Tesis para obtener el grado de Doctor en Ciencias Matemáticas

INFERENCIA ESTADÍSTICA BASADA EN PROCESOS EMPÍRICOS PARA MODELOS SEMIPARAMÉTRICOS DEL ANÁLISIS DE SUPERVIVENCIA

Angélica Hernández Quintero

Casa abierta al tiempo
UNIVERSIDAD AUTÓNOMA METROPOLITANA

UNIVERSIDAD AUTÓNOMA METROPOLITANA

Tesis para obtener el grado de Doctor en Ciencias
Matemáticas

Presentada por

Angélica Hernández Quintero

11 de junio de 2010

INFERENCIA ESTADÍSTICA BASADA EN PROCESOS EMPÍRICOS PARA MODELOS SEMIPARAMÉTRICOS DEL ANÁLISIS DE SUPERVIVENCIA

ASESORES

Dr. Gabriel Escarela Pérez, Universidad Autónoma Metropolitana, México.
Dr. Jean François Dupuy, Université de La Rochelle, Francia.

JURADO

Dr. Dauxois Jean-Yves, Université de Besançon, Francia.
Dr. Dupuy Jean-François, Université de La Rochelle, Francia.
Dr. Nieto Barajas Enrique, Instituto Tecnológico Autónomo, México.
Dr. Gordienko Illich Evgueni, Universidad Autónoma Metropolitana, México.
Dr. Montes de Oca Machorro José Raúl, Universidad Autónoma Metropolitana, México.

Agradecimientos

Antes que nada quiero agradecer a Gabriel y a Jean-François por permitirme trabajar con ustedes y brindarme su apoyo; sobre todo gracias por haber tenido confianza en mi, sin ustedes esta tesis no podría haber sido posible. Gracias por su tiempo, su dedicación y sus consejos que permitieron realizar este trabajo.

Agradezco a Jean-Yves Dauxois, Illich Gordienko, Raúl Montes de Oca y Luis Enrique Nieto por aceptar ser miembros del jurado. Principalmente, les agradezco su tiempo y los comentarios que permitieron enriquecer y complementar esta tesis.

Gracias Marco por todo tu apoyo y comprensión. Y sobre todo por toda la ayuda y consejos que me brindaste para desarrollar la tesis. Agradezco también a mi familia que siempre estuvo apoyándome.

Gracias CONACYT por brindarme el apoyo financiero para realizar el Doctorado.

A Marco por su apoyo y por siempre estar a mi lado,

A mi familia por estar presente en todo momento.

Resumen

Los datos de supervivencia se presentan en disciplinas tales como medicina, criminología, finanzas e ingeniería entre otras. En varias circunstancias el evento de interés puede clasificarse en varias causas de muerte o fallo y en algunas otras, el evento puede observarse sólo para una proporción de "susceptibles". Los datos correspondientes a estas dos situaciones son conocidos como de riesgos competitivos y de sobrevivientes a largo plazo respectivamente. Problemas relevantes en el análisis de estos dos tipos de datos incluyen a propiedades básicas tales como la estimación de los parámetros, la existencia, consistencia y normalidad asintótica de los estimadores, y su eficiencia cuando éstos siguen una estructura semiparamétrica. En este trabajo se investigan dichas propiedades en formulaciones semiparamétricas bien establecidas para el análisis de riesgos competitivos y de sobrevivientes a largo plazo. Se presenta una descripción general de herramientas matemáticas que permiten estudiar dichas propiedades básicas, y se describe cómo la teoría moderna de procesos empíricos y la teoría de eficiencia semiparamétrica facilitan la obtención de las demostraciones correspondientes. Adicionalmente se presentan estimaciones consistentes de la varianza para los componentes en los modelos respectivos.

Los resultados encontrados en esta investigación proveen los fundamentos teóricos para realizar inferencias de muestras grandes que incluyen el cálculo de bandas de confianza y la realización de pruebas de hipótesis. Los métodos se ilustran con bases de datos generadas a través de simulaciones.

Palabras claves. Análisis de supervivencia, eficiencia semiparamétrica, modelo de cura, modelo de mezcla, modelo de riesgos proporcionales, modelo de transformación, procesos empíricos, riesgos competitivos.

Índice general

Introducción

El análisis de supervivencia es un área de la estadística que comprende un conjunto de técnicas y procedimientos donde el interés es analizar el tiempo que transcurre entre un evento inicial y un evento final. Dichos procedimientos son hoy en día una pieza fundamental en gran parte de los ensayos clínicos, de estudios epidemiológicos y de muchas otras disciplinas como son la economía, la ciencia actuarial y la ingeniería. Las características principales del análisis de supervivencia radican en que la distribución asociada tiende a tener cola larga y además que se busca analizar datos que frecuentemente cuentan con una importante cantidad de observaciones censuradas, las cuales ocurren cuando los tiempos de supervivencia no se conocen con exactitud. Por ejemplo, en estudios clínicos puede ocurrir que algún individuo abandone el estudio antes de que ocurra el evento de interés, registrándose solo información parcial del paciente (el punto en que fue visto con vida por última vez), en este caso se dice que el paciente se encuentra censurado por la derecha.

El principal interés de los datos de supervivencia consiste en estimar la función de supervivencia, la cual es la probabilidad de experimentar un evento después de cierto periodo. Actualmente, existe teoría bien establecida para estimar la función de supervivencia en forma no-paramétrica, semiparamétrica y completamente paramétrica. En particular, el modelo semiparamétrico de Cox representa una forma flexible de modelar la función de supervivencia en presencia de variables explicativas pues su estructura de riesgos proporcionales permite ignorar la parte funcional de la función de supervivencia y por lo tanto es posible realizar inferencias sobre los parámetros asociados a las variables explicativas; en este caso la función de verosimilitud asociada tiene cualidades muy cercanas a las de una función de verosimilitud completamente especificada y permitiendo estimar la función de supervivenciacon.

En el contexto de datos de supervivencia se puede tener que la ocurrencia del evento puede ser clasificada en diferentes causas y además que tal ocurrencia es provocada por sólo una causa en particular. A este tipo de datos se les conoce como *riesgos competitivos*. Uno de los principales objetivos del estudio de estos datos es modelar y estimar la función de incidencia acumulada de un tipo de causa específica, la cual es la probabilidad de que ocurra el evento por cierta causa dentro de cierto periodo, en presencia de variables explicativas. Una forma de proceder para la estimación es considerar a los individuos que no experimentan el evento de interés como censurados. Sin embargo, un paciente que experimenta un evento de riesgo competitivo está censurado de una forma informativa; esto es, no puede ser excluido del estudio y por lo tanto los métodos usados para analizar un solo evento no son útiles. En este caso, la función de incidencia acumulada para un evento de interés debe de ser calculada tomando en cuenta la presencia de las demás causas competitivas.

Existen varios modelos para analizar riesgos competitivos. Entre ellos, el modelo de Larson y Dinse (1985) sobresale por su identificabilidad y fácil interpretación. Dicho modelo consiste en especificar a las funciones de incidencia acumulada en términos de las probabilidades de supervivencia condicionales de causa específica y de las probabilidades de que finalmente el evento sea de una causa específica. Larson y Dinse proponen una estructura completamente paramétrica donde las funciones de supervivencia condicionales tienen la forma paramétrica de riesgos proporcionales de Cox cuya función de riesgo base es construida con pedazos exponenciales y la probabilidad de causa específica sigue un modelo multinomial. Recientemente, Ng y McLachlan (2003) y Escarela y Bowater (2008) han propuesto un modelo semiparamétrico de la formulación planteada por Larson y Dinse de manera tal que las funciones condicionales de supervivencia de base se expresan a través del modelo semiparamétrico de riesgos proporcionales de Cox. De esta forma el modelo resultante provee una forma flexible y parsimoniosa para estudiar riesgos competitivos. Estos autores se han concentrado principalmente en el enfoque computacional asociado al método semiparamétrico de estimación de máxima verosimilitud, ignorando el estudio de las propiedades asintóticas de los estimadores resultantes.

Es común encontrar datos de riesgos competitivos donde uno de los riesgos es no observable y corresponde a un subgrupo *inmune* al evento de interés. Para este tipo de datos la gráfica de la función de supervivencia empírica muestra que los datos presentan una recta horizontal relativamente larga al final del seguimiento, lo que sugiere la existencia de una proporción de individuos que nunca experimentan el evento de interés. Un ejemplo de este tipo de datos es cuando en estudios clínicos, una proporción de individuos responden favorablemente a algún tratamiento y dejan de mostrar signos o síntomas de la enfermedad y, por tanto, son considerados como individuos completamente curados; el resto de los individuos pueden caer nuevamente enfermos. Berkson y Gage (1952), proponen utilizar una distribución exponencial para los tiempos de supervivencia y una constante para la fracción de cura para ajustar datos de cáncer de seno y de estómago. Kuk y Chen (1992) proponen el modelo llamado modelo de cura de riesgos proporcionales, en el cual el modelo de riesgos proporcionales de Cox es utilizado en el tiempo de supervivecia de los individuos susceptibles y un modelo de regresión logística para la fracción curable. Peng y Dear (2000) y Sy y Taylor (2000) han propuesto un modelo de mezclas noparamétrico en el cual la suposición de riesgos proporcionales es empleada para modelar efectos de covariables sobre los tiempos de fallo de los individuos que son no curados, y proponen un método de estimación basado en el algoritmo EM. Recientemente, una generalización a estos modelos ha sido propuesta por Lu y Ying (2004), la cual consiste en emplear los modelos de transformación para la especificación del tiempo de fallo de los individuos susceptibles y un modelo logístico para modelar la fracción curable.

Esta tesis tiene como finalidad estudiar las propiedades asintóticas de dos clases de modelos semiparamétricos de regresión importantes en el análisis de supervivencia: el modelo de mezclas semiparamétrico para riesgos competitivos y el modelo de cura semiparamétrico basado en modelos de transformación. La primera clase corresponde a la especificación propuesta por Escarela y Bowater (2008) cuya formulación permite especificar por separado la probabilidad de experimentar cualquier tipo de fallo y el riesgo condicional para cada tipo de fallo. El estudio del segundo modelo tiene como finalidad presentar una generalización del modelo propuesto de Lu y Ying (2004) el cual tiene la flexibilidad de incluir efectos de covariables dependientes del tiempo y combina una regresión logística para la probabilidad de ocurrencia del evento con la clase de modelos de transformación para los tiempos de ocurrencia. Dicha generalización extiende varios modelos de cura bien establecidos en la literatura tales como el modelo de riesgos proporcionales (Farewell, 1982; Kuk y Chen, 1992; Sy y Taylor, 2000; Peng y Dear, 2000) y el modelo momios-proporcionales (Lu y Ying, 2004). Específicamente este trabajo estudia las propiedades asintóticas de los estimadores de máxima verosimilitud noparamétricos de las dos clases de modelos expuestas. Se revisa el método de estimación de máxima verosimilitud noparamétrica (NPML) y se muestra que la formulación de los modelos se presta al uso de herramientas de la teoría moderna de los procesos empíricos para demostrar las propiedades asintóticas de los estimadores resultantes.

