observado (censurado o no), $\mathbf{Z}_i \in \mathbb{R}^p$ su vector de covariable asociado y Δ_i vale 0 si t_i es un tiempo de falla censurado a derecha y vale 1 en caso contrario. Entonces, para el tiempo de supervivencia t_i del individuo i con vector de variables independientes \mathbf{Z}_i , la función de intensidad es de la forma

$$\lambda(t_i, \mathbf{Z}_i) = \lambda_0(t_i) \exp(\boldsymbol{\beta}'_0 \mathbf{Z}_i) , \quad (3.1)$$

donde $\boldsymbol{\beta}_0$ es un p -vector de coeficientes de regresión desconocido y λ_0 , la función de riesgo para $\mathbf{Z}_i = \mathbf{0}$ (conocida como *baseline*), es también desconocida y es una función arbitraria no negativa del tiempo.

Para estimar $\boldsymbol{\beta}_0$ en (3.1), Cox (1975) por la falta de conocimiento de $\lambda_0(\cdot)$, propuso un método de estimación denominado *verosimilitud parcial* (VP). La función de verosimilitud parcial no es una verosimilitud en el sentido usual sino que fue motivada condicionando sobre los tiempos de fallas observados t_i , donde la probabilidad condicional de que haya una falla en t_i dado los casos que se encuentran en riesgo

de fallar en el tiempo t_i (es decir, aquellos para los cuales $t_k \geq t_i$) es

$$\frac{\lambda(t_i, \mathbf{Z}_i)}{\sum_{k \in R_i} \lambda(t_i, \mathbf{Z}_k)},$$

donde $R_i := \{1 \leq k \leq n : t_k \geq t_i\}$. Luego la función de VP es:

$$L_P(\boldsymbol{\beta}) = \prod_{i=1}^n \left[\frac{\lambda(t_i, \mathbf{Z}_i)}{\sum_{k \in R_i} \lambda(t_i, \mathbf{Z}_k)} \right]^{\Delta_i} = \prod_{i=1}^n \left[\frac{\exp(\boldsymbol{\beta}' \mathbf{Z}_i)}{\sum_{k \in R_i} \exp(\boldsymbol{\beta}' \mathbf{Z}_k)} \right]^{\Delta_i},$$

donde se canceló $\lambda_0(t)$, el parámetro de “nuisance”. Esta función es tratada como la verosimilitud usual. Notar que el numerador depende únicamente de la información del individuo que experimentó el evento, mientras que el denominador utiliza información sobre todos los individuos quienes no han experimentado el evento (incluyendo algunos individuos quienes serían luego censurados).

Entonces, el estimador de VP de $\boldsymbol{\beta}_0$ es solución de la siguiente ecuación de score

$$\mathbf{U}_P(\boldsymbol{\beta}) := \sum_{i=1}^n \Delta_i \left\{ \mathbf{Z}_i - \frac{\sum_{k \in R_i} \mathbf{Z}_k \exp(\boldsymbol{\beta}' \mathbf{Z}_k)}{\sum_{k \in R_i} \exp(\boldsymbol{\beta}' \mathbf{Z}_k)} \right\} = \mathbf{0}. \quad (3.2)$$

Bajo ciertas condiciones de regularidad, el estimador resultante es consistente y asintóticamente normal.

Andersen y Gill (1982) extienden el modelo de Cox para eventos recurrentes y para variables explicativas externas que dependen del tiempo. Consideran el proceso de conteo multivariado de n componentes $N = (N_1, \dots, N_n)$ de la vida de n individuos, donde N_i es la cantidad de eventos observados en la vida del i -ésimo individuo, $i = 1, \dots, n$, sobre el intervalo $[0, 1]$. Es decir que los tiempos de falla serán censurados por derecha. Ellos trabajan en este intervalo de tiempo por simplicidad pero hacen una discusión en su trabajo (Sección 4) de cómo extender los resultados a $[0, \infty)$. Asumen que N tiene un proceso de intensidad aleatorio $\lambda = (\lambda_1, \dots, \lambda_n)$ tal que

$$\lambda_i(t, \mathbf{Z}_i) = Y_i(t) \lambda_0(t) \exp(\boldsymbol{\beta}'_0 \mathbf{Z}_i(t)), \quad (3.3)$$

donde $Y_i(t)$ es el proceso $\{0, 1\}$ continuo a derecha, que indica con 1 cuando el i -ésimo individuo está bajo observación en el tiempo t , y que el vector de dimensión p de procesos de variables independientes $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{ip})$ es predecible y localmente acotado.

El estimador de VP de β_0 en (3.3) se define como la solución de la ecuación

$$U_{AG}(\beta) = \sum_{i=1}^n \int_0^1 \mathbf{Z}_i(s) dN_i(s) - \int_0^1 \frac{\sum_{i=1}^n \mathbf{Z}_i(s) Y_i(s) \exp(\beta' \mathbf{Z}_i(s))}{\sum_{i=1}^n Y_i(s) \exp(\beta' \mathbf{Z}_i(s))} d\bar{N}(s) = \mathbf{0},$$

donde $\bar{N} = \sum_{j=1}^n N_j$, que, bajo ciertas condiciones de regularidad y utilizando resultados de martingalas locales, es consistente y asintóticamente normal. Observar que en el caso especial en que las covariables \mathbf{Z}_i sean fijas en el tiempo, las funciones $U_P(\beta)$ y $U_{AG}(\beta)$ coinciden.

Sin embargo, el estimador de VP, $\hat{\beta}$, pierde eficiencia con respecto al estimador de máxima verosimilitud cuando β_0 se aleja del origen (por ejemplo, para $p = 1$, ver la Figura 3.1) y tiene función de influencia no acotada (Reid y Crépeau, 1985), haciéndose sensible a observaciones extremas. Uno de los autores que buscaron

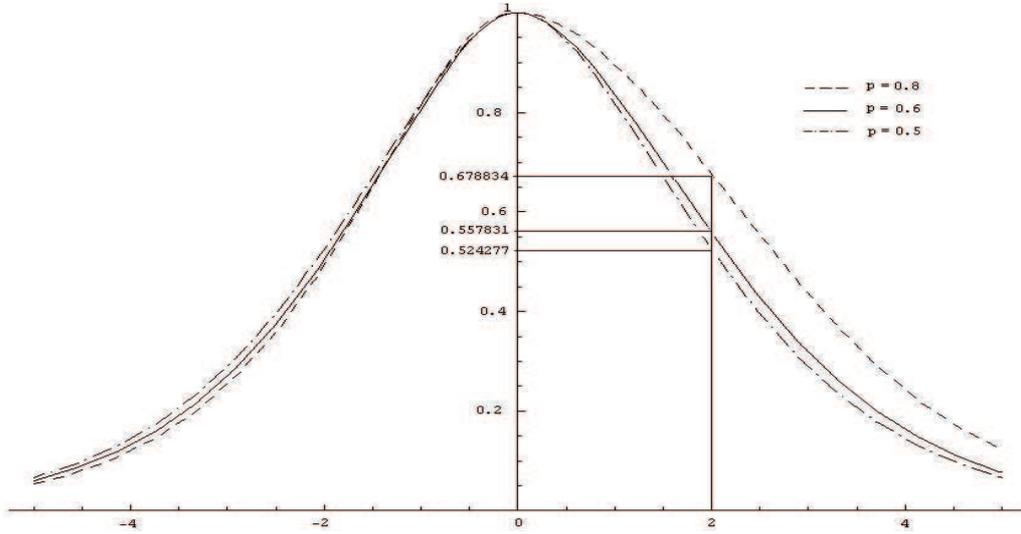


Figura 3.1: Para un problema de dos muestras ($Z = 0, 1$), con ausencia de censura y $\lambda(t, Z) = \lambda_0 \exp(\beta_0 Z)$ (λ_0 es constante, fija en el tiempo), Efron (1977, ecuación 4.11) calculó la eficiencia relativa asintótica (ARE) del estimador de VP comparada con el estimador de máxima verosimilitud. Esta es su gráfica donde p representa la proporción de 1 en la muestra.

estimadores robustos a partir del estimador de VP, fue Sasieni (1993a). Él propone una familia de estimadores que maximizan la VP “pesada” (nos referiremos a ellos como “estimadores pesados”), es decir, le asigna, a través de una función de peso, más importancia a lo que sucede en algunos tiempos de fallas que en otros. Esta

función es $w(t, \mathbb{P}_n)$, donde \mathbb{P}_n es la medida empírica del tiempo de falla con masa $1/n$ en cada punto, tal que $w(\cdot, \mathbb{P}_n)$ sea predecible, no negativa y localmente acotada (entre otras condiciones de regularidad). Formalmente, Sasieni propone el estimador pesado de β_0 como solución de

$$\mathcal{U}_S(\beta; w) := \sum_{i=1}^n w(t_i, \mathbb{P}_n) \left\{ \mathbf{Z}_i(t_i) - \frac{\sum_{j=1}^n Y_j(t_i) \mathbf{Z}_j(t_i) \exp(\beta' \mathbf{Z}_j(t_i))}{\sum_{j=1}^n Y_j(t_i) \exp(\beta' \mathbf{Z}_j(t_i))} \right\} \Delta_i = \mathbf{0}.$$

Este estimador resulta ser, bajo ciertas condiciones, consistente de tasa \sqrt{n} y asintóticamente normal.

Sasieni propuso que $w(t, \mathbb{P}_n) = \hat{S}(t)$, donde $\hat{S}(t)$ es el estimador de Kaplan-Meier (KM) (Kaplan y Meier, 1958) de la función de supervivencia marginal (es decir, ignorando las variables explicativas) y a este estimador, asumiendo que las covariables son acotadas, lo llama el estimador de Wilcoxon (por la analogía que tiene con respecto al test de Wilcoxon). De esta manera, Sasieni logra dar menos peso a observaciones grandes. La eficiencia relativa asintótica del estimador de Wilcoxon con respecto al de VP (que fue calculada por Sasieni (1993a, Corolario B, pag. 147) para el caso en que Z sea univariado, acotado y fijo en el tiempo) crece a medida que β_0 se aleja del 0 y además va aumentando a medida que se aumenta el porcentaje de censura. Esto se ve reflejado en la simulación que él realiza. Pero, por un lado, al utilizar un estimador preliminar de β_0 la función de peso $w(t, \mathbb{P}_n)$ dejará de ser predecible y, por otro lado, el estimador de KM no es robusto. En este sentido, Reid (1981) analizó cómo cambia la función de influencia del estimador de KM al agregar una observación, observando que es altamente sensible a observaciones atípicas.

Cuando el modelo de Cox se verifica, el estimador de VP para $w \equiv 1$ es eficiente (Begun, Hall, Huang y Wellner, 1983; Efron, 1977), es decir, tiene mínima varianza entre los estimadores pesados; si bien tiene función de influencia no acotada. En cualquier otro caso, cuando se consideran una función de peso $w(t, \mathbb{P}_n)$ que no depende de la muestra y que no es constantemente igual a 1, o bien, que depende de la muestra, los correspondientes estimadores pesados tendrán asintóticamente la misma eficiencia que el estimador VP pero serán menos eficientes que él en muestras finitas.

En cuanto a la función de influencia, Sasieni nota que el problema de que el estimador de VP tenga función de influencia alta (no acotada asintóticamente) es por dos razones: una, por los valores de \mathbf{Z} “extremos” y, la otra, por los individuos que

más sobrevivieron, es decir, por los valores grandes de T . La primera, sin importar la función de peso $w(t, \mathbb{P}_n)$ que se seleccione, no se soluciona ya que solo depende de los tiempos de falla y de su distribución empírica. Entonces asumiendo que \mathbf{Z} está acotado, cuando el estimador es el de Wilcoxon se obtiene un estimador con función de influencia acotada sin importar de que \mathbb{P}_n sea o no un miembro de Cox (una medida de probabilidad \mathbb{P} sobre (\mathbf{Z}, T, Δ) es un miembro del modelo de Cox con parámetro β_0 , y lo denotamos $\mathbb{P} \in \mathcal{P}(\beta_0)$, si existen T^u, T^c independientes condicionalmente dado \mathbf{Z} tal que $T = T^u \wedge T^c$, $\Delta = I(T^u \leq T^c)$ y el riesgo proporcional de T^u dado \mathbf{Z} en t es $\lambda(t, \mathbf{Z}) = \lambda_0(t) \exp(\beta_0' \mathbf{Z}(t))$).

Por lo tanto, para el método propuesto por Sasieni la selección de la función de peso $w(t, \mathbb{P}_n)$ a menudo será una compensación entre eficiencia y robustez.

Por otro lado, en otro trabajo de Sasieni (1993b) amplía la familia de los estimadores anteriormente expuesta permitiendo funciones de pesos que no solo dependan de los tiempos de falla y de su distribución empírica sino también que dependan de \mathbf{Z} . Esta familia más grande de estimadores la llamó *estimadores de clase K* y los definió de la siguiente manera: Sea K una función medible de $\mathbb{R}^p \times \mathbb{R}^+ \times \mathcal{Q}$ a \mathbb{R}^p , donde \mathcal{Q} es una extensión de $\mathcal{P}(\beta_0)$, que lo contiene y a todas las posibles distribuciones empíricas \mathbb{P}_n , entonces, bajo ciertas condiciones, el estimador de clase K , $\hat{\beta}_K$, es solución de:

$$\sum_{i=1}^n \left\{ K(\mathbf{Z}_i, T_i, \mathbb{P}_n) \mathbf{Z}_i - \frac{\sum_{k=1}^n K(\mathbf{Z}_k, T_k, \mathbb{P}_n) Y_k(T_i) \mathbf{Z}_k \exp(\beta' \mathbf{Z}_k)}{\sum_{k=1}^n Y_k(T_i) \exp(\beta' \mathbf{Z}_k)} \right\} \Delta_i = \mathbf{0}.$$

El estimador $\hat{\beta}_K$ es consistente de tasa \sqrt{n} y asintóticamente normal.

Observación 1. *Esta clase de estimadores fueron primero propuestos por Ritov y Wellner (1987) pero estos autores utilizaron funciones de peso que solo dependían de \mathbf{Z} y T , mientras que Sasieni además le incorporó la dependencia de las distribuciones empíricas de los tiempos de falla.*

Otro trabajo en el que se modificó la ecuación de estimación de score del riesgo proporcional (3.2) para obtener un estimador robusto, fue el de Bednarski (1993). Las modificaciones que realizó no solo producen estimadores consistentes y asintóticamente normales de β_0 para el modelo de riesgo proporcional sino que también para pequeños entornos del modelo. Estos son definidos como $\{G : \|G - F\|_\infty \leq \epsilon/\sqrt{n}\}$, donde F es la función de distribución acumulada “verdadera” de (T, \mathbf{Z}, C) del modelo de Cox, donde C es la variable de censura independiente de T dado \mathbf{Z} . La

herramienta importante que utiliza Bednarski es la diferenciabilidad Fréchet, con la que logra una función de influencia acotada y la norma del supremo del funcional del estimador conduce a un punto de ruptura no nulo.

La propuesta de Bednarski fue modificar la función de score $U_P(\boldsymbol{\beta})$ (ver (3.2)), introduciéndole una función de peso $A(t, \mathbf{z})$ con ciertas propiedades de regularidad, como sigue:

$$U_B(\boldsymbol{\beta}; A) := \sum_{j=1}^n A(t_j, \mathbf{Z}_{(j)}) \left[\mathbf{Z}_{(j)} - \frac{\sum_{k \in R_j} A(t_j, \mathbf{Z}_k) \mathbf{Z}_k \exp(\boldsymbol{\beta}' \mathbf{Z}_k)}{\sum_{k \in R_j} A(t_j, \mathbf{Z}_k) \exp(\boldsymbol{\beta}' \mathbf{Z}_k)} \right] \Delta_j = \mathbf{0}, \quad (3.4)$$

donde $R_j := \{1 \leq k \leq n : t_k \geq t_j\}$ y $\Delta_j := I(T_j \leq C_j)$. El efecto de la función $A(t, \mathbf{z})$ que está a la izquierda es para pesar hacia abajo las observaciones no censuradas con valores grandes de $t \exp(\boldsymbol{\beta}' \mathbf{z})$ y en las sumas del cociente, $A(t, \mathbf{z})$ es calculada para las “observaciones artificiales” ya que se combina el tiempo t_j con variables explicativas \mathbf{Z}_k , donde j y k son tales que $t_k \geq t_j$, que corresponden a tiempos distintos, así pesa hacia abajo todas las observaciones con valores relativamente grande de $\boldsymbol{\beta}' \mathbf{z}$ entre todas aquellas con $t_k \geq t_j$. Con esta “doble poda” logra dar consistencia al estimador.

Por ejemplo, Bednarski (1993), Minder y Bednarski (1996), Bednarski y Nowak (2003), Bednarski y Mocarska (2006) y Bednarski (2007) proponen las siguientes funciones $A(t, \mathbf{z})$:

$$\begin{aligned} M - \text{mín} [M, t \exp(\boldsymbol{\beta}' \mathbf{z})], \\ M - \text{mín} [M, \Lambda(t) \exp(\boldsymbol{\beta}' \mathbf{z})], \\ \exp \left[\frac{-\Lambda(t) \exp(\boldsymbol{\beta}' \mathbf{z})}{\alpha M} \right], \end{aligned} \quad (3.5)$$

donde M es una constante seleccionada apropiadamente, $\Lambda(t)$ es la función de riesgo acumulado y α es un factor de escala. Con estas funciones de peso, el estimador se calcula iterativamente y se estabiliza después de la tercera o cuarta iteración. Es decir, por ejemplo, primero se toma a $\boldsymbol{\beta} = \boldsymbol{\beta}_1$ como un estimador preliminar (que puede ser el de VP de Cox), se evalúa M como un percentil (el 80 o 90 %) de la muestra $T_1 \exp(\boldsymbol{\beta}'_1 \mathbf{Z}_1), T_2 \exp(\boldsymbol{\beta}'_1 \mathbf{Z}_2), \dots, T_n \exp(\boldsymbol{\beta}'_1 \mathbf{Z}_n)$ y luego tomando $A(t, \mathbf{z})$, como en (3.5), se calcula un nuevo estimador $\boldsymbol{\beta}_2$ utilizando (3.4). Esto se repite dos veces más y se considera a $\boldsymbol{\beta}_4$ como el estimador robusto final.

Por otro lado, cuando los datos provienen de un evento recurrente es usual modelarlos con intervalos de tiempo entre sus recurrencias. Es decir, sean $T_{i1} < \dots < T_{im_i}$

los tiempos continuos de fallas observados donde $m_i := N_i(C_i) = \sum_{j \geq 1} \Delta_{ij}$ (con $\Delta_{ij} := I(T_{ij} \leq C_i)$), es el número de fallas del individuo i en el intervalo de tiempo $[0, C_i]$, donde C_i es el tiempo de censura del individuo i (con $i = 1, \dots, n$). Luego se definen $X_{ij} := T_{i(j)} - T_{i(j-1)}$ como los tiempos de espera entre arribos o intervalos de tiempo entre la $(j-1)$ y j -ésima ocurrencia para el individuo i , y $j = 1, \dots, m_i + 1$, donde $T_{i0} := 0$ y $T_{i, m_i+1} := C_i$.

Dentro de este marco, Huang y Chen (2003) desarrollan un método que, si bien no es robusto, lo destacamos por ser una buena técnica para extender cualquier método de eventos simples a eventos recurrentes y es utilizado por varios autores. Entre ellos, en la Sección 3.3, se describirá el trabajo de Sun, Park y Sun (2006).

Huang y Chen argumentan (al igual que el trabajo de Wang y Chang (1999) en el que estiman la función de supervivencia marginal de los tiempos entre dos eventos sucesivos) que, bajos ciertas suposiciones, los intervalos de tiempo observados completos, X_{ij} ($j = 1, \dots, m_i$), es decir, aquellos intervalos de tiempo no censurados, están idénticamente distribuidos. Entonces, la intercambiabilidad de los intervalos de tiempo observados completos sugiere que un subconjunto de los datos observados puede ser tratado como datos observados de supervivencia en clases. Además, el procedimiento de regresión de Cox estándar puede ser aplicado a los datos del primer intervalo de tiempo de cada individuo, y este primer intervalo puede ser reemplazado por una selección aleatoria de la misma clase. Si bien con esto se logrará una estimación más eficiente, no obstante es una aproximación muy costosa computacionalmente. Entonces, Huang y Chen, proponen estimar β_0 utilizando la siguiente función de score:

$$U_{HCh}(\beta) := \int_0^\tau \left\{ d\hat{K}_1(s) - \frac{\hat{E}_{ij} \left[\mathbf{Z}_i \exp(\beta' \mathbf{Z}_i) \tilde{Y}_{ij}(s) \right]}{\hat{E}_{ij} \left[\exp(\beta' \mathbf{Z}_i) \tilde{Y}_{ij}(s) \right]} d\hat{K}_0(s) \right\},$$

donde τ es una constante (que por razones técnicas tiene cierta propiedad), $\tilde{Y}_{ij}(s) := I(X_{ij} \geq s)$, $\hat{K}_0(s) = \hat{E}_{ij} [\Delta_i I(X_{ij} \leq s)]$, $\hat{K}_1(s) = \hat{E}_{ij} [\mathbf{Z}_i \Delta_i I(X_{ij} \leq s)]$, $\hat{E}_{ij} := \hat{E}_i \hat{E}_j$ y \hat{E}_j representa el promedio empírico sobre $j = 1, \dots, m_i^*$, con $m_i^* := m_i \vee 0$, el máximo entre m_i y 0. Con esto logran un estimador consistente, asintóticamente normal y más eficiente que el que solo utiliza el primer intervalo.

Nosotros hemos realizado una simulación para comparar el estimador de Huang y Chen utilizando solo el primer intervalo de tiempo con el que utiliza todos los intervalos de tiempo. En la siguiente tabla mostramos los resultados de la siguiente

simulación: se generó una mezcla heterogénea de procesos de renovación individual, tal que el intervalo de tiempo tenga distribución exponencial estándar marginalmente. Las v.a. A (individuo) y B (episodio) son independientes con media cero y varianza ρ y $1 - \rho$ respectivamente, con $\rho \in [0, 1]$. En un individuo, el intervalo de tiempo es $-\ln(1 - \Phi(A + B))$, donde Φ es la función de distribución de la normal estándar. Así, el parámetro ρ indicará el nivel de heterogeneidad entre individuos, cuando $\rho = 0$ indica ausencia de heterogeneidad y cuando $\rho = 1$ determina que todos los episodios dentro de un individuo son iguales. Se han realizado 1000 réplicas, de $n = 50, 100$ individuos, con una covariable simple con distribución normal estándar, $\beta_0 = 1$, el tiempo de censura aleatorio con distribución uniforme $U(0, 1)$ y $U(0, 2)$, los valores tomados de ρ fueron 0.25, 0.50, 0.75 y $\lambda_0(t) = 1$.

Los resultados obtenidos se encuentran en la siguiente tabla. En la columna de $\hat{\beta}_1$ los tres números corresponden: al estimador utilizando solo el primer intervalo de tiempo, su sesgo (entre corchete) y su error estándar (entre paréntesis). Análogamente, en la columna $\hat{\beta}_2$ estos tres números fueron calculados para el estimador que proponen Huang y Chen (en el que se utiliza todos los intervalos de tiempo observados no censurados). Y el valor \bar{M} corresponde al promedio de intervalos que se utilizó para estimar $\hat{\beta}_2$.

n	C	ρ	$\hat{\beta}_1$	$\hat{\beta}_2$	\bar{M}
50	U(0,1)	0.25	1.67 [0.67] (1.93)	1.09 [0.09] (0.28)	2.13
		0.50	1.66 [0.66] (1.64)	1.10 [0.10] (0.30)	2.40
		0.75	1.62 [0.62] (1.77)	1.09 [0.09] (0.30)	3.21
	U(0,2)	0.25	1.82 [0.82] (2.12)	1.11 [0.11] (0.25)	3.08
		0.50	1.71 [0.71] (1.88)	1.10 [0.10] (0.25)	3.86
		0.75	1.52 [0.52] (1.56)	1.07 [0.07] (0.27)	5.56
100	U(0,1)	0.25	1.66 [0.66] (1.62)	1.08 [0.08] (0.19)	2.11
		0.50	1.59 [0.59] (1.60)	1.07 [0.07] (0.24)	2.85
		0.75	1.55 [0.55] (1.34)	1.06 [0.06] (0.21)	3.30
	U(0,2)	0.25	1.92 [0.92] (2.03)	1.09 [0.09] (0.18)	3.14
		0.50	1.77 [0.77] (2.00)	1.08 [0.08] (0.19)	4.28
		0.75	1.67 [0.67] (1.87)	1.06 [0.06] (0.18)	5.67

Como se ve claramente en la tabla, al utilizar todos los intervalos de tiempo observados no censurados se logra una mejor estimación que la que solo utiliza

el primer intervalo de tiempo observado. El sesgo varía entre 0.06 y 0.11 para el segundo estimador, mientras que el sesgo del primer estimador toma valores mucho más grandes, entre 0.52 y 0.92.

3.2. Modelo de tiempo de falla acelerado

En análisis de supervivencia los modelos de regresión de tiempo de falla acelerado son una útil alternativa al modelo de riesgo multiplicativo en algunos contextos. Ellos son ejemplos de modelos transformados del tiempo que puede ser utilizado tanto en el marco de eventos simples como en eventos recurrentes. Este modelo es un caso particular del modelo de tiempos transformados ya que el efecto de \mathbf{Z} es transformar la escala del tiempo t a $\exp(\beta'_0 \mathbf{Z})t$. Además, este modelo es log-lineal para T , ya que

$$\log T = \beta'_0 \mathbf{Z} + U, \quad (3.6)$$

donde U es la variable error. Luego $T = \exp(\beta'_0 \mathbf{Z})\tilde{T}$, donde $\tilde{T} = e^U > 0$, tiene función de riesgo $\lambda_0(\tilde{t})$, que es independiente de β_0 . Entonces la función de riesgo de T es de la forma

$$\lambda(t, \mathbf{Z}) = \lambda_0(t \exp(\beta'_0 \mathbf{Z})) \exp(\beta'_0 \mathbf{Z}). \quad (3.7)$$

La ecuación (3.6) permite aplicar procedimientos robustos de modelos de regresión para hallar los estimadores. Por ejemplo, el trabajo de Salibian-Barrera y Yohai (2008). Ellos proponen una clase de estimadores robustos de alto punto de ruptura cuando la respuesta contiene observaciones censuradas.

Consideran el modelo de regresión lineal:

$$y_i = \beta'_0 \mathbf{x}_i + u_i, \quad i = 1, \dots, n,$$

donde $\mathbf{x}_i \in \mathbb{R}^p$ es el vector de covariables, los errores u_i son independientes, idénticamente distribuidos (con distribución F , simétrica) e independientes de las covariables \mathbf{x}_i y β_0 es el vector de coeficientes desconocidos. Consideran censuras aleatorias a derecha, es decir, se observa $y_i^* = y_i \wedge c_i$ donde c_i son las v.a. de censura no observadas, independientes, idénticamente distribuidas e independientes de los errores u_i . Entonces se observa $(y_i^*, \mathbf{x}_i, \delta_i)$ donde $\delta_i = I(y_i \leq c_i)$. Ellos extienden la aproximación

de Buckley y James (1979) y Ritov (1990) para el caso de respuestas censuradas con una función de pérdida acotada. Proponen el M-estimador de regresión para observaciones censuradas definido por:

$$\hat{\beta}_n = \arg \min_{\beta \in \mathbb{R}^p} \frac{1}{n} \sum_{i=1}^n \mathbb{E}_{F_{\beta}} \left[\rho(u) | \mathbf{w}_i(\beta) \right],$$

donde la función $\rho : \mathbb{R} \rightarrow \mathbb{R}^+$ tiene ciertas propiedades de regularidad, F_{β} es la distribución de los residuos $r(\beta) = y - \beta' \mathbf{x}$, $\mathbf{w}_i(\beta) = (r_i^*(\beta), \delta_i)$, con $r_i^*(\beta) = y_i^* - \beta' \mathbf{x}_i$ los residuos censurados (con variable de censura $c_i - \beta' \mathbf{x}_i$, ya que $r_i^*(\beta) = (r_i(\beta)) \wedge (c_i - \beta' \mathbf{x}_i)$) y

$$\mathbb{E}_{F_{\beta}} \left[\rho(u) | \mathbf{w}_i(\beta) \right] = \begin{cases} \rho(r_i^*(\beta)) & \text{si } \delta_i = 1 \\ \int_{r_i^*(\beta)}^{\infty} \rho(u) dF_{\beta}(u) / [1 - F_{\beta}(r_i^*(\beta))] & \text{si } \delta_i = 0 \end{cases}$$

Como F_{β} es desconocida, se la puede estimar con el estimador de KM, $F_{n\beta}^*$ basado sobre $r_i^*(\beta)$. Luego, para garantizar la consistencia del estimador definido por

$$\hat{\beta}_n = \arg \min_{\beta \in \mathbb{R}^p} \frac{1}{n} \sum_{i=1}^n \mathbb{E}_{F_{n\beta}^*} \left[\rho(u) | \mathbf{w}_i(\beta) \right] \quad (3.8)$$

se requiere que $F_{n\beta}^*$ sea consistente para F_{β} , $\forall \beta \in \mathbb{R}^p$. Para esto dan algunas condiciones sobre las funciones de distribución de los errores y de las censuras. Pero, bajo estas condiciones, el estimador de KM resulta ser consistente si $\beta = \beta_0$ y para $\beta \neq \beta_0$ en general no lo es.

Por otra parte, el estimador $\hat{\beta}_n$ definido como solución de

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E}_{F_{n\beta}^*} \left[\psi(u) | \mathbf{w}_i(\beta) \right] \mathbf{x}_i = \mathbf{0}, \quad (3.9)$$

con $\psi(u) = \partial \rho(u) / \partial u$, es consistente en el sentido Fisher. Notar que las ecuaciones (3.8) y (3.9) no son equivalentes como sucede en el caso de regresión no censurada, ya que $F_{n\beta}^*$ depende de β , no se puede obtener (3.9) derivando (3.8).

Los M-estimadores con ψ monótona son solo robustos frente a outliers de alta palanca y la principal dificultad de utilizar un ψ redescendiente en (3.9) es que en general esta ecuación puede tener varias soluciones con diferentes propiedades

de robustez. La ecuación (3.8) no puede ser utilizada para obtener una solución consistente de (3.9). Por esta razón, ellos definen, para $\boldsymbol{\beta}, \boldsymbol{\gamma} \in \mathbb{R}^p$,

$$C_n(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}_{F^*} \left[\rho \left(\frac{u - \boldsymbol{\gamma}' \mathbf{x}_i}{s_n} \right) \middle| \mathbf{w}_i(\boldsymbol{\beta}) \right], \quad (3.10)$$

donde la función ρ cumple las mismas propiedades de regularidad que antes y s_n es un estimador robusto de escala del error. Luego definen para cada $\boldsymbol{\beta} \in \mathbb{R}^p$

$$\hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}) = \arg \min_{\boldsymbol{\gamma} \in \mathbb{R}^p} C_n(\boldsymbol{\beta}, \boldsymbol{\gamma}).$$

Entonces definen a un estimador de $\boldsymbol{\beta}_0$ por la ecuación $\hat{\boldsymbol{\gamma}}_n(\hat{\boldsymbol{\beta}}_n) = \mathbf{0}$ y, alternativa-mente, para evitar la existencia de problemas, ellos definen a $\hat{\boldsymbol{\beta}}_n$ como:

$$\hat{\boldsymbol{\beta}}_n = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \left[\hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta})' A_n \hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}) \right], \quad (3.11)$$

donde $A_n = A_n(\mathbf{x}_1, \dots, \mathbf{x}_n)$ es cualquier estimador equivariante de la matriz de covarianza de los \mathbf{x}_i . En (3.11) es necesario A_n para mantener la equivarianza afín del estimador. El estimador $\hat{\boldsymbol{\beta}}_n$ puede ser considerado una extensión de los M-estimadores de Ritov (1990) para datos censurados para el caso de funciones ρ acotadas. Además, este estimador tiene las mismas propiedades asintóticas como el estimador de Ritov.

También proponen otros estimadores alternativos robustos, como los S-estimadores propuestos por Rousseeuw y Yohai (1984). Esta propuesta consiste en reemplazar a s_n en (3.10) por el M-escala $S_n(\boldsymbol{\beta}, \boldsymbol{\gamma})$, que es definido como solución de $C_n(\boldsymbol{\beta}, \boldsymbol{\gamma}) = b$, con $b = \mathbb{E}_F(\rho(u))$. Sea

$$\hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}) = \arg \min_{\boldsymbol{\gamma} \in \mathbb{R}^p} S_n(\boldsymbol{\beta}, \boldsymbol{\gamma}).$$

Notar que $\hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta})$ es el S-estimador de regresión de los residuos $(r_i^*(\boldsymbol{\beta}), \mathbf{x}'_i)'$, $i = 1, \dots, n$. Entonces definen el S-estimador de regresión de las respuestas censuradas como el vector $\tilde{\boldsymbol{\beta}}_n$ tal que $\hat{\boldsymbol{\gamma}}_n(\tilde{\boldsymbol{\beta}}_n) = \mathbf{0}$ y, alternativamente, para evitar la existencia de problemas, también se puede definir a $\tilde{\boldsymbol{\beta}}_n$ como se hizo en (3.11). Además, un estimador de escala de los residuos robusto s_n puede ser definido como $s_n = S_n(\tilde{\boldsymbol{\beta}}_n, \hat{\boldsymbol{\gamma}}_n(\tilde{\boldsymbol{\beta}}_n))$. Como se necesita resolver un problema de optimización altamente complejo para calcular el estimador, ellos presentan en su trabajo un algoritmo computacional eficiente para calcularlo.

Si bien estos S-estimadores tienen un alto punto de ruptura, cuando los errores son normales no pueden alcanzar simultáneamente alta eficiencia y alto punto

de ruptura. Entonces para obtener estimadores con alta eficiencia y alto punto de ruptura, Salibian-Barrera y Yohai, realizan dos propuestas: los MM-estimadores y los τ -estimadores. Ambos son extensiones para respuestas censuradas de los MM-estimadores propuestos por Yohai (1987) y de los τ -estimadores propuestos por Yohai y Zamar (1988).