Las dos clases de modelos antes descritas poseen parámetros infinito-dimensionales, lo cual plantea retos teóricos para el análisis estadístico. En este trabajo se extienden y se aplican las técnicas desarrolladas por Murphy (1994,1995) y Parner (1998) para modelos *frailty*, y se hace uso de la teoría moderna de procesos empíricos establecida por van der Vaart y Wellner (1996) para crear un camino fácil en el estudio de la existencia, la consistencia y la normalidad asintótica de los estimadores resultantes de ambas clases; además, se demuestra que los estimadores de máxima verosimilitud para los parámetros de regresión son semiparamétricamente eficientes (Bickel *et al.*, 1993). Finalmente, se obtienen estimaciones consistentes de las varianzas asintóticas tanto para la estimación de los parámetros de regresión, así como para la estimación del parámetro infinito-dimensional.

La tesis está organizada en dos partes. La primera parte comprende cuatro capítulos y está dedicada a dar una breve introducción de temas que serán de utilidad durante el desarrollo del trabajo. El Capítulo 1 presenta definiciones básicas del análisis de supervivencia, describe algunas funciones paramétricas de supervivencia importantes e introduce la notación de censura por la derecha. Aquí también se revisa la inferencia paramétrica en presencia de censura en los datos y finalmente se presenta el modelo de Cox. El Capítulo 2 exhibe definiciones y propiedades importantes de los procesos contables y martingalas. En el Capítulo 3 se describe de manera general las principales técnicas de inferencia semiparamétrica, enfatizando la eficiencia semiparamétrica. El Capítulo 4 está enfocado a explicar la teoría de procesos empíricos la cual es una parte fundamental para el estudio de las propiedades asintóticas de los estimadores de cada clase de modelo que aquí se estudia.

La segunda parte comprende dos capítulos en los cuales las dos clases de modelos antes mencionados son analizadas. En el Capítulo 5 se describe el modelo de mezclas semiparamétrico para riesgos competitivos propuesto por Escarela y Bowater (2008). Condiciones de regularidad y el enfoque de máxima verosimilitud son presentados. La teoría de procesos empíricos es aplicada para demostrar la consistencia y la normalidad asintótica de la estimación de los parámetros de máxima verosimilitud. Estimaciones consistentes de la varianza para los parámetros finito e infinito dimensionales son presentadas. Por último se muestran dos simulaciones que permiten comparar el ajuste del modelo de mezclas semiparámetrico de riesgos competitivos con respecto a un modelo completamente paramétrico.

En el Capítulo 6 se estudia al modelo de cura semiparamétrico contruido a partir de modelos de transformación. Se da una breve introducción de los modelos de cura y de los modelos de transformación lineales. Posteriormente se presenta una especificación del modelo de cura utilizando modelos de transformación el cual permite incluir variables dependientes del tiempo. Se demuestran las propiedades de identificabilidad, existencia, consistencia, normalidad asintótica y eficiencia de lo estimadores de máxima verosimilitud noparamétricos. Finalmente, se presenta un estimador consistente de la varianza tanto para los parámetros de regresión como para el parámetro infinito-dimensional.

El Capítulo 7 presenta conclusiones y perspectivas derivados del estudio de ambas clases de modelos semiparamétricos.

Notaciones

Aquí se presentan algunas notaciones generales que serán utilizadas a lo largo del trabajo.

- Las variables aleatorias serán consideradas sobre un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbb{P})$.
- $\mathbb{R}$ denota el conjunto de números reales.
- $\mathbb{R}^k$ denota el espacio vectorial de dimensión k.
- $|\cdot|$ denota la norma euclidiana sobre $\mathbb{R}^k$.
- $a \wedge b$ denota el mínimo entre a y b.
- $a \vee b$ denota el máximo entre a y b.
- Sea C un conjunto,
  - $l^\infty(C)$ denota el espacio de funciones uniformemente acotadas en C.
  - $VB(C)$ es el espacio de funciones de C en $\mathbb{R}$, que son acotadas y de variación acotada.
  - Si f es una función acotada de C en $\mathbb{R}$, $\|f\|_\infty = \sup_{t\in C} |f(t)|$.
- L_2 funciones cuadrado integrables.
- Sea A una matriz de dimensión $n \times p$,
  - A' denota la matriz transpuesta de A.
  - Si $n = p$ y A es invertible, A^{-1} denota la inversa de A.
- Si a es un vector de dimensión p, $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$ y $a^{\otimes 2} = aa'$.
- c.s. denota casi seguramente.
- $\xrightarrow{d}$ denota la convergencia en distribución.
- $o_p(1)$ denota el símbolo de orden estocástico.
- función càdlàg es una función que es continua por la derecha y tiene límite por la izquierda.

Parte I

Preliminares

Capítulo 1

Análisis de Supervivencia

Introducción

El análisis de supervivencia es el termino usado para describir el análisis de datos que corresponden a la duración de un fenómeno. Dicha duración comienza en un *tiempo origen* bien determinado y concluye con la ocurrencia de un evento. Aquí, las respuestas consisten en tiempos de duración o ciclos de vida, como pueden ser el tiempo de recurrencia de una enfermedad, el tiempo que dura la eficacia de una intervención, el tiempo de un aprendizaje determinado, etc. En investigación médica, por ejemplo, el tiempo de origen generalmente se define como el momento en que se recluta a un individuo para ser observado en un estudio experimental; el evento puede definirse como la muerte del individuo en este caso, los datos son literalmente tiempos de supervivencia. En muchas disciplinas es posible encontrar fenómenos que no necesariamente son de naturaleza fatal, tales como la duración de huelgas o periodos de desempleo, o también en procesos de control de calidad en los cuales se estudia el tiempo hasta que cierto producto falla (tiempo de fallo), o el tiempo de espera hasta recibir un servicio (tiempo de espera), etc.

Dos características que presentan los datos de supervivencia son que -en general- los datos no son distribuidos simétricamente. Típicamente, un histograma construido de los datos de supervivencia tiende a ser de *cola larga* a la derecha del intervalo que contiene la mayoría de las observaciones. En consecuencia, no es razonable suponer que los datos provengan de una población de distribución normal. Esta dificultad podría ser resuelta al transformar los datos para obtener una distribución simétrica. Sin embargo, una metodología más satisfactoria es la de adoptar un modelo alternativo cuya distribución es más apropiada para los datos en términos de bondad de ajuste. Además, los tiempos de supervivencia se encuentran *censurados*, es decir, que se tiene información incompleta de los tiempos de supervivencia. Existen varias categorias de censura tales como, la censura por la derecha (que es cuando la observación cesa antes de que se observe el suceso), la censura por la izquierda (que es cuando la observación no comienza hasta después de que se haya producido el suceso) y la censura por intervalos (que significa conocer que la supervivencia verdadera solo se ha producido en algún momento dentro de un intervalo de tiempo conocido). En ésta tesis se enfocara en censura por la derecha.

La intención de este capítulo es dar una breve introducción al modelado estadístico de datos de supervivencia. En la Sección 1 se presentan algunas definiciones relevantes de la teoría del análisis de supervivencia. Posteriormente, lo correspondiente a la censura y a la construcción de la función de verosimilitud cuando hay datos censurados son explicados en la Sección 2. Finalmente, el modelo de riesgos proporcionales de Cox es estudiado en la Sección 3.

1.1. Función de supervivencia y fuerza de mortalidad

En el análisis de supervivencia existen dos funciones de gran interés: la *función de supervivencia* y la *fuerza de mortalidad* las cuales serán estudiadas en esta sección. Para comenzar con la introducción al modelado estadísticos de datos de supervivencia, es pertinente considerar algunas características relevantes de las distribuciones de probabilidad para el análisis; para esto se partirá de la suposición de que la población es homogénea. Denotemos por t al tiempo de supervivencia actual de un individuo, el cual es visto como un valor de una variable aleatoria continua positiva T la cual está asociada con el tiempo para que ocurra el evento. En ésta tesis nos enfocamos al análisis de supervivencia para el caso contínuo. Sin embargo, la variable T puede ser una variable aleatoria discreta positiva y por tanto las siguientes definiciones y propiedades pueden ser ajustadas al caso discreto.

Supóngase que la variable aleatoria T tiene una distribución de probabilidad cuya *función de densidad* se denota con $f_T(t)$. La *función de distribución* de T está dada por:

$$F_T(t) = \mathbb{P}\{T \leq t\} = \int_0^t f(u)du,$$

y representa la probabilidad de que el evento de interés ocurra antes del tiempo t.

La *función de supervivencia* es definida como la probabilidad de que el tiempo para que ocurra el evento sea mayor que el tiempo t y se denota por:

$$S_T(t) = \mathbb{P}\{T > t\} = 1 - F(t).$$

Otra interpretación que se le puede dar a $S_T(t)$ es la proporción de individuos que sobreviven al tiempo t.

La *fuerza de mortalidad*, denotada por $\lambda_T(t)$, es la probabilidad de que un individuo muera en el tiempo t dado que ha sobrevivido hasta ese tiempo; es decir, $\lambda_T(t)$ representa la tasa instantánea de mortalidad para un individuo que sobrevive al tiempo t y es definida como:

$$\lambda_T(t) = \lim_{\Delta t \to 0} \left[\frac{\mathbb{P}\{t \leq T < t + \Delta t | T \geq t\}}{\Delta t} \right].$$

La función $\lambda_T(t)$ es también conocida conocida como la *tasa de mortalidad*, la *tasa instantánea de muerte* , o *función de mortalidad* o *función de riesgo*.

Algunas relaciones que pueden derivarse de éstas definiciones son:

$$\lambda_T(t) = \frac{f_T(t)}{S_T(t)}$$

y

$$\lambda_T(t) = -\frac{d}{dt}\{\log S_T(t)\}$$

De las expresiones anteriores, la función de supervivencia puede ser escrita en términos de la fuerza de mortalidad de la manera siguiente:

$$S_T(t) = \exp\{-\Lambda_T(t)\},$$

donde,

$$\Lambda_T(t) = \int_0^t \lambda_T(u)du,$$

es conocida como la *fuerza de mortalidad integrada* o *función de riesgo acumulada.* Esta función tiene su propia interpretación: ella mide la cantidad total de riesgo que se ha acumulado hasta el momento t.

Una función $\lambda_T(t)$ es una fuerza de mortalidad sí y sólo sí satisface las siguientes propiedades:

1. $\lambda_T(x) \geq 0$, para toda x.
2. $\int_0^\infty \lambda_T(x)dx = \infty$.

Observación: Nótese que es suficiente conocer una de estas cinco funciones ($S_T(t)$, $\lambda_T(t)$, $\Lambda_T(t)$, $F_T(t)$ y $f_T(t)$) para poder deducir completamente las otras cuatro. En el análisis de supervivencia la función $\lambda_T(t)$ es importante de estimar debido a que abastece una descripción probabilista del futuro inmediato de un sujeto que continúa bajo seguimiento.

A continuación se ejemplifican las definiciones anteriores con dos modelos paramétricos del análisis de supervivencia más utlizados en la literatura.

- **Modelo exponencial.** Supóngase que T tiene una distribución exponencial con media $1/\gamma$, entonces

$$S_T(t) = \exp\{-\gamma t\}, \quad f_T = \gamma \exp\{-\gamma t\}, \quad \lambda_T(t) = \gamma, \quad \Lambda_T(t) = \gamma t,$$

  donde $\gamma > 0$. La fuerza de mortalidad constante refleja la propiedad de la distribución exponencial conocida como la falta de memoria; esto es, si la variable aleatoria T se distribuye exponencial con media $1/\gamma$, $T \sim \text{EXP}(\gamma)$ entonces,

$$\mathbb{P}\{T > a + t | T > a\} = \mathbb{P}\{T > t\} \qquad \text{para toda } a > 0 \text{ y } t > 0.$$

  El modelo exponencial es ampliamente utilizado en el dominio de confiabilidad (análisis estadístico para el estudio de duraciones de vida en el contexto industrial). Este modelo es estudiado con este fin por Meeker y Escobar (1998) y por Voivnov y Nikulin (1993). Este modelo fue históricamente uno de los primeros modelos en el análisis de supervivencia (ver Johonson *et al*, 1994; Klein y Moeschberger, 1997; Lawless, 1982).

- **Modelo Weibull.** Supóngase que T se distribuye Weibull con parámetro de escala γ y parámetro de forma α, i.e. $T \sim WEI(1/\gamma, \alpha)$, cuya función de densidad es

$$f_T(t) = \alpha \gamma^\alpha t^{\alpha - 1} \exp\{-(\gamma t)^\alpha\}, \qquad \alpha, \gamma > 0.$$

  La fuerza de mortalidad condicional y la función de supervivencia asociadas son

$$\lambda_T(t) = \alpha \gamma (\gamma t)^{\alpha - 1}$$
$$S_T(t) = \exp\{-(\gamma t)^\alpha\}.$$

  La conveniencia de utilizar el modelo Weibull en aplicaciones es por la flexibilidad de las funciones $S_T(t)$ y $\lambda_T(t)$. Nótese que cuando $\alpha = 1$ entonces se obtiene el modelo exponencial, mientras que si $\alpha < 1$ ($\alpha > 1$) entonces la función $\lambda_T(t)$ es decreciente (creciente) en el tiempo.
  Este modelo es utilizado en los dominio de confiabilidad y en el análisis de supervivencia en medicina, ejemplos de aplicación para este modelo podrán ser encontrados en Johonson *et al.* (1994).

Otros modelos paramétricos de supervivencia que pueden ser encontrados en la literatura son: el modelo del valor extremo, el modelo de Gompertz-Makeham, el modelo lognormal, el modelo de pedazos exponeciales, entre otros (ver, Cox y Oakes, 1984; Klein y Moeschberger, 1997; Johonson *et al.*,1994; Lawless, 1982; Meeker y Escobar,1998 y Voivnov y Nikulin, 1993).

1.2. Mécanismos de censura

El análisis de supervivencia posee problemas particulares debido a que las observaciones de duraciones de vida son censuradas la mayoría de las veces, es decir, privan una parte de información que ellas contienen. Los tipos de censura más comunes encontrados en la literatura son (ver Cox y Oakes, 1984):

- **Censura tipo I.** El suceso de interés se observa si ocurre antes de un instante de tiempo fijo predeterminado C . En este caso, C es una constante (de censura) prefijada por el investigador para todas las unidades muestrales.

- **Censura tipo II.** En este caso el investigador decide prolongar el periodo de observación hasta que ocurran k fallas de n posibles ($k < n$), registrando este último valor de falla para el resto de los individuos (censuras) que no observó. Una razón común para determinar el número de fallas a observar es la potencia que se requiere para el estudio. En estos dos casos, la censura está controlada por el investigador.

- **Censura aleatoria.** En este tipo de censura el investigador no tiene ningún control sobre la misma. Las censuras pueden ocurrir porque el individuo abandona el estudio, muere por una causa que no es de interés o permanece vivo al término del mismo.

Un individuo que entra a un estudio en el tiempo t_0, muere en el tiempo $t_0 + t$. Sin embargo, si el tiempo de supervivencia correspondiente es censurado, t es desconocido, ya sea porque el individuo sigue vivo o porque se le ha perdido de vista. Si el individuo fue visto con vida por última vez en el tiempo $t_0 + C$, el tiempo C es conocido como el tiempo de supervivencia censurado por la derecha. Los modelos de censura incluyen la variable T asociada a la ocurrencia del evento y una variable aleatoria C con valores reales en $\mathbb{R}^+ \cup \{\infty\}$ llamada censura. En el modelo de *censura por la derecha*, se observa $X = T \wedge C$, y por tanto se define la indicadora $\delta = 1_{\{T \leq C\}}$, la cual toma el valor 1 si $T \leq C$ y 0 en otro caso. Análogamente, en la *censura por la izquierda* se observa $Z = T \vee C$.

Otros tipos de censura pueden ser encontrados con mas detalle en Bagdonavičius y Nikulin (capítulo 8) y Lawless (capítulo 1).

Observación. En este estudio el principal supuesto de la censura es que el tiempo de censura C y el tiempo de ocurrencia T son idependientes. Es decir, se supone que los sujetos censurados tienen la misma curva de supervivencia subyacente después del tiempo de censura que los pacientes no censurados. Este estudio se restringe al estudio de mecanismos de censura por la derecha y entre éstas a censuras de tipo 1.

1.3. Máxima verosimilitud en un modelo de supervivencia paramétrico

En esta sección se presenta el método de máxima verosimilitud para un modelo de análisis de supervivencia con censura a la derecha. Para más detalles de este tema se pueden consultar, Bagdonavičius y Nikulin (capítulo 4), Kalbfleisch y Prentice (capítulo 3), Lawless (capítulo 3-6), Meeker y Escobar (capítulos 7, 8, 11), entre otros.

Sea T una variable aleatoria que representa la duración de vida. Supongamos que la medida $P_T = P_{\theta_0}$ de T pertenece a un conjunto de posibles medidas de probabilidad $\mathcal{P} = \{P_\theta : \theta \in \Theta \subseteq \mathbb{R}^p\}$, con $\theta_0 \in \Theta$. Aquí, suponemos que Θ es un abierto en $\mathbb{R}^p$. Sean $f_T(t;\theta)$, $F_T(t;\theta)$, $S_T(t;\theta)$ y $\lambda_T(t;\theta)$, la función de densidad, la función de distribución, la función de supervivencia y la función de riesgo instantáneo de T respectivamente indexadas por la medida P_θ.

Supóngase que los datos están constituidos por n parejas de observaciones, donde el par correspondiente a la i-ésima observación, $i = 1, ..., n$, es (t_i, δ_i), donde $t_i = T_i \wedge C_i$, C_i es la variable que representa el tiempo de censura y $\delta_i = 1_{\{T_i \leq C_i\}}$. Una observación con muerte en t contribuye a la verosimilitud con $f(t)$, la densidad evaluada en t, mientras que la contribución de una observación cuyo tiempo de supervivencia tiene censura C (en este caso $T > C$), y por tanto contribuye a la verosimilitud con $\mathbb{P}(T > C) = S(C)$, es decir, la probabilidad de supervivencia en C. La función de verosimilitud total para n observaciones independientes es entonces,

$$L_n(\theta) = \prod_{i=1}^{n} \{f_T(t_i;\theta)\}^{\delta_i} \{S_T(t_i;\theta)\}^{1-\delta_i}.$$

Estimar los parámetros desconocidos en esta función de verosimilitud son encontrados maximizando la función logaritmo de la función de verosimilitud. Una expresión alternativa para la función de verosimilitud puede ser obtenida reescribiendo la expresión anterior en la siguiente forma:

$$\begin{aligned} L_n(\theta) &= \prod_{i=1}^{n} \left\{ \frac{f_T(t_i;\theta)}{S_T(t_i;\theta)} \right\}^{\delta_i} S_T(t_i;\theta) \\ &= \prod_{i=1}^{n} \{\lambda_T(t_i)\}^{\delta_i} S_T(t_i;\theta). \end{aligned}$$

Esta versión de la función de verosimilitud es particularmente útil cuando la función de densidad tiene una forma complicada, como ocurre a menudo.

1.4. El modelo de riesgos proporcionales de Cox

En el modelado estadístico de datos de supervivencia no sólo la relación entre la tasa de supervivencia y el tiempo es importante, sino también su posible relación con diferentes variables explicativas para cada individuo. Una forma conveniente de hacerlo es expresar a la fuerza de mortalidad como función del tiempo y de variables explicativas. La idea fundamental es la misma que en cualquier modelo de regresión.

El modelo más utilizado para incluir información concomitante es el modelo de riesgos proporcionales de Cox (1972), el cual expresa a la fuerza de mortalidad como el producto de una fuerza de mortalidad base y una función liga que depende de las variables explicativas. Esto es, considérese el vector de variables explicativas $\mathbf{x}$ y el correspondiente vector de coeficientes β, entonces bajo esta especificación

$$\lambda(t; \beta, \mathbf{x}) = \psi(\beta, \mathbf{x})\lambda_0(t),$$

donde $\psi(\beta, \mathbf{x})$ es la función que liga las variables explicativas con el coeficiente de la fuerza de supervivencia y $\lambda_0(t)$ es una fuerza de mortalidad base, que corresponde a la de un grupo o subgrupo de referencia cuando $\psi(\cdot) = 1$. Obsérvese que la fuerza de mortalidad base no tiene una parametrización particular.

Como el riesgo relativo $\psi(\beta, \mathbf{x})$ no puede ser negativo, es conveniente definirlo como $\psi(\beta, \mathbf{x}) = \exp(\beta'\mathbf{x})$. De esta manera, el modelo de riesgos proporcionales de Cox es expresado como:

$$\lambda(t; \mathbf{x}) = \exp(\boldsymbol{\beta}'\mathbf{x})\lambda_0(t).$$

Con la especificación de riesgos proporcionales, se tiene

$$\log \frac{\lambda(t; \beta, \mathbf{x}}{\lambda_0(t)} = \beta'\mathbf{x}$$

es decir, el modelo plantea el logaritmo del riesgo relativo como una función lineal de las covariables. Por lo tanto, el riesgo relativo a diferencia del riesgo propiamente dicho, no depende del tiempo o, dicho de otra manera, es constante a lo largo del tiempo (de ahí el nombre de modelo de riesgo proporcional).

Utilizando el modelo de riesgos proporcionales la función de supervivencia correspondiente está dada por:

$$S(t; \mathbf{x}) = \exp\{-\Lambda_0(t) \exp(\boldsymbol{\beta}'\mathbf{x})\}$$

donde $\Lambda_0(t) = \int_0^t \lambda_0(u)du$ es la fuerza de mortalidad integrada base.

1.4.1. Estimación del modelo de riesgos proporcionales

La estimación del modelo de riesgos proporcionales de Cox dado un conjunto de observación de datos de supervivencia implica la estimación del vector de coeficientes desconocidos β en el componente lineal del modelo así como también de la fuerza de mortalidad base, $\lambda_0(t)$

En lo que corresponde al proceso de estimación para este modelo, Cox (1972) presenta el método de verosimilitud parcial para estimar β, el cual se basa en el producto de las verosimilitudes de todos los cambios ocurridos. Primero, el vector β es estimado y estas estimaciones son utilizadas para la estimación de la fuerza de mortalidad base.

Supóngase que los datos están compuestos por n individuos, de los cuales existen r tiempos de muertes diferentes. Por el momento, se supondrá que solo un individuo muere en cada tiempo. Los r tiempos de muerte ordenados se denotan con $t_1 < t_2 ... < t_r$. El conjunto de individuos que se encuentran en riesgo en el tiempo t_j se denotará con $R(t_j)$, así que $R(t_j)$ es el conjunto de individuos que se encuentran con vida y sin censura antes de t_j. La cantidad $R(t_j)$ es conocida como el conjunto en riesgo. Cox (1972) demostró que la función de verosimilitud parcial relevante para el modelo de riesgos proporcionales está dada por,

$$L(\beta) = \prod_{j=1}^{r} \frac{\exp(\beta' \mathbf{x}_j)}{\sum_{l \in R(t_j)} \exp(\beta' \mathbf{x}_l)},$$

donde $\mathbf{x}_j$ es el vector de variables explicativas para el individuo que muere en t_j. La suma del denominador de la función de verosimilitud corresponde a los valores de $\exp\{\beta' \mathbf{x}\}$ de todos los individuos que se encuentran en riesgo en el tiempo t_j. Los individuos con tiempos censurados no tienen contribución en el numerador, su contribución se encuentra en la suma de los individuos en riesgo del numerador.