Sobre eventos recurrentes, en los estudios a menudo el interés se centra en modelar la distribución del tiempo de falla entre la recurrencia de un evento (o intervalo de tiempo), o en la distribución de los tiempo de cada falla. Además existen procedimientos de inferencia basados en métodos marginales y métodos basados en intensidades. Los métodos marginales habitualmente se focalizan sobre la función de tasa acumulada o función de la media y no condicionan sobre la historia del evento completo. Un ejemplo de esto lo veremos en el trabajo que realizan Lin, Wei y Ying (1998). En cambio, los métodos de intensidad especifican como la probabilidad de recurrencia posterior dependerá de la historia del evento pasado. Aquí destacaremos el trabajo de Strawderman (2005).

Lin, Wei y Ying (1998) trabajan con los tiempos de falla recurrentes, es decir, para $i = 1, \dots, n$ y $j = 1, 2, \dots$, sea T_{ij} el j -ésimo tiempo de falla del evento para el sujeto i -ésimo. Asumen que los sujetos son independientes, pero no se impone ninguna estructura de dependencia sobre los tiempos de recurrencia del mismo sujeto. Definen a $N_i^*(t)$ como el número de fallas que han ocurrido sobre el sujeto i en el tiempo t en ausencia de censura, esto es $N_i^*(t) = \sum_{k \geq 1} I(T_{ik} \leq t)$. Además suponen que la función media del proceso de conteo $N_i^*(t)$ asociado al vector de variable explicativa $\mathbf{Z}_i \in \mathbb{R}^p$, que la suponen acotada, es de la forma:

$$\mathbb{E}(N_i^*(t)|\mathbf{Z}_i) = \mu_0(\exp(\boldsymbol{\beta}'_0 \mathbf{Z}_i) t),$$

donde $\boldsymbol{\beta}_0$ es un p -vector de parámetros de regresión desconocido y μ_0 es una función continua no especificada. De acuerdo a este modelo, el número esperado de eventos en el tiempo t bajo $\mathbf{Z}_i = \mathbf{z}$ es igual al número esperado de eventos en el tiempo $t \exp(\boldsymbol{\beta}'_0 \mathbf{z})$ bajo $\mathbf{Z}_i = \mathbf{0}$. En otras palabras, el conjunto de variables explicativas \mathbf{Z}_i afecta la frecuencia de recurrencia sobre el tiempo expandiendo o contrayendo la escala del tiempo en aquellas ocurrencias de eventos por un factor multiplicativo de $\exp(\boldsymbol{\beta}'_0 \mathbf{Z}_i)$ relativo a aquel de un vector de covariable cero.

Sea C_i el tiempo de censura del sujeto i , que lo asumen independiente de T_{ik} condicionado sobre \mathbf{Z}_i . Luego el proceso de conteo $N_i(t)$ de los tiempos de fallas censurados, se pueden expresar como $N_i(t) = \sum_{k \geq 1} I(T_{ik} \leq t \wedge C_i)$.

Motivados por la función de score de la verosimilitud parcial para el modelo de proceso de Poisson de intensidad proporcional (Andersen y Gill, 1982) y las funciones de estimación de rango pesado para el modelo log-lineal (3.6) (Prentice, 1978; Tsiatis, 1990; Wei, Ying y Ling, 1990), proponen la siguiente clase de funciones de estimaciones para β_0 :

$$\mathbf{U}(\beta) := \sum_{i=1}^n \int_0^{\infty} Q(t; \beta) [\mathbf{Z}_i - \bar{\mathbf{Z}}(t; \beta)] dN_i(t \exp(-\beta' \mathbf{Z}_i)), \quad (3.12)$$

donde $Q(t; \beta)$ tiene variación acotada y converge casi seguro a una función continua y

$$\bar{\mathbf{Z}}(t; \beta) = \frac{\sum_{j=1}^n I(t \exp(-\beta' \mathbf{Z}_j) \leq C_j) \mathbf{Z}_j}{\sum_{j=1}^n I(t \exp(-\beta' \mathbf{Z}_j) \leq C_j)}.$$

Ellos se refieren a $\mathbf{U}(\beta)$ como la función de estimación log-rango si $Q \equiv 1$ y como la función de estimación de Gehan si $Q(t; \beta) = \sum_{i=1}^n I(t \exp(-\beta' \mathbf{Z}_i) \leq C_i)/n$. Para la primer función de peso, como en el caso de estimación de rango para el modelo log-lineal (3.6), la función de estimación $\mathbf{U}(\beta)$ es una función constante a trozos de β , entonces definen el estimador $\hat{\beta}$ como un cero de $\mathbf{U}(\beta)$ o como un mínimo de $\|\mathbf{U}(\beta)\|_2$. Para la función de estimación de Gehan, (3.12) se convierte en

$$\mathbf{U}(\beta) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \sum_{k \geq 1} \Delta_{ik} (\mathbf{Z}_i - \mathbf{Z}_j) I[\log \tau \log T_{ik} \geq \beta'(\mathbf{Z}_i - \mathbf{Z}_j)],$$

y así obtienen $\hat{\beta}$ minimizando la función:

$$\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \sum_{k \geq 1} \Delta_{ik} \max \{ \log \tau - \log T_{ik} - \beta'(\mathbf{Z}_i - \mathbf{Z}_j), 0 \}.$$

El estimador resultante puede ser ligeramente diferente al mínimo de $\|\mathbf{U}(\beta)\|_2$, pero son asintóticamente equivalentes.

El estimador $\hat{\beta}$, solución de $\mathbf{U}(\beta) = \mathbf{0}$, resulta ser consistente y asintóticamente normal. Pero, resolver la ecuación de estimación $\mathbf{U}(\beta) = \mathbf{0}$ puede en general ser arduo cuando p (la dimensión de \mathbf{Z}) es grande. Entonces, proponen $\hat{\beta}^*$ como solución de

$$\mathbf{U}(\beta) = \sum_{i=1}^n D_i(\hat{\beta}) G_i,$$

donde $\hat{\beta}$ es un estimador inicial (que puede ser el de la estimación de Gehan),

$$D_i(\beta) := \int_0^\infty Q(t; \beta) [\mathbf{Z}_i - \bar{\mathbf{Z}}(t; \beta)] d \left[N_i(t \exp(-\beta' \mathbf{Z}_i)) - \int_0^t I(t \exp(-\beta' \mathbf{Z}_i) \leq C_i) d\hat{\mu}_0(s; \beta) \right],$$

con

$$\hat{\mu}_0(t; \beta) = \sum_{i=1}^n \int_0^t \frac{dN_i(t \exp(-\beta' \mathbf{Z}_i))}{\sum_{j=1}^n I(t \exp(-\beta' \mathbf{Z}_j) \leq C_j)}$$

y (G_1, \dots, G_n) son v.a. normales estándar independientes. También se puede obtener $\hat{\beta}^*$ como solución de $\mathbf{U}(\beta) = G$, donde G es normal con media cero y matriz de covarianza $\sum_{i=1}^n D_i(\hat{\beta}) D_i(\hat{\beta})'$. Luego $\sqrt{n}(\hat{\beta} - \hat{\beta}^*)$ tiene la misma distribución límite que $\sqrt{n}(\hat{\beta} - \beta_0)$ y además la matriz de covarianza de $\hat{\beta}$ puede ser estimada por la matriz de covarianza empírica de $\hat{\beta}^*$.

Cuando el modelo se ajusta razonablemente a los datos, Lin, Wei y Ying proponen que el estimador de la ecuación de estimación de Gehan (solución de $\mathbf{U}(\beta) = \mathbf{0}$ o $\mathbf{U}(\beta) = G$, que se puede resolver de manera eficiente), puede ser utilizado como un estimador inicial para estimaciones con funciones de peso más generales, ya que la solución de la ecuación de estimación de Gehan será similar a las soluciones de la ecuación de estimación con pesos más generales. Además, para la mayoría de los efectos prácticos, es suficiente hacer inferencias basadas sobre la estimación de Gehan.

El trabajo de Strawderman (2005) desarrolla un nuevo modelo semiparamétrico para el efecto de las variables explicativas, fijas en el tiempo, sobre la intensidad condicional de un proceso de conteo de eventos recurrentes. Su modelo es una extensión del modelo del tiempo de falla acelerado para datos de supervivencia univariado, intervalos de tiempo entre eventos, y la estimación del parámetro de regresión está motivada por las consideraciones de eficiencia semiparamétricas.

Primero considera un sujeto con un vector de covariable \mathbf{Z} de dimensión p , fija en el tiempo, que experimenta el evento recurrente en los tiempos $0 =: T_0 < T_1 < T_2 < \dots$. Define el j -ésimo intervalo de tiempo como $X_j = T_j - T_{j-1}$, $j \geq 1$. Sean V_1, V_2, \dots v.a. independientes e idénticamente distribuidas con función de distribución F_0 , tal que $F_0(t) = \int_0^t f_0(s) ds$ donde f_0 tiene primer derivada continua y segunda derivada acotada. Asume que, dado \mathbf{Z} , sus intervalos de tiempos, X_1, X_2, \dots son v.a. independientes donde

$$X_j = V_j \exp(-\beta_0' \mathbf{Z}),$$

siendo β_0 el vector de parámetros de regresión. Notar que \mathbf{Z} acelera o desacelera los intervalos basales de tiempo entre arribos V_j como sucede con el modelo de tiempo de falla acelerado.

Luego la función de riesgo de X_j , dado \mathbf{Z} es

$$\lambda_0(x \exp(\beta'_0 \mathbf{Z})) \exp(\beta'_0 \mathbf{Z}),$$

donde λ_0 es la función de riesgo asociada a F_0 . Entonces, en ausencia de censura y dado \mathbf{Z} , el proceso $N(t) := \max\{m : \sum_{j=1}^m X_j \leq t\}$ es un proceso de renovación.

Ahora supone n sujetos independientes, donde cada uno es observado en el intervalo de tiempo finito $[0, C_i]$. Entonces los datos observados son, para $i = 1, \dots, n$, $\{N_i(u \wedge C_i), \Delta_i(u), \mathbf{Z}_i, u \geq 0\}$ donde $\Delta_i(u) = I(u \leq C_i)$.

El trabajo de Strawderman fue motivado por los trabajos de Prentice (1978) (quien sugirió estimar β_0 del modelo semiparamétrico (3.7) invirtiendo una clase de estadísticos de rango lineal pesado), de Tsiatis (1990) (quien estableció las propiedades asintóticas de la clase de estimadores propuesto por Prentice) y de Ritov (1990) (quien estableció una correspondencia directa entre las funciones estimadas por Prentice y Tsiatis y las basadas sobre consideraciones de eficiencia semiparamétrica para modelos de regresión lineal censurados). Strawderman propone la clase pesada de funciones de estimación, mediante la siguiente función de score:

$$\bar{S}_W(\beta) := \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} W(\tilde{X}_{ij}(\beta)|\beta) \left[\mathbf{Z}_i - \frac{\sum_{k=1}^n \mathbf{Z}_k \sum_{r=1}^{m_k+1} \tilde{I}_{kr}(\tilde{X}_{ij}|\beta)}{\sum_{k=1}^n \sum_{r=1}^{m_k+1} \tilde{I}_{kr}(\tilde{X}_{ij}|\beta)} \right], \quad (3.13)$$

donde $\tilde{X}_{ij}(\beta) = X_{ij} \exp(\beta' \mathbf{Z}_i)$ para $j = 1, \dots, m_i + 1$, $m_i = N_i(C_i)$, $\tilde{I}_{kr}(t|\beta) := I(\tilde{X}_{kr}(\beta) \geq t)$ y $W(t|\beta)$ bajo ciertas propiedades de regularidad.

Como (3.13) es una función de estimación basada en rango, el estimador $\hat{\beta}$ puede ser definido como un cero de $\bar{S}_W(\beta)$ o un mínimo de $\|\bar{S}_W(\beta)\|_2$. Sin embargo, puede existir varios mínimos porque $\bar{S}_W(\beta)$ no es necesariamente monótona. Pero, si $W(u|\beta) = \sum_{i=1}^n \sum_{j=1}^{m_i+1} \tilde{I}_{ij}(u|\beta)/n$ (denominado peso de Gehan), esta dificultad desaparece. Y en este caso, (3.13) se reduce a

$$\bar{S}_G(\beta) := \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^n (\mathbf{Z}_i - \mathbf{Z}_k) \sum_{r=1}^{m_k+1} \tilde{I}_{kr}(\tilde{X}_{ij}|\beta),$$

que es el gradiente de la función objetivo convexa

$$L_G(\beta) := \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^n \sum_{r=1}^{m_k+1} \left\{ [\log \tilde{X}_{kr}(\beta) - \log \tilde{X}_{ij}(\beta)] \vee 0 \right\},$$

Los minimizadores de $L_G(\boldsymbol{\beta})$ y $\|\bar{S}_G(\boldsymbol{\beta})\|$ son asintóticamente equivalentes (Fygen-son y Ritov, 1994). Luego Strawderman toma $\hat{\boldsymbol{\beta}}_G = \arg \min L_G(\boldsymbol{\beta})$ y resulta ser consistente de tasa \sqrt{n} y asintóticamente normal.

Para este último estimador Strawderman comenta que “notablemente no se asume que el intervalo de tiempo sea acotado, es una útil consecuencia de la convexidad asociada con la función de peso de Gehan”. Además, da un algoritmo para calcular $\hat{\boldsymbol{\beta}}_G$ y $\hat{\Gamma}_G$, un estimador consistente de la covarianza de $\sqrt{n}(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}_0)$.

Para pesos generales, Strawderman propone un estimador $\hat{\boldsymbol{\beta}}_W$ de un paso a partir del $\hat{\boldsymbol{\beta}}_G$, que también resulta ser consistente de tasa \sqrt{n} y asintóticamente normal.

3.3. Modelo aditivo

Por último, veremos los resultados estadísticos del modelo de riesgo aditivo para la función de intensidad que es de la forma:

$$\lambda(t, \mathbf{Z}) = \lambda_0(t) + \boldsymbol{\beta}'_0 \mathbf{Z}(t) \quad (3.14)$$

donde $\boldsymbol{\beta}_0$ es el vector de parámetros de regresión y $\lambda_0(t) = \lambda(t, \mathbf{0})$ es la función baseline, desconocida, arbitraria y no negativa en función del tiempo. Este modelo propuesto por Lin y Ying (1994) es una alternativa al modelo de Aalen (1980) en el cual el vector de parámetros de regresión no está fijo en el tiempo, es decir, la función de intensidad es de la forma

$$\lambda_0(t) + \boldsymbol{\beta}_0(t)' \mathbf{Z}(t) = (\lambda_0(t), \boldsymbol{\beta}_0(t)') \begin{pmatrix} 1 \\ \mathbf{Z}(t) \end{pmatrix}.$$

Primero plantearemos aquí algunas de las distintas funciones de verosimilitud para el modelo aditivo solo para el caso de eventos simples.

La función de verosimilitud condicional sobre la historia del evento del sujeto i , $H_i(\tau)$, de n individuos con función de intensidad (3.14), donde $t_1 \leq \dots \leq t_n$ son los tiempos de falla observados en $[0, \tau]$ (τ tiempo fijo finito) y \mathbf{Z}_i corresponde al vector de covariable asociado a t_i , es

$$L_C(\boldsymbol{\beta}_0, \lambda_0) = \prod_{j=1}^n [\lambda_0(t_j) + \boldsymbol{\beta}'_0 \mathbf{Z}_j(t_j)] \exp \{ -[\Lambda_0(\tau) + \boldsymbol{\beta}'_0 \mathbf{Z}_j^*(\tau)] \},$$

donde $\mathbf{Z}_j^*(t) = \int_0^t \mathbf{Z}_j(u) du$ y $\Lambda_0(t) = \int_0^t \lambda_0(u) du$. Luego,

$$\frac{\partial}{\partial \boldsymbol{\beta}_0} \log L_C(\boldsymbol{\beta}_0, \lambda_0) = \sum_{j=1}^n \left\{ \frac{\mathbf{Z}_j(t_j)}{\lambda_0(t_j) + \boldsymbol{\beta}_0' \mathbf{Z}_j(t_j)} - \mathbf{Z}_j^*(\tau) \right\}. \quad (3.15)$$

Observemos que (3.15) depende del vector de parámetros de regresión $\boldsymbol{\beta}_0$ y de la función $\lambda_0(\cdot)$, lo cual complicaría la estimación de $\boldsymbol{\beta}_0$ ya que sería necesario estimar previamente o conjuntamente la función $\lambda_0(\cdot)$.

Por otra parte, la función de verosimilitud parcial es

$$L_P(\boldsymbol{\beta}_0, \lambda_0) = \prod_{j=1}^n \frac{\lambda_0(t_j) + \boldsymbol{\beta}_0' \mathbf{Z}_j(t_j)}{\sum_{k \in R_j} \lambda_0(t_j) + \boldsymbol{\beta}_0' \mathbf{Z}_k(t_j)}, \quad (3.16)$$

donde R_j es el conjunto de riesgo en el tiempo t_j . Entonces

$$\frac{\partial}{\partial \boldsymbol{\beta}_0} \log L_P(\boldsymbol{\beta}_0, \lambda_0) = \sum_{j=1}^n \left\{ \frac{\mathbf{Z}_j(t_j)}{\lambda_0(t_j) + \boldsymbol{\beta}_0' \mathbf{Z}_j(t_j)} - \frac{SZ_j}{r_j \lambda_0(t_j) + \boldsymbol{\beta}_0' SZ_j} \right\}, \quad (3.17)$$

donde r_j es el cardinal del conjunto R_j y $SZ_j := \sum_{k \in R_j} \mathbf{Z}_k(t_j)$.

Notar que aquí también, la función de verosimilitud parcial (3.16) no puede aplicarse como se hizo para el caso del modelo de riesgo multiplicativo (3.1), pues en este caso tampoco se eliminaría $\lambda_0(t)$ para la estimación de $\boldsymbol{\beta}_0$, (ver la función de score (3.17)).

Sin embargo, hay varios autores que han podido estimar $\boldsymbol{\beta}_0$ sin la necesidad de recurrir a las clásicas funciones de verosimilitud condicional y/o parcial antes descritas. Entre estos autores hemos considerado el trabajo de Lin y Ying (1994) en el cual imitaron la característica de martingala de la función de score de la verosimilitud parcial del modelo multiplicativo del parámetro $\boldsymbol{\beta}_0$ logrando construir una simple función de estimación permitiendo expresar en forma explícita al estimador $\hat{\boldsymbol{\beta}}_0$ (esto no sucede en los modelos multiplicativo y de falla acelerado).

Formalmente, Lin y Ying consideran n sujetos independientes y recogen en el proceso de conteo del sujeto i , $\{N_i(t); t \geq 0\}$ el número de eventos observados hasta el tiempo t . Bajo el modelo (3.14), la función de intensidad para $N_i(t)$ esta dada por

$$Y_i(t) \lambda(t, \mathbf{Z}_i) = Y_i(t) [\lambda_0(t) + \boldsymbol{\beta}_0' \mathbf{Z}_i(t)],$$

donde $Y_i(t)$ indica con 1 si el sujeto i está en riesgo en el tiempo t y en caso contrario con 0. Luego

$$Y_i(t) d\Lambda(t, \mathbf{Z}_i) = Y_i(t) [d\Lambda_0(t) + \boldsymbol{\beta}_0' \mathbf{Z}_i(t) dt], \quad (3.18)$$

donde $\Lambda_0(t) = \int_0^t \lambda_0(u) du$. Ellos proponen estimar β_0 imitando la función de score de la verosimilitud parcial del modelo multiplicativo, que bajo (3.18), es de la forma

$$\mathbf{U}(\beta) = \sum_{i=1}^n \int_0^\infty \mathbf{Z}_i(t) \left\{ dN_i(t) - Y_i(t) [d\hat{\Lambda}_0(\beta, t) + \beta' \mathbf{Z}_i(t) dt] \right\},$$

donde $\hat{\Lambda}_0$ es el estimador de Λ_0 del modelo (3.14) definido como

$$\hat{\Lambda}_0(\hat{\beta}, t) = \int_0^t \frac{\sum_{i=1}^n [dN_i(u) - Y_i(u) \hat{\beta}' \mathbf{Z}_i(u) du]}{\sum_{i=1}^n Y_i(u)},$$

siendo $\hat{\beta}$ un estimador consistente de β_0 . Luego, $\mathbf{U}(\beta)$ es equivalente a

$$\mathbf{U}_{LY}(\beta) := \sum_{i=1}^n \int_0^\infty [\mathbf{Z}_i(t) - \bar{\mathbf{Z}}(t)] [dN_i(t) - Y_i(t) \beta' \mathbf{Z}_i(t) dt], \quad (3.19)$$

donde $\bar{\mathbf{Z}}(t) = \sum_{j=1}^n Y_j(t) \mathbf{Z}_j(t) / \sum_{j=1}^n Y_j(t)$. Entonces el estimador queda definido explícitamente como

$$\hat{\beta} = \left(\sum_{i=1}^n \int_0^\infty Y_i(t) [\mathbf{Z}_i(t) - \bar{\mathbf{Z}}(t)]^{\otimes 2} dt \right)^{-1} \left(\sum_{i=1}^n \int_0^\infty [\mathbf{Z}_i(t) - \bar{\mathbf{Z}}(t)] dN_i(t) \right),$$

donde $\mathbf{a}^{\otimes 2} := \mathbf{a} \mathbf{a}'$ (matriz), siendo \mathbf{a} un vector. Este estimador resulta ser consistente de tasa \sqrt{n} y asintóticamente normal, bajo ciertas condiciones de regularidad.

Nosotros hemos realizado la siguiente simulación para ver el desempeño del estimador $\hat{\beta}$ frente a outliers. Para $n = 100, 200, 500, 1000$ individuos con $Z \sim U(0, 1)$, censuras aleatorias $C \sim U(0, 1)$, $\beta = 2$ y $\lambda_0(t) = 1$ se han contaminado a Z con masa puntual en $\delta = 0.01, 3$, y en un 5% y 10% de la muestra. En la siguiente tabla se muestra el resultado en 1000 réplicas de: el estimador, su sesgo (entre corchetes) y su error estándar (entre paréntesis).

%	n	Sin contaminar	$\delta = 0.01$	$\delta = 3$
5	100	2.00 [-0.00] (0.97)	1.78 [-0.22] (0.96)	0.17 [-1.83] (1.85)
	200	2.00 [0.00] (0.66)	1.78 [-0.22] (0.69)	0.13 [-1.87] (1.88)
	500	2.02 [0.02] (0.42)	1.79 [-0.21] (0.45)	0.11 [-1.90] (1.90)
	1000	2.00 [0.00] (0.30)	1.78 [-0.22] (0.35)	0.11 [-1.90] (1.90)
10	100	1.99 [-0.01] (0.95)	1.69 [-0.31] (0.97)	0.07 [-1.93] (1.94)
	200	1.99 [-0.01] (0.66)	1.62 [-0.38] (0.75)	0.06 [-1.94] (1.94)
	500	2.00 [0.00] (0.42)	1.59 [-0.41] (0.57)	0.05 [-1.95] (1.95)
	1000	2.00 [-0.00] (0.30)	1.60 [-0.40] (0.49)	0.05 [-1.95] (1.95)

Lo que hemos notado es que $\hat{\beta}$ está altamente influenciado, al contaminar los valores de Z (sobre todo para valores grandes de Z), lo que nos motivó en esta investigación a proponer estimadores alternativos robustos. Estas nuevas propuestas, junto con un estudio de simulación en donde se compara el comportamiento de $\hat{\beta}$ frente a outliers, se describirá en los siguientes capítulos.

Por último en el caso de eventos recurrentes destacamos el trabajo de Sun, Park y Sun (2006) en el que ajustan un modelo de riesgo aditivo $\lambda(t, \mathbf{Z}_i) = \lambda_0(t) + \beta'_0 \mathbf{Z}_i$ utilizando los intervalos de tiempo entre las ocurrencias de los tiempo de falla y toman a las covariables fijas en el tiempo. Ellos extienden la idea de Lin y Ying (1994) a eventos recurrentes aplicando los mismos argumentos y suposiciones que realizan Huang y Chen (2003) además de introducirle una función de peso a la función de score de estimación.

Siguiendo los mismos argumentos y las mismas notaciones que utilizamos al describir el trabajo de Huang y Chen (2003), Sun, Park y Sun proponen la ecuación de estimación $\mathbf{U}_{SPS}(\beta) = \mathbf{0}$ para estimar el vector de parámetros de regresión β_0 , donde

$$\mathbf{U}_{SPS}(\beta) := \int_0^\tau Q(s) \left\{ d\hat{K}_1(s) - \frac{\hat{G}_1(s)}{\hat{G}_0(s)} d\hat{K}_0(s) - \left[\hat{E}_{ij} [\mathbf{Z}'_i \mathbf{Z}_i I(X_{ij} \leq s)] - \frac{\hat{G}_1(s)' \hat{G}_1(s)}{\hat{G}_0(s)} \right] \beta ds \right\}$$

siendo $Q(s)$ un proceso de peso con ciertas propiedades de regularidad, $\tau \in (0, \infty)$ es una constante pre-especificada (en la práctica, τ es usualmente tomada como el tiempo de seguimiento más largo), $\hat{G}_0(t) = \hat{E}_{ij} [I(X_{ij} \geq t)]$ y $\hat{G}_1(t) = \hat{E}_{ij} [\mathbf{Z}_i I(X_{ij} \geq t)]$.

Luego el estimador de β_0 también se puede expresar de forma explícita como:

$$\hat{\beta} = \left\{ \int_0^\tau Q(s) \left[\hat{E}_{ij} [\mathbf{Z}'_i \mathbf{Z}_i I(X_{ij} \leq s)] - \frac{\hat{G}_1(s)' \hat{G}_1(s)}{\hat{G}_0(s)} \right] ds \right\}^{-1} \times \left\{ \int_0^\tau Q(s) \left[d\hat{K}_1(s) - \frac{\hat{G}_1(s)}{\hat{G}_0(s)} d\hat{K}_0(s) \right] \right\},$$

que es consistente de tasa \sqrt{n} y asintóticamente normal.

Ellos, al igual que Huang y Chen, también proponen un estimador que solo utiliza el primer intervalo de tiempo de cada individuo pero es menos eficiente que aquel que utiliza todos los intervalos (al igual que pasa en Huang y Chen). Además comentan

que “un problema que necesita ser estudiado a futuro es la selección de un proceso de peso $Q(t)$ que da el estimador más eficiente de β_0 para una situación particular”. Para el estudio de simulación que realizaron, obtuvieron resultados similares para

$$Q \equiv 1 \text{ y } Q(t) = \sum_{i=1}^n \frac{I(C_i \geq t)}{n}.$$

Capítulo 4

Propuestas robustas para el Modelo con Hazard Aditivo

El modelo con hazard o riesgo aditivo es una alternativa al modelo de riesgo multiplicativo común para el análisis de datos de supervivencia. En este capítulo se desarrollarán por completo las dos propuestas robustas que proponemos en esta investigación para la estimación del vector de regresión del modelo de riesgo aditivo. Estas dos propuestas consisten en introducir una función de peso a la función de score, como lo hicieron Sasieni (1993a, 1993b), en el que explota la estructura de martingala, y Bednarski (1993), en el que utiliza diferenciabilidad de Fréchet como herramienta principal, ambos para el modelo de riesgo proporcional. Estas dos alternativas son en las que nos basamos para adaptarlas al modelo con Hazard Aditivo.

A continuación se describirán y se darán los principales resultados, en el Apéndice se encuentran la demostraciones de los mismos y en el Capítulo 6 se harán las discusiones y conclusiones de estas dos propuestas robustas de estimadores. Además en el próximo capítulo se encuentra el estudio de simulación en donde se verá el desempeño de los estimadores propuestos frente al estimador clásico de Lin y Ying (1994).

4.1. Introducción

Recordemos que la v.a. continua T^* es no negativa y representa la duración hasta la ocurrencia de un evento para un individuo, que puede ser un tiempo no observado si está sujeto a censura. Asociado a este tiempo tenemos el vector de covariables $\mathbf{Z} \in \mathbb{R}^{p+}$ que es fijo en el estudio y por lo tanto, fijo en el tiempo. Dado el valor de la covariable $\{\mathbf{Z} = \mathbf{z}\}$, la duración T^* tiene función de riesgo aditivo

$$\lambda_*(t, \mathbf{z}) = \lambda_0(t) + \boldsymbol{\beta}'\mathbf{z} \quad (4.1)$$

donde λ_0 es una función de riesgo baseline no negativa y $\boldsymbol{\beta} \in \mathbb{R}^{p+}$ es un vector de parámetros reales no negativos.

Los datos se recogen para $i = 1, \dots, n$ individuos durante un intervalo de tiempo fijo $[0, \tau]$, sobre el cual algunas personas pueden experimentar la censura, ya sea porque $T_i^* > \tau$ ó porque $T_i^* > C_i$, donde C_i es una variable latente no observada aleatoria que indica un tiempo de censura. Este mecanismo supone una muestra $\mathcal{S}_n := \{(T_i, \mathbf{Z}_i, \Delta_i) : i = 1, \dots, n\}$ donde T_i es el tiempo del evento o el tiempo censurado, es decir, $T_i := T_i^* \wedge C_i$, y Δ_i es el indicador de que la observación no está censurada.

Como vimos en el capítulo anterior, Lin y Ying (1994), propusieron la siguiente ecuación para estimar $\boldsymbol{\beta}$ en (4.1):

$$\sum_{i=1}^n \int_0^\infty [\mathbf{z}_i - \bar{\mathbf{z}}_n(t)][dN_i(t) - Y_i(t)\boldsymbol{\beta}'\mathbf{z}_i] = \mathbf{0}, \quad (4.2)$$

donde $N_i(t) = I(T_i^* \leq t)$, $Y_i(t) = I(T_i \geq t)$ y

$$\bar{\mathbf{z}}_n(t) = \frac{\sum_{j=1}^n Y_j(t)\mathbf{z}_j}{\sum_{j=1}^n Y_j(t)} = \frac{\sum_{j=1}^n I(t_j \geq t)\mathbf{z}_j}{\sum_{j=1}^n I(t_j \geq t)}.$$

Con esto queda definido explícitamente el estimador del parámetro de regresión como:

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^n \int_0^\infty Y_i(t)[\mathbf{z}_i - \bar{\mathbf{z}}_n(t)]^{\otimes 2} dt \right)^{-1} \left(\sum_{i=1}^n \int_0^\infty [\mathbf{z}_i - \bar{\mathbf{z}}_n(t)] dN_i(t) \right), \quad (4.3)$$

recordemos que $\mathbf{a}^{\otimes 2} := \mathbf{a}\mathbf{a}'$ es una matriz.

En el estudio de simulación que se realizó en el capítulo anterior, se vio que este estimador es muy sensible a observaciones atípicas y esto fue lo que nos motivó a proponer las siguientes dos familias de estimadores robustos.

4.2. Estimadores con función de influencia acotada (EFIA)

Nuestra primer propuesta consiste en utilizar la teoría de martingalas continuas como se ha realizado en el trabajo de Lin y Ying (1994). Para esto primero nos basamos en expresiones alternativas de (4.2) y (4.3), basadas en la distribuciones empíricas $H_n(t, \mathbf{z}, \Delta)$ y $H_n(t, \mathbf{z})$ de (T, \mathbf{Z}, Δ) y (T, \mathbf{Z}) , respectivamente, de la siguiente manera:

$$\iint [\mathbf{z} - \bar{\mathbf{z}}_n(t)] dH_n(t, \mathbf{z}, \Delta = 1) - \iint \int_0^t [\mathbf{z} - \bar{\mathbf{z}}_n(u)] du \mathbf{z}' dH_n(t, \mathbf{z}) \boldsymbol{\beta} = \mathbf{0},$$

donde

$$\bar{\mathbf{z}}_n(u) = \left[\iint I(\tilde{t} \geq u) dH_n(\tilde{t}, \mathbf{x}) \right]^{-1} \left[\iint I(\tilde{t} \geq u) \mathbf{x} dH_n(\tilde{t}, \mathbf{x}) \right].$$

Luego el estimador del vector de parámetros queda definido explícitamente como:

$$\hat{\boldsymbol{\beta}}_n = \left[\iint \left(\int_0^t [\mathbf{z} - \bar{\mathbf{z}}_n(u)]^{\otimes 2} du \right) dH_n(t, \mathbf{z}) \right]^{-1} \times \left[\iint [\mathbf{z} - \bar{\mathbf{z}}_n(t)] dH_n(t, \mathbf{z}, \Delta = 1) \right]. \quad (4.4)$$

Como ya vimos, Lin y Ying prueban que $\hat{\boldsymbol{\beta}}_n$ en (4.4) es consistente y asintóticamente normal, y dan una fórmula para la estimación de la varianza asintótica. Notar que el estimador $\hat{\boldsymbol{\beta}}_n$ es un funcional no lineal de la función de distribución conjunta empírica.