Observación. Esta función de verosimilitud parcial tiene todas las propiedades (estimaciones consistentes y eficientes) que posee la función de verosimilitud ordinaria (e.g. Efron, 1977, Andersen y Gill, 1982). La principal justificación para su uso es que si uno está interesado en la estimación de los efectos de las covariables, la información que para ello hay entre dos tiempos de muerte consecutivos es despreciable. Una condición para utilizar la función de verosimilitud parcial es que los riesgos mantengan la propiedad de proporcionalidad.

Capítulo 2

Teoría Probabilistica

2.1. Preliminares

Los tiempos de supervivencia pueden ser representados a través de ciertos procesos estocásticos. Los datos en sí pueden ser descritos como un proceso contable, el cual es simplemente una función aleatoria $N(t)$. Esta función es cero en el tiempo inicial y constante en el tiempo excepto en los tiempo de ocurrencia en donde hace saltos de tamaño 1. La razón de usar procesos contables y martingalas es porque éstos métodos proporcionan formas directas de estudiar la propiedades de muestras grandes de los estimadores tanto para modelos no-paramétricos así como también para modelos semiparamétricos.

Antes de dar las definiciones y propiedades más importantes de los procesos contables y de las martingalas, se hará una breve introducción a la teoría de procesos estocásticos.

Toda la teoría presentada en este capítulo es desarrollada sobre un espacio medible $(\Omega, \mathcal{F}, \mathbb{P})$, donde $\mathcal{F}$ es una σ-álgebra y $\mathbb{P}$ es una medida de probabilidad definida en $\mathcal{F}$.

Definición 2.1.1. *Un procesos estocástico es una familia de variables aleatorias indexadas por el tiempo* $\{X(t) : t \in \Gamma)$ *las cuales son definidas sobre el mismo espacio de probabilidad* $(\Omega, \mathcal{F}, P)$.

El mapeo $t \to X(t, \omega)$, donde $\omega \in \Omega$, es llamado una dirección simple o trayectoria de X. El proceso estocástico X induce una familia creciente de sub-σ-álgebras,

$$\mathcal{F}_t^X = \sigma\{X(s) : 0 \leq s \leq t\}$$

llamada la *historia interna* de X.

Definición 2.1.2. *Una historia o filtración* $(\mathcal{F}_t; t \geq 0)$ *es una familia de sub-*σ*-álgebras tal que, para todo* $s \leq t, \mathcal{F}_s \subset \mathcal{F}_t$*, lo cual significa que si* $A \in \mathcal{F}_s$ *entonces* $A \in \mathcal{F}_t$*.*

Algunas veces, las filtraciones pueden ser combinadas; es decir, sean dos filtraciones $(\mathcal{F}_t^1)$ y $(\mathcal{F}_t^2)$, entonces $\mathcal{F}_t^1 \vee \mathcal{F}_t^2$ denota la filtración más pequeña que contiene a $\mathcal{F}_t^1$ y $\mathcal{F}_t^2$.

Definición 2.1.3. *Un proceso estocástico X es adaptado a la filtración* $(\mathcal{F}_t)$ *si, para cada* $t \geq 0$*,* $X(t)$ *es* $\mathcal{F}_t$*-medible, y en este caso* $\mathcal{F}_t^X \subset \mathcal{F}_t$*.*

Una variable aleatoria nonegativa T es llamada *tiempo de paro* con respecto a la filtración $(\mathcal{F}_t)$ si $(T \leq t) \in \mathcal{F}_t$ para todo $t \geq 0$. Para un proceso estocástico X y un tiempo de paro T, el proceso de paro X^Tes definido $X(t) = X(t \wedge T)$.

2.2. Martingalas

Las martingalas juegan un papel importante en aplicaciones estadísticas. En esta sección se presentan definiciones básicas de martingalas que serán utilizadas más adelante.

Definición 2.2.1. *Una martingala respecto a la filtración* $(\mathcal{F}_t)$ *es un proceso estocástico continuo por la derecha M con límite por la izquierda que satisface las siguientes condiciones,*

- *M es adaptado a* $\mathcal{F}_t$
- $E|M(t)| < \infty$ *para todo t*
- $$E(M(t)|\mathcal{F}_s) = M(s) \qquad \textit{para todo } s \leq t. \tag{2.1}$$

Obsérvese que la ecuación (2.1) es equivalente a

$$E(dM(t)|\mathcal{F}_{t-}) = 0 \qquad \text{para todo } t \geq 0. \tag{2.2}$$

donde $\mathcal{F}_{t-}$ es la σ-álgebra más pequeña que contiene todas las $\mathcal{F}_s$, $s < t$ y $dM(t) = M((t + dt)-) - M(t-)$. Por lo tanto, una martingala tiene media cero en los incrementos dado el pasado. La segunda condición de la definición anterior garantiza que M es *integrable*.

Si M satisface

$$E(M(t)|\mathcal{F}_s) \geq M(s) \qquad \text{para todo } s \leq t,$$

en vez de la condición (2.1), entonces M es una submartingala. De manera similar, si M satisface

$$E(M(t)|\mathcal{F}_s) \leq M(s) \qquad \text{para todo } s \leq t,$$

en vez de la condición (2.1), entonces M es una supermartingala.

Una martingala es llamada *cuadrado integrable* si $\sup_t E(M(t)^2) < \infty$. Una *martingala local* M es un proceso en el que existe una sucesión de tiempos de paro (T_n) tal que para cada n, M^{T_n} es una martingala. Si además, M^{T_n} es una martingala cuadrado integrable, entonces M es llamada una *martingala cuadrado integrable local*.

Para ser capaces de formular la descomposición de Doob-Meyer, se necesita introducir la definición de proceso previsible.

Definición 2.2.2. *Un proceso X es previsible sí y sólo si $X(T)$ es $\mathcal{F}_t$-medible para todo tiempo de paro T.*

Sea X un proceso adaptado càdlàg, entonces A es llamado *compensador* de X si A es un proceso previsible, càdlàg y de variación finita tal que $X - A$ es una martingala local de media cero. Si un compensador existe, este es único.

2.3. Procesos contables

Un acercamiento alternativo para desarrollar los procedimientos de inferencia para datos censurados es utilizar la teoría de procesos contables. Este acercamiento fue desarrollado primero por Aalen (1975) y consiste en combinar elementos de integración estocástica, la teoría de martingalas continuas y la teoría de procesos contables dentro de una metodología que permite desarrollar fácilmente las técnicas de inferencia para el análisis de supervivencia con datos cesurados. Para más detalles de éste tema véase Andersen *et al.* (1993) y Fleming y Harrington (1991).

Definición 2.3.1. *Un proceso contable $\{N(t)\}$ es un proceso estocástico el cual es adaptado a la filtración $(\mathcal{F}_t)$, càdlàg, con $N(0) = 0$ y $N(t) < \infty$ c.s. y cuyas trayectorias son constantes entre los eventos y saltos de tamaño 1 en los tiempos de ocurrencia.*

Dada una muestra aleatoria con censura a la derecha, los procesos $N_i(t) = I\{T_i \leq t, \delta_i = 1\}$ son procesos contables, los cuales son ceros hasta que el individuo i muere y entonces tiene un salto de tamaño 1. El proceso $N(t) = \sum_{i=1}^{n} N_i(t) = \sum_{t_i \leq t} \delta_i$ es también un proceso contable el cual cuenta el número de muertes al tiempo anterior o igual a t.

En el caso de datos con censura a la derecha, la filtración al tiempo t, $\mathcal{F}_t$, consiste en el conocimiento de los pares (T_i, δ_i) siempre que $T_i \leq t$ para los individuos que continuan en el estudio al tiempo t. Aquí se denotará la filración a un instante antes de t por $(\mathcal{F}_{t-})$. La filtración $\{\mathcal{F}_t, t \geq 0\}$ para un problema dado depende de las observaciones del proceso contable.

Para datos censurados a la derecha, si los tiempos de muerte T_i y los tiempos de censura C_i son independientes, entonces el cambio de un evento al tiempo t, dada la historia antes de t, está dado por

$$\mathbb{P}[t \leq T_i \leq t + dt, \delta_i = 1 | \mathcal{F}_{t-}]$$

$$= \begin{cases} \mathbb{P}[t \leq T_i \leq t + dt, C_i > t + dt_i | T_i \geq t, C_i \geq t] = \lambda(t)dt & \text{si } T_i \geq t \\ 0 & \text{si } T_i < t \end{cases}.$$

Para un proceso contable dado, defínase $dN(t)$ como el cambio en el proceso $N(t)$ sobre un intervalo de tiempo corto $[t, t + dt)$; aquí, $dN(t)$ es uno si la muerte ocurre en t o 0, en otro caso. Si definimos el proceso $Y(t)$ como el número de individuos con un tiempo de estudio $T_i \geq t$, entonces

$$E(dN(t)|\mathcal{F}_{t-}) = Y(t)\lambda(t)dt.$$

El proceso $h(t) = Y(t)\lambda(t)$ es llamado el *proceso intensidad* de un proceso contable. $h(t)$ es en sí un proceso estocástico que depende de la información contenida en la historia $\mathcal{F}_t$ a través de $Y(t)$. El proceso estocástico $Y(t)$ proporciona el número de individuos en riesgo en un momento dado.

Para el caso contínuo, defínase el proceso

$$H(t) = \int_0^t h(s)ds,$$

el cual es llamado *proceso de intensidad acumulativo* y que tiene la siguiente propiedad:

$$E(N(t)|\mathcal{F}_{t-}) = E(H(t)|\mathcal{F}_{t-}) = H(t).$$

Esta última igualdad se cumple porque una vez que conocemos la historia justo antes de t, el valor de $Y(t)$ es fijo y entonces $H(t)$ es no aleatoria.

Con las definiciones dadas en las secciones anteriores, podemos establecer el Teorema de descomposición de Doob-Meyer. Este teorema establece que para cualquier submartingala N continua por la derecha existe un único proceso previsible continuo por la derecha H tal que $H(0) = 0$ y $M + H$ es una martingala.

Teorema 2.3.1 (Descomposición de Doob-Meyer)**.** *(Fleming y Harrington, 1991 p. 37) Sea N una submartingala no negativa y contínua por la derecha con respecto a la base estocástica* $(\Omega, \mathcal{F}, \{\mathcal{F} : t \geq 0\}, \mathbb{P})$*, entonces existe una martingala M contínua por la derecha y un proceso previsible creciente y contínuo por la derecha H tal que $E(H(t)) < \infty$ y*

$$N(t) = M(t) + H(t) \qquad c.s.$$

para cada $t \geq 0$. *Si* $H(0) = 0$ *c.s. y si* $N = M' + H'$ *es otra descomposición tal que* $H'(0) = 0$, *entonces para cualquier* $t \geq 0$,

$$P\{M'(t) \neq M(t)\} = 0 = P\{H'(t) \neq H(t)\}.$$

Si además N *es acotado, entonces* M *es uniformemente integrable y* H *es integrable.*

Capítulo 3

Modelos semiparamétricos

En este capítulo se presentan las principales ideas y técnicas de la inferencia semiparamétrica, haciendo énfasis en la eficiencia semiparamétrica. Para más detalles de la teoría aquí presentada véase Bickel *et. al.* (1993),Tsiatis (2006) and van der Vaart (1998).

Un *modelo semiparamétrico* $\mathcal{P} = \{P_\theta\}$ es un modelo estadístico el cual tiene uno o más parámetros de dimensión finita y uno o más parámetros de dimensión infinita (por ejemplo, una función real-valuada). Tales modelos pueden ser paramatrizados de la forma $\theta = (\eta, \Lambda(\cdot)) \rightarrow P_{\eta,\Lambda(\cdot)}$, donde η corresponde al parámetro finito-dimensional y $\Lambda(\cdot)$ se encuentra en un conjunto infinito-dimensional y es considerado como un parámetro de ruido. Denótese por $\hat{\theta}_n$ la estimación de θ.

Por ejemplo, considérese el modelo de riesgos proporcionales de Cox. Éste modelo especifica la función de distribución de la variable T (el tiempo de fallo) como,

$$F(t) = 1 - \exp\left(-\int_0^t \lambda_0(u) e^{\beta'\mathbf{x}} \mathrm{d}u\right),$$

donde $\mathbf{x}$ es el vactor de covariables de dimensión q. En este caso $\theta = (\beta, \lambda_0(u))$. Notése que β es el vector de parámetros desconocidos de dimension q y $\lambda_0(u)$ es una función desconocida no negativa la cual es infinito-dimensional.

Obsérvese que la estimación del modelo semiparamétrico $\mathcal{P}$ es más dificil que la estimación bajo cualquier submodelo paramétrico. Se dice que $\mathcal{P}_0$ es un submodelo paramétrico del modelo semiparamétrico $\mathcal{P}$ si satisface las siguientes condiciones:

- $\mathcal{P}_0 \subset \mathcal{P}$.
- El sumbodelo paramétrico $\mathcal{P}_0$ contiene el valor verdadero del parámetro.

Para cada submodelo paramétrico se puede calcular la información de Fisher de la estimación de θ. Por tanto la información de $\hat{\theta}_n$ no es más grande que el infimo de la información sobre todos los submodelos parametricos y por tanto a $\hat{\theta}_n$ se dice que es *semiparametricamente eficiente*.

sualmente en los modelos semiparamétricos es suficiente considerar submodelos paramétricos unidemsionales de la forma $\{P_{\eta+ta,\Lambda_t}\}$, los cuales son diferenciable en media cuadrática. Entonces, se puede obtener que

$$\frac{\partial}{\partial t} P_{\eta+ta,\Lambda_t}\Big|_{t=0} = a' S_{\eta,\Lambda} + S_{\Lambda}$$

donde $S_{\eta,\Lambda}$ es la función de score para η y S_{Λ} es la función de score para Λ y es considerada como el conjunto tangente $\dot{\mathcal{P}}$ para Λ.

Sea $\prod_{\eta,\Lambda}$ la proyección ortognal de $\dot{\mathcal{P}}$ en L_2. Entonces la función de score eficiente para η es,

$$\tilde{S}_{\eta,\Lambda} = S_{\eta,\Lambda} - \prod_{\eta,\Lambda} S_{\eta,\Lambda}$$

y su matriz de covarianza es definida por

$$\tilde{I}_{\eta,\Lambda} = P_{\eta,\Lambda} \tilde{S}_{\eta,\Lambda} \tilde{S}'_{\eta,\Lambda}.$$

Denótese por $\psi_{\eta,\Lambda} = \tilde{I}^{-1}_{\eta,\Lambda} \tilde{S}_{\eta,\Lambda}$ la función de influencia. Entonces para una muestra aleatoria, $X_1, ..., X_n$ independientes e identicamentes distribuidas, se dice que el estimador es *(*asintóticamente eficiente) para η si $\hat{\eta}_n$ satisface

$$\begin{aligned} \sqrt{n}(\hat{\eta}_n - \eta) &= \sqrt{n}\mathbb{P}_n \psi_{\eta,\Lambda} + o_P(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{\eta,\Lambda}(X_i) + o_P(1). \end{aligned} \tag{3.1}$$

Capítulo 4

Procesos Empíricos

En este capítulo se da una breve introducción de los resultados básicos de la teoría de los procesos empíricos. Las técnicas aquí presentadas son parte fundamental para demostrar las propiedades asintóticas de los estimadores máxima verosimilitud de modelos semiparamétricos. Para mas detalles de la teoría aquí presentada véase, Huber y Lecoutre (1989), van der Vaart y Wellner (1996) y van der Vaart (1998).

4.1. Introducción a los procesos empíricos

Un *proceso empírico* es un proceso estocástico basado en una muestra aleatoria $X_1, ..., X_n$ i.i.d sobre un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbb{P})$.

Definición 4.1.1. *Para cada $\omega \in \Omega$ y $n \geq 1$, la medida empírica es definida como*

$$\mathbb{P}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i(\omega)},$$

donde δ_x es la medida de Dirac en el punto a esto es,

$$\delta_x(A) = \begin{cases} 1 & \textit{if } x \in A \\ 0 & \textit{if } x \notin A \end{cases},$$

para cualquier conjunto medible A.

Sea $\mathbb{P}_n(A)$ la medida empírica sobre el conjunto de borel A. La medida empírica de una muestra de elementos aleatorios $X_1, ..., X_n$ en un espacio medible $(\Omega, \mathcal{A})$ es la medida discreta dada por

$$\mathbb{P}_n(A) = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}(A) = \frac{card\{X_i \in A : i = 1, \ldots, n\}}{n}.$$

Sea $\mathbb{F}_n$ la función de distribución (aleatoria) definida por

$$\mathbb{F}_n(x) = \mathbb{F}_n(x)(\omega) = \mathbb{P}_n(]-\infty, x])(\omega),$$

for all $x \in \mathbb{R}$ and $\omega \in \Omega$. Obsérvese que

$$\mathbb{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1\{X_i \leq x\}, \quad x \in \mathbb{R}.$$

La función $\mathbb{F}_n$ es llamada la función de distribución empírica y el correspondiente proceso empírico es $\sqrt{n}(\mathbb{F}_n - F)$. La función de distribución empírica es el estimador natural de F si esta es completamente desconocida.

Para cada $x \in \mathbb{R}$ y por la Ley de los Grandes Números se tiene que

$$\mathbb{F}_n(x) \overset{c.s.}{\to} F(x).$$

Y el teorema del límite central permite demostrar que el estimador es asintóticamente normal, esto es

$$\sqrt{n}(\mathbb{F}_n(x) - F(x)) \to N\left(0, F(x)(1 - F(x))\right).$$

Dos resultados básicos que conciernen a los términos $\mathbb{F}$ y $\sqrt{n}(\mathbb{F}_n(x) - F(x))$ son el *teorema de Glivenko-Cantelli* y el *teorema de Donsker*. El primer resultado extiende la Ley de los Grandes Números y da la convergencia uniforme.

Teorema 4.1.1 (Glivenko-Cantelli)**.** *(Ver van der Vaart 1998, p.266). Sean $X_1, \ldots, X_n$ variables aleatorias real valuadas i.i.d. definidas sobre un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbb{P})$ con función de distribución F. Entonces*

$$||\mathbb{F}_n - F||_\infty = \sup_{x\in\mathbb{R}} |\mathbb{F}_n(x) - F(x)| \overset{c.s.}{\to} 0.$$

El teorema de Donsker da la convergencia en distribución del proceso empírico $\sqrt{n}(\mathbb{F}_n - F)$ a un proceso gaussiano.

Teorema 4.1.2 (Donsker)**.** *(Ver van der Vaart 1998, p.266). Sean $X_1, \ldots, X_n$ variables aleatorias real valuadas i.i.d. definidas sobre un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbb{P})$ con función de distribución F. Entonces la sucesión de procesos empíricos $\sqrt{n}(\mathbb{F}_n - F)$ converge en distribución en el espacio $D[-\infty, \infty]$ de funciones càdlàg a un proceso gausiano centrado $\mathbb{G}_F$, donde la función de covarianza en el punto (s, t) está dada por*

$$F(s \wedge t) - F(s)F(t).$$

El proceso $\mathbb{G}_F$ es conocido como puente Browniano. La notación de convergencia en distribución del proceso $\sqrt{n}(\mathbb{F}_n - F)$ a $\mathbb{G}_F$ depende de la topología del espacio. Para más detalle del estudio de la convergencia en distribución del proceso empírico real se pueden consultar los artículos de Donsker (1951, 1952) y Doob (1949) y el libro de Billingsley (1968). Este último estudia la convergencia en distribución en espacios métricos.

Existen generalizaciones de los teoremas 4.1.1 y 4.1.2 para un conjunto de funciones medibles, lo que permite obtener las *Clases Glivenko-Cantelli* y las *Clases Donsker* de funciones. Ambas clases intervienen en el estudio de procesos indexados por un conjunto de funciones. A continuación se presentan algunas notaciones de éstas clases de funciones que seran usadas a través de los siguientes capítulos (ver van der Vaart and Wellner, 1996).

Sea $X_1, ..., X_n$ una muestra aleatoria de una distribución de probabilidad P en un espacio medible (χ, ε). Para una función medible real-valuada definida en (χ, ε), se escribe

$$\mathbb{P}_n f = \frac{1}{n}\sum_{i=1}^{n} f(X_i).$$

Entonces $\{\mathbb{P}_n(f) : f \in \mathcal{F}\}$ es la medida empírica indexada por $\mathcal{F}$, mientras que $\{\mathbb{G}_n(f) : f \in \mathcal{F}\}$ es el proceso empírico indexado por $\mathcal{F}$, donde

$$\mathbb{G}_n f = \sqrt{n}(\mathbb{P}_n - P)f = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} f(X_i) - Pf\right).$$

Si $f \in L_1(P)$, entonces $Pf = \int f dP < \infty$ y por la Ley de los Grandes Números se sique que

$$\mathbb{P}_n(f) \xrightarrow{c.s.} P(f). \tag{4.1}$$

Supóngase que $\mathcal{F}$ es una colección de funciones real-valuadas tales que $f : \chi \to \mathbb{R}$. Si la convergencia en la ecuación (4.1) se satisface uniformemente sobre $f \in \mathcal{F}$, es decir,

$$||\mathbb{P}_n f - Pf||_{\mathcal{F}} = \sup_{f\in\mathcal{F}} |\mathbb{P}_n f - Pf| \xrightarrow{c.s.} 0,$$

entonces $\mathcal{F}$ es una *clase Glivenko-Cantelli*.