Ahora nos centramos en la descripción de la primer propuesta robusta para el parámetro de regresión del modelo con Hazard Aditivo. Para datos de supervivencia censurados de tamaño n con funciones de riesgo acumulados $\Lambda_i(t, \mathbf{z}_i)$, con los procesos de conteo asociados $N_i(t)$ y con los procesos asociados “en riesgo” $Y_i(t)$, las funciones de intensidad son $Y_i(t)\Lambda_i(t, \mathbf{z})$ y los procesos pueden ser compensados para formar martingalas restando

$$M_i(t) := N_i(t) - \int_0^t Y_i(u) d\Lambda_i(u, \mathbf{z}).$$

Recordemos que, por el Teorema 2.5, las martingalas $M_i(t)$ son MCCIL (martingalas con cuadrado integrable localmente). En el caso del modelo de Hazard Aditivo

$$d\Lambda_i(t, \mathbf{z}_i) = \lambda_i(t, \mathbf{z}_i) dt = [\lambda_0(t) + \boldsymbol{\beta}' \mathbf{z}_i] dt,$$

y esto implica que

$$dM_i(t) = dN_i(t) - Y_i(t) [\lambda_0(t) + \boldsymbol{\beta}' \mathbf{z}_i] dt.$$

Por los resultados clásicos (por ejemplo, Andersen y otros 1993, cap. VI), la función de score de *verosimilitud parcial* está dada por

$$\sum_{i=1}^n \int_0^\infty \left(\frac{\partial}{\partial \boldsymbol{\beta}} \log(\lambda_i(t, \mathbf{z}_i; \boldsymbol{\beta})) \right) \{dN_i(t) - Y_i(t) d\Lambda_i(t, \mathbf{z}_i; \boldsymbol{\beta})\},$$

y bajo el modelo de Hazard Aditivo es de la forma:

$$\mathbf{U}_{\text{PL}}(\boldsymbol{\beta}) := \sum_{i=1}^n \int_0^\infty \left(\frac{\mathbf{z}_i}{\lambda_0(t) + \boldsymbol{\beta}' \mathbf{z}_i} \right) \{dN_i(t) - Y_i(t) [\lambda_0(t) dt + \boldsymbol{\beta}' \mathbf{z}_i dt]\}. \quad (4.5)$$

En contraste con el modelo de riesgo proporcional de Cox, la función de score parcial en la ecuación (4.5) no elimina la función estorbo (nuisance) de riesgo baseline, $\lambda_0(t)$. Generalizando esta expresión, manteniendo en la ecuación de estimación la medida de martingala intacta pero reemplazando el integrando por otro término más general, tenemos

$$\mathbf{U}_{\text{K}}(\boldsymbol{\beta}) = \sum_{i=1}^n \int_0^\infty K_i(t, \mathbf{z}_i, \boldsymbol{\beta}) \{dN_i(t) - Y_i(t) [\lambda_0(t) dt + \boldsymbol{\beta}' \mathbf{z}_i dt]\}.$$

La verosimilitud parcial determina $K_i(t, \mathbf{z}_i, \boldsymbol{\beta}) = [\lambda_0(t) + \boldsymbol{\beta}' \mathbf{z}_i]^{-1} \mathbf{z}_i$ y la propuesta de Lin y Ying consiste en tomar $K_i(t, \mathbf{z}_i, \boldsymbol{\beta}) = \mathbf{z}_i$, el mismo que da la verosimilitud parcial bajo el modelo de riesgo multiplicativo (o modelo de Cox). Como ya se ha mencionado en el capítulo anterior, el estimador resultante de Lin y Ying es consistente y asintóticamente normal, como así también tiene una expresión explícita, una característica que no posee el estimador en el modelo de riesgo multiplicativo. Basados sobre estas mismas ideas nosotros proponemos

$$K_i(t, \mathbf{z}_i, \boldsymbol{\beta}) = w_1(t) \mathbf{w}_2(\mathbf{z}_i) \mathbf{z}_i, \quad (4.6)$$

donde w_1 y \mathbf{w}_2 son las funciones de peso, con las siguientes condiciones:

A1 La función $w_1(\cdot) : \mathbb{R}^+ \mapsto [0, 1]$ es acotada, no creciente (es decir, para cualquier $u > v$, $w_1(u) \leq w_1(v)$), predecible y $\int w_1(t) dt < \infty$.

A2 La función $\mathbf{w}_2(\cdot) : \mathbb{R}^{p^+} \mapsto \mathbb{M}_p([0, 1])$, donde $\mathbb{M}_p([0, 1])$ es el conjunto de matrices simétricas definidas con elementos en $[0, 1]$, es definida positiva y es uniformemente continua para casi todo \mathbf{z} . Además, $\int \mathbf{w}_2(\mathbf{z}) \mathbf{z}^{\otimes l} f_Z(\mathbf{z}) d\mathbf{z} < \infty$, para $l = 0, 1$, donde $\mathbf{a}^{\otimes 1} := \mathbf{a}$ (vector) y $\mathbf{a}^{\otimes 0} := 1$ (constante), y , para $u \geq 0$, $\det(\int \exp(-\boldsymbol{\beta}' \mathbf{z} u) \mathbf{w}_2(\mathbf{z}) f_Z(\mathbf{z}) d\mathbf{z}) \neq 0$.

Estas condiciones permiten el control de observaciones atípicas en la muestra. Algunas de las selecciones especiales posibles de $w_1(\cdot)$ son las siguientes:

1. $w_1(u) = I(u \leq a)$ donde a es una constante positiva conocida. Esta selección es equivalente a censurar artificialmente las observaciones donde los u son más grandes que alguna constante dada a .
2. Si ρ es cualquier función- ρ como las utilizadas en estadística robusta (por ejemplo, Maronna, Martin y Yohai 2006), $w_1(u) = \rho''(u)/u$ con $u \geq 0$. Por ejemplo, si tomamos la función de Huber

$$w_1(t) = \min \left\{ 1, \frac{a}{t} \right\}, \quad t \geq 0,$$

y a es una constante que controla el equilibrio entre eficiencia y robustez.

Supuestos adicionales sobre w_1 se añadirán más adelante. En particular, será de especial interés el caso cuando las funciones w dependerán de los datos. En concreto, se propondrá

$$w_{1,n}(t) = \mathbb{F}_n(t)^a [1 - \mathbb{F}_n(t)]^b, \quad (4.7)$$

donde \mathbb{F}_n es la distribución empírica de los tiempos observados (censurados o no) y las constantes predeterminadas a y b son las que regulan la robustez contra pequeños o grandes valores de t respectivamente.

Con respecto a la otra función de peso, $\mathbf{w}_2(\mathbf{z})$, es una matriz de $p \times p$ con entradas escalares $w_2^{(l,m)}$ dadas por

$$\mathbf{w}_2(\mathbf{z}_i) = \begin{pmatrix} w_2^{(1)}(z_{i1}) & w_2^{(1,2)}(z_{i1}, z_{i2}) & \dots & w_2^{(1,p)}(z_{i1}, z_{ip}) \\ w_2^{(2,1)}(z_{i2}, z_{i1}) & w_2^{(2)}(z_{i2}) & \dots & w_2^{(2,p)}(z_{i2}, z_{ip}) \\ \vdots & \vdots & \ddots & \vdots \\ w_2^{(p,1)}(z_{ip}, z_{i1}) & w_2^{(p,2)}(z_{ip}, z_{i2}) & \dots & w_2^{(p)}(z_{ip}) \end{pmatrix}.$$

Por simplicidad, tomaremos $w_2^{(l,m)} = 0$ para $l \neq m$, y para la diagonal principal

$$w_2^{(m)}(z_{im}) = \min \left\{ 1, \frac{a}{z_{im}} \tilde{s} \right\} \quad \text{para } z_{im} > 0, \quad (4.8)$$

o uno en caso contrario, y donde

$$\tilde{s} = \frac{\text{median}\{z_{1m}, \dots, z_{nm}\}}{\log(2)}$$

es un estimador de escala que es igual a la unidad en la distribución exponencial estandar, y a es una constante de corte que regula la robustez. Más general, los elementos que no están en las diagonales de la matriz de peso \mathbf{w}_2 son para posibles multicolinealidad entre las variables independientes. No perseguimos ese interés aquí. Es de notar, sin embargo, que en la selección de la función de peso dada en las ecuaciones (4.7) y (4.8) los pesos depende de los datos, es decir,

$$w_1(t) = w_1(t, \mathbb{P}_n) \quad \text{y} \quad \mathbf{w}_2(\mathbf{z}) = \mathbf{w}_2(\mathbf{z}, \mathbb{Q}_n),$$

donde \mathbb{P}_n y \mathbb{Q}_n son, respectivamente, las distribuciones empíricas del tiempo t y de la covariables \mathbf{z} en la muestra.

Nuestra propuesta en (4.6) produce la ecuación de estimación

$$\mathbf{U}_R^{(n)}(\boldsymbol{\beta}) = \sum_{i=1}^n \int_0^\infty w_1(t) \mathbf{w}_2(\mathbf{z}_i) [\mathbf{z}_i - \tilde{\mathbf{z}}_n(t)] [dN_i(t) - Y_i(t) \boldsymbol{\beta}' \mathbf{z}_i dt] = \mathbf{0}, \quad (4.9)$$

donde $\tilde{\mathbf{z}}_n(t)$ es una función vectorial en columna dada por

$$\begin{aligned} \tilde{\mathbf{z}}_n(t) &= \left[\sum_{j=1}^n Y_j(t) \mathbf{w}_2(\mathbf{z}_j) \right]^{-1} \left[\sum_{j=1}^n Y_j(t) \mathbf{w}_2(\mathbf{z}_j) \mathbf{z}_j \right] \\ &= \left[\sum_{j=1}^n I(t_j \geq t) \mathbf{w}_2(\mathbf{z}_j) \right]^{-1} \left[\sum_{j=1}^n I(t_j \geq t) \mathbf{w}_2(\mathbf{z}_j) \mathbf{z}_j \right]. \end{aligned}$$

Claramente, cuando $w_1 = 1$ y \mathbf{w}_2 es la identidad, la ecuación de estimación anterior coincide con la de Lin y Ying.

A partir de la estimación de la ecuación (4.9) podemos obtener un estimador de $\boldsymbol{\beta}$ en forma explícita dado por

$$\begin{aligned} \tilde{\boldsymbol{\beta}}_n &= \left[\sum_{i=1}^n \int_0^\infty Y_i(t) w_1(t) \mathbf{w}_2(\mathbf{z}_i) [\mathbf{z}_i - \tilde{\mathbf{z}}_n(t)]^{\otimes 2} dt \right]^{-1} \times \\ &\quad \left[\sum_{i=1}^n \int_0^\infty w_1(t) \mathbf{w}_2(\mathbf{z}_i) [\mathbf{z}_i - \tilde{\mathbf{z}}_n(t)] dN_i(t) \right], \quad (4.10) \end{aligned}$$

que sin la notación del proceso de conteo se puede expresar en una fórmula más conveniente para el cálculo como

$$\begin{aligned} \tilde{\beta}_n = & \left[\frac{1}{n} \sum_{i=1}^n \mathbf{w}_2(\mathbf{z}_i) \left(\int_0^{t_i} w_1(u) [\mathbf{z}_i - \tilde{\mathbf{z}}_n(u)]^{\otimes 2} du \right) \right]^{-1} \times \\ & \left[\frac{1}{n} \sum_{i=1}^n \Delta_i w_1(t_i) \mathbf{w}_2(\mathbf{z}_i) [\mathbf{z}_i - \tilde{\mathbf{z}}_n(t_i)] \right]. \end{aligned} \quad (4.11)$$

Además, si $\hat{\beta}$ es un estimador consistente de β obtenemos un estimador de la función de riesgo acumulado baseline $\Lambda_0(t)$, dada por

$$\hat{\Lambda}_0(t) = \int_0^t \frac{\sum_{j=1}^n \mathbf{w}_2(\mathbf{z}_j) \left[dN_j(u) - Y_j(u) \hat{\beta}' \mathbf{z}_j du \right]}{\sum_{j=1}^n \mathbf{w}_2(\mathbf{z}_j) Y_j(u)}.$$

Con el fin de establecer la consistencia en el sentido Fisher a continuación, primero expresamos el estimador de $\tilde{\beta}_n$ a partir de la ecuación (4.11) en notación funcional como

$$\tilde{\beta}_n = \mathcal{Q}_w(H_n) = [\mathcal{D}_w(H_n)]^{-1} [\mathcal{N}_w(H_n)],$$

donde

$$\begin{aligned} \mathcal{D}_w(H_n) &= \iint \mathbf{w}_2(\mathbf{z}) \left(\int_0^t w_1(u) [\mathbf{z} - \tilde{\mathbf{z}}_n(u)]^{\otimes 2} du \right) dH_n(t, \mathbf{z}), \\ \mathcal{N}_w(H_n) &= \iint w_1(t) \mathbf{w}_2(\mathbf{z}) [\mathbf{z} - \tilde{\mathbf{z}}_n(t)] dH_n(t, \mathbf{z}, \Delta = 1), \end{aligned}$$

y

$$\tilde{\mathbf{z}}_n(u) = \left[\iint I(\tilde{t} \geq u) \mathbf{w}_2(\mathbf{x}) dH_n(\tilde{t}, \mathbf{x}) \right]^{-1} \left[\iint I(\tilde{t} \geq u) \mathbf{w}_2(\mathbf{x}) \mathbf{x} dH_n(\tilde{t}, \mathbf{x}) \right].$$

Notar que el estimador $\tilde{\beta}_n$ es un funcional no lineal de la función de distribución conjunta empírica.

El siguiente teorema, en el que se establece la *consistencia en el sentido Fisher*, será el punto de partida para el cálculo de la función de influencia que se encuentra más adelante.

Teorema 4.1. *Bajo las condiciones A1 y A2, el estimador $\tilde{\beta}_n$ propuesto en la ecuación (4.10) es consistente en el sentido Fisher en la Familia de Riesgo Aditiva. Es decir, sea $H \in \mathcal{A}_\beta$ para algún $\beta \in \mathbb{R}^{p^+}$, entonces $\mathcal{Q}_w(H) = \beta$.*

A continuación establecemos la normalidad asintótica. En primer lugar, daremos un argumento heurístico y la demostración rigurosa se encuentra en el Apéndice. Notar que la ecuación (4.11) es un producto de una matriz $p \times p$ por un vector columna p -dimensional. Ese vector es un promedio de términos $V_i := w_1(t_i)\mathbf{w}_2(\mathbf{z}_i)[\mathbf{z}_i - \tilde{\mathbf{z}}_n(t_i)]$ que son “casi” independientes e idénticamente distribuidos (iid). Lo que hace la dependencia de los V_i es la presencia del factor $\tilde{\mathbf{z}}_n(t)$, que depende de la totalidad de la muestra. Si sustituimos el $\tilde{\mathbf{z}}_n(t)$ por su límite probabilístico, $\tilde{\mathbf{z}}_\infty(t)$, entonces los sumandos modificados serían iid y la normalidad asintótica seguiría por el Teorema Central del Límite multivariado estándar. Bajo una regularidad adecuada, es de esperar que exista la convergencia en probabilidad de $\tilde{\mathbf{z}}_n(t)$ a $\tilde{\mathbf{z}}_\infty(t)$, entonces se espera heurísticamente que nuestro estimador sea asintóticamente normal. Para esto, necesitamos por lo tanto las siguientes condiciones adicionales sobre las funciones de peso:

A3 Existen una función matricial $\mathcal{S}^{(0)} : \mathbb{R}^+ \mapsto \mathbb{M}_p([0, 1])$ y una función vectorial $\mathcal{S}^{(1)} : \mathbb{R}^+ \mapsto \mathbb{R}^{p+}$ tales que para $l = 0, 1$

$$\sup_{t \geq 0} \|\mathcal{S}_n^{(l)}(t) - \mathcal{S}^{(l)}(t)\|_\infty \xrightarrow{\text{Prob}} 0, \quad \text{cuando } n \rightarrow \infty,$$

donde

$$\mathcal{S}_n^{(l)}(t) := \frac{1}{n} \sum_{i=1}^n \mathbf{w}_2(\mathbf{z}_i) \mathbf{z}_i^{\otimes l} Y_i(t).$$

Además, las funciones $\mathcal{S}^{(l)}$ ($l = 0, 1$) son acotadas y uniformemente continuas y $\mathcal{S}^{(0)}$ es acotada fuera del cero.

A4 Sea $X_i(t) := w_1(t)\mathbf{w}_2(\mathbf{z}_i)\mathbf{z}_i$. Existe $\delta > 0$ tal que, cuando $n \rightarrow \infty$

$$\frac{1}{\sqrt{n}} \sup_{\substack{t \geq 0 \\ 1 \leq i \leq n}} \|X_i(t)\|_2 Y_i(t) I\left(\lambda_0(t) + \boldsymbol{\beta}'_0 \mathbf{z}_i > \exp(-\delta \|X_i(t)\|_\infty - t)\right) \xrightarrow{\text{Prob}} 0.$$

A5 Para $p = 1$, existe una constante $\vartheta > 0$ tal que $w_2(z)$ es del orden de $z^{-(2+\vartheta)}$ cuando $z \rightarrow \infty$.

La condición **A3** asegura la convergencia en probabilidad de $\tilde{\mathbf{z}}_n(t)$. La condición **A4**, junto con las anteriores, asegura la normalidad asintótica del estimador y la convergencia de su matriz de covarianza. La condición **A5** nos garantiza que la selección de función de peso w_2 conduce a una función de influencia acotada.

Teorema 4.2. Sea $\tilde{\beta}_n$ el estimador propuesto en la ecuación (4.10) con funciones de peso que satisfacen las condiciones **A1** a **A4**, y sea $H \in \mathcal{A}_{\beta_0}$. Entonces $\sqrt{n}(\tilde{\beta}_n - \beta_0)$ converge débilmente al vector aleatorio normal p -dimensional con vector de media cero y con matriz de covarianzas

$$\Sigma := \left[\int_0^\infty w_1(u) L(u) du \right]^{-1} \tilde{\Sigma} \left[\int_0^\infty w_1(u) L(u) du \right]^{-1},$$

donde

$$L(u) := \mathbb{E} \left(\mathbf{w}_2(\mathbf{Z}) \left[\mathbf{Z} - (\mathcal{S}^{(0)}(u))^{-1} \mathcal{S}^{(1)}(u) \right]^{\otimes 2} Y(u) \right)$$

$$\tilde{\Sigma} := \int_0^\infty w_1(u)^2 \left(L_0(u) \lambda_0(u) + L_1(u, \beta_0) \right) du$$

$$L_0(u) := \mathbb{E} \left(\left[\mathbf{w}_2(\mathbf{Z}) \left[\mathbf{Z} - (\mathcal{S}^{(0)}(u))^{-1} \mathcal{S}^{(1)}(u) \right] \right]^{\otimes 2} Y(u) \right)$$

$$L_1(u, \beta_0) := \mathbb{E} \left(\left[\mathbf{w}_2(\mathbf{Z}) \left[\mathbf{Z} - (\mathcal{S}^{(0)}(u))^{-1} \mathcal{S}^{(1)}(u) \right] \right]^{\otimes 2} Y(u) \beta_0' \mathbf{Z} \right)$$

Corolario 1. En las mismas condiciones del teorema anterior, la matriz de covarianzas Σ es estimada consistentemente por $\hat{\Sigma} = D^{-1} B D^{-1}$, donde

$$B = \frac{1}{n} \sum_{i=1}^n \int_0^\infty [w_1(t) \mathbf{w}_2(\mathbf{z}_i) [\mathbf{z}_i - \tilde{\mathbf{z}}_n(t)]]^{\otimes 2} dN_i(t)$$

y

$$D = \frac{1}{n} \sum_{i=1}^n \int_0^\infty Y_i(t) w_1(t) \mathbf{w}_2(\mathbf{z}_i) [\mathbf{z}_i - \tilde{\mathbf{z}}_n(t)]^{\otimes 2} dt.$$

4.2.1. Función de Influencia

Ahora desarrollaremos, por simplicidad notacional, la función de influencia para $p = 1$. Primero observemos que en este caso, unidimensional, nuestro estimador

puede ser expresado como

$$\tilde{\beta}_n = \frac{\sum_{i=1}^n \Delta_i w_1(t_i) w_2(z_i) [z_i - \tilde{z}_n(t_i)]}{\sum_{i=1}^n w_2(z_i) \int_0^{t_i} w_1(u) [z_i - \tilde{z}_n(u)]^2 du} \quad (4.12)$$

$$= \frac{\iint w_1(t) w_2(z) [z - \tilde{z}_n(t)] dH_n(t, z, \Delta = 1)}{\iint \left\{ w_2(z) \int_0^t w_1(u) [z - \tilde{z}_n(u)]^2 du \right\} dH_n(t, z)}, \quad (4.13)$$

donde $H_n(t, z)$ en el denominador es la distribución conjunta empírica de los t_i y z_i , $H_n(t, z, \Delta = 1)$ en el numerador es la función de distribución empírica de solo el t_i no censurado y z_i , y

$$\tilde{z}_n(t) = \frac{\sum_{j=1}^n I(t_j \geq t) w_2(z_j) z_j}{\sum_{j=1}^n I(t_j \geq t) w_2(z_j)} = \frac{\iint I(\tilde{t} \geq t) w_2(x) \tilde{z} dH_n(\tilde{t}, x)}{\iint I(\tilde{t} \geq t) w_2(x) dH_n(\tilde{t}, x)}.$$

Para escribir $\tilde{\beta}_n$ como un funcional de la distribución empírica expresamos, como antes,

$$\tilde{\beta}_n = \mathcal{Q}_w(H_n) = \frac{\mathcal{N}_w(H_n)}{\mathcal{D}_w(H_n)},$$

donde $\mathcal{N}_w(H_n)$ y $\mathcal{D}_w(H_n)$ son el numerador y denominador de la ecuación (4.13) respectivamente. Ya que establecimos la consistencia en el sentido Fisher de $\tilde{\beta}_n$, para la Función de Influencia, busquemos

$$\mathcal{IF}_w(t_0, z_0, \Delta_0, \beta) := \left. \frac{d}{d\varepsilon} \mathcal{Q}_w(H_\varepsilon) \right|_{\varepsilon=0},$$

donde H_ε representa la contaminación del miembro $H \in \mathcal{A}_\beta$, con el parámetro “verdadero” $\beta \in \mathbb{R}^+$ con masa puntual en algún (t_0, z_0) que podría ser censurado, cuando $\Delta_0 = 0$, o completamente observado, en caso contrario. Esta distribución contaminada es

$$\begin{aligned} H_\varepsilon(t, z, \Delta) &:= (1 - \varepsilon) H(t, z, \Delta) + \varepsilon \delta_{(t_0, z_0, \Delta_0)}(t, z, \Delta) \\ &= H(t, z, \Delta) + \varepsilon [\delta_{(t_0, z_0, \Delta_0)}(t, z, \Delta) - H(t, z, \Delta)]. \end{aligned} \quad (4.14)$$

Teorema 4.3. *Bajo las suposiciones A1, A2 y A5, la función de influencia del estimador $\tilde{\beta}_n$ definido en la ecuación (4.12) es*

$$\mathcal{IF}_w(t_0, z_0, \Delta_0, \beta) = \left[\int w_1(t) \tilde{S}(t) \left(\Upsilon_2(t) - \frac{\Upsilon_1(t)^2}{\Upsilon_0(t)} \right) dt \right]^{-1} \nu(z_0) \left[\Delta_0 w_1(t_0) \times \right. \\ \left. \left[z_0 - \frac{\Upsilon_1(t_0)}{\Upsilon_0(t_0)} \right] - \int_0^{t_0} w_1(t) \lambda_0(t) \left(z_0 - \frac{\Upsilon_1(t)}{\Upsilon_0(t)} \right) dt + \beta \left(z_0 \gamma_1(t_0) - z_0^2 \gamma_0(t_0) \right) \right],$$

donde $\Upsilon_l(u) := \mathbb{E} [w_2(Z) \exp(-\beta u Z) Z^l] / \mathbb{E} [w_2(Z)]$, $\nu(z) = w_2(z) / \mathbb{E} [w_2(Z)]$ y $\gamma_l(t) = \int_0^t w_1(u) (\Upsilon_1(u) / \Upsilon_0(u))^l du$.

Además para cualquier $\beta \in \mathbb{R}^+$, la función $\mathcal{IF}_w(t_0, z_0, \Delta_0, \beta)$ es acotada en (t_0, z_0) .

Observación 2. *Cuando $w_1 = w_2 = 1$ obtenemos la función de influencia para el estimador clásico de Lin y Ying, que resulta no acotado.*

Para ilustrar esto veamos el siguiente ejemplo.

Ejemplo 4.1: Consideremos el caso *más suave*, es decir, tomar $\lambda_0(t) = 1$, que representa el riesgo baseline exponencial, asumir que existe una censura tal que $S_C(t) = 1$ y tomar $Z \sim \mathcal{E}(1)$ que denota la variable exponencial estándar. Además, suponer que el valor real de β es 1.

Observar que para el estimador de Lin y Ying, es decir, para el caso en que $w_1(\cdot) = w_2(\cdot) = 1$, la función de influencia es

$$\mathcal{IF}_1(t_0, z_0, \Delta_0, 1) = \frac{\Delta_0 \left(z_0 - \frac{1}{t_0+1} \right) - (z_0^2 + z_0)t_0 + (z_0 + 1) \ln(t_0 + 1)}{\int_0^\infty e^{-t} (t+1)^{-3} dt}. \quad (4.15)$$

En la Figura 4.1 se realizan las gráficas de $\mathcal{IF}_1(t_0, z_0, \Delta_0, 1)$ para $\Delta_0 = 0, 1$, respectivamente. Se puede ver claramente que no están acotadas. Esto hace que el estimador clásico de Ling y Ying sea influenciado por valores atípicos, como se mostró en la simulación de la Sección 3.3 y se mostrará nuevamente en el próximo capítulo.

□

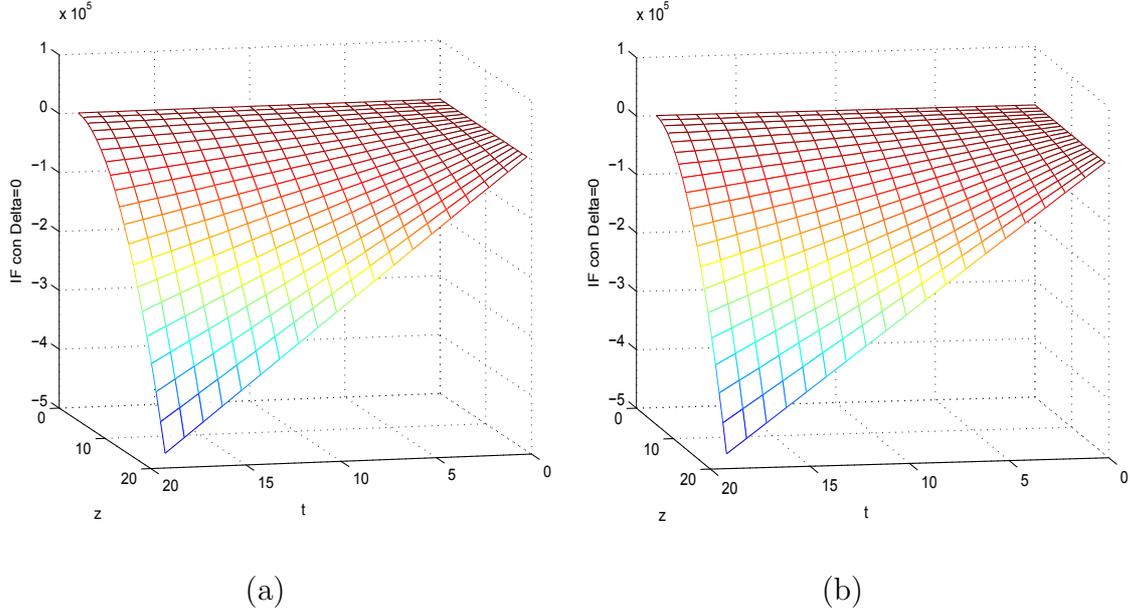


Figura 4.1: Las gráficas corresponden a la función de influencia (4.15), en (a) $\Delta_0 = 0$ y en (b) $\Delta_0 = 1$.

4.3. Estimadores con punto de ruptura no nulo (EPRNN)

Aquí proponemos una nueva familia de estimadores robustos que son consistentes, asintóticamente normales y con punto de ruptura no nulo, no solo bajo el modelo aditivo sino en un entorno pequeño. Una manera formal para describir entornos pequeños del modelo es posible a través de una aproximación de la distribución. Estos entornos pueden ser construidos de varias maneras, aquí lo definimos como: sea $H \in \mathcal{A}_{\beta_0}$, un miembro de la Familia de Riesgo Aditivo con parámetro β_0 (esto fue definido en el Capítulo 2) y sea G una distribución contaminada, se dice que G está en un ε -entorno de H si

$$\|G - H\|_{\infty} := \sup_{(t, \mathbf{z}, \Delta)} |G(t, \mathbf{z}, \Delta) - H(t, \mathbf{z}, \Delta)| \leq \varepsilon,$$

y lo denotaremos como $G \in \mathcal{B}(H, \varepsilon)$.

Para esta nueva familia de estimadores proponemos la siguiente función de score:

$$L(H_n, \boldsymbol{\beta}, W) = \left[\iint W(t, \mathbf{z}) [\mathbf{z} - \mathbf{z}_{H_n}(t)] dH_n(t, \mathbf{z}, \Delta = 1) \right] - \left[\iint W(t, \mathbf{z}) \left(\int_0^t [\mathbf{z} - \mathbf{z}_{H_n}(u)] du \right) \mathbf{z}' dH_n(t, \mathbf{z}) \right] \boldsymbol{\beta}, \quad (4.16)$$

donde la función $W : \mathbb{R}^+ \times \mathbb{R}^{p^+} \mapsto \mathbb{R}^+$ es una función de peso perteneciente a una familia \mathcal{W} , con características que enunciaremos más adelante, y

$$\mathbf{z}_{H_n}(t) = \frac{\iint I(u \geq t) W(t, \mathbf{x}) \mathbf{x} dH_n(u, \mathbf{x})}{\iint I(u \geq t) W(t, \mathbf{x}) dH_n(u, \mathbf{x})}.$$

Observar que esta nueva familia de estimadores coincide con la anterior propuesta cuando $W(u, \mathbf{z}) = w_1(u) \mathbf{w}_2(\mathbf{z})$, pero allí, w_1 y \mathbf{w}_2 eran procesos estocásticos determinísticos o predecibles. Eso llevó a un estimador en forma exacta, pero, más importante aún, nos permitió hacer todas las demostraciones que utilizan la teoría de martingalas. Además para $W(u, \mathbf{z}) = 1$ se obtiene el estimador clásico de Lin y Ying.

Denotaremos con $\boldsymbol{\beta}(G)$ a la solución de $L(G, \boldsymbol{\beta}(G), W) = \mathbf{0}$.

En busca de hallar una aproximación lineal de $\boldsymbol{\beta}(G) - \boldsymbol{\beta}_0$ es necesario definir alguna diferenciabilidad. Nosotros aquí utilizaremos la diferenciabilidad Fréchet (o diferenciabilidad fuerte). Esta diferenciabilidad sobre \mathcal{W} implica más que continuidad, que existe una aproximación lineal funcional. Es decir, para $H \in \mathcal{A}_{\boldsymbol{\beta}_0}$

$$\lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \sup_{\substack{G \in \mathcal{B}(H, \varepsilon) \\ W \in \mathcal{W}}} \|\boldsymbol{\beta}(G) - \boldsymbol{\beta}_0\|_2 = 0,$$

donde FD es un funcional lineal designado como el término “derivada de Fréchet”. La diferenciabilidad Fréchet es uniforme en todo \mathcal{W} . Esto es importante para posibles opciones de adaptación de la función de peso W . Este aspecto será discutido más adelante.

Para evitar la notación excesiva, haremos las demostraciones siguientes para el caso de sin censura.

Sea $\boldsymbol{\beta}_0 \in B$, B un subconjunto abierto y acotado de \mathbb{R}^{p^+} y sea $H \in \mathcal{A}_{\boldsymbol{\beta}_0}$. Sean

$$\mathcal{W}^* = \{W^* : \mathbb{R}^+ \times \mathbb{R}^{p^+} \mapsto \mathbb{R}^+\}$$

y $W_0 : \mathbb{R}^+ \mapsto \mathbb{R}^+$ una función continua con soporte acotado, $K := [a, b]$, $0 < a < b < \infty$. Definimos, ahora al conjunto \mathcal{W} como

$$\mathcal{W} = \{W_0 W^* : W^* \in \mathcal{W}^*\}$$

Para alcanzar la diferenciabilidad Fréchet del estimador es necesario asumir condiciones fuertes. Por lo tanto, se asumen las siguientes condiciones:

B1 Para todo $W^* \in \mathcal{W}^*$ y $u \in K$, la integral $\iint W^*(u, \mathbf{z}) I(t \geq u) dH(t, \mathbf{z}) > \varepsilon$, para algún $\varepsilon > 0$.

B2 Todas las funciones de \mathcal{W}^* se anulan fuera de un conjunto acotado, son absolutamente continuas y tienen variación conjunta acotada. El conjunto \mathcal{W}^* es compacto para la norma supremo.

Las condiciones **B1** y **B2** aseguran la diferenciabilidad de las funciones de \mathcal{W}^* y restringimos las consideraciones para funciones con soporte acotado. La compacidad asumida en **B2** tiene como objetivo permitir la adaptación en la selección de W en función de β .

La parte de “conjuntamente acotada” de los supuestos es necesaria para la afirmación de que la diferenciabilidad Fréchet es uniforme en \mathcal{W} .

El objetivo de los siguientes resultados es hallar una expresión lineal de $\beta(G) - \beta_0$ en términos de $G - H$ donde $H \in \mathcal{A}_{\beta_0}$ y $G \in \mathcal{B}(H, \epsilon)$. Para esto, necesitamos considerar:

$$\begin{aligned} L(G, \beta, W) - L(H, \beta_0, W) &= [L(G, \beta, W) - L(H, \beta, W)] + \\ &\quad [L(H, \beta, W) - L(H, \beta_0, W)] \end{aligned} \quad (4.17)$$

Para la primer diferencia, el siguiente lema proporciona una aproximación.

Lema 1. *Bajo las suposiciones B1 y B2,*

$$\sup_{\substack{\boldsymbol{\beta} \in B \\ W \in \mathcal{W}}} \|L(G, \boldsymbol{\beta}, W) - L(H, \boldsymbol{\beta}, W) - \text{Lin}_{W, \boldsymbol{\beta}}(G - H)\|_2 = o(\|G - H\|_\infty^2),$$

donde

$$\begin{aligned} \text{Lin}_{W, \boldsymbol{\beta}}(G - H) = & \iint W(t, \mathbf{z}) [\mathbf{z} - \mathbf{z}_H(t)] d(G - H)(t, \mathbf{z}) - \\ & \left\{ \iint W(t, \mathbf{z}) \left(\int_0^t [\mathbf{z} - \mathbf{z}_H(u)] du \right) \mathbf{z}' d(G - H)(t, \mathbf{z}) \right\} \boldsymbol{\beta} + \\ & \iint W(t, \mathbf{z}) [\mathbf{z}_H(t) - \mathbf{z}_G(t)] dH(t, \mathbf{z}) - \\ & \left\{ \iint W(t, \mathbf{z}) \left(\int_0^t [\mathbf{z}_H(u) - \mathbf{z}_G(u)] du \right) \mathbf{z}' dH(t, \mathbf{z}) \right\} \boldsymbol{\beta}. \end{aligned}$$

Más aún, $\text{Lin}_{W, \boldsymbol{\beta}_0}(H) = \mathbf{0}$.

Para la segunda diferencia de (4.17) tenemos:

Lema 2. *Para cualquier $W \in \mathcal{W}$ y $\boldsymbol{\beta} \in B$*

$$L(H, \boldsymbol{\beta}, W) - L(H, \boldsymbol{\beta}_0, W) = \mathcal{L}(H, W)(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$$

donde $\mathcal{L}(H, W)$ es la matriz de derivadas

$$\mathcal{L}(H, W) := \frac{\partial L(H, \boldsymbol{\beta}, W)}{\partial \boldsymbol{\beta}} = - \left[\iint W(t, \mathbf{z}) \left(\int_0^t [\mathbf{z} - \mathbf{z}_H(u)] du \right) \mathbf{z}' dH(t, \mathbf{z}) \right].$$

El siguiente resultado permite una acotación de $\text{Lin}_{W, \boldsymbol{\beta}}(G - H)$ en términos de $G - H$:

Lema 3. *Bajo las condiciones B1 y B2 existe constantes $c > 0$ y $\varepsilon > 0$, tales que $\forall W \in \mathcal{W}$ y $\boldsymbol{\beta} \in B$, si $G \in \mathcal{B}(H, \varepsilon)$ entonces*

$$\|\text{Lin}_{W, \boldsymbol{\beta}}(G - H)\|_2 \leq c \|G - H\|_\infty.$$

Para el próximo resultado necesitamos incorporar la siguiente condición:

B3 Supongamos que para algunos números $a_1, a_0 > 0$, tenemos $\forall W \in \mathcal{W}$

$$a_0 \leq |\det\{\mathcal{L}(H, W)\}| \leq a_1.$$

Esto nos garantiza la existencia de la inversa de $\mathcal{L}(H, W)$.