Si $f \in L_2(P)$, es decir, $Pf^2 = \int f^2 dP < \infty$, entonces

$$\sqrt{n}(\mathbb{P}_n - P)(f) \xrightarrow{d} N\left(0, P(f - Pf)^2)\right), \tag{4.2}$$

por el Teorema del Límite Central. Supóngase que $\mathcal{F}$ es una colección de funciones real-valuadas tales que $f : \chi \to \mathbb{R}$. Si la convergencia en la ecuación (4.2) se satisface uniformemente sobre $f \in \mathcal{F}$, entonces

$$\sqrt{n}(\mathbb{P}_n - P)(f) \Rightarrow G(f) \quad \text{in } \ell^{\infty}(\mathcal{F})$$

donde G es un procesos de puente Browniano con media cero y función de covarianza

$$cov(G(f), G(g)) = Pfg - PfPg$$

y ℓ^∞ es definida como

$$\ell^\infty(\mathcal{F}) = \left\{ x : \mathcal{F} \to \mathbb{R} | \|x\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |x(f)| < \infty \right\}.$$

Entonces $\mathcal{F}$ es una *clase Donsker*.

4.2. Ejemplo de Clases Donsker

Existen diversos métodos que permiten determinar si una clase de funciones es Donsker, van der Vaart (1998) y van der Vaart y Wellner (1996) presentan los siguientes ejemplos.

Ejemplo 1. Si $\mathcal{F}$ es igual a la colección de funciones indicadoras de la forma $f_t = 1(-\infty, t]$ con $t \in \mathbb{R}$. Entonces $\mathcal{F}$ es una clase Donsker (ver ejemplo 19.6 en van der Vaart, 1998).

Ejemplo 2. El conjunto de funciones uniformemente acotadas y de variación uniformente acotada es Donsker (ver Teorema 2.7.1 en van der Vaart y Wellner, 1996).

Ejemplo 3. La clase de funciones cuyas derivadas de orden k existen y son uniformemente acotadas por constantes M_k es Donsker (ver ejemplo 19.9 en van der Vaart, 1998).

Sin embargo, se puede construir clases Donsker a partir de otras clases Donsker. Por ejemplo,

Ejemplo 4. Si $\mathcal{F}$ es una clase Donsker tal que $\|P\|_{\mathcal{F}}$ y sea ϕ una función Lipschitz, entonces la clase de funciones de la forma $\phi(f)$ es una clase Donsker (ver Teorema 2.10.6 en van der Vaart y Wellner, 1996).

Ejemplo 5. Si $\mathcal{F}$ y $\mathcal{G}$ son clases Donsker y $\|P\|_{\mathcal{F} \cup \mathcal{G}} < \infty$, entonces los pares $\mathcal{F} \wedge \mathcal{G}$, $\mathcal{F} \vee \mathcal{G}$ y $\mathcal{F} + \mathcal{G}$ son clases Donsker (ver ejemplo 2.10.7 in van der Vaart y Wellner, 1996).

Ejemplo 6. Si la clase $\mathcal{F}$ y $\mathcal{G}$ son clases Donsker y uniformemente acotadas, entonces $\{fg : f \in \mathcal{F}, g \in \mathcal{G}$ es una clase Donsker (ver ejemplo 19.20 en van der der Vaart, 1998).

Ejemplo 7. Si $\mathcal{F}$ es Donsker con $\|P\|_{\mathcal{F}} < \infty$ y $f > 0$ para cada $f \in \mathcal{F}$, entonces $1/\mathcal{F} = \{1/f : f \in \mathcal{F}\}$ es Donsker (ver ejemplo 2.10.9 en van der Vaart y Wellner, 1996).

Ejemplo 8. Si Z es un proceso càglàd en $[0, t]$, $t \in \mathbb{R}$ de variación uniformemente acotada, entonces $Z(\cdot)$ es Donsker (ver Lema 2 de Parner, 1998).

Ejemplo 9. Si $\mathcal{F}$ es Donsker, entonces es Glivenko Cantelli.

Parte II

Modelo de Mezclas Semiparamétrico para Riesgos Competitivos y Modelo de Cura Semiparamétrico utilizando Modelos de Transformación

Capítulo 5

Modelo semiparamétrico para riesgos competitivos

Actualmente el modelado y el ajuste de datos de riesgos competitivos se presentan en un gran número de disciplinas tales como la investigación clínica y la epidemiológica, las finanzas, la criminología y la ingeniería entre otras, véase por ejemplo, Holt (1978), Whitmore (1986), Spivey y Gross (1991), Gaynor *et al.* (1993), Klein y Moeschberger (1997) y Klein y Bajorunaite (2004). Aquí el interés se centra en analizar duraciones de tiempo de ciertos eventos, comenzando por un tiempo de origen establecido y terminando con la ocurrencia del evento el cual es clasificado en varias causas. Por ejemplo, cuando una empresa contrata seguros de vida para sus trabajadores; aquí los montos de las indemnizaciones varían de acuerdo a la causa de muerte del trabajador; generalmente una muerte relacionada con las condiciones laborales tiende a ser más cuantiosa que alguna de otro tipo, por lo que es imperativo estimar las probabilidades asociadas a cada una de las causas en presencia de variables auxiliares importantes, tales como edad y género, para así calcular las primas adecuadas. En muchas situaciones prácticas, los datos se presentan con información concomitante de la cual se cree que depende la ocurrencia de los eventos.

Los datos de riesgos competitivos tienen su propia metodología de análisis debido a la presencia de distintas causas de fallo y además que la ocurrencia de una causa no permite observar la ocurrencia de otra; además, es común encontrar que el tiempo de ocurrencia del evento de interés no ha sido observado al final del seguimiento de algunas unidades experimentales en la muestra.

5.1. Marco téorico del modelo de riesgos competitivos

Considérese el siguiente marco teórico para desarrollar una metodología adecuada para el análisis de riesgos competitivos (ver e.g. Eldant-Johnson y Johnson, 1980):

- Cada muerte es provocada por una causa en particular.
- Cada individuo es susceptible de fallecer de cualquier causa.

Supóngase que existen J tipos de muerte actuando simultáneamente. Dada la primera suposición del marco teórico, que cada muerte es provocada por una causa particular, no se puede observar $(T_1, ..., T_J)$ conjuntamente. En cambio, se puede observar el tiempo de ocurrencia $T = \text{mín}(T_1, ..., T_J)$ El objetivo para este tipo de datos se centra en modelar la función de supervivencia multivariada del vector aleatorio de los tiempos de supervivencia $(T_1, ..., T_J)$, la cual es definida como,

$$S(t_1, ..., t_J) = \mathbb{P}\{T_1 > t_1, ..., T_J > t_J\}.$$

Se asume que $S(t_1, ..., t_J)$ es absolutamente continua. La *fuerza de mortalidad de causa específica*, es la probabilidad instantánea de que ocurra un evento de tipo j en el tiempo t, dado que el individuo ha sobrevivido a todas las causas y se expresa de la siguiente manera,

$$\lambda_j^*(t) = \lim_{\Delta t \to 0} \frac{\mathbb{P}\left\{t < T_j < t + \Delta t | T_j \geq t\right\}}{\Delta t}.$$

En términos de $S(t_1, ..., t_J)$, la fuerza de mortalidad de causa específica es expresada como

$$\lambda_j^*(t) = -\left.\frac{\partial \log S(t_1, ..., t_J)}{\partial t_j}\right|_{t_i = t}.$$

Por lo que la función de supervivencia total es expresada como:

$$S_T(t) = \mathbb{P}\{T > t\} = \mathbb{P}\left(\cap_{j=1}^{J}\{T_j > t\}\right) = S(t, .., t). \tag{5.1}$$

Defínase H como una variable que indica la causa de fallo,

$$H = \sum_{j=1}^{J} jI(T = T_J).$$

Si se observa el par de valores (T, H), entonces el tiempo de ocurrencia y la causa de ocurrencia son identificadas. A partir de aquí se deduce la distribución del par aleatorio (T, H).

La probabilidad condicional de morir por la causa j en el intervalo $(t, t+\Delta t)$ dado que se vive hasta el instante t, y en presencia de que todas las causas afectan a la población simultáneamente es, $\lambda_j^*(t)dt$, mientras que la probabilidad incondicional de que se muera por la causa j en $(t, t+\Delta t)$ está dada por $S_T(t)\lambda_j^*(t)dt$. Así, las probabilidades de fallo de causa específica, también conocidas como funciones de incidencia (véase Gaynor et al. 1993) son definidas como:

$$\begin{aligned} F_j^*(t) &= \mathbb{P}\{\text{fallo por la causa } j \text{ en un tiempo menor o igual a} t\} \\ &= \mathbb{P}\{T \leq t, H = j\} \\ &= \int_0^t \lambda_j^*(u) S_T(u) du, \end{aligned}$$

para $j = 1, \ldots, J$. Denótese por p_j la proporción de muertes esperadas por causa j, entonces,

$$p_j = \mathbb{P}\{H = j\} = \int_0^\infty \lambda_j^*(u) S_T(u) du = F_j^*(\infty),$$

con $p_1 + ... + p_J = 1$.

5.2. Contribuciones al modelo de riesgos competitivos

Un problema que se presenta en el modelado de los riesgos competitivos es que no necesariamente existe una fuerza de mortalidad de causa específica que de manera única identifique a la función de supervivencia conjunta de los riesgos. Para este problema de indentificabilidad (ver Tsiatis, 1975), varios modelos suponen que los eventos son mutuamente independientes, lo cual resuelve el problema de identificabilidad; sin embargo, esta suposición puede ser muy cuestionable. Carrière (1995) propone utilizar una función cópula para modelar la función de supervivencia conjunta la cual permite una estructura de dependencia entre los riesgos y hace que el modelo de riesgos competitivos correspondiente sea identificable. Un problema que se encuentra en la formulación dada por Carrière, es que existen pocas cópulas multivariadas, lo cual restringe el modelado a dos riesgos; mas aún, no existe una metodología para determinar qué cópula y qué marginales son adecuadas para el ajuste de los datos.

En la literatura existen varias extensiones del modelo de Cox para riesgos competitivos. La propuesta por Kalbfleish y Prentice (1973), la cual es rutinariamente usada, consiste en modelar las probabilidades de ocurrencia de las causas y las funciones de supervivencia condicionales para cada causa en forma separada, tratando los otros tipos de muerte como datos censurados. El método de estimación correspondiente no considera a todas las causas que actúan en el proceso y en consecuencia, el modelo puede tener bandas de confianza muy amplias para las probabilidades de supervivencia de causa específica (ver e.g. Fine,1999).

Otra extensión al modelo proporcional de Cox es la presentada por Larson y Dinse (1985), la cual ha sido también usada frecuentemente en años recientes. Esta formulación es capaz de ajustar información concomitante a través de una mezcla de un modelo multinomial generalizado para las probabilidades de la causa y el modelo de Cox para las funciones de supervivencia condicionales de cada causa. La desventaja de su modelo es que, al ser completamente paramétrico, es poco flexible; escoger las formas paramétricas incorrectas de las funciones de supervivencia condicionales implica tener inferencias erroneas. Larson y Dinse, desarrollan la estimación de máxima verosimilitud para su modelo, mientras Choi y Zhou (2002) y Maller y Zhou (2002) investigan las propiedades asintóticas de los estimadores resultantes que incluye existencia, consistencia y normalidad asintótica.

Una especificación alternativa a la de Larson y Dinse (1985) es el modelo de Kuk (1992), el cual consiste en especificar a las funciones de supervivencia condicionales con el modelo semiparamétrico de Cox. Aunque este modelo cuenta con la flexibilidad que la especificación semiparamétrica ofrece a los datos de un sólo tipo de evento, no es posible ignorar la forma funcional de cada función de supervivencia condicional de base en el proceso de inferencia, además su modelo queda pobremente especificado y es entonces susceptible a dar estimaciones incorrectas.