La consistencia del estimador en un entorno de $H \in \mathcal{A}_{\beta_0}$ esta avalada por el siguiente resultado:

Teorema 4.4. *Si la clase \mathcal{W} satisface las condiciones **B1** a **B3**. Entonces existen $\varepsilon > 0$ y $M > 0$ tal que si $G \in \mathcal{B}(H, \varepsilon)$ entonces la ecuación $L(G, \beta, W) = \mathbf{0}$ tiene una solución en la bola $\|\beta - \beta_0\|_2 < M\|G - H\|_\infty, \forall W \in \mathcal{W}$.*

La diferenciabilidad Fréchet del estimador esta asegurada por el siguiente resultado

Teorema 4.5. *Denotamos por $\beta(G, W)$ una solución de la ecuación $L(G, \beta, W) = \mathbf{0}$. Si la clase \mathcal{W} satisface las condiciones **B1** a **B3**. Entonces*

$$\lim_{c \rightarrow 0} \sup_{\substack{W \in \mathcal{W} \\ G \in \mathcal{B}(H, c)}} \frac{1}{c} \left\| [\beta(G, W) - \beta_0] + \mathcal{L}^{-1}(H, W) \text{Lin}_{W, \beta_0}(G - H) \right\|_2 = 0.$$

Esto implica que la derivada de Fréchet del funcional $\beta(\cdot, W)$ evaluada en H en la dirección de $G - H$ es:

$$\text{FD}(G - H) := -\mathcal{L}^{-1}(H, W) \text{Lin}_{W, \beta_0}(G - H)$$

En el siguiente Teorema investigamos la convergencia en distribución bajo alternativas cercanas a la “verdadera” distribución en la Familia de Riesgo Aditivo, $H \in \mathcal{A}_{\beta_0}$.

Teorema 4.6. *Sea H_n la distribución empírica de una muestra de tamaño n de una distribución $G \in \mathcal{B}(H, c_0/\sqrt{n})$ para algún $c_0 > 0$. La clase \mathcal{W} satisface las condiciones **B1** a **B3**. Entonces $\forall \varepsilon > 0, \delta > 0$ existe n_0 tal que $\forall n \geq n_0$, $G \in \mathcal{B}(H, c_0/\sqrt{n})$ y $W \in \mathcal{W}$ tenemos*

$$G^n \left(\sup_{W \in \mathcal{W}} \sqrt{n} \left\| \beta(H_n, W) - \beta_0 - \iint \mathcal{J}\mathcal{F} dH_n \right\|_2 > \varepsilon \right) < \delta.$$

donde G^n es la distribución de una muestra aleatoria independiente x_1, x_2, \dots, x_n , idénticamente distribuída (su distribución es G), $\mathcal{J}\mathcal{F}(x, \beta_0, W) := -\mathcal{L}^{-1}(H, W) \text{Lin}_{W, \beta_0}(\delta_x - H)$ y δ_x tiene masa uno en el punto $x := (t, \mathbf{z}, \Delta)$.

La importancia de este teorema es que los estimadores son no solamente asintóticamente normales en el modelo sino que además en un entorno infenitesimal.

4.3.1. En busca del Punto de Ruptura

El punto de ruptura finito de un estimador $\hat{\theta}$ del parámetro θ es la proporción más grande que puede tener la muestra de outliers tal que $\hat{\theta}$ siga estimado bien a θ . Por ejemplo, la mediana muestral tiene punta de ruptura del 0,5 mientras que la media muestral tiene un punto de ruptura del 0. Con lo cual, es importante que un estimador tenga punto de ruptura no nulo, así la estimación no se altera en el caso en que la muestra contenga outliers o en el caso en que se presente un leve desvío del modelo.

Como demostramos en la sección anterior, la diferenciabilidad Fréchet conduce a propiedades muy deseables del estimador del parámetro β , dando más fiabilidad a las inferencias paramétricas. Como esta diferenciabilidad es probada para la norma supremo sobre entornos de la función de distribución, implica continuidad, y en esa norma, lo cual conduce a un punto de ruptura no nulo. O sea, que el funcional de estimación sea diferenciable Fréchet implica que sea continuo (respecto de la norma del supremo), es decir que no solo la estimación sigue siendo buena en presencia de outliers en la muestra, sino también cuando no se verifican algún supuesto del modelo (independencia sobre los tiempos de falla y censura, el modelo en sí, etc.). Esto permite que bajo modelos de contaminación más generales que simples outliers, la estimación sigue siendo óptima. El valor del punto de ruptura es una cuestión más

complicada de determinar teóricamente y el mismo depende de elementos usualmente desconocidos como la distribución de la censura y de las variables independientes. En el siguiente capítulo se lo trata de hallar en forma empírica. En estos estudios de simulación este dio bastante alto.

Capítulo 5

Análisis de datos y simulación

5.1. Simulación para los EFIA

En esta sección desarrollaremos un pequeño estudio de simulación para evaluar el desempeño del estimador propuesto. Específicamente tomamos $p = 1$, con función de riesgo verdadero de la forma $\lambda_*(t, z) = 1 + 0,5z$, que implica una distribución exponencial para los tiempos del evento. Para la distribución de la covariable tomamos $Z \sim \chi_1^2$, es decir, una distribución Chi-cuadrado con un grado de libertad. Optamos por una censura fija $C = 1,4$ (seleccionada para censurar el 15% de los tiempos observados). Realizamos dos estimadores para esta familia que llamaremos “Pesado” y “Robusto”. En ambos la función de peso $w_1(t) = 1$ es la misma pero cambiamos la función $w_2(z)$. Definimos

$$w_2(z) = \min \left\{ 1, \frac{a}{z} \tilde{s} \right\} \quad \text{y} \quad \tilde{s} = \frac{\text{median}\{z_1, \dots, z_n\}}{\log(2)}, \quad (5.1)$$

donde $a = 2,3026$. Este valor de a es tal que si Z tiene una distribución exponencial estandar, el 90% de las observaciones se mantendrían sin modificaciones. Entonces para el estimador Pesado se toma $w_2(z)$ como la definida en (5.1) y para el Robusto su cuadrado, es decir $w_2^2(z)$, que tiene una penalización más fuerte para valores grandes de z . Llamamos a este estimador Robusto porque cumple el supuesto **A2** en los Teoremas 4.2 y 4.3, ya que $zw_2^2(z) \rightarrow 0$ cuando $z \rightarrow \infty$. Esta propiedad no se satisface para el estimador Pesado.

En el Cuadro 5.1 comparamos el desempeño de las estimaciones de β , aumentando el tamaño de la muestra, a través del procedimiento clásico (el estimador de Lin y

n	Clásico	Pesado (con w_2)	Robusto (con w_2^2)
50	0.56 [0.06] (0.33)	0.52 [0.02] (0.30)	0.50 [-0.00] (0.34)
200	0.51 [0.01] (0.15)	0.49 [-0.01] (0.14)	0.47 [-0.03] (0.16)
1000	0.48 [-0.02] (0.07)	0.48 [-0.02] (0.07)	0.47 [-0.03] (0.07)
10000	0.48 [-0.02] (0.03)	0.48 [-0.02] (0.03)	0.48 [-0.02] (0.03)

Cuadro 5.1: Comparación de los estimadores en muestras no contaminadas.

Ying dado en (4.3)) y de las dos propuestas nuestras: Pesado y Robusto. Allí, sobre 1000 réplicas, se encuentra en cada casillero el estimador, su sesgo (entre corchetes) y su error estándar (entre paréntesis).

Observamos un muy buen desempeño en los estimadores. Además observamos que el desempeño del estimador Robusto es todavía muy buena, pero los errores estándares son ligeramente más grandes que los otros dos estimadores en muestras chicas.

En una segunda simulación, se compara el desempeño de los estimadores clásicos y nuestra propuesta, en presencia de contaminación. Basados sobre una muestra de tamaño $n = 50$ de un modelo generado como en la simulación anterior, reemplazamos un valor de la variable independiente z_i por los valores extremos $z^* = 5, 10, 50, 100, 1000$ respectivamente.

z^*	Clásico	Pesado (con w_2)	Robusto (con w_2^2)
5	0.45 [-0.05] (0.32)	0.46 [-0.04] (0.29)	0.48 [-0.02] (0.32)
10	0.31 [-0.19] (0.37)	0.37 [-0.13] (0.29)	0.45 [-0.05] (0.31)
50	0.10 [-0.40] (0.47)	0.19 [-0.31] (0.36)	0.44 [-0.06] (0.30)
100	0.08 [-0.42] (0.51)	0.14 [-0.36] (0.40)	0.44 [-0.06] (0.31)
1000	0.03 [-0.47] (0.79)	0.03 [-0.47] (0.48)	0.45 [-0.05] (0.31)

Cuadro 5.2: Comparación de los estimadores en muestras contaminadas.

En el Cuadro 5.2 mostramos los resultados sobre las 1000 réplicas, donde se encuentra en cada casillero el estimador, su sesgo (entre corchetes) y su error estándar (entre paréntesis). Se calcularon los siguientes estimadores: el estimador clásico, el Pesado y Robusto (definidos como en la simulación anterior). Observamos que el estimador Robusto con la selección de w_2^2 tiene un muy buen desempeño en todos

los casos. Alternativamente, la selección de w_2 sigue siendo muy sensible a la contaminación elegida y solo exhibe una mejora muy pequeña con respecto a la del estimador clásico.

5.2. Simulación para los EPRNN

En esta sección se evalúa el desempeño de los EPRNN. Para esto se realizaron tres simulaciones unidimensionales y en cada una se tomó como función de riesgo aditivo verdadera a $\lambda_*(t, z) = 1 + 0.5 z$, con $Z \sim \chi_1^2$ y, para no introducir censura, $C = 1000$. En todos los casos, en la ecuación de estimación (4.16), se seleccionó

$$W(t_i, z_i) = 1 - \text{mín}(1, z_i/M), \quad (5.2)$$

donde M es el percentil $m\%$ de los Z .

En la primera simulación se muestra el desempeño de nuestro estimador en varios tamaños de muestra. Estos valores fueron $n = 50, 200, 500, 1000$ y 10000 y en la función $W(t_i, z_i)$ se seleccionó $m = 90, 95$ y 99 .

n	Clásico	EPRNN		
		$m = 90$	$m = 95$	$m = 99$
50	0.59 [0.09] (0.36)	0.54 [0.04] (0.48)	0.54 [0.04] (0.39)	0.55 [0.05] (0.35)
200	0.52 [0.02] (0.15)	0.52 [0.02] (0.22)	0.51 [0.01] (0.18)	0.51 [0.01] (0.15)
500	0.51 [0.01] (0.09)	0.50 [0.00] (0.14)	0.50 [0.00] (0.11)	0.50 [0.00] (0.09)
1000	0.50 [0.00] (0.06)	0.50 [0.00] (0.09)	0.50 [0.00] (0.08)	0.50 [0.00] (0.06)
10000	0.50 [0.00] (0.02)	0.50 [0.00] (0.03)	0.50 [0.00] (0.02)	0.50 [0.00] (0.02)

Cuadro 5.3: Comparación del estimador aumentando el tamaño de la muestra bajo el modelo.

En el Cuadro 5.3 comparamos el desempeño del EPRNN frente al estimador clásico (el estimador de Lin y Ying). Se hicieron 1000 réplicas. Los tres números que figuran en cada columna son, igual que en los casos anteriores, el estimador, su sesgo (entre corchetes) y su error estándar (entre paréntesis).

Observamos en las tres opciones de EPRNN un muy buen desempeño en la estimación del parámetro y un leve aumento en el error estándar. Este error disminuye con el aumento de la muestra y el aumento del m .

En la segunda simulación se compara el desempeño del estimador clásico y EFIA vs. EPRNN frente a una contaminación en la muestra. Esta contaminación consiste en reemplazar un porcentaje de las covariables por el valor 10. Como la función de peso W , en (5.2), anula para los valores de las covariables mayores al valor de M , se obtienen resultados similares si en lugar de 10 se reemplazaría por un valor más alto. Se han considerado 1000 réplicas de una muestra unidimensional de tamaño 200.

%	Sin Contaminar	Contaminando las covariables		
	Clásico	Clásico	EFIA	EPRNN
0	0.52 [0.02] (0.15)	0.52 [0.02] (0.15)	0.52 [0.02] (0.16)	0.53 [0.03] (0.21)
0.5	0.53 [0.03] (0.15)	0.41 [-0.10] (0.20)	0.50 [-0.00] (0.15)	0.51 [0.01] (0.21)
5	0.53 [0.03] (0.15)	0.10 [-0.40] (0.40)	0.37 [-0.13] (0.17)	0.52 [0.02] (0.18)
15	0.52 [0.02] (0.15)	0.04 [-0.46] (0.46)	0.19 [-0.31] (0.31)	0.51 [0.01] (0.15)
25	0.52 [0.02] (0.16)	0.02 [-0.48] (0.48)	0.09 [-0.41] (0.41)	0.52 [0.02] (0.17)

Cuadro 5.4: Comparación de los estimadores contaminando la muestra bajo el modelo.

Los métodos de estimación calculados en el Cuadro 5.4 fueron el “Clásico” (método clásico de estimación, el de Ling y Ying), “EFIA” (con $w_1(t) = 1$ y $w_2^2(z)$ definida en (5.1)) y “EPRNN” (se utilizó la función (5.2) con $m = 90$). Notemos que para esta simulación el estimador EFIA no soporta un alto porcentaje de contaminación (con un 5 % no es buena la estimación) mientras que EPRNN sí (con un 25 % de contaminación la estimación sigue siendo muy buena). Además, el estimador EPRNN aumenta su error estándar comparado con el clásico si no contaminamos la muestra).

El objetivo de la última simulación es buscar el punto de ruptura cuando se contamina el modelo con otro modelo. Para esto, se contaminó un porcentaje de la muestra (T, Z) con el modelo lineal: $T = 4 + 0.25 Z$, donde T es el tiempo de falla asociado a la covariable Z . Un ejemplo de esto se encuentra en la Figura 5.1.

En el Cuadro 5.5 se encuentra los resultados de 200 réplicas a muestras de tamaño 500. Observar que para los EPRNN seleccionados aumentan su punto de ruptura si disminuimos el valor de m . Para $m = 90$ el punto de ruptura estaría entre 3 y 10 y para $m = 70$ estaría entre 25 y 30.

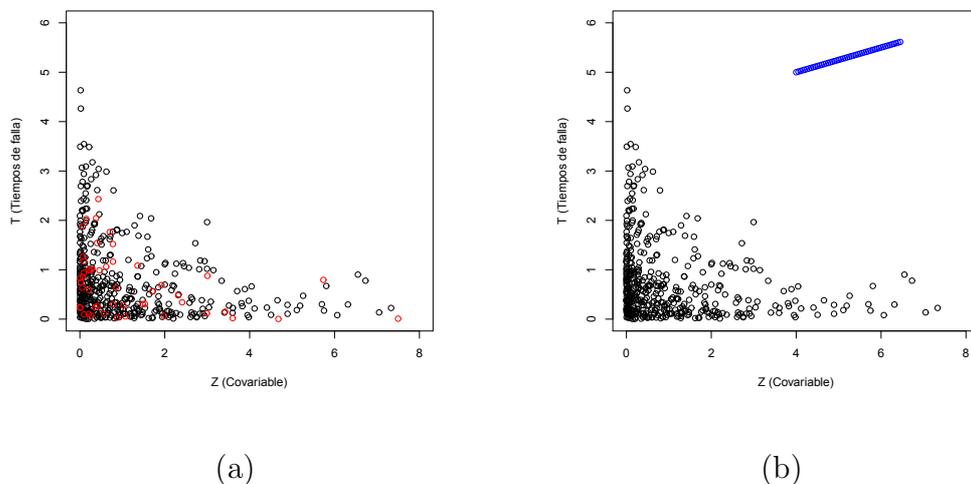


Figura 5.1: Las gráficas corresponden a una muestra (T, Z) del modelo de riesgo aditivo. En (a) es la muestra original (sin contaminar) donde los puntos de color rojo son los que se cambiarán en (b). En (b) se reemplazó los puntos rojos (que representan el 10%) por los puntos azules.

5.3. Análisis de datos

En esta sección aplicamos nuestro método a datos reales: conjunto de datos del sur de Gales de una refinería de níquel. Estos datos se introdujeron por primera vez por Doll y otros (1970) y posteriormente se analizaron con un modelo de riesgo multiplicativo por Breslow y Day (1987), y con un modelo de riesgo aditivo (no robusto) por Lin y Ying (1994).

El objetivo del estudio, que se realizó entre los años 1934 y 1981, fue determinar el riesgo de desarrollar cáncer pulmonar o nasal para los trabajadores mineros en una instalación de refinería de níquel situada en el sur de Gales, ya que ambas enfermedades en 1949 fueron preinscriptas como enfermedades ocupacionales. Las muertes observadas se registraron para los trabajadores y su causa de muerte determinada a partir del certificado de defunción. La censura se produjo a partir de los trabajadores que aún estaban con vida al final del estudio, los trabajadores cuyo nivel de vida era desconocido, así como los trabajadores muertos cuya causa de muerte no se pudo determinar claramente del certificado de defunción. Adicionalmente, en el conjunto de datos original, los autores decidieron considerar censurado todas las muertes de los trabajadores a los 85 años o más, ya que consideran que para las personas de esas

%	Sin Contaminar		Contaminando	
	$m = 70$	$m = 90$	$m = 70$	$m = 90$
0.2	0.55 [0.05] (0.08)	0.52 [0.02] (0.02)	0.55 [0.05] (0.08)	0.52 [0.02] (0.02)
3	0.50 [0.00] (0.07)	0.51 [0.01] (0.02)	0.50 [0.00] (0.06)	0.51 [0.01] (0.02)
10	0.50 [-0.00] (0.08)	0.52 [0.02] (0.02)	0.51 [0.01] (0.05)	0.21 [-0.30] (0.18)
20	0.50 [-0.00] (0.07)	0.50 [0.00] (0.02)	0.50 [0.00] (0.02)	-0.13 [-0.63] (0.80)
25	0.48 [-0.02] (0.07)	0.51 [0.01] (0.02)	0.51 [0.01] (0.02)	-0.16 [-0.66] (0.87)
30	0.51 [0.01] (0.07)	0.53 [0.03] (0.02)	0.22 [-0.28] (0.16)	-0.17 [-0.67] (0.89)

Cuadro 5.5: Estimación del parámetro, su sesgo y su error cuadrático medio para una muestra contaminada con otro modelo.

edades podrían haber clasificado mal la causa de la muerte debido a la vejez. Nuestra preocupación era que esta decisión podría haber introducido los valores atípicos en los datos. Los valores extremos son particularmente peligrosas para la inferencia, ya que todos ellos corresponden a supervivencias de gran tamaño.

En nuestro modelo, la variable de supervivencia T corresponde a la duración desde el primer empleo hasta la muerte para aquellos trabajadores que fallecieron debido al cáncer pulmonar o nasal o hasta el fin del estudio para lo que sobrevivieron. Incluimos tres covariables:

$$\begin{aligned} Z_1 &= \log(\text{AFE}), \\ Z_2 &= (\text{YFE} - 1900)/10, \\ Z_3 &= \log(\text{EXP} + 1), \end{aligned}$$

donde AFE es la edad del trabajador minero que tenía al comenzar a trabajar en la refinería, YFE es el año en que comenzó a trabajar en la refinería y EXP es el nivel de exposición en la refinería.

En el Cuadro 5.6 a continuación se presentan las estimaciones de los parámetros de los estimadores clásicos y robustos en dos marcos:

- i)* con el conjunto de datos completos,
- ii)* con la reducción del conjunto de datos donde las muertes en edad de 85 o más fueron omitidas, que es la alternativa natural para el tratamiento de los valores atípicos potenciales.

Coord. de $\hat{\beta}$	Conjunto de datos completos		Conjunto de datos reducidos	
	Clásico	Robusto	Clásico	Robusto
β_1	8.60 (1.65)	7.27 (1.95)	11.16 (1.86)	8.35 (2.17)
β_2	2.33 (0.83)	2.27 (0.99)	1.84 (0.91)	2.11 (1.09)
β_3	4.60 (0.74)	3.49 (0.88)	4.70 (0.80)	3.42 (0.95)

Cuadro 5.6: Estimación del parámetro y error estándar para el conjunto de datos Niquel. Todos los resultados fueron multiplicados por 1000.

Los estimadores robustos se calcularon con las funciones de peso

$$w_1(t) = [1 - F_n(t)]^{1.5},$$

donde $F_n(t)$ es la distribución empírica de los t , y $\mathbf{w}_2(\mathbf{z}_i)$ una matriz diagonal con entradas de

$$w_2(z_{j,i}) = \min(1, d_j s_i / z_{j,i}),$$

con $s_i = \text{median}(z_{\cdot,i}) / \log 2$ y $d = (0.82, 1.005, 2.735)$.

Las estimaciones robustas se ven menos afectadas por la eliminación de los valores atípicos potenciales que el clásico, ya que cambian mucho menos en ambos casos. Además, la estimación robusta se asemeja a la clásica en el conjunto de datos reducidos. Estos resultados, por lo tanto, parecen validar desde el punto de vista de la robustez las conclusiones de la aplicación de los estimadores clásicos con el conjunto de datos completos.

Capítulo 6

Conclusiones

Como hemos dicho antes, el objetivo principal de este trabajo fue proponer estimadores robustos para el parámetro de regresión para el Modelo de Hazard Aditivo (MHA).

A lo largo del estudio sobre los modelo de riesgo han surgido dos familias de estimadores robusta: estimadores con función de influencia acotada (EFIA) y estimadores con punto de ruptura no nulo (EPRNN).

Cabe destacar que hasta el momento no hay procedimientos de estimación robustos para MHA en el caso de eventos simples.

Como punto de partida para la robustificación del MHA tomamos la representación de la función de score como un problema de M-estimación. Recordemos que esta propuesta, realizada por Andersen y otros (1993), expresa la función de score como

$$\sum_{i=1}^n \int_0^{\infty} \frac{\partial}{\partial \boldsymbol{\beta}} [\log(\lambda(t, \mathbf{z}_i; \boldsymbol{\beta}))] [dN_i(t) - Y_i(t)\lambda(t, \mathbf{z}_i; \boldsymbol{\beta})dt] \quad (6.1)$$

donde el término $dN_i(t) - Y_i(t)\lambda_i(t; \boldsymbol{\beta})dt$ es una martingala. Como ya lo mencionamos en la Sección 4.2, la generalización de (6.1) es:

$$\sum_{i=1}^n \int_0^{\infty} K_i(t, \mathbf{z}_i, \boldsymbol{\beta}) [dN_i(t) - Y_i(t)\lambda(t, \mathbf{z}_i; \boldsymbol{\beta})dt].$$

La primer propuesta fue motivada por analogía con los trabajos de Sasieni (1993a, 1993b) que modifica la función de score del Modelo de Hazard Multiplicativo (MHM),

introduciendo una función de peso, como se lo desarrolló en la Sección 3.1. Creemos que esta modificación en la función de score funciona bien porque mientras modifica el funcional de estimación y utiliza esta misma modificación para la estimación de la función de riesgo acumulado baseline, su método mantiene intacta la estructura de martingala y la dependencia de la estimación de la función de riesgo acumulado en la especificación del modelo. Nuestra modificación en la función de score en el MHA introduciendo funciones de peso también mantiene esas mismas propiedades.

Además, como vimos en la Sección 3.3, en la estimación para MHA sería necesario utilizar el integrando

$$K_i(t, \mathbf{z}_i, \boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\beta}} \log[\lambda(t, \mathbf{z}_i; \boldsymbol{\beta})] = \frac{\mathbf{z}_i}{\lambda_0(t) + \boldsymbol{\beta}' \mathbf{z}_i} \quad (6.2)$$

ya sea en la estimación de máxima verosimilitud clásica (3.15) o de máxima verosimilitud parcial (3.17). Observar que en contraste con el MHM, la estimación de máxima verosimilitud parcial

$$K_i(t, \mathbf{z}_i, \boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\beta}} \log[\lambda(t, \mathbf{z}_i; \boldsymbol{\beta})] = \frac{\mathbf{z}_i}{\lambda_0(t) \exp(\boldsymbol{\beta}' \mathbf{z}_i)} = \mathbf{z}_i, \quad (6.3)$$

elimina la función baseline.

La propuesta de Lin y Ying (1994) para MHA, sustituye el lado derecho de (6.2) por \mathbf{z}_i (el mismo que resulta para MHM, (6.3)), obteniendo un estimador implícito (esto no sucede para el MHM). Sin embargo, nosotros conjeturamos que esta sustitución disminuye robustez incluso con respecto al estimador de máxima verosimilitud, en el sentido que la estimación es más sensible a valores grandes de \mathbf{z}_i o $\lambda_0(t)$, ya que al no estar el denominador de (6.2) desaparece el peso hacia abajo sobre la ecuación de score. La incorporación nuestra de funciones de peso en los EFIA a propósito, están destinadas a aumentar su robustez. Además, su buen desempeño frente a outliers esta asegurado con la función de influencia acotada, esto no sucede con el estimador clásico de Lin y Ying.

Por estas dos razones concluimos que nuestra primer propuesta de estimadores robusta funciona como esperábamos y esto se vió claramente reflejado en los estudios de simulación.

La segunda propuesta surge a partir del trabajo de Bednarski (1993) quien utilizó la diferenciabilidad funcional como herramienta crucial para establecer la consistencia y la normalidad asintótica, no solo bajo el modelo MHA sino también bajo pequeños entornos del mismo.

Como se mostro en la Sección 5.2, en la simulación del Cuadro 5.4, hay inconsistencia en el estimador clásico bajo pequeñas perturbaciones del modelo, es decir que el estimador clásico no se comporta bien sin importar cuan cerca estamos del modelo. Esto se debe a que el estimador clásico no es continuo con respecto a la distribución del modelo. Por ejemplo, considerar el Ejemplo 4.1 y

$$G_n(t, z, \delta) = \left(1 - \frac{1}{n}\right) H_1(t, z, \delta) + \frac{1}{n} \nu_{(n,0,1)}(t, z, \delta)$$

donde $\nu_{(n,0,1)}(t, z, \delta)$ es la masa puntual en $(n, 0, 1)$. Como vimos en la Sección 4.2.1, la función de influencia es no acotada entonces es fácil ver que $|G_n - H_1| \rightarrow 0$ pero $|\mathcal{Q}(G_n) - \mathcal{Q}(H_1)| \rightarrow \infty$.

En contraste, nuestra modificación en el funcional de estimación no solo es continuo sino que también es diferenciable Fréchet.

Finalmente concluimos con respecto al desempeño de ambas propuestas robustas que la diferencia en la estimación bajo la presencia de valores atípicos o perturbaciones del modelo es que la segunda es más dañina que la primera, ya que se perturba la dependencia del tiempo de supervivencia y del vector de covariables. Por esto, nuestra segunda propuesta es mejor que la primera, pero ambas son mejores que el estimador clásico.

La mejora alcanzada para los EPRNN es gracias a la diferenciable Fréchet que conduce a la estabilidad asintótica más allá del modelo.

Capítulo 7

Apéndice

Recordemos que si \mathbf{a} es un vector entonces $\mathbf{a}^{\otimes 0} = 1$ (constante), $\mathbf{a}^{\otimes 1} = \mathbf{a}$ (vector) y $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}'$ (matriz). Además, $\mathcal{S}_n^{(l)}(t)$ es definido para $l = 0, 1$ como:

$$\mathcal{S}_n^{(l)}(t) := \frac{1}{n} \sum_{j=1}^n \mathbf{w}_2(\mathbf{z}_j) \mathbf{z}_j^{\otimes l} Y_j(t). \quad (7.1)$$

Demostración del Teorema 4.1

Recordemos que por la ecuación (4.10), podemos expresar al estimador como $\tilde{\boldsymbol{\beta}}_n = \mathcal{Q}_w(H_n) = [\mathcal{D}_w(H_n)]^{-1} [\mathcal{N}_w(H_n)]$ donde

$$\begin{aligned} \mathcal{D}_w(H_n) &= \iint \mathbf{w}_2(\mathbf{z}) \left(\int_0^t w_1(u) [\mathbf{z} - \tilde{\mathbf{z}}_n(u)]^{\otimes 2} du \right) dH_n(t, \mathbf{z}), \\ \mathcal{N}_w(H_n) &= \iint w_1(t) \mathbf{w}_2(\mathbf{z}) [\mathbf{z} - \tilde{\mathbf{z}}_n(t)] dH_n(t, \mathbf{z}, \Delta = 1), \end{aligned}$$

y

$$\tilde{\mathbf{z}}_n(u) = \left[\iint I(s \geq u) \mathbf{w}_2(\mathbf{x}) dH_n(s, \mathbf{x}) \right]^{-1} \left[\iint I(s \geq u) \mathbf{w}_2(\mathbf{x}) \mathbf{x} dH_n(s, \mathbf{x}) \right].$$

Antes de comenzar con la demostración notemos que

$$\iint I(t \geq u) \mathbf{w}_2(\mathbf{z}) [\mathbf{z} - \tilde{\mathbf{z}}(u)] dH(t, \mathbf{z}) = \mathbf{0}, \quad \forall u. \quad (7.2)$$

Ya que

$$\begin{aligned}
\iint I(t \geq u) \mathbf{w}_2(\mathbf{z}) [\mathbf{z} - \tilde{\mathbf{z}}(u)] dH(t, \mathbf{z}) &= \iint I(t \geq u) \mathbf{w}_2(\mathbf{z}) \mathbf{z} dH(t, \mathbf{z}) - \\
&\quad \iint I(t \geq u) \mathbf{w}_2(\mathbf{z}) dH(t, \mathbf{z}) \tilde{\mathbf{z}}(u) \\
&= \iint I(t \geq u) \mathbf{w}_2(\mathbf{z}) \mathbf{z} dH(t, \mathbf{z}) - \\
&\quad \iint I(s \geq u) \mathbf{w}_2(\mathbf{x}) \mathbf{x} dH(s, \mathbf{x}) = \mathbf{0}.
\end{aligned}$$

Esto implica que:

$$\begin{aligned}
\mathcal{D}_w(H) &= \iint \mathbf{w}_2(\mathbf{z}) \left(\int_0^t w_1(u) [\mathbf{z} - \tilde{\mathbf{z}}_n(u)]^{\otimes 2} du \right) dH(t, \mathbf{z}) \\
&= \iint \mathbf{w}_2(\mathbf{z}) \left(\int_0^t w_1(u) [\mathbf{z} - \tilde{\mathbf{z}}(u)] [\mathbf{z} - \tilde{\mathbf{z}}_n(u)]' du \right) dH(t, \mathbf{z}) \\
&= \int w_1(u) \iint I(t \geq u) \mathbf{w}_2(\mathbf{z}) [\mathbf{z} - \tilde{\mathbf{z}}(u)] \mathbf{z}' dH(t, \mathbf{z}) du - \\
&\quad \int w_1(u) \left(\iint I(t \geq u) \mathbf{w}_2(\mathbf{z}) [\mathbf{z} - \tilde{\mathbf{z}}(u)] dH(t, \mathbf{z}) \right) \tilde{\mathbf{z}}_n(u)' du \\
&\stackrel{(7.2)}{=} \int w_1(u) \iint I(t \geq u) \mathbf{w}_2(\mathbf{z}) [\mathbf{z} - \tilde{\mathbf{z}}(u)] \mathbf{z}' dH(t, \mathbf{z}) du \\
&= \iint \mathbf{w}_2(\mathbf{z}) \left(\int_0^t w_1(u) [\mathbf{z} - \tilde{\mathbf{z}}(u)] du \right) \mathbf{z}' dH(t, \mathbf{z}). \tag{7.3}
\end{aligned}$$

Hagamos unas cuentas previas antes de la demostración. Pero necesitamos primero definir dos funciones para simplificar los resultados. La primer función que definimos es:

$$\Gamma_l(u) := \int \mathbf{w}_2(\mathbf{z}) \mathbf{z}^{\otimes l} \exp(-\boldsymbol{\beta}' \mathbf{z} u) f_Z(\mathbf{z}) d\mathbf{z}, \tag{7.4}$$

para $l = 0, 1, 2$ y $u \geq 0$. Observar que $\Gamma_0(u)$ es una matriz de $p \times p$ invertible (por la condición **A2**), $\Gamma_1(u)$ es un vector de $p \times 1$, $\Gamma_2(u)$ es una matriz de $p \times p$. Además, por la condición **A2**,

$$\Gamma_l(u) \leq \int \mathbf{w}_2(\mathbf{z}) \mathbf{z}^{\otimes l} f_Z(\mathbf{z}) d\mathbf{z} < \infty, \quad \forall u \geq 0. \tag{7.5}$$

La segunda función que necesitamos definir es

$$\gamma_l(t) := \int_0^t w_1(u) [\Gamma_0^{-1}(u) \Gamma_1(u)]^{\otimes l} du, \tag{7.6}$$

para $l = 0, 1$ y $t \geq 0$. Notar que $\gamma_0(t)$ es un real positivo no negativo y $\gamma_1(t)$ es un vector de $p \times 1$. Además, $\gamma_0(0) = 0$, $\gamma_1(0) = \mathbf{0}$, $d\gamma_l(t)/dt = w_1(t) [\Gamma_0^{-1}(t)\Gamma_1(t)]^{\otimes l}$,

$$\lim_{t \rightarrow \infty} \gamma_0(t) = \int_0^\infty w_1(t) dt < \infty \quad \text{y} \quad \lim_{t \rightarrow \infty} \gamma_1(t) = \int_0^\infty w_1(t) [\Gamma_0^{-1}(t)\Gamma_1(t)] dt < \infty, \quad (7.7)$$

(por la condición **A1**).

Recordar también que $\tilde{S}(t) := S_0(t)S_C(t) = [1 - F_0(t)][1 - F_C(t)]$, entonces $\tilde{S}(0) = 1$, $\lim_{t \rightarrow \infty} \tilde{S}(t) = 0$ y

$$\begin{aligned} \frac{d}{dt} \tilde{S}(t) &= (-f_0(t))S_C(t) + S_0(t)(-f_C(t)) = -S_0(t)S_C(t) \left(\frac{f_0(t)}{S_0(t)} + \frac{f_C(t)}{S_C(t)} \right) \\ &\stackrel{(2.2)}{=} -\tilde{S}(t)(\lambda_0(t) + \lambda_C(t)) = -\tilde{S}(t)\tilde{\lambda}(t). \end{aligned} \quad (7.8)$$

Por último, integremos por partes las siguientes dos integrales:

$$D_0 := \int \tilde{S}(t)\gamma_0(t) \exp(-\boldsymbol{\beta}'\mathbf{z}t)(\boldsymbol{\beta}'\mathbf{z}) dt \quad \text{y} \quad D_1 := \int \tilde{S}(t)\gamma_1(t) \exp(-\boldsymbol{\beta}'\mathbf{z}t)(\boldsymbol{\beta}'\mathbf{z}) dt.$$

En la primera tenemos que:

$$\begin{aligned} D_0 &= \int \tilde{S}(t) \gamma_0(t) d(-\exp(-\boldsymbol{\beta}'\mathbf{z}t)) \\ &\stackrel{(7.8)}{=} \tilde{S}(t)\gamma_0(t)(-\exp(-\boldsymbol{\beta}'\mathbf{z}t)) \Big|_0^\infty + \int \exp(-\boldsymbol{\beta}'\mathbf{z}t) \left[-\tilde{S}(t)\tilde{\lambda}(t)\gamma_0(t) + \tilde{S}(t)w_1(t) \right] dt \\ &\stackrel{(7.7)}{=} \int \exp(-\boldsymbol{\beta}'\mathbf{z}t) \left[\tilde{S}(t)w_1(t) - \tilde{S}(t)\tilde{\lambda}(t)\gamma_0(t) \right] dt \\ &= \int \tilde{S}(t)w_1(t) \exp(-\boldsymbol{\beta}'\mathbf{z}t) dt - \int \tilde{S}(t)\tilde{\lambda}(t)\gamma_0(t) \exp(-\boldsymbol{\beta}'\mathbf{z}t) dt. \end{aligned} \quad (7.9)$$

Y en la segunda, tenemos que:

$$\begin{aligned} D_1 &= \int \tilde{S}(t) \gamma_1(t) d(-\exp(-\boldsymbol{\beta}'\mathbf{z}t)) \\ &\stackrel{(7.8)}{=} \tilde{S}(t)\gamma_1(t)(-\exp(-\boldsymbol{\beta}'\mathbf{z}t)) \Big|_0^\infty + \int \exp(-\boldsymbol{\beta}'\mathbf{z}t) \left[-\tilde{S}(t)\tilde{\lambda}(t)\gamma_1(t) + \tilde{S}(t)w_1(t)\Gamma_0^{-1}(t)\Gamma_1(t) \right] dt \\ &\stackrel{(7.7)}{=} \int \exp(-\boldsymbol{\beta}'\mathbf{z}t) \left[\tilde{S}(t)w_1(t)\Gamma_0^{-1}(t)\Gamma_1(t) - \tilde{S}(t)\tilde{\lambda}(t)\gamma_1(t) \right] dt \\ &= \int \tilde{S}(t)w_1(t)\Gamma_0^{-1}(t)\Gamma_1(t) \exp(-\boldsymbol{\beta}'\mathbf{z}t) dt - \int \tilde{S}(t)\tilde{\lambda}(t)\gamma_1(t) \exp(-\boldsymbol{\beta}'\mathbf{z}t) dt. \end{aligned} \quad (7.10)$$

Demostración: Sea $H \in \mathcal{A}_\beta$, para algún $\beta \in \mathbb{R}^{p+}$. Entonces tenemos que:

$$dH(t, \mathbf{z}, \Delta) = \tilde{S}(t) \exp(-\boldsymbol{\beta}'\mathbf{z}t) [\lambda_0(t) + \boldsymbol{\beta}'\mathbf{z}] f_Z(\mathbf{z}) dz dt, \quad (7.11)$$

y

$$dH(t, \mathbf{z}) = \tilde{S}(t) \exp(-\boldsymbol{\beta}' \mathbf{z} t) [\tilde{\lambda}(t) + \boldsymbol{\beta}' \mathbf{z}] f_Z(\mathbf{z}) d\mathbf{z} dt, \quad (7.12)$$

donde $\tilde{S}(t) = S_0(t)S_C(t)$ y $\tilde{\lambda}(t) = \lambda_0(t) + \lambda_C(t)$.

Luego por un lado, tenemos que, para $l = 0, 1$:

$$\begin{aligned} \iint I(s \geq u) \mathbf{w}_2(\mathbf{x}) \mathbf{x}^{\otimes l} dH(s, \mathbf{x}) &= \iint I(s \geq u) \mathbf{w}_2(\mathbf{x}) \mathbf{x}^{\otimes l} dF_{T|Z}(s, \mathbf{x}) dF_Z(\mathbf{x}) \\ &= \int \left[\int I(s \geq u) dF_{T|Z}(s, \mathbf{x}) \right] \mathbf{w}_2(\mathbf{x}) \mathbf{x}^{\otimes l} f_Z(\mathbf{x}) d\mathbf{x} \\ &= \int \text{Prob}(T \geq u | \mathbf{Z} = \mathbf{x}) \mathbf{w}_2(\mathbf{x}) \mathbf{x}^{\otimes l} f_Z(\mathbf{x}) d\mathbf{x} \\ &\stackrel{(2.12)}{=} \int \tilde{S}(u) \exp(-\boldsymbol{\beta}' \mathbf{x} u) \mathbf{w}_2(\mathbf{x}) \mathbf{x}^{\otimes l} f_Z(\mathbf{x}) d\mathbf{x} \\ &= \tilde{S}(u) \int \exp(-\boldsymbol{\beta}' \mathbf{x} u) \mathbf{w}_2(\mathbf{x}) \mathbf{x}^{\otimes l} f_Z(\mathbf{x}) d\mathbf{x} \\ &\stackrel{(7.4)}{=} \tilde{S}(u) \Gamma_l(u). \end{aligned} \quad (7.13)$$

Entonces

$$\begin{aligned} \tilde{\mathbf{z}}(u) &= \left[\iint I(s \geq u) \mathbf{w}_2(\mathbf{x}) dH(s, \mathbf{x}) \right]^{-1} \left[\iint I(s \geq u) \mathbf{w}_2(\mathbf{x}) \mathbf{x} dH(s, \mathbf{x}) \right] \\ &\stackrel{(7.13)}{=} \left[\tilde{S}(u) \Gamma_0(u) \right]^{-1} \tilde{S}(u) \Gamma_1(u) = \Gamma_0^{-1}(u) \Gamma_1(u). \end{aligned} \quad (7.14)$$

Por lo tanto:

$$\begin{aligned} \mathcal{N}_w(H) &= \iint w_1(t) \mathbf{w}_2(\mathbf{z}) [\mathbf{z} - \tilde{\mathbf{z}}(t)] dH(t, \mathbf{z}, \Delta = 1) \\ &\stackrel{(7.11)}{=} \iint w_1(t) \mathbf{w}_2(\mathbf{z}) [\mathbf{z} - \tilde{\mathbf{z}}(t)] \tilde{S}(t) \exp(-\boldsymbol{\beta}' \mathbf{z} t) [\lambda_0(t) + \boldsymbol{\beta}' \mathbf{z}] f_Z(\mathbf{z}) d\mathbf{z} dt \\ &= \int \tilde{S}(t) w_1(t) \int \mathbf{w}_2(\mathbf{z}) \mathbf{z} \exp(-\boldsymbol{\beta}' \mathbf{z} t) [\lambda_0(t) + \boldsymbol{\beta}' \mathbf{z}] f_Z(\mathbf{z}) d\mathbf{z} dt - \\ &\quad \int \tilde{S}(t) w_1(t) \int \mathbf{w}_2(\mathbf{z}) \tilde{\mathbf{z}}(t) \exp(-\boldsymbol{\beta}' \mathbf{z} t) [\lambda_0(t) + \boldsymbol{\beta}' \mathbf{z}] f_Z(\mathbf{z}) d\mathbf{z} dt \end{aligned}$$

$$\begin{aligned}
&= \int \tilde{S}(t) \lambda_0(t) w_1(t) \int \mathbf{w}_2(\mathbf{z}) \mathbf{z} \exp(-\boldsymbol{\beta}' \mathbf{z} t) f_Z(\mathbf{z}) d\mathbf{z} dt + \\
&\quad \int \tilde{S}(t) w_1(t) \int \mathbf{w}_2(\mathbf{z}) \mathbf{z} \exp(-\boldsymbol{\beta}' \mathbf{z} t) \boldsymbol{\beta}' \mathbf{z} f_Z(\mathbf{z}) d\mathbf{z} dt - \\
&\quad \int \tilde{S}(t) \lambda_0(t) w_1(t) \int \mathbf{w}_2(\mathbf{z}) \tilde{\mathbf{z}}(t) \exp(-\boldsymbol{\beta}' \mathbf{z} t) f_Z(\mathbf{z}) d\mathbf{z} dt - \\
&\quad \int \tilde{S}(t) w_1(t) \int \mathbf{w}_2(\mathbf{z}) \tilde{\mathbf{z}}(t) \exp(-\boldsymbol{\beta}' \mathbf{z} t) \boldsymbol{\beta}' \mathbf{z} f_Z(\mathbf{z}) d\mathbf{z} dt \\
&= \int \tilde{S}(t) \lambda_0(t) w_1(t) \left(\int \mathbf{w}_2(\mathbf{z}) \mathbf{z} \exp(-\boldsymbol{\beta}' \mathbf{z} t) f_Z(\mathbf{z}) d\mathbf{z} \right) dt + \\
&\quad \int \tilde{S}(t) w_1(t) \left(\int \mathbf{w}_2(\mathbf{z}) \mathbf{z} \mathbf{z}' \exp(-\boldsymbol{\beta}' \mathbf{z} t) f_Z(\mathbf{z}) d\mathbf{z} \right) dt \boldsymbol{\beta} - \\
&\quad \int \tilde{S}(t) \lambda_0(t) w_1(t) \left(\int \mathbf{w}_2(\mathbf{z}) \exp(-\boldsymbol{\beta}' \mathbf{z} t) f_Z(\mathbf{z}) d\mathbf{z} \right) \tilde{\mathbf{z}}(t) dt - \\
&\quad \int \tilde{S}(t) w_1(t) \left(\int \mathbf{w}_2(\mathbf{z}) \tilde{\mathbf{z}}(t) \mathbf{z}' \exp(-\boldsymbol{\beta}' \mathbf{z} t) f_Z(\mathbf{z}) d\mathbf{z} \right) dt \boldsymbol{\beta} \\
&\stackrel{(7.4)}{=} \int \tilde{S}(t) \lambda_0(t) w_1(t) \Gamma_1(t) dt - \int \tilde{S}(t) \lambda_0(t) w_1(t) \Gamma_0(t) \tilde{\mathbf{z}}(t) dt + \\
&\quad \int \tilde{S}(t) w_1(t) \left(\int \mathbf{w}_2(\mathbf{z}) \mathbf{z}^{\otimes 2} \exp(-\boldsymbol{\beta}' \mathbf{z} t) f_Z(\mathbf{z}) d\mathbf{z} \right) dt \boldsymbol{\beta} - \\
&\quad \int \tilde{S}(t) w_1(t) \left(\int \mathbf{w}_2(\mathbf{z}) \tilde{\mathbf{z}}(t) \mathbf{z}' \exp(-\boldsymbol{\beta}' \mathbf{z} t) f_Z(\mathbf{z}) d\mathbf{z} dt \right) \boldsymbol{\beta} \\
&\stackrel{(7.14)}{=} \int \tilde{S}(t) \lambda_0(t) w_1(t) \left[\Gamma_1(t) - \Gamma_0(t) \left(\Gamma_0^{-1}(t) \Gamma_1(t) \right) \right] dt + \\
&\quad \int \tilde{S}(t) w_1(t) \int \mathbf{w}_2(\mathbf{z}) [\mathbf{z}^{\otimes 2} - \tilde{\mathbf{z}}(t) \mathbf{z}'] \exp(-\boldsymbol{\beta}' \mathbf{z} t) f_Z(\mathbf{z}) d\mathbf{z} dt \boldsymbol{\beta} \\
&= \left[\int \tilde{S}(t) w_1(t) \int \mathbf{w}_2(\mathbf{z}) [\mathbf{z}^{\otimes 2} - \tilde{\mathbf{z}}(t) \mathbf{z}'] \exp(-\boldsymbol{\beta}' \mathbf{z} t) f_Z(\mathbf{z}) d\mathbf{z} dt \right] \boldsymbol{\beta}. \quad (7.15)
\end{aligned}$$

Por otro lado,

$$\begin{aligned}
\mathcal{D}_w(H) &\stackrel{(7.3)}{=} \iint \mathbf{w}_2(\mathbf{z}) \left(\int_0^t w_1(u) [\mathbf{z} - \tilde{\mathbf{z}}(u)] du \right) \mathbf{z}' dH(t, \mathbf{z}) \\
&\stackrel{(7.12)}{=} \iint \mathbf{w}_2(\mathbf{z}) \int_0^t w_1(u) [\mathbf{z} - \tilde{\mathbf{z}}(u)] du \mathbf{z}' \tilde{S}(t) \exp(-\boldsymbol{\beta}' \mathbf{z} t) [\tilde{\lambda}(t) + \boldsymbol{\beta}' \mathbf{z}] f_Z(\mathbf{z}) d\mathbf{z} dt
\end{aligned}$$

$$\begin{aligned}
& \stackrel{(7.14)}{=} \int \mathbf{w}_2(\mathbf{z}) \mathbf{z} \mathbf{z}' \int \tilde{S}(t) [\tilde{\lambda}(t) + \boldsymbol{\beta}' \mathbf{z}] \left(\int_0^t w_1(u) du \right) \exp(-\boldsymbol{\beta}' \mathbf{z} t) dt f_Z(\mathbf{z}) d\mathbf{z} - \\
& \quad \int \mathbf{w}_2(\mathbf{z}) \int \tilde{S}(t) \left(\int_0^t w_1(u) \Gamma_0^{-1}(u) \Gamma_1(u) du \right) [\tilde{\lambda}(t) + \boldsymbol{\beta}' \mathbf{z}] \exp(-\boldsymbol{\beta}' \mathbf{z} t) dt \mathbf{z}' f_Z(\mathbf{z}) d\mathbf{z} \\
& \stackrel{(7.6)}{=} \int \mathbf{w}_2(\mathbf{z}) \mathbf{z}^{\otimes 2} \int \tilde{S}(t) [\tilde{\lambda}(t) + \boldsymbol{\beta}' \mathbf{z}] \gamma_0(t) \exp(-\boldsymbol{\beta}' \mathbf{z} t) dt f_Z(\mathbf{z}) d\mathbf{z} - \\
& \quad \int \mathbf{w}_2(\mathbf{z}) \int \tilde{S}(t) \gamma_1(t) [\tilde{\lambda}(t) + \boldsymbol{\beta}' \mathbf{z}] \exp(-\boldsymbol{\beta}' \mathbf{z} t) dt \mathbf{z}' f_Z(\mathbf{z}) d\mathbf{z} \\
& = \int \mathbf{w}_2(\mathbf{z}) \mathbf{z}^{\otimes 2} \left[\int \tilde{S}(t) \tilde{\lambda}(t) \gamma_0(t) \exp(-\boldsymbol{\beta}' \mathbf{z} t) dt \right] f_Z(\mathbf{z}) d\mathbf{z} + \\
& \quad \int \mathbf{w}_2(\mathbf{z}) \mathbf{z}^{\otimes 2} \left[\int \tilde{S}(t) \gamma_0(t) \exp(-\boldsymbol{\beta}' \mathbf{z} t) (\boldsymbol{\beta}' \mathbf{z}) dt \right] f_Z(\mathbf{z}) d\mathbf{z} - \\
& \quad \int \mathbf{w}_2(\mathbf{z}) \left[\int \tilde{S}(t) \tilde{\lambda}(t) \gamma_1(t) \exp(-\boldsymbol{\beta}' \mathbf{z} t) dt \right] \mathbf{z}' f_Z(\mathbf{z}) d\mathbf{z} - \\
& \quad \int \mathbf{w}_2(\mathbf{z}) \left[\int \tilde{S}(t) \gamma_1(t) \exp(-\boldsymbol{\beta}' \mathbf{z} t) (\boldsymbol{\beta}' \mathbf{z}) dt \right] \mathbf{z}' f_Z(\mathbf{z}) d\mathbf{z} \\
& \stackrel{(7.9), (7.10), (7.14)}{=} \int \mathbf{w}_2(\mathbf{z}) \mathbf{z}^{\otimes 2} \left[\int \tilde{S}(t) \tilde{\lambda}(t) \gamma_0(t) \exp(-\boldsymbol{\beta}' \mathbf{z} t) dt \right] f_Z(\mathbf{z}) d\mathbf{z} + \\
& \quad \int \mathbf{w}_2(\mathbf{z}) \mathbf{z}^{\otimes 2} \left[\int \tilde{S}(t) w_1(t) \exp(-\boldsymbol{\beta}' \mathbf{z} t) dt - \int \tilde{S}(t) \tilde{\lambda}(t) \gamma_0(t) \exp(-\boldsymbol{\beta}' \mathbf{z} t) dt \right] f_Z(\mathbf{z}) d\mathbf{z} - \\
& \quad \int \mathbf{w}_2(\mathbf{z}) \left[\int \tilde{S}(t) \tilde{\lambda}(t) \gamma_1(t) \exp(-\boldsymbol{\beta}' \mathbf{z} t) dt \right] \mathbf{z}' f_Z(\mathbf{z}) d\mathbf{z} - \\
& \quad \int \mathbf{w}_2(\mathbf{z}) \left[\int \tilde{S}(t) w_1(t) \tilde{\mathbf{z}}(t) \exp(-\boldsymbol{\beta}' \mathbf{z} t) dt - \int \tilde{S}(t) \tilde{\lambda}(t) \gamma_1(t) \exp(-\boldsymbol{\beta}' \mathbf{z} t) dt \right] \mathbf{z}' f_Z(\mathbf{z}) d\mathbf{z} \\
& = \int \mathbf{w}_2(\mathbf{z}) \mathbf{z}^{\otimes 2} \int \tilde{S}(t) w_1(t) \exp(-\boldsymbol{\beta}' \mathbf{z} t) dt f_Z(\mathbf{z}) d\mathbf{z} - \\
& \quad \int \mathbf{w}_2(\mathbf{z}) \int \tilde{S}(t) w_1(t) \tilde{\mathbf{z}}(t) \exp(-\boldsymbol{\beta}' \mathbf{z} t) dt \mathbf{z}' f_Z(\mathbf{z}) d\mathbf{z} \\
& = \int \tilde{S}(t) w_1(t) \int \mathbf{w}_2(\mathbf{z}) [\mathbf{z}^{\otimes 2} - \tilde{\mathbf{z}}(t) \mathbf{z}'] \exp(-\boldsymbol{\beta}' \mathbf{z} t) f_Z(\mathbf{z}) d\mathbf{z} dt. \tag{7.16}
\end{aligned}$$

Por (7.15) y (7.16) se desprende fácilmente que

$$\mathcal{Q}_w(H) = [\mathcal{D}_w(H)]^{-1} [\mathcal{N}_w(H)] = \boldsymbol{\beta}, \tag{7.17}$$

como queríamos. □

Demostración del Teorema 4.2

Primero, recordemos que el correspondiente *proceso de intensidad* para los procesos de conteo $N_i(t) := I(T_i^* \leq t)$ del individuo i está dada por

$$Y_i(t)\lambda_*(t; \mathbf{z}_i) = Y_i(t) [\lambda_0(t) + \boldsymbol{\beta}'_0 \mathbf{z}_i]. \quad (7.18)$$

Por los resultados expuestos en la Sección 2.2, los procesos

$$M_i(t) := N_i(t) - \int_0^t Y_i(u) \lambda_*(u; \mathbf{z}_i) du, \quad (7.19)$$

son MCCIL. Reemplazando (7.18) en (7.19) obtenemos:

$$M_i(t) = N_i(t) - \int_0^t Y_i(u) [\lambda_0(u) + \boldsymbol{\beta}'_0 \mathbf{z}_i] du.$$

Luego

$$dM_i(t) = dN_i(t) - Y_i(t) [\lambda_0(t) + \boldsymbol{\beta}'_0 \mathbf{z}_i] dt, \quad (7.20)$$

o equivalentemente,

$$dN_i(t) - Y_i(t) \boldsymbol{\beta}'_0 \mathbf{z}_i dt = dM_i(t) + Y_i(t)\lambda_0(t) dt. \quad (7.21)$$

Recordemos también que nuestro propósito es estimar $\boldsymbol{\beta}$ de la ecuación de estimación

$$\mathbf{U}_R^{(n)}(\boldsymbol{\beta}) := \sum_{i=1}^n \int_0^\infty w_1(t) \mathbf{w}_2(\mathbf{z}_i) [\mathbf{z}_i - \tilde{\mathbf{z}}_n(t)] [dN_i(t) - Y_i(t) \boldsymbol{\beta}' \mathbf{z}_i dt] = \mathbf{0}, \quad (7.22)$$

donde $\tilde{\mathbf{z}}_n(t) \stackrel{(7.1)}{=} (\mathcal{S}_n^{(0)}(t))^{-1} \mathcal{S}_n^{(1)}(t)$.

Además, los siguientes resultados se utilizarán en esta demostración.

Teorema 7.1. Sean X_i v.a. independientes entre sí. Si $\sum \text{Var}(X_n)/b_n^2 < \infty$ con $b_n \nearrow \infty$, entonces

$$\frac{1}{b_n} \sum_{k=1}^n [X_k - E(X_k | X_1, \dots, X_{k-1})] \xrightarrow{c.s.} 0.$$

Observar que si $\{X_n : n = 0, 1, \dots\}$ es una martingala, $E(X_k | X_1, \dots, X_{k-1}) = X_{k-1}$ entonces $\frac{1}{b_n} \sum_{k=1}^n (X_k - X_{k-1}) = \frac{1}{b_n} (X_n - X_0)$.

Proposición 1. *Bajo las condiciones **A1** y **A2**, el proceso $U_R^{(n)}(\beta_0)$ es una MCCIL.*

Demostración: Primero notar que $\forall t \geq 0$,

$$\sum_{i=1}^n \mathbf{w}_2(\mathbf{z}_i)[\mathbf{z}_i - \tilde{\mathbf{z}}_n(t)]Y_i(t) = \mathbf{0}. \quad (7.23)$$

Ya que

$$\begin{aligned} \sum_{i=1}^n \mathbf{w}_2(\mathbf{z}_i)[\mathbf{z}_i - \tilde{\mathbf{z}}_n(t)]Y_i(t) &= \sum_{i=1}^n \mathbf{w}_2(\mathbf{z}_i)\mathbf{z}_i Y_i(t) - \sum_{i=1}^n \mathbf{w}_2(\mathbf{z}_i)Y_i(t) \left[(\mathcal{S}_n^{(0)}(t))^{-1} \mathcal{S}_n^{(1)}(t) \right] \\ &\stackrel{(7.1)}{=} n \mathcal{S}_n^{(1)}(t) - n \mathcal{S}_n^{(0)}(t) \left[(\mathcal{S}_n^{(0)}(t))^{-1} \mathcal{S}_n^{(1)}(t) \right] = \mathbf{0}. \end{aligned}$$

Entonces

$$\begin{aligned} U_R^{(n)}(\beta_0) &\stackrel{(7.22)}{=} \sum_{i=1}^n \int_0^\infty w_1(t) \mathbf{w}_2(\mathbf{z}_i)[\mathbf{z}_i - \tilde{\mathbf{z}}_n(t)] [dN_i(t) - Y_i(t)\beta'_0 \mathbf{z}_i dt] \\ &\stackrel{(7.21)}{=} \sum_{i=1}^n \int_0^\infty w_1(t) \mathbf{w}_2(\mathbf{z}_i)[\mathbf{z}_i - \tilde{\mathbf{z}}_n(t)] [dM_i(t) + Y_i(t)\lambda_0(t) dt] \\ &= \sum_{i=1}^n \int_0^\infty w_1(t) \mathbf{w}_2(\mathbf{z}_i)[\mathbf{z}_i - \tilde{\mathbf{z}}_n(t)] dM_i(t) + \\ &\quad \int_0^\infty w_1(t) \left(\sum_{i=1}^n \mathbf{w}_2(\mathbf{z}_i)[\mathbf{z}_i - \tilde{\mathbf{z}}_n(t)]Y_i(t) \right) \lambda_0(t) dt \\ &\stackrel{(7.23)}{=} \sum_{i=1}^n \int_0^\infty w_1(t) \mathbf{w}_2(\mathbf{z}_i)[\mathbf{z}_i - \tilde{\mathbf{z}}_n(t)] dM_i(t). \quad (7.24) \end{aligned}$$

Como $M_i(t)$ es una MCCIL y el proceso integrado $w_1(t)\mathbf{w}_2(\mathbf{z}_i)[\mathbf{z}_i - \tilde{\mathbf{z}}_n(t)]$ es predecible y localmente acotado (por las condiciones **A1** y **A2**), por el Teorema 2.5, el proceso $U_R^{(n)}(\beta_0)$ es una MCCIL, como queríamos.

□

Proposición 2. *Bajo las condiciones A1 a A4, el funcional estimado $n^{-1/2}\mathbf{U}_R^{(n)}(\boldsymbol{\beta}_0)$ converge débilmente a una distribución normal p -variada con media cero y matriz de covarianza $\tilde{\Sigma}$, donde*

$$\tilde{\Sigma} := \int_0^\infty w_1(u)^2 \left(L_0(u) \lambda_0(u) + L_1(u, \boldsymbol{\beta}_0) \right) du,$$

con

$$\begin{aligned} L_0(u) &:= \mathbb{E} \left(\left[\mathbf{w}_2(\mathbf{Z}) [\mathbf{Z} - (\mathcal{S}^{(0)}(u))^{-1} \mathcal{S}^{(1)}(u)] \right]^{\otimes 2} Y(u) \right) \\ L_1(u, \boldsymbol{\beta}_0) &:= \mathbb{E} \left(\left[\mathbf{w}_2(\mathbf{Z}) [\mathbf{Z} - (\mathcal{S}^{(0)}(u))^{-1} \mathcal{S}^{(1)}(u)] \right]^{\otimes 2} Y(u) \boldsymbol{\beta}'_0 \mathbf{Z} \right) \end{aligned}$$

Demostración: Primero expresamos a $n^{-1/2}\mathbf{U}_R^{(n)}(\boldsymbol{\beta}_0)$ de una manera conveniente, para poder utilizar el Teorema 2.6, como sigue:

$$\begin{aligned} \frac{1}{\sqrt{n}} \mathbf{U}_R^{(n)}(\boldsymbol{\beta}_0) &\stackrel{(7.24)}{=} \sum_{l=1}^n \int_0^\infty w_1(t) \mathbf{w}_2(\mathbf{z}_l) [\mathbf{z}_l - \tilde{\mathbf{z}}_n(t)] dM_l(t) \\ &= \int_0^\infty \sum_{l=1}^n \frac{1}{\sqrt{n}} w_1(t) \mathbf{w}_2(\mathbf{z}_l) [\mathbf{z}_l - \tilde{\mathbf{z}}_n(t)] dM_l(t) \\ &= \int_0^\infty \sum_{l=1}^n J_l^{(n)}(t), \end{aligned}$$

donde

$$J_l^{(n)}(t) := n^{-1/2} w_1(t) \mathbf{w}_2(\mathbf{z}_l) [\mathbf{z}_l - \tilde{\mathbf{z}}_n(t)] \quad (7.25)$$

es un proceso en \mathbb{R}^p , es decir, $J_{il}^{(n)}(t)$ es la coordenada i del integrando $n^{-1/2} w_1(t) \mathbf{w}_2(\mathbf{z}_l) [\mathbf{z}_l - \tilde{\mathbf{z}}_n(t)]$ para $i = 1, \dots, p$.

Recordemos que, por (7.18), el proceso de intensidad del proceso de conteo $N_l(t)$, está dado por $Y_l(t) [\lambda_0(t) + \boldsymbol{\beta}'_0 \mathbf{z}_l]$. Luego

$$Y_l(t) [\lambda_0(t) + \boldsymbol{\beta}'_0 \mathbf{z}_l] dt \stackrel{(7.20)}{=} dN_l(t) - dM_l(t). \quad (7.26)$$

Introducimos además la martingala $W^{(n)}(t) := (W_1^{(n)}(t), \dots, W_p^{(n)}(t))$, donde

$$W_i^{(n)}(t) := \int_0^t \sum_{l=1}^n J_{il}^{(n)}(u) dM_l^{(n)}(u).$$

Observamos que el vector $W^{(n)}(\infty) = n^{-1/2} \mathbf{U}_R^{(n)}(\boldsymbol{\beta}_0)$, que, por la Proposición 1, es una MCCIL.

Como, para $i, j = 1, \dots, p$, $l = 1, \dots, n$ y $u \geq 0$, tenemos que

$$\begin{aligned} nJ_{il}^{(n)}(u)J_{jl}^{(n)}(u) &= \left[w_1(u) \mathbf{w}_2(\mathbf{z}_l) [\mathbf{z}_l - \tilde{\mathbf{z}}_n(u)] \right]_i \left[w_1(u) \mathbf{w}_2(\mathbf{z}_l) [\mathbf{z}_l - \tilde{\mathbf{z}}_n(u)] \right]_j \\ &= w_1(u)^2 \left[\mathbf{w}_2(\mathbf{z}_l) [\mathbf{z}_l - \tilde{\mathbf{z}}_n(u)] \right]_i \left[\mathbf{w}_2(\mathbf{z}_l) [\mathbf{z}_l - \tilde{\mathbf{z}}_n(u)] \right]_j \\ &= w_1(u)^2 \left[\mathbf{w}_2(\mathbf{z}_l) [\mathbf{z}_l - \tilde{\mathbf{z}}_n(u)] \right]_{ij}^{\otimes 2} \end{aligned} \quad (7.27)$$

$$= \left[w_1(u) \mathbf{w}_2(\mathbf{z}_l) [\mathbf{z}_l - \tilde{\mathbf{z}}_n(u)] \right]_{ij}^{\otimes 2}. \quad (7.28)$$

Luego, $\forall i, j = 1, \dots, n$ y $t \geq 0$ tenemos

$$\begin{aligned} \langle W_i^{(n)}, W_j^{(n)} \rangle(t) &= \int_0^t \sum_{l=1}^n J_{il}^{(n)}(u) J_{jl}^{(n)}(u) \lambda_l^{(n)}(u) du \\ &\stackrel{(7.26)}{=} \int_0^t \sum_{l=1}^n J_{il}^{(n)}(u) J_{jl}^{(n)}(u) Y_l(u) [\lambda_0(u) + \boldsymbol{\beta}'_0 \mathbf{z}_l] du \quad (7.29) \\ &\stackrel{(7.27)}{=} \int_0^t \frac{1}{n} \sum_{l=1}^n w_1(u)^2 \left[\mathbf{w}_2(\mathbf{z}_l) [\mathbf{z}_l - \tilde{\mathbf{z}}_n(u)] \right]_{ij}^{\otimes 2} Y_l(u) \lambda_0(u) du + \\ &\quad \int_0^t \frac{1}{n} \sum_{l=1}^n w_1(u)^2 \left[\mathbf{w}_2(\mathbf{z}_l) [\mathbf{z}_l - \tilde{\mathbf{z}}_n(u)] \right]_{ij}^{\otimes 2} Y_l(u) \boldsymbol{\beta}'_0 \mathbf{z}_l du \\ &= \int_0^t w_1(u)^2 \left(\frac{1}{n} \sum_{l=1}^n \left[\mathbf{w}_2(\mathbf{z}_l) [\mathbf{z}_l - \tilde{\mathbf{z}}_n(u)] \right]_{ij}^{\otimes 2} Y_l(u) \right) \lambda_0(u) du + \\ &\quad \int_0^t w_1(u)^2 \left(\frac{1}{n} \sum_{l=1}^n \left[\mathbf{w}_2(\mathbf{z}_l) [\mathbf{z}_l - \tilde{\mathbf{z}}_n(u)] \right]_{ij}^{\otimes 2} Y_l(u) \boldsymbol{\beta}'_0 \mathbf{z}_l \right) du \\ &= \int_0^t w_1(u)^2 (L_0^{(n)}(u))_{ij} \lambda_0(u) du + \int_0^t w_1(u)^2 (L_1^{(n)}(u, \boldsymbol{\beta}_0))_{ij} du, \end{aligned}$$

donde, para $k = 0, 1$,

$$L_k^{(n)}(u) := \frac{1}{n} \sum_{l=1}^n \left[\mathbf{w}_2(\mathbf{z}_l) [\mathbf{z}_l - \tilde{\mathbf{z}}_n(u)] \right]_{ij}^{\otimes 2} Y_l(u) (\boldsymbol{\beta}'_0 \mathbf{z}_l)^{\otimes k}.$$

Luego para cada $u \in [0, t]$ obtenemos, por las condiciones **A2** y **A3**, que, cuando $n \rightarrow \infty$:

$$L_0^{(n)}(u) \xrightarrow{\text{Prob}} \mathbb{E} \left(\left[\mathbf{w}_2(\mathbf{Z}) [\mathbf{Z} - (\mathcal{S}^{(0)}(u))^{-1} \mathcal{S}^{(1)}(u)] \right]_{ij}^{\otimes 2} Y(u) \right) = L_0(u)$$

y

$$L_1^{(n)}(u, \boldsymbol{\beta}_0) \xrightarrow{\text{Prob}} \mathbb{E} \left(\left[\mathbf{w}_2(\mathbf{Z}) [\mathbf{Z} - (\mathcal{S}^{(0)}(u))^{-1} \mathcal{S}^{(1)}(u)] \right]^{\otimes 2} Y(u) \boldsymbol{\beta}'_0 \mathbf{Z} \right) = L_1(u, \boldsymbol{\beta}_0).$$

Entonces, cuando $n \rightarrow \infty$ y $\forall t \geq 0$,

$$\langle W_i^{(n)}, W_j^{(n)} \rangle (t) \xrightarrow{\text{Prob}} \Sigma_{ij}(t),$$

que es la condición (2.16) del Teorema 2.6.