Otras generalizaciones semiparámetricas del modelo de Larson y Dinse han sido recientemente propuestas e investigadas. Por ejemplo, Ng y McLachlan(2003) y Escarela y Bowater (2008) consideran un modelo semiparmétrico que permite incluir covariables, donde las marginales condicionales de cada causa siguen una forma multinomial logística y la distribución condicional del tiempo de fallo dado las covariables y la causa de fallo es especificada a través del modelo de riesgos proporcionales de Cox (1972) donde las funciones de supervivencia de base toman la forma del estimador de prodcuto límite. Naskar *et al.* (2005) consideran un modelo similar para el caso en el que los datos correspondientes a los tiempos de fallo son agrupados y desarrollan el algoritmo de Monte Carlo para efectuar el proceso de estimación de los parámetros.

Ng y McLanclan (2003) y Escarela y Bowater (2008) han investigado el enfoque computacional asociado con el método de estimación de máxima verosimilitud noparamétrica para modelos de mezclas para riesgos competitivos; sin embargo el enfoque teórico ha sido pobremente estudiado. Dupuy y Escarela (2007) dan una demostración de la consitencia para los estimadores de máxima verosimilitud. Lu y Peng (2008) construyen una estimación de parámetros de la versión semiparamétrica de Larson y Dinse. Esta estimación se basa en ecuaciones en martingalas las cuales permiten establecer la consistencia y la normalidad asintótica de sus estimadores obtenidos.

El objetivo de éste capítulo es estudiar las propiedades de los estimadores de máxima verosimilitud del modelo semiparamétrico generalizado de Larson y Dinse desarrollado por Escarela y Bowater (2008). Específicamente, se da un estudio riguroso de las propiedades asintóticas de los estimadores. Esto es alcanzado, siguiendo los enfoques y técnicas desarrolladas por Murphy (1994, 1995) y Parner (1998) para modelos frailty. Otros enfoques de utilidad pueden ser encontrados en Fang *et al.* (2005), Dupuy *et al.* (2006), Kosorok y Song (2007) y Lu (2008), entre otros. Aquí se estudia la existencia, consistencia y normalidad asintótica de los estimadores. También se demuestra que el estimador propuesto para el parámetro de interés en la regresión es semiparamétricamente eficiente y finalmente se obtienen los estimadores consistentes de la varianza.

5.3. El modelo de mezclas semiparamétrico

Supóngase que todas las variables son definidas sobre un espacio de probabilidad $(\Omega, \mathcal{C}, \mathbb{P})$ y que la muestra aleatoria es de tamaño n. Se considerarán J tipos de fallos. Para $i = 1, ..., n$, T_i^0 denota la variable aleatoria que representa el tiempo de fallo de interés. Supóngase que la variable T_i^0 se encuentra censurada por la derecha por una variable aleatoria positiva C_i. $\mathbf{Z}_i$ y $\mathbf{X}_i$ son los vectores de covariables de dimensiónes p y q respectivamente del i-ésimo individuo. La variable $\mathbf{X}_i$ contiene la ordenada y los vectores $\mathbf{X}_i$ y $\mathbf{Z}_i$ pueden tener comoponentes en común. H_i es la variable que representa la causa de ocurrencia y $\mathcal{J} = \{1, ..., J\}$ es el conjunto de los posibles valores de H_i. Para cada $j \in \mathcal{J}$ se define la variable indicadora $\Gamma^j = 1\{H = j\}$. En riesgos competitivos tanto la causa de fallo H y el indicador Γ^j son observados solamente si el tiempo de supervivencia no se encuentra censurado.

El modelo de mezclas propuesto por Larson y Dinse (1985) supone que la causa de muerte de un individuo es escogido por un mecanismo estocástico de J posibles causas. Sea

$$p_j = \mathbb{P}\{H = j\}$$

la proporción de individuos que finalmente fallecen por la causa j. Defínase la función de supervivencia condicional como

$$S_j(t) = \mathbb{P}\{T > t|H = j\}, \qquad j = 1, ...J$$

donde $S_j(t)$ es una distribución de supervivencia propia, es decir, $S_j(0) = 1$ y $S_j(\infty) = 0$. Se sigue que las funciones de incidencia pueden ser calculadas utilizando la expresión $F_j(t) = p_j[1 - S_j(t)]$. De esta manera, la probabilidad a sobrevivir todas las causas al tiempo t, definida por $S_T(t) = \mathbb{P}\{\bigcap_{i=1}^{J}(T_j > t)\}$, puede ser expresada por

$$S_T(t) = \mathbb{P}\{T > t\} = \sum_{j=1}^{J} p_j S_j(t).$$

Considerar el efecto de covariables sobre la función de supervivencia condicional $S_j(t)$, es conveniente hacer uso del modelo de riesgos proporcionales de Cox. Más específicamente, es conveniente asumir que el componente de riesgo es caracterizado por la siguiente función de supervivencia:

$$S_j(t; \mathbf{Z}) = \exp\left\{-\Lambda_j(t) e^{\beta_j' \mathbf{Z}}\right\}, \quad j = 1, ..., J$$

donde $\Lambda_j(t) = \int_0^t \lambda_j(u)du$ es la función base de la fuerza de mortalidad integrada condicional para el tipo de falla j, β_j denota el vector de los coeficientes de regresión. En términos de la fuerza de mortalidad, la suposición de riesgos proporcionales implica que

$$\begin{aligned}\lambda_j(t; \mathbf{Z}) &= \lim_{h\downarrow 0} \frac{1}{h}\mathbb{P}(t < T^0 \leq t + h | T^0 > t, H = j, \mathbf{Z}) \\ &= \lambda_j(t) \exp(\beta_j' \mathbf{Z}), \quad j = 1, ..., J. \end{aligned} \tag{5.2}$$

De la misma manera, el efecto de covariables en las probabilidades de causa específica de muerte pueden ser modeladas uilizando el *modelo logístico generalizado* (ver Cox y Snell, 1989, pp 155-157), por lo que el modelo de probabilidad sigue la siguiente forma,

$$p_j = \mathbb{P}(H = j|\mathbf{X}) = \frac{\exp(\gamma_j' \mathbf{X})}{\sum_{k=1}^{J} \exp(\gamma_k' \mathbf{X})}, \quad j = 1, ..., J, \tag{5.3}$$

donde γ_j es el vector de coeficientes de dimensión q correspodiente a la variable $\mathbf{X}$. Para propósitos de identificabilidad se supone que γ_J es igual a 0. Observe que el vector $\mathbf{X}$ puede contener algunas o todas las componentes del vector $\mathbf{Z}$.

El modelo semiparamétrico para riesgos competitivos ha sido empleado por diferentes autores. Por ejemplo, Kuk (1992) analiza datos de trasplantes de corazón, usando una simplificación de los modelos (5.3)-(5.2), (tomando $\mathbf{X} = \mathbf{Z}$), mientras que, Ng y McLachlan (2003) y Escarela y Bowater (2008) ajustan el modelo para datos de cáncer de próstata. Obsérvese que el modelo (5.2)-(5.3) está relacionado con los modelos de mezclas semiparamétricos de cura, véase entre otros, Kuk and Chen (1992), Taylor (1995), Sy and Taylor (2000), Peng (2003) quienes investigan la parte computacional de esta clase de modelos. Ver también Fang *et al.* (2005) y Lu (2008), quienes estudian la propiedades asintóticas de los estimadores de máxima verosimilitud para modelos de cura.

5.4. Notaciones y suposiciones del modelo

En esta sección se establecen algunas notaciones y suposiciones del modelo de mezclas semiparmétrico para riesgos competitivos propuesto por Escarela y Bowater (2008) que serán utilizadas en secciones posteriores.

Los datos consisten de n vectores independientes $(T_i, \Delta_i, \mathbf{Z}_i, \mathbf{X}_i, \Delta_i H_i)$ $(i = 1, \ldots, n)$, donde $T_i = \text{mín}\{T_i^0, \text{mín}(C_i, \tau)\}$, $\Delta_i = 1\{T_i^0 \leq \text{mín}(C_i, \tau)\}$, y $\tau < \infty$ es una constante fija que denota el tiempo final de observación en el estudio. Denotémos por $\mathbf{O}$ y $\mathbf{O}_i$ la abreviación del vector observado $(T, \Delta, \mathbf{Z}, \mathbf{X}, \Delta H)$ y sus n replicas independientes $(T_i, \Delta_i, \mathbf{Z}_i, \mathbf{X}_i, \Delta_i H_i)$, respectivamente. El problema estadístico del modelo es la estimación de los parámetros β_j, γ_j y las funciones de fuerza acumulada $\Lambda_j(t) = \int_0^t \lambda_j$ para los vectores $\mathbf{O}_i$, $i = 1, \ldots, n$. En práctica, los coeficientes β_j ($j \in \mathcal{J}$) son a menudo los parámetros de interés en el modelo de mezclas (5.2)-(5.3).

Denótese por $\mathbf{G} = (\gamma_1' \ldots \gamma_{J-1}')'$, $p_{\mathbf{G}}^{j,X} = \mathbb{P}(H = j|X)$, y $p_{\mathbf{G},i}^{j,X} = \mathbb{P}(H = j|X_i)$. Sean $N(t) = 1\{T \le t\}\Delta$ y $Y(t) = 1\{T \ge t\}$ con $t \in [0, \tau]$ el proceso de conteo de ocurrencia y el proceso de riesgo, respectivamente. Para $j \in \mathcal{J}$, defínase el proceso contable $N^j(t) = 1\{T \le t\}\Delta^j$, donde $\Delta^j = \Delta\Gamma^j$. Observe que $N^j(t)$ es igual a 1 si la ocurrencia surge por la j-ésima causa. Cantidades correspondientes para el i-ésimo individuo serán denotadas por N_i, N_i^j, y Δ_i^j.

Considérense las siguientes condiciones de regularidad:

(C1) Condicional en $\mathbf{Z}$, H y $\mathbf{X}$, el tiempo de censura C es independiente al tiempo de ocurrencia T^0. Condicional en $\mathbf{Z}$ y $\mathbf{X}$, C es independiente de H.

(C2) Existe una constante positiva c_0 tal que, $\mathbb{P}(C \ge \tau|\mathbf{Z}, \mathbf{X}) > c_0$ casi seguramente.

(C3) La fuerza de mortalidad dados $\mathbf{Z}$ y $\mathbf{X}$, $\lambda(s|\mathbf{Z}, \mathbf{X})$, es uniformemente acotada casi seguramente.

(C4) Sea $\mathbf{B} = (\beta_1' \ldots \beta_J')' \in \mathbb{R}^{pJ} \equiv \mathbb{R}^P$ y $\mathbf{G} = (\gamma_1' \ldots \gamma_{J-1}')' \in \mathbb{R}^{q(J-1)} \equiv \mathbb{R}^Q$. Denotemos por $\mathbf{B}_0$ y $\mathbf{G}_0$ los verdaderos valores de $\mathbf{B}$ y de $\mathbf{G}$ respectivamente, los cuales pertenecen al interior de conjuntos compactos conocidos $\mathcal{B} \subset \mathbb{R}^P$ y $\mathcal{G} \subset \mathbb{R}^Q$ respectivamente.

(C5) Para cada $j \in \mathcal{J}$, el verdadero valor de la fuerza de mortalidad condicional acumulada denotada por $\Lambda_{j,0}$ es una función estrictamente creciente en $[0, \tau]$ con $\Lambda_{j,0}(0) = 0$ y $\Lambda_{j,0}(\tau) < \infty$. $\Lambda_{j,0}$ es continuamente diferenciable en $[0, \tau]$ tal que $\lambda_{j,0}(t) = d\Lambda_{j,0}(t)/dt$.