Ahora necesitamos probar la condición (2.17) del Teorema 2.6, es decir, tenemos que probar que $\forall i = 1, \dots, p$ y para cualquier $\epsilon > 0$,

$$I_n := \int_0^\infty \sum_{l=1}^n \left[J_{il}^{(n)}(t) \right]^2 I \left(\left| J_{il}^{(n)}(t) \right| > \epsilon \right) \lambda_l^{(n)}(t) dt \xrightarrow{\text{Prob}} 0 \quad \text{cuando } n \rightarrow \infty.$$

Para ver esto, expresamos

$$\begin{aligned} \sqrt{n} J_{il}^{(n)}(t) &= w_1(t) [\mathbf{w}_2(\mathbf{z}_l)] [\mathbf{z}_l - \tilde{\mathbf{z}}_n(t)]_i \\ &= w_1(t) [\mathbf{w}_2(\mathbf{z}_l) \mathbf{z}_l]_i - w_1(t) [\mathbf{w}_2(\mathbf{z}_l) \tilde{\mathbf{z}}_n(t)]_i \\ &= [w_1(t) \mathbf{w}_2(\mathbf{z}_l) \mathbf{z}_l]_i - [w_1(t) \mathbf{w}_2(\mathbf{z}_l) (\mathcal{S}_n^{(0)}(t))^{-1} \mathcal{S}_n^{(1)}(t)]_i \\ &= X_{li}(t) - V_{li}(t), \end{aligned}$$

donde

$$V_{li}(t) = [w_1(t) \mathbf{w}_2(\mathbf{z}_l) (\mathcal{S}_n^{(0)}(t))^{-1} \mathcal{S}_n^{(1)}(t)]_i. \quad (7.30)$$

Ahora nos basamos en la siguiente desigualdad, que se comprueba fácilmente para cualesquiera números reales a y b ,

$$|a - b|^2 I(|a - b| > \epsilon) \leq 4|a|^2 I(|a| > \epsilon/2) + 4|b|^2 I(|b| > \epsilon/2),$$

para obtener

$$\begin{aligned} I_n &\leq 4 \int_0^\infty \sum_{l=1}^n \left[\frac{X_{li}(t)}{\sqrt{n}} \right]^2 I \left(\left| \frac{X_{li}(t)}{\sqrt{n}} \right| > \frac{\epsilon}{2} \right) \lambda_l^{(n)}(t) dt + \\ &\quad 4 \int_0^\infty \sum_{l=1}^n \left[\frac{V_{li}(t)}{\sqrt{n}} \right]^2 I \left(\left| \frac{V_{li}(t)}{\sqrt{n}} \right| > \frac{\epsilon}{2} \right) \lambda_l^{(n)}(t) dt \\ &= 4I_n^{(1)} + 4I_n^{(2)}, \end{aligned} \quad (7.31)$$

donde

$$I_n^{(1)} := \int_0^\infty \sum_{l=1}^n \left[\frac{X_{li}(t)}{\sqrt{n}} \right]^2 I \left(\left| \frac{X_{li}(t)}{\sqrt{n}} \right| > \frac{\epsilon}{2} \right) \lambda_l^{(n)}(t) dt$$

y

$$I_n^{(2)} := \int_0^\infty \sum_{l=1}^n \left[\frac{V_{li}(t)}{\sqrt{n}} \right]^2 I \left(\left| \frac{V_{li}(t)}{\sqrt{n}} \right| > \frac{\epsilon}{2} \right) \lambda_l^{(n)}(t) dt.$$

Ahora, por la condición **A2** y **A3** tenemos que, $\forall t \geq 0$,

$$(\mathcal{S}_n^{(0)}(t))^{-1} \mathcal{S}_n^{(1)}(t) \xrightarrow{\text{Prob}} (\mathcal{S}^{(0)}(t))^{-1} \mathcal{S}^{(1)}(t) \quad \text{cuando } n \rightarrow \infty.$$

Además, los tres términos $w_1(t)$, $\mathbf{w}_2(\mathbf{z}_l)$ y $(\mathcal{S}^{(0)}(t))^{-1} \mathcal{S}^{(1)}(t)$ son acotados para cualquier $t \geq 0$ y $\forall \mathbf{z}_l$. Por lo tanto para valores de n suficientemente grandes, tenemos que

$$I \left(\left| \frac{V_{li}(t)}{\sqrt{n}} \right| > \frac{\epsilon}{2} \right) \stackrel{(7.30)}{=} I \left(\left| \left[w_1(t) \mathbf{w}_2(\mathbf{z}_l) (\mathcal{S}_n^{(0)}(t))^{-1} \mathcal{S}_n^{(1)}(t) \right]_i \right| > \frac{\epsilon \sqrt{n}}{2} \right) = 0.$$

Así

$$I_n^{(2)} = \int_0^\infty \sum_{l=1}^n \left[\frac{V_{li}(t)}{\sqrt{n}} \right]^2 I \left(\left| \frac{V_{li}(t)}{\sqrt{n}} \right| > \frac{\epsilon}{2} \right) \lambda_l^{(n)}(t) dt \xrightarrow{\text{Prob}} 0, \quad \text{cuando } n \rightarrow \infty.$$

Para el primer término en (7.31), notar que

$$I_n^{(1)} \stackrel{(7.18)}{=} \int_0^\infty \sum_{l=1}^n \left[\frac{X_{li}(t)}{\sqrt{n}} \right]^2 I \left(\left| \frac{X_{li}(t)}{\sqrt{n}} \right| > \frac{\epsilon}{2} \right) Y_l(t) (\lambda_0(t) + \boldsymbol{\beta}'_0 \mathbf{z}_l) dt,$$

y se puede descomponer como $I_n^{(1)} = I_n^{(1,1)} + I_n^{(1,2)}$ siendo

$$I_n^{(1,1)} := \int_0^\infty \sum_{l=1}^n \left[\frac{X_{li}(t)}{\sqrt{n}} \right]^2 I \left(\left| \frac{X_{li}(t)}{\sqrt{n}} \right| > \frac{\epsilon}{2} \right) C_l(t) (\lambda_0(t) + \boldsymbol{\beta}'_0 \mathbf{z}_l) dt$$

y

$$I_n^{(1,2)} := \int_0^\infty \sum_{l=1}^n \left[\frac{X_{li}(t)}{\sqrt{n}} \right]^2 I \left(\left| \frac{X_{li}(t)}{\sqrt{n}} \right| > \frac{\epsilon}{2} \right) C_l^c(t) (\lambda_0(t) + \boldsymbol{\beta}'_0 \mathbf{z}_l) dt,$$

con $C_l(t) := I(\lambda_0(t) + \boldsymbol{\beta}'_0 \mathbf{z}_l > \exp(-\delta \|X_l(t)\|_\infty - t)) Y_l(t)$, $C_l^c(t) := R_l(t) Y_l(t)$ y $R_l(t) = I(\lambda_0(t) + \boldsymbol{\beta}'_0 \mathbf{z}_l \leq \exp(-\delta \|X_l(t)\|_\infty - t))$.

Considerar el conjunto $\Omega_n := \{1 \leq l \leq n : C_l(t) = 1\}$. Entonces,

$$\begin{aligned} I_n^{(1,1)} &= \int_0^\infty \sum_{l \in \Omega_n} \left[\frac{X_{li}(t)}{\sqrt{n}} \right]^2 I \left(\left| \frac{X_{li}(t)}{\sqrt{n}} \right| > \frac{\epsilon}{2} \right) C_l(t) (\lambda_0(t) + \beta'_0 \mathbf{z}_l) dt \\ &= \int_0^\infty \sum_{l \in \Omega_n} \left[\frac{X_{li}(t)C_l(t)}{\sqrt{n}} \right]^2 I \left(\left| \frac{X_{li}(t)C_l(t)}{\sqrt{n}} \right| > \frac{\epsilon}{2} \right) (\lambda_0(t) + \beta'_0 \mathbf{z}_l) dt. \end{aligned}$$

La condición **A4** implica que, $\forall i = 1, \dots, p, l = 1, \dots, n$ y $t \geq 0$

$$\frac{X_{li}(t)C_l(t)}{\sqrt{n}} \xrightarrow{\text{Prob}} 0, \quad \text{cuando } n \rightarrow \infty.$$

Por lo tanto,

$$I_n^{(1,1)} \xrightarrow{\text{Prob}} 0, \quad \text{cuando } n \rightarrow \infty.$$

Por otra parte, ya que $\|X_l(t)\|_\infty \geq |X_{li}(t)|, \forall i = 1, \dots, p$, y $Y_l(t) \leq 1$,

$$\begin{aligned} I_n^{(1,2)} &\leq \frac{1}{n} \int_0^\infty \sum_{l=1}^n X_{li}(t)^2 \exp(-\delta|X_{li}(t)| - t) I \left(|X_{li}(t)| > \frac{\sqrt{n}\epsilon}{2} \right) R_l(t) dt \\ &= \frac{1}{n} \int_0^\infty \sum_{l=1}^n X_{li}(t)^2 \exp(-\delta|X_{li}(t)|) I \left(|X_{li}(t)| > \frac{\sqrt{n}\epsilon}{2} \right) |R_l(t) \exp(-t) dt. \end{aligned}$$

Y como $\lim_{x \rightarrow \infty} x^2 \exp(-\delta|x|) = 0$, existe algún $\eta > 0$ tal que para n suficientemente grande

$$X_{li}(t)^2 \exp(-\delta|X_{li}(t)|) I \left(|X_{li}(t)| > \frac{\sqrt{n}\epsilon}{2} \right) I(\lambda_0(t) + \beta'_0 \mathbf{z}_l \leq \exp(-\delta\|X_l(t)\|_\infty - t)) \leq \eta$$

así $I_n^{(1,2)} \leq \eta$, con lo cual asegura que $I_n^{(1,2)} \xrightarrow{\text{Prob}} 0$ cuando $n \rightarrow \infty$. Y, por lo tanto, concluimos que

$$I_n^{(1)} \xrightarrow{\text{Prob}} 0, \quad \text{cuando } n \rightarrow \infty.$$

Entonces podemos afirmar que $I_n \xrightarrow{\text{Prob}} 0$ cuando $n \rightarrow \infty$, como queríamos.

Finalmente, por el Teorema 2.6, podemos asegurar que $n^{-1/2} \mathbf{U}_R^{(n)}(\beta_0)$ converge débilmente a una distribución normal p -variada con media cero y matriz de covarianza $\tilde{\Sigma}$, como queríamos. □

Corolario 2. *Bajo las mismas condiciones que la proposición anterior, la matriz de covarianza $\tilde{\Sigma}$ puede ser estimada consistentemente por*

$$B := \frac{1}{n} \sum_{l=1}^n \int_0^\infty [w_1(t) \mathbf{w}_2(\mathbf{z}_l) [\mathbf{z}_l - \tilde{\mathbf{z}}_n(t)]]^{\otimes 2} dN_l(t).$$

Demostración: Utilizando los mismos argumentos y cuentas que en la demostración anterior, tenemos que:

$$\begin{aligned} \langle W_i^{(n)}, W_j^{(n)} \rangle (t) &\stackrel{(7.29)}{=} \int_0^t \sum_{l=1}^n J_{il}^{(n)}(u) J_{jl}^{(n)}(u) Y_l(u) [\lambda_0(u) + \beta'_0 \mathbf{z}_l] du \\ &\stackrel{(7.26)}{=} \int_0^t \sum_{l=1}^n J_{il}^{(n)}(u) J_{jl}^{(n)}(u) [dN_l(u) - dM_l(u)] \\ &\stackrel{(7.28)}{=} \int_0^t \frac{1}{n} \sum_{l=1}^n [w_1(u) \mathbf{w}_2(\mathbf{z}_l) [\mathbf{z}_l - \tilde{\mathbf{z}}_n(u)]]_{ij}^{\otimes 2} [dN_l(u) - dM_l(u)] \\ &= \frac{1}{n} \sum_{l=1}^n \int_0^t [w_1(u) \mathbf{w}_2(\mathbf{z}_l) [\mathbf{z}_l - \tilde{\mathbf{z}}_n(u)]]_{ij}^{\otimes 2} dN_l(u) - \\ &\quad \frac{1}{n} \sum_{l=1}^n \int_0^t [w_1(u) \mathbf{w}_2(\mathbf{z}_l) [\mathbf{z}_l - \tilde{\mathbf{z}}_n(u)]]_{ij}^{\otimes 2} dM_l(u). \end{aligned}$$

En el segundo término, la integral es una MCCIL, por Teorema 2.5, ya que su integrando es predecible y localmente acotado (por las condiciones **A1** y **A2**). Luego por el Teorema 7.1, el segundo término converge a cero en probabilidad.

Por lo tanto, la matriz de covarianza $\tilde{\Sigma}$ puede ser estimada consistentemente por

$$B = \frac{1}{n} \sum_{l=1}^n \int_0^\infty [w_1(t) \mathbf{w}_2(\mathbf{z}_l) [\mathbf{z}_l - \tilde{\mathbf{z}}_n(t)]]^{\otimes 2} dN_l(t),$$

como queríamos. □

Ahora sí demostraremos el Teorema 4.2.

Demostración: Por un lado, la matriz

$$\begin{aligned} D_n &:= \frac{1}{n} \sum_{l=1}^n \int_0^\infty w_1(t) \mathbf{w}_2(\mathbf{z}_l) [\mathbf{z}_l - \tilde{\mathbf{z}}_n(t)]^{\otimes 2} Y_l(t) dt \\ &= \int_0^\infty w_1(t) \left[\frac{1}{n} \sum_{l=1}^n \mathbf{w}_2(\mathbf{z}_l) [\mathbf{z}_l - \tilde{\mathbf{z}}_n(t)]^{\otimes 2} Y_l(t) \right] dt =: \int_0^\infty w_1(t) L^{(n)}(t) dt \end{aligned}$$

es simétrica y definida positiva, por las condiciones **A1** y **A2**, por lo tanto es invertible.

Luego, por las condiciones **A2** y **A3**, tenemos que

$$L^{(n)}(t) \xrightarrow{\text{Prob}} \mathbb{E} \left(\mathbf{w}_2(\mathbf{Z}) [\mathbf{Z} - (\mathcal{S}^{(0)}(t))^{-1} \mathcal{S}^{(1)}(t)]^{\otimes 2} Y(t) \right) = L(t),$$

cuando $n \rightarrow \infty$.

La matriz $L(t)$ también resulta ser simétrica y definida positiva, luego es invertible.

Entonces,

$$D_n \xrightarrow{\text{Prob}} \int_0^\infty w_1(t) L(t) dt, \quad \text{cuando } n \rightarrow \infty$$

que es una constante no nula.

Por otro lado, sea $\tilde{\boldsymbol{\beta}}_n$ solución de $\mathbf{U}_R^{(n)}(\boldsymbol{\beta}) = \mathbf{0}$, que se puede expresar explícitamente como en (4.10). Sea $\tilde{J}_l^{(n)}(t) := n^{1/2} J_l^{(n)}(t)$, donde $J_l^{(n)}(t)$ está definido en (7.25). Entonces

$$\begin{aligned} \tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0 &= \left(\sum_{l=1}^n \int_0^\infty \tilde{J}_l^{(n)}(t) [\mathbf{z}_l - \tilde{\mathbf{z}}_n(t)]' Y_l(t) dt \right)^{-1} \left(\sum_{l=1}^n \int_0^\infty \tilde{J}_l^{(n)}(t) dN_l(t) \right) - \boldsymbol{\beta}_0 \\ &= \left(\sum_{l=1}^n \int_0^\infty \tilde{J}_l^{(n)}(t) [\mathbf{z}_l - \tilde{\mathbf{z}}_n(t)]' Y_l(t) dt \right)^{-1} \left[\sum_{l=1}^n \int_0^\infty \tilde{J}_l^{(n)}(t) dN_l(t) - \right. \\ &\quad \left. \sum_{l=1}^n \int_0^\infty \tilde{J}_l^{(n)}(t) [\mathbf{z}_l - \tilde{\mathbf{z}}_n(t)]' Y_l(t) dt \boldsymbol{\beta}_0 \right] \\ &= \left(\sum_{l=1}^n \int_0^\infty \tilde{J}_l^{(n)}(t) [\mathbf{z}_l - \tilde{\mathbf{z}}_n(t)]' Y_l(t) dt \right)^{-1} \times \\ &\quad \left[\sum_{l=1}^n \int_0^\infty \tilde{J}_l^{(n)}(t) \left(dN_l(t) - [\mathbf{z}_l - \tilde{\mathbf{z}}_n(t)]' Y_l(t) dt \boldsymbol{\beta}_0 \right) \right] \\ &= \left(\sum_{l=1}^n \int_0^\infty \tilde{J}_l^{(n)}(t) [\mathbf{z}_l - \tilde{\mathbf{z}}_n(t)]' Y_l(t) dt \right)^{-1} \left[\sum_{l=1}^n \int_0^\infty \tilde{J}_l^{(n)}(t) \left(dN_l(t) - \boldsymbol{\beta}_0' \mathbf{z}_l Y_l(t) dt \right) + \right. \\ &\quad \left. \sum_{l=1}^n \int_0^\infty \tilde{J}_l^{(n)}(t) \tilde{\mathbf{z}}_n(t)' Y_l(t) dt \boldsymbol{\beta}_0 \right] \end{aligned}$$

$$\begin{aligned}
& \stackrel{(7.25)}{=} \left(\sum_{l=1}^n \int_0^\infty \tilde{J}_l^{(n)}(t) [\mathbf{z}_l - \tilde{\mathbf{z}}_n(t)]' Y_l(t) dt \right)^{-1} \left[\sum_{l=1}^n \int_0^\infty \tilde{J}_l^{(n)}(t) (dN_l(t) - \boldsymbol{\beta}'_0 \mathbf{z}_l Y_l(t) dt) + \right. \\
& \quad \left. \int_0^\infty n^{1/2} w_1(t) \left(\sum_{l=1}^n \mathbf{w}_2(\mathbf{z}_l) [\mathbf{z}_l - \tilde{\mathbf{z}}_n(t)] Y_l(t) \right) \tilde{\mathbf{z}}_n(t)' dt \boldsymbol{\beta}_0 \right] \\
& \stackrel{(7.23)}{=} \left(\sum_{l=1}^n \int_0^\infty \tilde{J}_l^{(n)}(t) [\mathbf{z}_l - \tilde{\mathbf{z}}_n(t)]' Y_l(t) dt \right)^{-1} \left[\sum_{l=1}^n \int_0^\infty \tilde{J}_l^{(n)}(t) (dN_l(t) - \boldsymbol{\beta}'_0 \mathbf{z}_l Y_l(t) dt) \right] \\
& \stackrel{(7.20)}{=} \left(\sum_{l=1}^n \int_0^\infty \tilde{J}_l^{(n)}(t) [\mathbf{z}_l - \tilde{\mathbf{z}}_n(t)]' Y_l(t) dt \right)^{-1} \left[\sum_{l=1}^n \int_0^\infty w_1(t) \mathbf{w}_2(\mathbf{z}_l) [\mathbf{z}_l - \tilde{\mathbf{z}}_n(t)] dM_l(t) \right] \\
& \stackrel{(7.24)}{=} \left(\sum_{l=1}^n \int_0^\infty w_1(t) \mathbf{w}_2(\mathbf{z}_l) [\mathbf{z}_l - \tilde{\mathbf{z}}_n(t)]^{\otimes 2} Y_l(t) dt \right)^{-1} \mathbf{U}_R^{(n)}(\boldsymbol{\beta}_0) \\
& = \left(\frac{1}{n} \sum_{l=1}^n \int_0^\infty w_1(t) \mathbf{w}_2(\mathbf{z}_l) [\mathbf{z}_l - \tilde{\mathbf{z}}_n(t)]^{\otimes 2} Y_l(t) dt \right)^{-1} \frac{\mathbf{U}_R^{(n)}(\boldsymbol{\beta}_0)}{n} \\
& = D_n^{-1} \frac{1}{n} \mathbf{U}_R^{(n)}(\boldsymbol{\beta}_0).
\end{aligned}$$

Por lo tanto,

$$\sqrt{n}(\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) = D_n^{-1} n^{-1/2} \mathbf{U}_R^{(n)}(\boldsymbol{\beta}_0)$$

y, por la Proposición 2, converge débilmente a una distribución normal p -variada con vector de media nula y matriz de covarianza

$$\Sigma = \left[\int_0^\infty w_1(u) L(u) du \right]^{-1} \tilde{\Sigma} \left[\int_0^\infty w_1(u) L(u) du \right]^{-1},$$

como queríamos. □

Demostración del Corolario 1

Demostración: Por el Corolario 2, la matriz Σ puede ser estimada consistentemente por $D_n^{-1} B D_n^{-1}$, como queríamos. □

Demostración del Teorema 4.3

Recordemos nuevamente que si $H \in \mathcal{A}_\beta$, con $\beta \in \mathbb{R}^+$, entonces

$$dH(t, z, \Delta) = \tilde{S}(t) \exp(-\beta z t) (\lambda_0(t) + \beta z) f_Z(z) dz dt \quad (7.32)$$

y

$$dH(t, z) = \tilde{S}(t) \exp(-\beta z t) [\tilde{\lambda}(t) + \beta z] f_Z(z) dz dt, \quad (7.33)$$

donde $\tilde{S}(t) = S_0(t) S_C(t)$ y $\tilde{\lambda}(t) = \lambda_0(t) + \lambda_C(t)$.

Además, recordemos que una distribución contaminada de H es H_ε , tal que:

$$H_\varepsilon(t, z, \Delta) \stackrel{(4.14)}{=} H(t, z, \Delta) + \varepsilon [\delta_{(t_0, z_0, \Delta_0)}(t, z, \Delta) - H(t, z, \Delta)]$$

y

$$H_\varepsilon(t, z) := (1 - \varepsilon)H(t, z) + \varepsilon \delta_{(t_0, z_0)}(t, z) = H(t, z) + \varepsilon [\delta_{(t_0, z_0)}(t, z) - H(t, z)]. \quad (7.34)$$

Observar que cuando $\varepsilon = 0$, $H_0(t, z, \Delta) = H(t, z, \Delta)$ y $H_0(t, z) = H(t, z)$.

Y, por último, recordemos la definición de $\Upsilon_l(u)$ para $l = 0, 1, 2, 3$:

$$\Upsilon_l(u) := \frac{\mathbb{E} [w_2(Z) \exp(-\beta Z u) Z^l]}{\mathbb{E} [w_2(Z)]} =: \mathbb{E}_w [\exp(-\beta u Z) Z^l], \quad (7.35)$$

donde \mathbb{E}_w denota la esperanza de la distribución pesada de Z , extendemos la definición de (7.6) para $l = 0, 1, 2$:

$$\gamma_l(t) := \int_0^t w_1(u) \left(\frac{\Upsilon_1(u)}{\Upsilon_0(u)} \right)^l du, \quad (7.36)$$

y, finalmente,

$$\nu(z) := w_2(z) / \mathbb{E} [w_2(Z)]. \quad (7.37)$$

Observar que, para $l = 0, 1, 2$, $\Upsilon_l(u) = \Gamma_l(u) / \mathbb{E} [w_2(Z)]$, por (7.4) y cuando $p = 1$.

Por último recordemos que

$$\mathcal{Q}_w(H_\varepsilon) = \frac{\mathcal{N}_w(H_\varepsilon)}{\mathcal{D}_w(H_\varepsilon)},$$

donde

$$\mathcal{N}_w(H_\varepsilon) := \iint w_1(t)w_2(z)[z - \tilde{z}_\varepsilon(t)] dH_\varepsilon(t, z, \Delta = 1)$$

y

$$\mathcal{D}_w(H_\varepsilon) := \iint w_2(z) \int_0^t w_1(u)[z - \tilde{z}_\varepsilon(u)]^2 du dH_\varepsilon(t, z)$$

con

$$\tilde{z}_\varepsilon(u) = \frac{\iint I(\tilde{t} \geq u)w_2(x)x dH_\varepsilon(\tilde{t}, x)}{\iint I(\tilde{t} \geq u)w_2(x) dH_\varepsilon(\tilde{t}, x)}.$$

Observar que $\tilde{z}_0(u) = \tilde{z}(u)$.

Antes de comenzar con la demostración hagamos unas cuentas previas.

Para $l = 0, 1, 2, 3$, tenemos que:

$$\frac{d}{du} \Upsilon_l(u) = \frac{d}{du} \mathbb{E}_w[\exp(-\beta u Z) Z^l] = -\beta \mathbb{E}_w[Z \exp(-\beta u Z) Z^l] = -\beta \Upsilon_{l+1}(u). \quad (7.38)$$

Para $l = 0, 1, 2$, tenemos que:

$$\frac{d}{du} \gamma_l(u) = \frac{d}{du} \int_0^u w_1(t) \left(\frac{\Upsilon_1(t)}{\Upsilon_0(t)} \right)^l dt = w_1(u) \left(\frac{\Upsilon_1(u)}{\Upsilon_0(u)} \right)^l. \quad (7.39)$$

El resultado obtenido en (7.13) es válido en particular cuando $p = 1$, es decir, tenemos que

$$\iint I(t \geq u)w_2(z)z^l dH(t, z) = \tilde{S}(u) \Gamma_l(u), \quad \text{para } l = 0, 1.$$

Por lo tanto,

$$\iint I(t \geq u)w_2(z)z^l dH(t, z) = \mathbb{E}[w_2(Z)] \tilde{S}(u) \Upsilon_l(u). \quad (7.40)$$

Luego, para $l = 0, 1$, tenemos que:

$$\begin{aligned} \iint I(s \geq u)w_2(x)x^l dH_\varepsilon(s, x) &\stackrel{(7.34)}{=} \iint I(s \geq u)w_2(x)x^l d[H(s, x) + \varepsilon(\delta_{(t_0, z_0)}(s, x) - H(s, x))] \\ &= \iint I(s \geq u)w_2(x)x^l dH(s, x) + \varepsilon \left[I(t_0 \geq u)w_2(z_0)z_0^l - \right. \\ &\quad \left. \iint I(s \geq u)w_2(x)x^l dH(s, x) \right] \\ &\stackrel{(7.40)}{=} \mathbb{E}[w_2(Z)] \tilde{S}(u) \Upsilon_l(u) + \varepsilon \left[I(t_0 \geq u)w_2(z_0)z_0^l - \right. \\ &\quad \left. \mathbb{E}[w_2(Z)] \tilde{S}(u) \Upsilon_l(u) \right]. \end{aligned}$$

Entonces

$$\begin{aligned} \frac{\iint I(s \geq u) w_2(x) x^l dH_\varepsilon(s, x)}{\mathbb{E}[w_2(Z)]} &\stackrel{(7.37)}{=} \tilde{S}(u) \Upsilon_l(u) + \varepsilon \left[I(t_0 \geq u) \nu(z_0) z_0^l - \tilde{S}(u) \Upsilon_l(u) \right] \\ &=: v_l(\varepsilon). \end{aligned} \quad (7.41)$$

Por lo tanto

$$\tilde{z}_\varepsilon(u) = \frac{\iint I(s \geq u) w_2(x) x dH_\varepsilon(s, x)}{\iint I(s \geq u) w_2(x) dH_\varepsilon(s, x)} \stackrel{(7.41)}{=} \frac{v_1(\varepsilon)}{v_0(\varepsilon)}.$$

En particular, evaluando en $\varepsilon = 0$, $v_l(0) = \tilde{S}(u) \Upsilon_l(u)$, y entonces

$$\tilde{z}(u) = \frac{\Upsilon_1(u)}{\Upsilon_0(u)}. \quad (7.42)$$

Por otro lado,

$$\frac{\partial}{\partial \varepsilon} v_l(\varepsilon) = I(t_0 \geq u) \nu(z_0) z_0^l - \tilde{S}(u) \Upsilon_l(u),$$

entonces

$$\begin{aligned} \tilde{z}'(u) &:= \left. \frac{\partial}{\partial \varepsilon} \tilde{z}_\varepsilon(u) \right|_{\varepsilon=0} = \left. \frac{\partial}{\partial \varepsilon} \left(\frac{v_1(\varepsilon)}{v_0(\varepsilon)} \right) \right|_{\varepsilon=0} \\ &= \frac{\left[I_0(u) \nu(z_0) z_0 - \tilde{S}(u) \Upsilon_1(u) \right] \tilde{S}(u) \Upsilon_0(u) - \tilde{S}(u) \Upsilon_1(u) \left[I_0(u) \nu(z_0) - \tilde{S}(u) \Upsilon_0(u) \right]}{\left[\tilde{S}(u) \Upsilon_0(u) \right]^2} \\ &= \frac{I_0(u) \nu(z_0) z_0 \Upsilon_0(u) - I_0(u) \nu(z_0) \Upsilon_1(u)}{\tilde{S}(u) \Upsilon_0(u)^2} \stackrel{(7.42)}{=} I_0(u) \nu(z_0) \frac{[z_0 - \tilde{z}(u)]}{\tilde{S}(u) \Upsilon_0(u)}, \end{aligned} \quad (7.43)$$

donde $I_0(u) := I(t_0 \geq u)$.

Sea FT una función que solo depende de t . Entonces, para $l = 0, 1$,

$$\begin{aligned} \iint FT(t) w_2(z) z^l dH(t, z, \Delta = 1) &\stackrel{(7.32)}{=} \iint FT(t) w_2(z) z^l \tilde{S}(t) \exp(-\beta z t) (\lambda_0(t) + \beta z) f_Z(z) dz dt \\ &= \int FT(t) \tilde{S}(t) \left[\lambda_0(t) \int w_2(z) \exp(-\beta z t) z^l f_Z(z) dz + \right. \\ &\quad \left. \beta \int w_2(z) \exp(-\beta z t) z^{l+1} f_Z(z) dz \right] dt \\ &\stackrel{(7.35)}{=} \mathbb{E}[w_2(Z)] \int FT(t) \tilde{S}(t) [\lambda_0(t) \Upsilon_l(t) + \beta \Upsilon_{l+1}(t)] dt. \end{aligned}$$

Entonces, en particular, tenemos que:

\hookrightarrow si $FT(t) = w_1(t)$ y $l = 1$:

$$\begin{aligned} T_1 &:= \iint w_1(t) w_2(z) z dH(t, z, \Delta = 1) \\ &= \mathbb{E}[w_2(Z)] \int w_1(t) \tilde{S}(t) \left[\lambda_0(t) \Upsilon_1(t) + \beta \Upsilon_2(t) \right] dt. \end{aligned} \quad (7.44)$$

\hookrightarrow si $FT(t) = w_1(t) \tilde{z}(t)$ y $l = 0$:

$$\begin{aligned} T_2 &:= \iint w_1(t) \tilde{z}(t) w_2(z) dH(t, z, \Delta = 1) \\ &= \mathbb{E}[w_2(Z)] \int w_1(t) \tilde{z}(t) \tilde{S}(t) \left[\lambda_0(t) \Upsilon_0(t) + \beta \Upsilon_1(t) \right] dt \\ &\stackrel{(7.42)}{=} \mathbb{E}[w_2(Z)] \int w_1(t) \tilde{S}(t) \left[\lambda_0(t) \Upsilon_1(t) + \beta \frac{\Upsilon_1(t)^2}{\Upsilon_0(t)} \right] dt. \end{aligned} \quad (7.45)$$

\hookrightarrow si $FT(t) = w_1(t) \frac{I_0(t) \nu(z_0) z_0}{\tilde{S}(t) \Upsilon_0(t)}$ y $l = 0$:

$$\begin{aligned} T_3 &:= \iint \frac{w_1(t) w_2(z) I_0(t) \nu(z_0) z_0}{\tilde{S}(t) \Upsilon_0(t)} dH(t, z, \Delta = 1) \\ &= \mathbb{E}[w_2(Z)] \int \frac{w_1(t) I_0(t) \nu(z_0) z_0 \tilde{S}(t)}{\tilde{S}(t) \Upsilon_0(t)} \left[\lambda_0(t) \Upsilon_0(t) + \beta \Upsilon_1(t) \right] dt \\ &= \mathbb{E}[w_2(Z)] \nu(z_0) z_0 \int_0^{t_0} w_1(t) \left[\lambda_0(t) + \beta \frac{\Upsilon_1(t)}{\Upsilon_0(t)} \right] dt \\ &= \mathbb{E}[w_2(Z)] \nu(z_0) z_0 \left[\int_0^{t_0} w_1(t) \lambda_0(t) dt + \beta \int_0^{t_0} w_1(t) \frac{\Upsilon_1(t)}{\Upsilon_0(t)} dt \right] \\ &\stackrel{(7.36)}{=} \mathbb{E}[w_2(Z)] \nu(z_0) z_0 \left[\int_0^{t_0} w_1(t) \lambda_0(t) dt + \beta \gamma_1(t_0) \right]. \end{aligned} \quad (7.46)$$

\hookrightarrow si $FT(t) = w_1(t) \frac{I_0(t) \nu(z_0) \tilde{z}(t)}{\tilde{S}(t) \Upsilon_0(t)}$ y $l = 0$:

$$\begin{aligned} T_4 &:= \iint \frac{w_1(t) w_2(z) I_0(t) \nu(z_0) \tilde{z}(t)}{\tilde{S}(t) \Upsilon_0(t)} dH(t, z, \Delta = 1) \\ &= \mathbb{E}[w_2(Z)] \int \frac{w_1(t) I_0(t) \nu(z_0) \tilde{z}(t) \tilde{S}(t)}{\tilde{S}(t) \Upsilon_0(t)} \left[\lambda_0(t) \Upsilon_0(t) + \beta \Upsilon_1(t) \right] dt \\ &= \mathbb{E}[w_2(Z)] \nu(z_0) \int_0^{t_0} w_1(t) \tilde{z}(t) \left[\lambda_0(t) + \beta \frac{\Upsilon_1(t)}{\Upsilon_0(t)} \right] dt \\ &\stackrel{(7.42)}{=} \mathbb{E}[w_2(Z)] \nu(z_0) \int_0^{t_0} w_1(t) \frac{\Upsilon_1(t)}{\Upsilon_0(t)} \left[\lambda_0(t) + \beta \frac{\Upsilon_1(t)}{\Upsilon_0(t)} \right] dt \end{aligned}$$

$$\begin{aligned}
&= \mathbb{E}[w_2(Z)]\nu(z_0) \left[\int_0^{t_0} w_1(t) \lambda_0(t) \frac{\Upsilon_1(t)}{\Upsilon_0(t)} dt + \beta \int_0^{t_0} w_1(t) \left(\frac{\Upsilon_1(t)}{\Upsilon_0(t)} \right)^2 dt \right] \\
&\stackrel{(7.36)}{=} \mathbb{E}[w_2(Z)]\nu(z_0) \left[\int_0^{t_0} w_1(t) \lambda_0(t) \frac{\Upsilon_1(t)}{\Upsilon_0(t)} dt + \beta \gamma_2(t_0) \right]. \tag{7.47}
\end{aligned}$$

Luego, restando (7.44) y (7.45), obtenemos:

$$\begin{aligned}
T_1 - T_2 &= \iint w_1(t) w_2(z) [z - \tilde{z}(t)] dH(t, z, \Delta = 1) \\
&= \mathbb{E}[w_2(Z)] \int w_1(t) \tilde{S}(t) \left[\lambda_0(t) \Upsilon_1(t) + \beta \Upsilon_2(t) - \lambda_0(t) \Upsilon_1(t) - \beta \frac{\Upsilon_1(t)^2}{\Upsilon_0(t)} \right] dt \\
&= \beta \mathbb{E}[w_2(Z)] \int w_1(t) \tilde{S}(t) \left[\Upsilon_2(t) - \frac{\Upsilon_1(t)^2}{\Upsilon_0(t)} \right] dt, \tag{7.48}
\end{aligned}$$

y, restando (7.46) y (7.47):

$$\begin{aligned}
T_3 - T_4 &= \iint \frac{w_1(t) w_2(z) I_0(t) \nu(z_0) [z_0 - \tilde{z}(t)]}{\tilde{S}(t) \Upsilon_0(t)} dH(t, z, \Delta = 1) \\
&= \mathbb{E}[w_2(Z)]\nu(z_0) \left[z_0 \int_0^{t_0} w_1(t) \lambda_0(t) dt + z_0 \beta \gamma_1(t_0) - \int_0^{t_0} w_1(t) \lambda_0(t) \frac{\Upsilon_1(t)}{\Upsilon_0(t)} dt - \beta \gamma_2(t_0) \right] \\
&= \mathbb{E}[w_2(Z)]\nu(z_0) \left[\int_0^{t_0} w_1(t) \lambda_0(t) \left[z_0 - \frac{\Upsilon_1(t)}{\Upsilon_0(t)} \right] dt + \beta (z_0 \gamma_1(t_0) - \gamma_2(t_0)) \right]. \tag{7.49}
\end{aligned}$$

Además, por un lado, por (7.38), para $l = 0, 1, 2$ tenemos que:

$$\beta \int \tilde{S}(t) \gamma_l(t) \Upsilon_{3-l}(t) dt = - \int \tilde{S}(t) \gamma_l(t) d\Upsilon_{2-l}(t).$$

Por el otro lado, integrando por partes esta última integral, tenemos que:

$$- \int \tilde{S}(t) \gamma_l(t) d\Upsilon_{2-l}(t) = \int \Upsilon_{2-l}(t) \left[\frac{d}{dt} \tilde{S}(t) \gamma_l(t) + \tilde{S}(t) \frac{d}{dt} \gamma_l(t) \right] dt \tag{7.50}$$

$$\stackrel{(7.8), (7.39)}{=} \int \Upsilon_{2-l}(t) \left[-\tilde{S}(t) \tilde{\lambda}(t) \gamma_l(t) + \tilde{S}(t) w_1(t) \left(\frac{\Upsilon_1(t)}{\Upsilon_0(t)} \right)^l \right] dt$$

$$\stackrel{(7.42)}{=} \int \tilde{S}(t) \Upsilon_{2-l}(t) \left[w_1(t) \tilde{z}(t)^l - \tilde{\lambda}(t) \gamma_l(t) \right] dt, \tag{7.51}$$

en (7.50) el primer término se anula por las condiciones **A1** y **A2**, por (7.5) y (7.7).

Ahora, sea $FU(u)$ una función de u y $l = 0, 1, 2$, entonces:

$$\begin{aligned}
\iint w_2(z) \int_0^t FU(u) z^l du dH(t, z) &= \iint_0^t FU(u) du \int w_2(z) z^l dH(t, z) \\
&\stackrel{(7.33)}{=} \iint_0^t FU(u) du \int w_2(z) z^l \tilde{S}(t) \exp(-\beta zt) [\tilde{\lambda}(t) + \beta z] f_Z(z) dz dt \\
&= \iint_0^t FU(u) du \tilde{S}(t) \left[\tilde{\lambda}(t) \int w_2(z) z^l \exp(-\beta zt) f_Z(z) dz + \right. \\
&\quad \left. \beta \int w_2(z) z^{l+1} \exp(-\beta zt) f_Z(z) dz \right] dt \\
&\stackrel{(7.35)}{=} \mathbb{E}[w_2(Z)] \int \left(\int_0^t FU(u) du \right) \tilde{S}(t) [\tilde{\lambda}(t) \Upsilon_l(t) + \beta \Upsilon_{l+1}(t)] dt
\end{aligned}$$

En particular tenemos que:

\hookrightarrow si $FU(u) = w_1(u)$ y $l = 2$:

$$\begin{aligned}
T_5 &:= \iint w_2(z) \int_0^t w_1(u) z^2 du dH(t, z) \\
&= \mathbb{E}[w_2(Z)] \int \left(\int_0^t w_1(u) du \right) \tilde{S}(t) [\tilde{\lambda}(t) \Upsilon_2(t) + \beta \Upsilon_3(t)] dt \\
&\stackrel{(7.36)}{=} \mathbb{E}[w_2(Z)] \int \gamma_0(t) \tilde{S}(t) [\tilde{\lambda}(t) \Upsilon_2(t) + \beta \Upsilon_3(t)] dt \\
&= \mathbb{E}[w_2(Z)] \left[\int \gamma_0(t) \tilde{S}(t) \tilde{\lambda}(t) \Upsilon_2(t) dt + \beta \int \tilde{S}(t) \gamma_0(t) \Upsilon_3(t) dt \right] \\
&\stackrel{(7.38)}{=} \mathbb{E}[w_2(Z)] \left[\int \gamma_0(t) \tilde{S}(t) \tilde{\lambda}(t) \Upsilon_2(t) dt - \int \tilde{S}(t) \gamma_0(t) d\Upsilon_2(t) \right] \\
&\stackrel{(7.51)}{=} \mathbb{E}[w_2(Z)] \left[\int \tilde{S}(t) \Upsilon_2(t) \tilde{\lambda}(t) \gamma_0(t) dt + \int \tilde{S}(t) \Upsilon_2(t) [w_1(t) - \tilde{\lambda}(t) \gamma_0(t)] dt \right] \\
&= \mathbb{E}[w_2(Z)] \int \tilde{S}(t) \Upsilon_2(t) w_1(t) dt. \tag{7.52}
\end{aligned}$$

\hookrightarrow si $FU(u) = w_1(u) \tilde{z}(u)$ y $l = 1$:

$$\begin{aligned}
T_6 &:= \iint w_2(z) \int_0^t w_1(u) z \tilde{z}(u) du dH(t, z) \\
&= \mathbb{E}[w_2(Z)] \int \left(\int_0^t w_1(u) \tilde{z}(u) du \right) \tilde{S}(t) [\tilde{\lambda}(t) \Upsilon_1(t) + \beta \Upsilon_2(t)] dt \\
&\stackrel{(7.14)}{=} \mathbb{E}[w_2(Z)] \int \left(\int_0^t w_1(u) \frac{\Upsilon_1(u)}{\Upsilon_0(u)} du \right) \tilde{S}(t) [\tilde{\lambda}(t) \Upsilon_1(t) + \beta \Upsilon_2(t)] dt
\end{aligned}$$

$$\begin{aligned}
& \stackrel{(7.36)}{=} \mathbb{E} [w_2(Z)] \int \gamma_1(t) \tilde{S}(t) \left[\tilde{\lambda}(t) \Upsilon_1(t) + \beta \Upsilon_2(t) \right] dt \\
& = \mathbb{E} [w_2(Z)] \left[\int \gamma_1(t) \tilde{S}(t) \tilde{\lambda}(t) \Upsilon_1(t) dt + \beta \int \tilde{S}(t) \gamma_1(t) \Upsilon_2(t) dt \right] \\
& \stackrel{(7.38)}{=} \mathbb{E} [w_2(Z)] \left[\int \gamma_1(t) \tilde{S}(t) \tilde{\lambda}(t) \Upsilon_1(t) dt - \int \tilde{S}(t) \gamma_1(t) d\Upsilon_1(t) \right] \\
& \stackrel{(7.51)}{=} \mathbb{E} [w_2(Z)] \left[\int \tilde{S}(t) \Upsilon_1(t) \gamma_1(t) \tilde{\lambda}(t) dt + \int \tilde{S}(t) \Upsilon_1(t) \left[w_1(t) \tilde{z}(t) - \tilde{\lambda}(t) \gamma_1(t) \right] dt \right] \\
& = \mathbb{E} [w_2(Z)] \int \tilde{S}(t) \Upsilon_1(t) w_1(t) \tilde{z}(t) dt \stackrel{(7.42)}{=} \mathbb{E} [w_2(Z)] \int \tilde{S}(t) \Upsilon_1(t) w_1(t) \frac{\Upsilon_1(t)}{\Upsilon_0(t)} dt \\
& = \mathbb{E} [w_2(Z)] \int \tilde{S}(t) w_1(t) \frac{\Upsilon_1(t)^2}{\Upsilon_0(t)} dt. \tag{7.53}
\end{aligned}$$

\hookrightarrow si $FU(u) = w_1(u) \tilde{z}(u)^2$ y $l = 0$:

$$\begin{aligned}
T_7 & := \iint w_2(z) \int_0^t w_1(u) \tilde{z}(u)^2 du dH(t, z) \\
& = \mathbb{E} [w_2(Z)] \int \left(\int_0^t w_1(u) \tilde{z}(u)^2 du \right) \tilde{S}(t) \left[\tilde{\lambda}(t) \Upsilon_0(t) + \beta \Upsilon_1(t) \right] dt \\
& \stackrel{(7.14)}{=} \mathbb{E} [w_2(Z)] \int \left(\int_0^t w_1(u) \frac{\Upsilon_1(u)^2}{\Upsilon_0(u)} du \right) \tilde{S}(t) \left[\tilde{\lambda}(t) \Upsilon_0(t) + \beta \Upsilon_1(t) \right] dt \\
& \stackrel{(7.36)}{=} \mathbb{E} [w_2(Z)] \int \gamma_2(t) \tilde{S}(t) \left[\tilde{\lambda}(t) \Upsilon_0(t) + \beta \Upsilon_1(t) \right] dt \\
& = \mathbb{E} [w_2(Z)] \left[\int \gamma_2(t) \tilde{S}(t) \tilde{\lambda}(t) \Upsilon_0(t) dt + \beta \int \tilde{S}(t) \gamma_2(t) \Upsilon_1(t) dt \right] \\
& \stackrel{(7.38)}{=} \mathbb{E} [w_2(Z)] \left[\int \gamma_2(t) \tilde{S}(t) \tilde{\lambda}(t) \Upsilon_0(t) dt - \int \tilde{S}(t) \gamma_2(t) d\Upsilon_0(t) \right] \\
& \stackrel{(7.51)}{=} \mathbb{E} [w_2(Z)] \left[\int \gamma_2(t) \tilde{S}(t) \tilde{\lambda}(t) \Upsilon_0(t) dt + \int \tilde{S}(t) \Upsilon_0(t) \left[w_1(t) \tilde{z}(t)^2 - \tilde{\lambda}(t) \gamma_2(t) \right] dt \right] \\
& = \mathbb{E} [w_2(Z)] \int \tilde{S}(t) \Upsilon_0(t) w_1(t) \tilde{z}(t)^2 dt \\
& \stackrel{(7.42)}{=} \mathbb{E} [w_2(Z)] \int \tilde{S}(t) \Upsilon_0(t) w_1(t) \left(\frac{\Upsilon_1(t)}{\Upsilon_0(t)} \right)^2 dt \\
& = \mathbb{E} [w_2(Z)] \int \tilde{S}(t) w_1(t) \frac{\Upsilon_1(t)^2}{\Upsilon_0(t)} dt. \tag{7.54}
\end{aligned}$$

Luego, por (7.52), (7.53) y (7.54)

$$\begin{aligned}
T_5 - 2 T_6 + T_7 &= \iint w_2(z) \int_0^t w_1(u) [z^2 - 2 z \tilde{z}(u) + \tilde{z}(u)^2] du dH(t, z) \\
&= \mathbb{E} [w_2(Z)] \int \tilde{S}(t) w_1(t) \left\{ \Upsilon_2(t) - 2 \frac{\Upsilon_1(t)^2}{\Upsilon_0(t)} + \frac{\Upsilon_1(t)^2}{\Upsilon_0(t)} \right\} dt \\
&= \mathbb{E} [w_2(Z)] \int \tilde{S}(t) w_1(t) \left\{ \Upsilon_2(t) - \frac{\Upsilon_1(t)^2}{\Upsilon_0(t)} \right\} dt \quad (7.55)
\end{aligned}$$

Ahora sí, la demostración.

Demostración: La función de Influencia es

$$\begin{aligned}
\mathcal{IF}_w(t_0, z_0, \Delta_0, \beta) &:= \left. \frac{\partial}{\partial \varepsilon} \mathcal{Q}_w(H_\varepsilon) \right|_{\varepsilon=0} = \left. \frac{\mathcal{N}'_w(H_\varepsilon) \mathcal{D}_w(H_\varepsilon) - \mathcal{N}_w(H_\varepsilon) \mathcal{D}'_w(H_\varepsilon)}{\mathcal{D}_w^2(H_\varepsilon)} \right|_{\varepsilon=0} \\
&= \frac{\mathcal{N}' \mathcal{D} - \mathcal{N} \mathcal{D}'}{\mathcal{D}^2} = \frac{\mathcal{N}' \mathcal{N}}{\mathcal{N} \mathcal{D}} - \frac{\mathcal{N} \mathcal{D}'}{\mathcal{D} \mathcal{D}} = \left[\frac{\mathcal{N}'}{\mathcal{N}} - \frac{\mathcal{D}'}{\mathcal{D}} \right] \frac{\mathcal{N}}{\mathcal{D}} \\
&= \left[\frac{\mathcal{N}'}{\mathcal{N}} - \frac{\mathcal{D}'}{\mathcal{D}} \right] \mathcal{Q}_w(H) \stackrel{(7.17)}{=} \left[\frac{\mathcal{N}'}{\mathcal{N}} - \frac{\mathcal{D}'}{\mathcal{D}} \right] \beta
\end{aligned}$$

donde $\mathcal{N}'_w(H_\varepsilon) := \partial \mathcal{N}_w(H_\varepsilon) / \partial \varepsilon$, $\mathcal{N}' := \mathcal{N}'_w(H)$, $\mathcal{N} := \mathcal{N}_w(H)$, $\mathcal{D}'_w(H_\varepsilon) := \partial \mathcal{D}_w(H_\varepsilon) / \partial \varepsilon$, $\mathcal{D}' := \mathcal{D}'_w(H)$ y $\mathcal{D} := \mathcal{D}_w(H)$.

Por un lado, tenemos que

$$\begin{aligned}
\mathcal{N} &= \iint w_1(t) w_2(z) [z - \tilde{z}(t)] dH(t, z, \Delta = 1) \\
&\stackrel{(7.48)}{=} \beta \mathbb{E} [w_2(Z)] \int w_1(t) \tilde{S}(t) \left[\Upsilon_2(t) - \frac{\Upsilon_1(t)^2}{\Upsilon_0(t)} \right] dt.
\end{aligned}$$

Luego

$$\begin{aligned}
\mathcal{N}' &= \left. \frac{\partial}{\partial \varepsilon} \left(\mathcal{N}_w(H_\varepsilon) \right) \right|_{\varepsilon=0} = \left. \frac{\partial}{\partial \varepsilon} \left(\iint w_1(t) w_2(z) [z - \tilde{z}_\varepsilon(t)] dH_\varepsilon(t, z, \Delta = 1) \right) \right|_{\varepsilon=0} \\
&= - \iint w_1(t) w_2(z) \tilde{z}'(t) dH(t, z, \Delta = 1) + \Delta_0 w_1(t_0) w_2(z_0) [z_0 - \tilde{z}(t_0)] - \\
&\quad \iint w_1(t) w_2(z) [z - \tilde{z}(t)] dH(t, z, \Delta = 1) \\
&\stackrel{(7.42)}{=} \Delta_0 w_1(t_0) w_2(z_0) \left[z_0 - \frac{\Upsilon_1(t_0)}{\Upsilon_0(t_0)} \right] - \mathcal{N} - \iint w_1(t) w_2(z) \tilde{z}'(t) dH(t, z, \Delta = 1) \\
&\stackrel{(7.43)}{=} \Delta_0 w_1(t_0) w_2(z_0) \left[z_0 - \frac{\Upsilon_1(t_0)}{\Upsilon_0(t_0)} \right] - \mathcal{N} -
\end{aligned}$$

$$\begin{aligned}
& \iint w_1(t) w_2(z) I_0(t) \nu(z_0) \frac{[z_0 - \tilde{z}(t)]}{\tilde{S}(t) \Upsilon_0(t)} dH(t, z, \Delta = 1) \\
(7.49) \quad & \stackrel{=}{=} -\mathcal{N} + \Delta_0 w_1(t_0) w_2(z_0) \left[z_0 - \frac{\Upsilon_1(t_0)}{\Upsilon_0(t_0)} \right] - \\
& \mathbb{E}[w_2(Z)] \nu(z_0) \left[\int_0^{t_0} w_1(t) \lambda_0(t) \left[z_0 - \frac{\Upsilon_1(t)}{\Upsilon_0(t)} \right] dt + \beta(z_0 \gamma_1(t_0) - \gamma_2(t_0)) \right] \\
(7.37) \quad & \stackrel{=}{=} \mathbb{E}[w_2(Z)] \nu(z_0) \left[\Delta_0 w_1(t_0) \left[z_0 - \frac{\Upsilon_1(t_0)}{\Upsilon_0(t_0)} \right] - \int_0^{t_0} w_1(t) \lambda_0(t) \left[z_0 - \frac{\Upsilon_1(t)}{\Upsilon_0(t)} \right] dt - \right. \\
& \left. \beta(z_0 \gamma_1(t_0) - \gamma_2(t_0)) \right] - \mathcal{N}.
\end{aligned}$$

Por lo tanto, $\frac{\mathcal{N}'}{\mathcal{N}}$ es igual a

$$\frac{\nu(z_0) \left[\Delta_0 w_1(t_0) \left[z_0 - \frac{\Upsilon_1(t_0)}{\Upsilon_0(t_0)} \right] - \int_0^{t_0} w_1(t) \lambda_0(t) \left[z_0 - \frac{\Upsilon_1(t)}{\Upsilon_0(t)} \right] dt - \beta(z_0 \gamma_1(t_0) - \gamma_2(t_0)) \right]}{\beta \int w_1(t) \tilde{S}(t) \left(\Upsilon_2(t) - \frac{\Upsilon_1(t)^2}{\Upsilon_0(t)} \right) dt} - 1.$$

Por otro lado, tenemos que

$$\begin{aligned}
\mathcal{D} &= \iint w_2(z) \left(\int_0^t w_1(u) [z - \tilde{z}(u)]^2 du \right) dH(t, z) \\
&= \iint w_2(z) \left(\int_0^t w_1(u) [z^2 - 2z\tilde{z}(u) + \tilde{z}(u)^2] du \right) dH(t, z) \\
(7.55) \quad & \stackrel{=}{=} \mathbb{E}[w_2(Z)] \int \tilde{S}(t) w_1(t) \left\{ \Upsilon_2(t) - \frac{\Upsilon_1(t)^2}{\Upsilon_0(t)} \right\} dt.
\end{aligned}$$

Luego

$$\begin{aligned}
\mathcal{D}' &= \frac{\partial}{\partial \varepsilon} \left(\mathcal{D}_w(H_\varepsilon) \right) \Big|_{\varepsilon=0} = \frac{\partial}{\partial \varepsilon} \left(\iint w_2(z) \int_0^t w_1(u) [z - \tilde{z}_\varepsilon(u)]^2 du dH_\varepsilon(t, z) \right) \Big|_{\varepsilon=0} \\
&= \iint w_2(z) \left(\int_0^t w_1(u) 2[z - \tilde{z}(u)](-\tilde{z}'(u)) du \right) dH(t, z) + \\
& \quad w_2(z_0) \int_0^{t_0} w_1(u) [z_0 - \tilde{z}(u)]^2 du - \iint w_2(z) \left(\int_0^t w_1(u) [z - \tilde{z}(u)]^2 du \right) dH(t, z) \\
&= w_2(z_0) \int_0^{t_0} w_1(u) [z_0 - \tilde{z}(u)]^2 du - \mathcal{D} - \\
& \quad 2 \int w_1(u) \left[\iint I(t \geq u) w_2(z) [z - \tilde{z}(u)] dH(t, z) \right] \tilde{z}'(u) du
\end{aligned}$$

$$\stackrel{(7.2)}{=} w_2(z_0) \int_0^{t_0} w_1(u) [z_0 - \tilde{z}(u)]^2 du - \mathcal{D}$$

$$\stackrel{(7.37)}{=} \mathbb{E} [w_2(Z)] \nu(z_0) \int_0^{t_0} w_1(u) [z_0^2 - 2z_0\tilde{z}(u) + \tilde{z}(u)^2] du - \mathcal{D}$$

$$= \mathbb{E} [w_2(Z)] \nu(z_0) \left[z_0^2 \int_0^{t_0} w_1(u) du - 2z_0 \int_0^{t_0} w_1(u) \tilde{z}(u) du + \int_0^{t_0} w_1(u) \tilde{z}(u)^2 du \right] - \mathcal{D}$$

$$\stackrel{(7.42)}{=} \mathbb{E} [w_2(Z)] \nu(z_0) \left[z_0^2 \left(\int_0^{t_0} w_1(u) du \right) - 2z_0 \left(\int_0^{t_0} w_1(u) \frac{\Upsilon_1(u)}{\Upsilon_0(u)} du \right) + \right. \\ \left. \left(\int_0^{t_0} w_1(u) \left(\frac{\Upsilon_1(u)}{\Upsilon_0(u)} \right)^2 du \right) \right] - \mathcal{D}$$

$$\stackrel{(7.36)}{=} \mathbb{E} [w_2(Z)] \nu(z_0) [z_0^2 \gamma_0(t_0) - 2z_0 \gamma_1(t_0) + \gamma_2(t_0)] - \mathcal{D}$$

Por lo tanto,

$$\frac{\mathcal{D}'}{\mathcal{D}} = \frac{\nu(z_0) [z_0^2 \gamma_0(t_0) - 2z_0 \gamma_1(t_0) + \gamma_2(t_0)]}{\int \tilde{S}(t) w_1(t) \left(\Upsilon_2(t) - \frac{\Upsilon_1^2(t)}{\Upsilon_0(t)} \right) dt} - 1.$$

Por último,

$$\frac{\mathcal{N}'}{\mathcal{N}} - \frac{\mathcal{D}'}{\mathcal{D}} = \left(\beta \int w_1(t) \tilde{S}(t) \left(\Upsilon_2(t) - \frac{\Upsilon_1(t)^2}{\Upsilon_0(t)} \right) \right)^{-1} \nu(z_0) \left\{ \Delta_0 w_1(t_0) \left[z_0 - \frac{\Upsilon_1(t_0)}{\Upsilon_0(t_0)} \right] - \right. \\ \left. \int_0^{t_0} w_1(t) \lambda_0(t) \left[z_0 - \frac{\Upsilon_1(t)}{\Upsilon_0(t)} \right] dt - \beta \left(z_0 \gamma_1(t_0) - \gamma_2(t_0) \right) - \right. \\ \left. \beta \left[z_0^2 \gamma_0(t_0) - 2z_0 \gamma_1(t_0) + \gamma_2(t_0) \right] \right\}$$

$$= \left(\beta \int w_1(t) \tilde{S}(t) \left(\Upsilon_2(t) - \frac{\Upsilon_1(t)^2}{\Upsilon_0(t)} \right) \right)^{-1} \nu(z_0) \left\{ \Delta_0 w_1(t_0) \left[z_0 - \frac{\Upsilon_1(t_0)}{\Upsilon_0(t_0)} \right] - \right. \\ \left. \int_0^{t_0} w_1(t) \lambda_0(t) \left[z_0 - \frac{\Upsilon_1(t)}{\Upsilon_0(t)} \right] dt + \beta \left(z_0 \gamma_1(t_0) - z_0^2 \gamma_0(t_0) \right) \right\}$$

Para finalizar tenemos que

$$\mathcal{JF}_w(t_0, z_0, \Delta_0, \beta) = \left[\frac{\mathcal{N}'}{\mathcal{N}} - \frac{\mathcal{D}'}{\mathcal{D}} \right] \beta = \left[\int w_1(t) \tilde{S}(t) \left(\Upsilon_2(t) - \frac{\Upsilon_1(t)^2}{\Upsilon_0(t)} \right) \right]^{-1} \nu(z_0) \left\{ \Delta_0 w_1(t_0) \times \right. \\ \left. \left[z_0 - \frac{\Upsilon_1(t_0)}{\Upsilon_0(t_0)} \right] - \int_0^{t_0} w_1(t) \lambda_0(t) \left[z_0 - \frac{\Upsilon_1(t)}{\Upsilon_0(t)} \right] dt + \beta \left(z_0 \gamma_1(t_0) - z_0^2 \gamma_0(t_0) \right) \right\}$$

como queríamos. □

Demostración del Lema 1

Demostración: Para simplificar la notación sean $L(G) := L(G, \beta, W)$, $L(H) := L(H, \beta, W)$ y $\text{Lin} := \text{Lin}_{W, \beta}(G - H)$ donde $\beta \in B$ y $W \in \mathcal{W}$.

Observar que, como $W \in \mathcal{W}$, $W = W_0 W^*$, donde W_0 es una función continua y con soporte acotado K y $W^* \in \mathcal{W}^*$, la siguiente función que definimos, por **B1**, verifica que, para $u \in K$:

$$d(u) := W_0(u) \iint W^*(u, \mathbf{z}) I(t \geq u) dH(t, \mathbf{z}) > W_0(u) \varepsilon > 0. \quad (7.56)$$

Además, por (7.56), podemos definir las siguientes funciones:

$$\begin{aligned} \mathbf{a}(u) &:= \frac{\iint W(u, \mathbf{z}) \mathbf{z} I(t \geq u) dH(t, \mathbf{z})}{\iint W(u, \mathbf{z}) I(t \geq u) dH(t, \mathbf{z})} && \text{(función vectorial)} \\ b(u) &:= \frac{\iint W(u, \mathbf{z}) I(t \geq u) d(G - H)(t, \mathbf{z})}{\iint W(u, \mathbf{z}) I(t \geq u) dH(t, \mathbf{z})} && \text{(función real)} \\ \mathbf{c}(u) &:= \frac{\iint W(u, \mathbf{z}) \mathbf{z} I(t \geq u) d(G - H)(t, \mathbf{z})}{\iint W(u, \mathbf{z}) I(t \geq u) dH(t, \mathbf{z})} && \text{(función vectorial)}. \end{aligned}$$

Notar que, como $b(u) \geq 0$, entonces $\left| \frac{1}{b(u)+1} \right| \leq 1$.

Ahora, para continuar simplificando la notación, definimos: $W := W(t, \mathbf{z})$, $dH := dH(t, \mathbf{z})$, $dG := dG(t, \mathbf{z})$ y las funciones $d := d(t)$, $\mathbf{a} := \mathbf{a}(t)$, $b := b(t)$, y $\mathbf{c} := \mathbf{c}(t)$.

Entonces $\mathbf{z}_H(t) = \mathbf{a}$ y $\mathbf{z}_G(t) = \frac{\mathbf{a} + \mathbf{c}}{b+1}$, por lo tanto

$$\mathbf{z}_H(t) - \mathbf{z}_G(t) = \mathbf{a} - \frac{\mathbf{a} + \mathbf{c}}{b+1} = \frac{\mathbf{a}b - \mathbf{c}}{b+1} = \frac{\mathbf{a}b}{b+1} - \frac{\mathbf{c}}{b+1}.$$

En consecuencia, tenemos que:

$$\begin{aligned}
L(G) - L(H) &= \iint W[z - z_G(t)]dG - \iint W\left(\int_0^t [z - z_G(u)]du\right) z' dG \beta - \\
&\quad \iint W[z - z_H(t)]dH + \iint W\left(\int_0^t [z - z_H(u)]du\right) z' dH \beta \\
&= \iint W[z - z_G(t)]dG - \iint W[z - z_H(t)]dG + \iint W[z - z_H(t)]dG - \\
&\quad \iint W[z - z_H(t)]dH - \iint W\left(\int_0^t [z - z_G(u)]du\right) z' dG \beta + \\
&\quad \iint W\left(\int_0^t [z - z_H(u)]du\right) z' dG \beta - \iint W\left(\int_0^t [z - z_H(u)]du\right) z' dG \beta + \\
&\quad \iint W\left(\int_0^t [z - z_H(u)]du\right) z' dH \beta \\
&= \iint W[z_H(t) - z_G(t)]dG + \iint W[z - z_H(t)]d(G - H) - \\
&\quad \iint W\left(\int_0^t [z_H(u) - z_G(u)]du\right) z' dG \beta - \\
&\quad \iint W\left(\int_0^t [z - z_H(u)]du\right) z' d(G - H) \beta
\end{aligned}$$

Luego

$$\begin{aligned}
L(G) - L(H) - \text{Lin} &= \iint W[z_H(t) - z_G(t)]dG + \iint W[z - z_H(t)]d(G - H) - \\
&\quad \iint W\left(\int_0^t [z_H(u) - z_G(u)]du\right) z' dG \beta - \\
&\quad \iint W\left(\int_0^t [z - z_H(u)]du\right) z' d(G - H) \beta - \\
&\quad \iint W[z - z_H(t)]d(G - H) + \\
&\quad \iint W\int_0^t [z - z_H(u)]du z' d(G - H) \beta - \\
&\quad \iint W[z_H(t) - z_G(t)] dH + \iint W\left(\int_0^t [z_H(u) - z_G(u)]du\right) z' dH \beta \\
&= \iint W[z_H(t) - z_G(t)]d(G - H) - \\
&\quad \iint W\left(\int_0^t [z_H(u) - z_G(u)]du\right) z' d(G - H) \beta
\end{aligned}$$

$$\begin{aligned}
&= \iint W \left[\frac{\mathbf{a}b}{b+1} - \frac{\mathbf{c}}{b+1} \right] d(G-H) - \\
&\quad \iint W \left(\int_0^t \left[\frac{\mathbf{a}b}{b+1} - \frac{\mathbf{c}}{b+1} \right] du \right) \mathbf{z}' d(G-H)\boldsymbol{\beta} \\
&= L_1 - L_2 - L_3 + L_4,
\end{aligned}$$

donde

$$\begin{aligned}
L_1 &:= \iint W_0(t) W^*(t, \mathbf{z}) \mathbf{a}(t) \frac{b(t)}{b(t)+1} d(G-H)(t, \mathbf{z}) \\
L_2 &:= \iint W_0(t) W^*(t, \mathbf{z}) \mathbf{c}(t) \frac{1}{b(t)+1} d(G-H)(t, \mathbf{z}) \\
L_3 &:= \iiint I(t \geq u) W_0(t) W^*(t, \mathbf{z}) \mathbf{a}(u) \frac{b(u)}{b(u)+1} \boldsymbol{\beta}' \mathbf{z} du d(G-H)(t, \mathbf{z}) \\
L_4 &:= \iiint I(t \geq u) W_0(t) W^*(t, \mathbf{z}) \mathbf{c}(u) \frac{1}{b(u)+1} \boldsymbol{\beta}' \mathbf{z} du d(G-H)(t, \mathbf{z}).
\end{aligned}$$

Por la condición **B2**, para valores suficientemente grandes de M_0, M_1, \dots, M_p , sea $\tilde{K} := \prod_{i=0}^p [0, M_i]$ el conjunto acotado en $\mathbb{R}^+ \times \mathbb{R}^{p+}$ que contenga al soporte acotado de las funciones de W^* . Observar que $[0, M_0]$ contiene a K .

Sea $u \in [0, M]$ y $W \in \mathcal{W}$. Entonces

$$\begin{aligned}
\iint W(u, \mathbf{z}) \mathbf{z} I(t \geq u) dH(t, \mathbf{z}) &= \iint_{\tilde{K}} W_0(u) W^*(u, \mathbf{z}) \mathbf{z} I(t \geq u) dH(t, \mathbf{z}) \\
&= W_0(u) \int_0^{M_p} \dots \int_0^{M_1} W^*(u, \mathbf{z}) \mathbf{z} \left[\int_0^M I(t \geq u) \frac{\partial}{\partial t} \left(\frac{\partial^p}{\partial \mathbf{z}} H(t, \mathbf{z}) \right) dt \right] d\mathbf{z},
\end{aligned}$$

donde $\frac{\partial^p}{\partial \mathbf{z}} = \frac{\partial^p}{\partial z_1 \dots \partial z_p}$ y $d\mathbf{z} = dz_1 \dots dz_p$. La integral del corchete

$$\begin{aligned}
\int_0^{M_0} I(t \geq u) \frac{\partial}{\partial t} \left(\frac{\partial^p}{\partial \mathbf{z}} H(t, \mathbf{z}) \right) dt &= I(M_0 > u) \int_u^{M_0} \frac{\partial}{\partial t} \left(\frac{\partial^p}{\partial \mathbf{z}} H(t, \mathbf{z}) \right) dt \\
&= I(M_0 > u) \left[\frac{\partial^p}{\partial \mathbf{z}} H(M_0, \mathbf{z}) - \frac{\partial^p}{\partial \mathbf{z}} H(u, \mathbf{z}) \right]
\end{aligned}$$

Para simplificar la notación definimos el siguiente operador:

$$\Delta_{[u, M_0]} f(t) := I(M_0 > u) [f(M_0) - f(u)].$$

Entonces

$$\int_0^{M_0} I(t \geq u) \frac{\partial}{\partial t} \left(\frac{\partial^p}{\partial \mathbf{z}} H(t, \mathbf{z}) \right) dt = \Delta_{[u, M_0]} \frac{\partial^p}{\partial \mathbf{z}} H(t, \mathbf{z}).$$

Luego, si notamos al conjunto $\prod_{i=1}^p [0, M_i]$ por K_z y a las integrales $\int_0^{M_p} \cdot \int_0^{M_1}$ por \int_{K_z} , tenemos que:

$$\int_{K_z} W^*(u, \mathbf{z}) \mathbf{z} \int_0^{M_0} I(t \geq u) \frac{\partial}{\partial t} \frac{\partial^p}{\partial \mathbf{z}} H(t, \mathbf{z}) dt d\mathbf{z} = \Delta_{[u, M_0]} \int_{K_z} W^*(u, \mathbf{z}) \mathbf{z} \frac{\partial^p}{\partial \mathbf{z}} H(t, \mathbf{z}) d\mathbf{z}.$$

Ahora, para cualquier $i = 0, 1, \dots, p$ y como los M_i son suficientemente grandes como para que

$$W^*(\dots, M_i, \dots) = W^*(\dots, 0, \dots) = 0,$$

aplicando iterativamente integración por partes obtenemos que:

$$\Delta_{[u, M_0]} \int_{K_z} W^*(u, \mathbf{z}) \mathbf{z} \frac{\partial^p}{\partial \mathbf{z}} H(t, \mathbf{z}) d\mathbf{z} = (-1)^p \Delta_{[u, M_0]} \int_{K_z} H(t, \mathbf{z}) \frac{\partial^p}{\partial \mathbf{z}} (W^*(u, \mathbf{z}) \mathbf{z}) d\mathbf{z}.$$

En consecuencia tenemos que:

$$\begin{aligned} \mathbf{a}(u) &= \frac{\iint W(u, \mathbf{z}) \mathbf{z} I(t \geq u) dH(t, \mathbf{z})}{\iint W(u, \mathbf{z}) I(t \geq u) dH(t, \mathbf{z})} \\ &= \frac{W_0(u) (-1)^p \Delta_{[u, M_0]} \int_{K_z} H(t, \mathbf{z}) \frac{\partial^p}{\partial \mathbf{z}} (W^*(u, \mathbf{z}) \mathbf{z}) d\mathbf{z}}{W_0(u) \iint W^*(u, \mathbf{z}) I(t \geq u) dH(t, \mathbf{z})} \\ &= \frac{(-1)^p \Delta_{[u, M_0]} \int_{K_z} H(t, \mathbf{z}) \frac{\partial^p}{\partial \mathbf{z}} (W^*(u, \mathbf{z}) \mathbf{z}) d\mathbf{z}}{\iint W^*(u, \mathbf{z}) I(t \geq u) dH(t, \mathbf{z})}. \end{aligned}$$

Entonces, $\forall u \in K$:

$$\|\mathbf{a}(u)\|_2^2 \leq \frac{\|H\|_\infty^2 \left\| \int_{K_z} \frac{\partial^p}{\partial \mathbf{z}} (W^*(u, \mathbf{z}) \mathbf{z}) d\mathbf{z} \right\|_2^2}{\iint W^*(u, \mathbf{z}) I(t \geq u) dH(t, \mathbf{z})},$$

que está acotado por **B1** y **B2**. Es decir, existe una constante finita c_a tal que $\|\mathbf{a}(u)\|_2 \leq c_a$.

Luego, siguiendo el mismo procedimiento anterior (integración por partes), tenemos que:

$$\begin{aligned}
\iint_{\tilde{K}} W(u, \mathbf{z}) I(t \geq u) d(G-H)(t, \mathbf{z}) &= \iint_{\tilde{K}} W_0(u) W^*(u, \mathbf{z}) I(t \geq u) d(G-H)(t, \mathbf{z}) \\
&= W_0(u) \int_{K_z} W^*(u, \mathbf{z}) \left[\int_0^{M_0} I(t \geq u) \frac{\partial}{\partial t} \left(\frac{\partial^p}{\partial \mathbf{z}} (G-H)(t, \mathbf{z}) \right) dt \right] d\mathbf{z} \\
&= \Delta_{[u, M_0]} W_0(u) \int_{K_z} W^*(u, \mathbf{z}) \frac{\partial^p}{\partial \mathbf{z}} (G-H)(t, \mathbf{z}) d\mathbf{z} \\
&= W_0(u) (-1)^p \Delta_{[u, M_0]} \int_{K_z} (G-H)(t, \mathbf{z}) \frac{\partial^p}{\partial \mathbf{z}} W^*(u, \mathbf{z}) d\mathbf{z}.
\end{aligned}$$

Entonces

$$\begin{aligned}
b(u) &= \frac{\iint W(u, \mathbf{z}) I(t \geq u) d(G-H)(t, \mathbf{z})}{\iint W(u, \mathbf{z}) I(t \geq u) dH(t, \mathbf{z})} \\
&= \frac{W_0(u) (-1)^p \Delta_{[u, M_0]} \int_{K_z} (G-H)(t, \mathbf{z}) \frac{\partial^p}{\partial \mathbf{z}} W^*(u, \mathbf{z}) d\mathbf{z}}{\iint W_0(u) W^*(u, \mathbf{z}) I(t \geq u) dH(t, \mathbf{z})} \\
&= \frac{(-1)^p \Delta_{[u, M_0]} \int_{K_z} (G-H)(t, \mathbf{z}) \frac{\partial^p}{\partial \mathbf{z}} W^*(u, \mathbf{z}) d\mathbf{z}}{\iint W^*(u, \mathbf{z}) I(t \geq u) dH(t, \mathbf{z})}
\end{aligned}$$

y por lo tanto

$$|b(u)| \leq \frac{\|G-H\|_\infty}{\varepsilon} \int_{K_z} \left| \frac{\partial^p}{\partial \mathbf{z}} W^*(u, \mathbf{z}) \right| d\mathbf{z}.$$

La integral de arriba, por la condición **B2**, está acotada. Luego, $\forall u \in K$ $|b(u)| \leq c_b \|G-H\|_\infty$, para alguna constante finita c_b .

Nuevamente integrando por partes, tenemos que:

$$\begin{aligned}
\iint_{\tilde{K}} W(u, \mathbf{z}) \mathbf{z} I(t \geq u) d(G-H)(t, \mathbf{z}) &= \iint_{\tilde{K}} W_0(u) W^*(u, \mathbf{z}) \mathbf{z} I(t \geq u) d(G-H)(t, \mathbf{z}) \\
&= W_0(u) \int_{K_z} W^*(u, \mathbf{z}) \mathbf{z} \left[\int_0^{M_0} I(t \geq u) \frac{\partial}{\partial t} \left(\frac{\partial^p}{\partial \mathbf{z}} (G-H)(t, \mathbf{z}) \right) dt \right] d\mathbf{z} \\
&= \Delta_{[u, M_0]} W_0(u) \int_{K_z} W^*(u, \mathbf{z}) \mathbf{z} \frac{\partial^p}{\partial \mathbf{z}} (G-H)(t, \mathbf{z}) d\mathbf{z} \\
&= W_0(u) (-1)^p \Delta_{[u, M_0]} \int_{K_z} (G-H)(t, \mathbf{z}) \frac{\partial^p}{\partial \mathbf{z}} (W^*(u, \mathbf{z}) \mathbf{z}) d\mathbf{z}.
\end{aligned}$$

Entonces

$$\begin{aligned}
\mathbf{c}(u) &= \frac{\iint W(u, \mathbf{z}) \mathbf{z} I(t \geq u) d(G - H)(t, \mathbf{z})}{\iint W(u, \mathbf{z}) I(t \geq u) dH(t, \mathbf{z})} \\
&= \frac{W_0(u) (-1)^p \Delta_{[u, M_0]} \int_{K_z} (G - H)(t, \mathbf{z}) \frac{\partial^p}{\partial \mathbf{z}} (W^*(u, \mathbf{z}) \mathbf{z}) d\mathbf{z}}{\iint W_0(u) W^*(u, \mathbf{z}) I(t \geq u) dH(t, \mathbf{z})},
\end{aligned}$$

por lo tanto,

$$\begin{aligned}
\|\mathbf{c}(u)\|_2 &\leq \frac{\|G - H\|_\infty \left\| \int_{K_z} \frac{\partial^p}{\partial \mathbf{z}} (W^*(u, \mathbf{z}) \mathbf{z}) d\mathbf{z} \right\|_2}{\iint W^*(u, \mathbf{z}) I(t \geq u) dH(t, \mathbf{z})} \\
&< \frac{\|G - H\|_\infty}{\varepsilon} \left\| \int_{K_z} \frac{\partial^p}{\partial \mathbf{z}} (W^*(u, \mathbf{z}) \mathbf{z}) d\mathbf{z} \right\|_2,
\end{aligned}$$

donde la norma de la integral es acotada por **B2**. Por la tanto, $u \in K$ $\|\mathbf{c}(u)\|_2 \leq c_c \|G - H\|_\infty$, para alguna constante finita c_c .

Por último definimos las integrales:

$$S_1 := \iint W_0(t) W^*(t, \mathbf{z}) d(G - H)(t, \mathbf{z})$$

y

$$S_2 := \iint W_0(t) W^*(t, \mathbf{z}) \int I(t \geq u) du d(G - H)(t, \mathbf{z}).$$

Integrando por partes estas integrales obtenemos que:

$$\begin{aligned}
S_1 &= \iint_{\tilde{K}} W_0(t) W^*(t, \mathbf{z}) d(G - H)(t, \mathbf{z}) \\
&= (-1)^{p+1} \iint_{\tilde{K}} (G - H)(t, \mathbf{z}) \frac{\partial^p}{\partial \mathbf{z}} \left(\frac{\partial}{\partial t} W_0(t) W^*(t, \mathbf{z}) dt \right) d\mathbf{z}
\end{aligned}$$

y

$$\begin{aligned}
S_2 &= \iint_{\tilde{K}} W_0(t) W^*(t, \mathbf{z}) t d(G - H)(t, \mathbf{z}) \\
&= (-1)^{p+1} \iint_{\tilde{K}} (G - H)(t, \mathbf{z}) \frac{\partial^p}{\partial \mathbf{z}} \left(\frac{\partial}{\partial t} W_0(t) W^*(t, \mathbf{z}) t dt \right) d\mathbf{z}.
\end{aligned}$$

Notar que

$$|S_1| \leq \|G - H\|_\infty \iint_{\tilde{K}} \left| \frac{\partial^p}{\partial \mathbf{z}} \left(\frac{\partial}{\partial t} W_0(t) W^*(t, \mathbf{z}) \right) \right| dt d\mathbf{z}$$

y

$$|S_2| \leq \|G - H\|_\infty \iint_{\tilde{K}} \left| \frac{\partial^p}{\partial \mathbf{z}} \left(\frac{\partial}{\partial t} W_0(t) W^*(t, \mathbf{z}) t \right) \right| dt d\mathbf{z},$$

además, ambas integrales son acotados por **B2**. O sea, existen constantes finitas s_1 y s_2 , tales que

$$|S_1| \leq s_1 \|G - H\|_\infty \quad \text{y} \quad |S_2| \leq s_2 \|G - H\|_\infty \quad (7.57)$$

Por último, notar que $\mathbf{z} \in K_z = \prod_{i=1}^p [0, M_i]$ (subconjunto cerrado y acotado de \mathbb{R}^{p+}) tenemos que $\|\boldsymbol{\beta}' \mathbf{z}\|_\infty < \infty$.

Ahora sí podemos comenzar a acotar las normas de los L_i :

$$\begin{aligned} \|L_1\|_2 &= \left\| \iint W_0(t) W^*(t, \mathbf{z}) \mathbf{a}(t) \frac{b(t)}{b(t) + 1} d(G - H)(t, \mathbf{z}) \right\|_2 \\ &\leq c_a |b(t)| |S_1| \sqrt{p} \\ &\leq c_a c_b \|G - H\|_\infty s_1 \|G - H\|_\infty \sqrt{p} = o(\|G - H\|_\infty^2) \\ \|L_2\|_2 &= \left\| \iint W_0(t) W^*(t, \mathbf{z}) \mathbf{c}(t) \frac{1}{b(t) + 1} d(G - H)(t, \mathbf{z}) \right\|_2 \\ &\leq c_c |S_1| \sqrt{p} \\ &\leq c_c \|G - H\|_\infty s_1 \|G - H\|_\infty \sqrt{p} = o(\|G - H\|_\infty^2) \\ \|L_3\|_2 &= \left\| \iiint I(t \geq u) W_0(t) W^*(t, \mathbf{z}) \mathbf{a}(u) \frac{b(u)}{b(u) + 1} \boldsymbol{\beta}' \mathbf{z} du d(G - H)(t, \mathbf{z}) \right\|_2 \\ &\leq c_a |b(u)| \|\boldsymbol{\beta}' \mathbf{z}\|_\infty |S_2| \sqrt{p} \\ &\leq c_a c_b \|G - H\|_\infty \|\boldsymbol{\beta}' \mathbf{z}\|_\infty s_2 \|G - H\|_\infty \sqrt{p} = o(\|G - H\|_\infty^2) \\ \|L_4\|_2 &= \left\| \iiint I(t \geq u) W_0(t) W^*(t, \mathbf{z}) \mathbf{c}(u) \frac{1}{b(u) + 1} \boldsymbol{\beta}' \mathbf{z} du d(G - H)(t, \mathbf{z}) \right\|_2 \\ &\leq c_c \|\boldsymbol{\beta}' \mathbf{z}\|_\infty |S_2| \sqrt{p} \\ &\leq c_c \|G - H\|_\infty \|\boldsymbol{\beta}' \mathbf{z}\|_\infty s_2 \|G - H\|_\infty \sqrt{p} = o(\|G - H\|_\infty^2) \end{aligned}$$

Entonces podemos concluir que

$$\sup_{\substack{\boldsymbol{\beta} \in B \\ W \in \mathcal{W}}} \left\| L(G, \boldsymbol{\beta}, W) - L(H, \boldsymbol{\beta}, W) - \text{Lin}_{W, \boldsymbol{\beta}}(G - H) \right\|_2 = o(\|G - H\|_\infty^2).$$

Para terminar, si reemplazamos $G-H$ por H en las funciones $b(u)$ y $\mathbf{c}(u)$ anteriormente definidas obtenemos que $b(u) = 1$ y $\mathbf{c}(u) = \mathbf{a}(u)$. Entonces $\mathbf{z}_H(u) - \mathbf{z}_G(u) = \mathbf{0}$, con lo cual solo quedan los dos primeros términos de $\text{Lin}_{W, \beta_0}(H)$. Que por la consistencia en el sentido Fisher, tenemos que $\text{Lin}_{W, \beta_0}(H) = L(H, \beta_0, W) = \mathbf{0}$.

□

Demostración del Lema 2

Demostración: Para cualquier $W \in \mathcal{W}$ fijo y $\beta \in B$ tenemos que

$$\begin{aligned} \frac{\partial L(H, \beta, W)}{\partial \beta} &= \frac{\partial}{\partial \beta} \left[\iint W(t, \mathbf{z}) [\mathbf{z} - \mathbf{z}_H(t)] dH(t, \mathbf{z}, \Delta=1) - \right. \\ &\quad \left. \iint W(t, \mathbf{z}) \left(\int_0^t [\mathbf{z} - \mathbf{z}_H(u)] du \right) \mathbf{z}' dH(t, \mathbf{z}) \beta \right] \\ &= - \left[\iint W(t, \mathbf{z}) \left(\int_0^t [\mathbf{z} - \mathbf{z}_H(u)] du \right) \mathbf{z}' dH(t, \mathbf{z}) \right] = \mathcal{L}(H, W). \end{aligned}$$

Entonces podemos expresar a $L(H, \beta, W)$ como:

$$L(H, \beta, W) = \iint W(t, \mathbf{z}) [\mathbf{z} - \mathbf{z}_H(t)] dH(t, \mathbf{z}, \Delta=1) + \mathcal{L}(H, W) \beta,$$

y $L(H, \beta_0, W)$ como:

$$L(H, \beta_0, W) = \iint W(t, \mathbf{z}) [\mathbf{z} - \mathbf{z}_H(t)] dH(t, \mathbf{z}, \Delta=1) + \mathcal{L}(H, W) \beta_0.$$

Por lo tanto, restando, obtenemos:

$$L(H, \beta, W) - L(H, \beta_0, W) = \mathcal{L}(H, W) (\beta - \beta_0).$$

□

Demostración del Lema 3

Demostración: Seguiremos las mismas notaciones y procedimientos que se utilizaron en la demostración del Lema 1.

Expresemos a $\text{Lin}_{W,\beta}(G - H)$ de la siguiente manera:

$$\begin{aligned} \text{Lin}_{W,\beta}(G - H) &= \iint W[\mathbf{z} - \mathbf{a}]d(G - H) - \iint W \left(\int_0^t [\mathbf{z} - \mathbf{a}]du \right) \mathbf{z}'d(G - H)\beta + \\ &\quad \iint W \left[\frac{\mathbf{a}b}{b+1} - \frac{\mathbf{c}}{b+1} \right] dH - \iint W \left(\int_0^t \left[\frac{\mathbf{a}b}{b+1} - \frac{\mathbf{c}}{b+1} \right] du \right) \mathbf{z}'dH\beta \\ &= R_1 - R_2 - R_3 + R_4 + \tilde{L}_1 - \tilde{L}_2 - \tilde{L}_3 + \tilde{L}_4, \end{aligned}$$

donde

$$\begin{aligned} R_1 &:= \iint W_0(t)W^*(t, \mathbf{z}) \mathbf{z} d(G - H)(t, \mathbf{z}) \\ R_2 &:= \iint W_0(t)W^*(t, \mathbf{z})\mathbf{a}(t) d(G - H)(t, \mathbf{z}) \\ R_3 &:= \iint W_0(t)W^*(t, \mathbf{z}) \left(\int_0^t \mathbf{z} du \right) \beta' \mathbf{z} d(G - H)(t, \mathbf{z}) \\ R_4 &= \iint W_0(t)W^*(t, \mathbf{z}) \left(\int I(t \geq u)\mathbf{a}(u) du \right) \beta' \mathbf{z} d(G - H)(t, \mathbf{z}) \\ \tilde{L}_1 &:= \iint W_0(t)W^*(t, \mathbf{z}) \frac{\mathbf{a}(t)b(t)}{b(t)+1} dH(t, \mathbf{z}) \\ \tilde{L}_2 &:= \iint W_0(t)W^*(t, \mathbf{z}) \frac{\mathbf{c}(t)}{b(t)+1} dH(t, \mathbf{z}) \\ \tilde{L}_3 &:= \iint W_0(t)W^*(t, \mathbf{z}) \left(\int I(t \geq u) \frac{\mathbf{a}(u)b(u)}{b(u)+1} du \right) \beta' \mathbf{z} dH(t, \mathbf{z}) \\ \tilde{L}_4 &:= \iint W_0(t)W^*(t, \mathbf{z}) \left(\int I(t \geq u) \frac{\mathbf{c}(u)}{b(u)+1} du \right) \beta' \mathbf{z} dH(t, \mathbf{z}). \end{aligned}$$

Observar que la expresión de los \tilde{L}_i tiene el mismo integrando que los L_i (definidos en la demostración del Lema 1) pero difieren en el diferencial, allí teníamos $d(G - H)(t, \mathbf{z})$ y aquí tenemos $dH(t, \mathbf{z})$.

Ahora siguiendo el mismo procedimiento y argumento que allí, tenemos que existe una constante positiva $c_1 < \infty$ tal que

$$\left\| \tilde{L}_1 - \tilde{L}_2 - \tilde{L}_3 + \tilde{L}_4 \right\|_2 \leq c_1 \|G - H\|_\infty.$$

Veamos entonces la cota de la norma de los términos restantes.

Integrando por partes, por un lado, tenemos que:

$$R_1 = (-1)^{p+1} \iint_{\tilde{K}} (G - H)(t, \mathbf{z}) \frac{\partial^p}{\partial \mathbf{z}} \left(\frac{\partial}{\partial t} W_0(t)W^*(t, \mathbf{z})dt \right) \mathbf{z} d\mathbf{z}$$

entonces

$$\|R_1\|_2 \leq \|G - H\|_\infty \left\| \iint_{\tilde{K}} \frac{\partial^p}{\partial \mathbf{z}} \left(\frac{\partial}{\partial t} W_0(t) W^*(t, \mathbf{z}) \right) \mathbf{z} \, d\mathbf{z} \right\|_2,$$

que por **B2** la norma de la integral es finita. Es decir que existe una constante positiva c_2 tal que $\|R_1\|_2 \leq c_2 \|G - H\|_\infty$. Y además, obtenemos:

$$\iint W(t, \mathbf{z}) t \mathbf{z} d(G - H)(t, \mathbf{z}) = (-1)^{p+1} \iint_{\tilde{K}} (G - H)(t, \mathbf{z}) \frac{\partial^p}{\partial \mathbf{z}} \left(\frac{\partial}{\partial t} W_0(t) W^*(t, \mathbf{z}) t \right) \mathbf{z} \, d\mathbf{z}.$$

Con lo cual:

$$\|R_3\|_2 \leq \|\beta' \mathbf{z}\|_\infty \|G - H\|_\infty \iint_{\tilde{K}} \left\| \frac{\partial^p}{\partial \mathbf{z}} \frac{\partial}{\partial t} W_0(t) W^*(t, \mathbf{z}) t \mathbf{z} \right\|_2 \, dt \, d\mathbf{z},$$

por **B2**, la última integral es finita, entonces existe una constante positiva c_4 tal que $\|R_3\|_2 \leq c_4 \|G - H\|_\infty$.

Por otro lado:

$$\begin{aligned} \|R_2\|_2 &= \left\| \iint W_0(t) W^*(t, \mathbf{z}) \mathbf{a}(t) d(G - H)(t, \mathbf{z}) \right\|_2 \\ &\leq c_a |S_1| \sqrt{p} \stackrel{(7.57)}{\leq} c_a s_1 \|G - H\|_\infty \sqrt{p} \\ &\leq c_3 \|G - H\|_\infty \end{aligned}$$

para alguna constante finita c_3 . Y además

$$\begin{aligned} \|R_4\|_2 &= \left\| \iint W_0(t) W^*(t, \mathbf{z}) \int I(t \geq u) \mathbf{a}(u) du (\beta' \mathbf{z}) d(G - H)(t, \mathbf{z}) \right\|_2 \\ &\leq c_a \|\beta' \mathbf{z}\|_\infty |S_2| \sqrt{p} \stackrel{(7.57)}{\leq} c_a \|\beta' \mathbf{z}\|_\infty s_2 \|G - H\|_\infty \sqrt{p} \\ &\leq c_5 \|G - H\|_\infty, \end{aligned}$$

para alguna constante finita c_5 .

En conclusión tenemos que:

$$\begin{aligned} \left\| \text{Lin}_{W, \beta}(G - H) \right\|_2 &\leq \|R_1\|_2 + \|R_2\|_2 + \|R_3\|_2 + \|R_4\|_2 + \left\| \tilde{L}_1 - \tilde{L}_2 - \tilde{L}_3 + \tilde{L}_4 \right\|_2 \\ &\leq c \|G - H\|_\infty. \end{aligned}$$

donde $c := c_1 + c_2 + c_3 + c_4 + c_5$.

□

Demostración del Teorema 4.4

Demostración: Por los Lemas 1 y 3, $\forall W \in \mathcal{W}$ y $\beta \in B$, existe una constante positiva M_0 tal que

$$\|L(G, \beta, W) - L(H, \beta, W)\|_2 \leq M_0 \|G - H\|_\infty. \quad (7.58)$$

Por el Lema 2

$$L(H, \beta, W) - L(H, \beta_0, W) - \mathcal{L}(H, W)(\beta - \beta_0) = \mathbf{0},$$

y si lo sumamos a (7.58), obtenemos que

$$\|L(G, \beta, W) - \mathcal{L}(H, W)(\beta - \beta_0)\|_2 \leq M_0 \|G - H\|_\infty,$$

ya que $L(H, \beta_0, W) = \mathbf{0}$. Esto asegura que

$$\|\mathcal{L}^{-1}(H, W)L(G, \beta, W) - (\beta - \beta_0)\|_2 \leq M_0 \|G - H\|_\infty a^{-1/p}, \quad (7.59)$$

donde $a := |\det\{\mathcal{L}(H, W)\}|$, que por **B4** es no nulo y finito.

Sea $M := M_0 a^{-1/p} > 0$.

Ahora definimos la función continua de B en B como:

$$g(\beta) := \beta - \mathcal{L}^{-1}(H, W)L(G, \beta, W),$$

para $H \in \mathcal{A}_{\beta_0}$, $G \in \mathcal{B}(H, \varepsilon)$ y $W \in \mathcal{W}$ fijos. Entonces por (7.59) tenemos que:

$$\|\beta_0 - g(\beta)\|_2 \leq M \|G - H\|_\infty \quad (7.60)$$

Luego para $\|G - H\|_\infty$ suficientemente chico, la bola $\mathcal{B} := \{\beta : \|\beta - \beta_0\|_2 \leq M \|G - H\|_\infty\}$ es un subconjunto de B . También, por la ecuación (7.60), si algún $\beta \in \mathcal{B}$, su imagen $g(\beta) \in \mathcal{B}$. Por el Teorema de punto fijo de Brouwer, existe $\beta^* \in \mathcal{B}$ para el cual $g(\beta^*) = \beta^*$, es decir,

$$\beta^* - \mathcal{L}^{-1}(H, W)L(G, \beta^*, W) = \beta^*,$$

lo que implica $L(G, \beta^*, W) = \mathbf{0}$.

□

Demostración del Teorema 4.5

Demostración: Como $L(H, \beta_0, W) = \mathbf{0}$, por nuestras hipótesis, tenemos que

$$L(G, \beta, W) = L(G, \beta, W) - L(H, \beta, W) + L(H, \beta, W) - L(H, \beta_0, W). \quad (7.61)$$

Luego, por los Lemas 1 y 2, respectivamente, $\forall W \in \mathcal{W}$ y $\beta \in B$, tenemos que

$$\begin{aligned} L(G, \beta, W) - L(H, \beta, W) &= \text{Lin}_{W, \beta_0}(G - H) + o(\|G - H\|_\infty) \\ L(H, \beta, W) - L(H, \beta_0, W) &= \mathcal{L}(H, W)(\beta - \beta_0). \end{aligned}$$

Utilizando estas dos ecuaciones en (7.61) tenemos que:

$$L(G, \beta, W) = \text{Lin}_{W, \beta_0}(G - H) + \mathcal{L}(H, W)(\beta - \beta_0) + o(\|G - H\|_\infty)$$

Notar que por el Teorema 4.4, para $\|G - H\|_\infty$ suficientemente chico, existe $\beta(G, W)$ tal que $\beta(G, W) - \beta_0 = o(\|G - H\|_\infty)$ y $L(G, \beta(G, W), W) = \mathbf{0}$. Es decir,

$$\mathcal{L}(H, W)[\beta(G, W) - \beta_0] + \text{Lin}_{W, \beta_0}(G - H) = o(\|G - H\|_\infty).$$

Ya que por **B4** $\det \mathcal{L}(H, W)$ está acotado y es no nulo sobre \mathcal{W} , premultiplicando por la inversa obtenemos

$$\beta(G, W) - \beta_0 + \mathcal{L}(H, W)^{-1} \text{Lin}_{W, \beta_0}(G - H) = o(\|G - H\|_\infty).$$

□

Demostración del Teorema 4.6

Demostración:

$$\|H_n - H\|_\infty = \|H_n - G\|_\infty + \|G - H\|_\infty$$

Por suposición $\|G - H\|_\infty = o(n^{-1/2})$ y por Glivenko-Cantelli $\|H_n - G\|_\infty = o_p(n^{-1/2})$, así por Teorema 4.5,

$$\begin{aligned} \beta(H_n, W) - \beta_0 + \mathcal{L}(H, W)^{-1} \text{Lin}_{W, \beta_0}(H_n - H) &= o_p(n^{-1/2}) \\ n^{1/2} \left[\beta(H_n, W) - \beta_0 + \mathcal{L}(H, W)^{-1} \text{Lin}_{W, \beta_0}(H_n - H) \right] &= o_p(1) \end{aligned}$$

Como $\text{Lin}_{W, \beta_0}(H_n - H)$ es lineal y $\text{Lin}_{W, \beta_0}(H - H) = \mathbf{0}$, se sigue la conclusión del Teorema Central del Límite.

□

Capítulo 8

Referencias

- Aalen, O.O. (1980)** *A Model for Nonparametric Regression Analysis of Counting Processes*. Lecture Notes in Statistics, Ed. N. Klonecki, A. Koesh and J. Rosinski, Springer, New York. **2**: 1-25.
- Álvarez, E.E. y Ferrario, J. (2012)** *Revisión de la estimación robusta en Modelos Semiparamétricos de Supervivencia*. Revista Estadística del Instituto Interamericano de Estadística. **64**: 182-183, 85-106.
- Andersen, P. K. and Gill, R.D. (1982)** *Cox's regression model for counting processes: a large sample study*. The Annals of Statistics. **10**: 4, 1110-1120.
- Andersen, P.K., Borgan, O., Gill, R.D. and Keiding, N. (1993)** *Statistical Models Based on Counting Processes*. Springer-Verlag, New York.
- Bednarski, T. (1993)** *Robust Estimation in Cox's Regression Model*. Scandinavian Journal Statistics. **20**: 213-225.
- Bednarski, T. and Nowak, M. (2003)** *Robustness and efficiency of Sasieni-type estimators in the Cox model*. Journal of Statistical Planning and Inference. **115**: 1, 261-272.
- Bednarski, T. and Mocarska, E. (2006)** *On robust model selection within the Cox model*. Ecometrics Journal. **9**: 179-290.
- Bednarski, T. (2007)** *On a robust modification of Breslow's cumulated hazard estimator*. Computational Statistics and Data Analysis. **52**: 234-238.

- Begun, J. M., Hall, W. J., Huang, W. M. and Wellner, J. A. (1983)** *Information and Asymptotic Efficiency in Parametric-Nonparametric Models*. The Annals of Statistics. **11**: 2, 432-452.
- Beran, R. (1981)** *Efficient Robust Estimates in Parametric Models*. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete. **55**: 91-108.
- Breslow, N.E., and Day, N.E. (1987)** *Statistical Models in Cancer Research 2, The Design and Analysis of Cohort Studies*. Lyon: IARC.
- Buckley, J. and James, I. (1979)** *Linear regression with censored data*. Biometrika. **66**: 429-436.
- Cox, D.R. (1972)** *Regression Models and Life-Tables (with discussion)*. Journal of the Royal Statistical Society. Series B (Methodological). **34**: 2, 187-220.
- Cox, D.R. (1975)** *Partial Likelihood*. Biometrika. **62**: 262-276.
- Doll, R. , Morgan, L.G., and Speizer, F.E. (1970)** *Cancers of the Lung and Nasal Sinuses in Nickel Workers*. British Journal of Cancer. **24**: 623-632.
- Efron, B. (1977)** *Efficiency of Cox's Likelihood Function for Censored Data*. Journal of the American Statistical Association. **72**: 359, 557-565.
- Fleming, T.R. and Harrington, D.P. (1991)** *Counting Processes and Survival Analysis*. Wiley, New York.
- Fyngenson, T.R. and Ritov, Y. (1994)** *Monotone estimating equations for censored data*. The Annals of Statistics. **22**: 2, 732-746.
- Huang, Y. and Chen, Y. Q. (2003)** *Marginal regression of gaps between recurrent events*. Lifetime Data Analysis. **9**: 293-303.
- Huber, P. J. (1981)** *Robust Statistics*. Wiley, New York.
- Kaplan, E. L. and Meier, P. (1958)** *Nonparametric estimation from incomplete observations*. Journal of the American Statistical Association. **53**: 457-481.
- Lin, D.Y. and Ying, Z. (1994)** *Semiparametric Analysis of the Additive Risk Model*. Biometrika. **81**: 1, 61-71.

- Lin, D. Y., Wei, L. J., and Ying, Z. (1998)** *Accelerate failure time models for counting processes*. *Biometrika*. **85**: 3, 605-618.
- Loève, M (1960)** *Probability Theory*. Springer-Verlag, New York. Second Edition.
- Maronna, R.A., Martin, R.D. and Yohai, V.J. (2006)** *Robust Statistics: Theory and Methods*. *Wiley Series in Probability and Statistics*. John Wiley & Sons. **855**.
- Martinussen, T. and Scheike, T.H. (2006)** *Dynamic Regression Models for Survival Data*. *Statistics for Biology and Health*. Springer Series in Statistics Series.
- Minder, C. E. and Bednarski, T. (1996)** *A robust method for proportional hazards regression*. *Statistics in Medicine*. **15**: 1033-1047.
- Prentice, R. L. (1978)** *Linear rank tests with right censored data*. *Biometrika*. **65**: 167-179.
- Reid, N. (1981)** *Influence functions for censored data*. *The Annals of Statistics*. **9**: 1, 78-92.
- Reid, N. and Crépeau, H. (1985)** *Influence Functions for Proportional Hazards Regression*. *Biometrika*. **72**: 1, 1-9.
- Ritov, Y. and Wellner, J. A. (1987)** *Censoring, Martingala and The Cox Model*. Technical Report, University of Washington, Department of Statistics. **108**.
- Ritov, Y. (1990)** *Estimation in a linear regression model with censored data*. *The Annals of Statistics*. **18**: 1, 303-328.
- Rousseeuw, P. J. and Yohai, V. J. (1984)** *Robust regression by means of S-estimators*. *Robust and Nonlinear Time Series Analysis* (J. Franke, W. Hardle and R. D. Martin, eds.) *Lecture Notes in Statist.* **26**: 256-276.
- Salibian-Barrera, M. and Yohai, V. J. (2008)** *High breakdown point robust regression with censored data*. *Annals of Statistics*. **36**: 118-146.

- Sasieni, P. (1993a)** *Maximum Weighted Partial Likelihood Estimators for the Cox Model*. Journal of the American Statistical Association. **88**: 421, 144-152.
- Sasieni, P. (1993b)** *Some new Estimators for Cox Regression*. Annals of Statistics. **21**: 4, 1721-1759.
- Strawderman, R. L. (2005)** *The accelerated gap times model*. Biometrika. **92**: 647-666.
- Sun, L., Park, D. and Sun, J. (2006)** *The additive hazards model for recurrent gap times*. Statistica Sinica. **16**: 919-932.
- Tsiatis, A.A. (1990)** *Estimating regression parameters using linear rank tests for censored data*. The Annals of Statistics. **18**: 1, 354-372.
- Wang, M.C. and Chang, S.H. (1999)** *Nonparametric estimation of a recurrent survival function*. Journal of the American Statistical Association. **94**: 146-153.
- Wei, L.J., Ying, Z. and Ling, D.Y. (1990)** *Linear regression analysis of censored survival data based on rank tests*. Biometrika. **7**: 845-851.
- Yohai, V. J. (1987)** *High breakdown-point and high efficiency robust estimates for regression*. Annals of Statistics. **15**: 642-656.
- Yohai, V. J. and Zamar, R. H. (1988)** *High breakdown point and high efficiency estimates of regression by means of the minimization of an efficient scale*. Journal of the American Statistical Association. **83**: 406-413.