(C6) Los vectores de covariables $\mathbf{Z}$ y $\mathbf{X}$ son acotados; es decir, $\|\mathbf{X}\| < c_1$ y $\|\mathbf{Z}\| < c_1$ para alguna constante $0 < c_1 < \infty$ (donde, $\|\cdot\|$ denota la norma Euclideana). Las matrices de covarianza de $\mathbf{Z}$ y $\mathbf{X}$ son positivas definidas. Denotemos por

$$c_2 = \min_{\beta_j, j \in \mathcal{J}, \|\mathbf{Z}\| < c_1} \exp(\beta_j' \mathbf{Z}) \quad \text{y} \quad c_3 = \max_{\beta_j, j \in \mathcal{J}, \|\mathbf{Z}\| < c_1} \exp(\beta_j' \mathbf{Z}).$$

Sea $\mathcal{L}$ el conjunto de funciones que satisfacen las condiones de **C5**, sea θ el vector de parámetros igual a $(\mathbf{B}, \mathbf{G}, \Lambda_j; j \in \mathcal{J})$, $\theta_0 = (\mathbf{B}_0, \mathbf{G}_0, \Lambda_{j,0}; j \in \mathcal{J})$ el valor verdadero de θ, denótese por $\Theta = \mathcal{B} \times \mathcal{G} \times \mathcal{L}^{\otimes J}$ el espacio de parámetros. Bajo los verdaderos valores θ_0, la esperanza de una variable aleatoria será denotada por P_{θ_0}.

(C7) Existe una constante positiva c_4 tal que para cada $j \in \mathcal{J}$, $P_{\theta_0}[Y(\tau)\Gamma^j] > c_4$.

(C8) Existe una constante positiva c_5 tal que para cada $j \in \mathcal{J}$,

$$P_{\theta_0}[\Delta^j|T, \mathbf{Z}, \mathbf{X}] > c_5.$$

(C9) La distribución de la causa de fallo H condicionada a $\mathbf{X}$ y $\mathbf{Z}$ no involucra términos de $\mathbf{Z}$ que no están en $\mathbf{X}$. La distribución de C, $\mathbf{Z}$, y $\mathbf{X}$ no depende de θ.

Observación. La condición **C1** garantiza que ninguna información acerca de θ se pierde por la eliminación de términos referentes a la censura. La condition **C2** garantiza que el seguimiento es lo suficientemente largo para determinar las funciones base del riesgo acumulado $\Lambda_{j,0}$ sobre el intervalo $[0, \tau]$. Las condiciones **C3**-**C8** son utilizadas para la identificabilidad de θ_0 y para las propiedades asintóticas de los estimadores propuestos. La condición **C7** garantiza que el seguimiento es lo suficientemente largo (para cada causa de fallo) para realizar la estimación de $\Lambda_{j,0}$ sobre todo el intervalo $[0, \tau]$. La condición **C8** garantiza que para cada causa los fallos pueden ocurrir y su causa es observada en cualquier momento para cualquier valor de las covariables. La condición **C9** asegura que ninguna información acerca de θ se pierde por la eliminación de los términos en las distribuciones marginales de $\mathbf{Z}$ y $\mathbf{X}$ en la verosimilitud.

5.5. Estimación de máxima verosimiltud noparamétrica

Para comenzar, se asume que no existen tiempos de fallos empatados (esta suposición es hecha para una fácil presentación, pero los resultados pueden ser fácilmente adaptados para incorporar empates). Bajo los modelos establecidos en (5.2) y (5.3) y utilizando las condiciones establecidas en **C1**-**C9**, la función de verosimilitud para el parámetro θ, de las observaciones $\mathbf{O}_i$ $(i = 1, \ldots, n)$, es proporcional a,

$$\prod_{i=1}^{n} \left\{ \prod_{j \in \mathcal{J}} \left[\lambda_j(T_i) e^{\beta_j' \mathbf{Z}_i} \exp\left(-e^{\beta_j' \mathbf{Z}_i} \Lambda_j(T_i) \right) p_{\mathbf{G},i}^{j,\mathbf{X}} \right]^{\Delta_i^j} \left[\sum_{j \in \mathcal{J}} \exp\left(-e^{\beta_j' \mathbf{Z}_i} \Lambda_j(T_i) \right) p_{\mathbf{G},i}^{j,\mathbf{X}} \right]^{1-\Delta_i} \right\}. \tag{5.4}$$

Es natural calcular el estimador de máxima verosimilitud (MLE) de θ_0 maximizando el logarítmo de la función 5.4. Sin embargo, el máximo de esta función es infinito cuando las funciones Λ_j $(j \in \mathcal{J})$ pertenecen al conjunto $\mathcal{L}$ de funciones de fuerza de mortalidad acumulada base que son absolutamente continuas. Esto es, porque se pueden escoger funciones Λ_j $(j \in \mathcal{J})$ con valores fijos en el tiempo de fallo T_i y sea $d\Lambda_{j,0}(T_i)/dT_i = \lambda_j(T_i)$, obsérvese que este término tiende a infinito para algún T_i con $\Delta_i^j = 1$.

Para resolver este problema, se pueden restringir las funciones Λ_j $(j \in \mathcal{J})$ a funciones continuas por la derecha, tal que cada Λ_j tiene saltos en los tiempo de fallo observados T_i. Sea $\Lambda_j\{t\}$ el tamaño de salto de Λ_j en t donde se busca maximizar la función $L_n(\mathbf{B}, \mathbf{G}, \Lambda_j; j \in \mathcal{J}) =$

$$\prod_{i=1}^{n} \left\{ \prod_{j \in \mathcal{J}} \left[\Lambda_j\{T_i\} e^{\beta_j' \mathbf{Z}_i} \exp\left(-e^{\beta_j' \mathbf{Z}_i} \Lambda_j(T_i) \right) p_{\mathbf{G},i}^{j,\mathbf{X}} \right]^{\Delta_i^j} \left[\sum_{j \in \mathcal{J}} \exp\left(-e^{\beta_j' \mathbf{Z}_i} \Lambda_j(T_i) \right) p_{\mathbf{G},i}^{j,\mathbf{X}} \right]^{1-\Delta_i} \right\}$$

sobre el espacio $\Theta_n =$

$$\left\{(\mathbf{B}, \mathbf{G}, \Lambda_j) : \mathbf{B} \in \mathcal{B}, \mathbf{G} \in \mathcal{G}, \Lambda_j \text{ es una función creciente continua por la derecha en } [0, \tau], j \in \mathcal{J}\right\}.$$

Si el estimador existe, entonces es conocido como el estimador de máxima verosimilitud no-paramétrico (NPMLE). Sea $\widehat{\theta}_n = (\widehat{\mathbf{B}}_n, \widehat{\mathbf{G}}_n, \widehat{\Lambda}_{j,n}; j \in \mathcal{J})$, donde $\widehat{\mathbf{B}}_n = (\widehat{\beta}'_{1,n} \ldots \widehat{\beta}'_{J,n})'$ y $\widehat{\mathbf{G}}_n = (\widehat{\gamma}'_{1,n} \ldots \widehat{\gamma}'_{J-1,n})'$. La siguiente proposición establece la existencia del NPMLE:

Proposición 5.5.1. *Bajo las condiciones **C1-C9**, el máximo $\widehat{\theta}_n$ de L_n sobre Θ_n existe y es alcanzado.*

Demostración: Para comenzar con la demostración es necesario identificar la forma de un posible maximizador de $\widehat{\Lambda}_{j,n}$ de L_n en el espacio Θ_n.

Sea $j \in \mathcal{J}$. Defínase $\mathcal{S}_n^j = \{i \in \{1, \ldots, n\} | \Delta_i^j = 1\}$ como el conjunto de individuos de la muestra que son observados por morir por la j-ésima causa. Para cada $j \in \mathcal{J}$ y cualquier función Λ_j en Θ_n se puede construir una función creciente de saltos Λ_j^*, cuyos saltos son efectuados solamente en los tiempos de fallo $\{T_i, i \in \mathcal{S}_n^j\}$, tal que $\Lambda_j^*(T_i) = \Lambda_j(T_i)$. Claramente en cada uno de estos tiempos de fallo $\Lambda_j^*\{T_i\} \geq \Lambda_j\{T_i\}$, lo cual implica que $L_n(\mathbf{B}, \mathbf{G}, \Lambda_j; j \in \mathcal{J}) \leq L_n(\mathbf{B}, \mathbf{G}, \Lambda_j^*; j \in \mathcal{J})$. Por lo tanto, el máximo $\widehat{\Lambda}_{j,n}$ (si existiera) tendría que ser una función de saltos con saltos positivos en el tiempo de fallo T_i tal que $\Delta_i^j = 1$. Esto restringe el problema de maximización de L_n al siguiente subespacio de Θ_n:

$$\left\{\left(\mathbf{B}, \mathbf{G}, \Lambda_j\{t_k^j\}\right) : \mathbf{B} \in \mathcal{B}, \mathbf{G} \in \mathcal{G}, \Lambda_j\{t_k^j\} \in [0, \infty), k = 1, \ldots, |\mathcal{S}_n^j|, j \in \mathcal{J}\right\}, \tag{5.5}$$

donde para cada $j \in \mathcal{J}$, $|\mathcal{S}_n^j|$ denota la cardinalidad del conjunto $\mathcal{S}_n^j$. Sean $t_1^j < \ldots < t_{|\mathcal{S}_n^j|}^j$ los tiempos de fallo ordenados en el conjunto $\{T_i, i \in \mathcal{S}_n^j\}$. Esto es, se maximiza la función $L_n(\mathbf{B}, \mathbf{G}, (\Lambda_j\{t_k^j\})_{j,k}) =$

$$\prod_{i=1}^{n} \left\{ \prod_{j \in \mathcal{J}} \left[\Lambda_j\{T_i\} e^{\beta_j' \mathbf{Z}_i} \exp\left(-e^{\beta_j' \mathbf{Z}_i} \sum_{k=1}^{|\mathcal{S}_n^j|} \Lambda_j\{t_k^j\} 1\{t_k^j \leq T_i\} \right) p_{\mathbf{G},i}^{j,\mathbf{X}} \right]^{\Delta_i^j} \right.$$
$$\left. \times \left[\sum_{j \in \mathcal{J}} \exp\left(-e^{\beta_j' \mathbf{Z}_i} \sum_{k=1}^{|\mathcal{S}_n^j|} \Lambda_j\{t_k^j\} 1\{t_k^j \leq T_i\} \right) p_{\mathbf{G},i}^{j,\mathbf{X}} \right]^{1-\Delta_i} \right\} \tag{5.6}$$

con respecto a los vectores β_j, γ_j y $\Lambda_j\{t_k^j\}$.

Ahora se demostrá que tal maximizador existe. Supóngase que $\Lambda_j\{t_k^j\} \leq L < \infty$ para cada $k = 1, \ldots, |\mathcal{S}_n^j|$ y $j \in \mathcal{J}$. L_n es una función continua de β_j, γ_j, y $\Lambda_j\{t_k^j\}$ sobre el conjunto compacto $\mathcal{B} \times \mathcal{G} \times [0, L]^{s_n}$, donde $s_n = \sum_{j \in \mathcal{J}} |\mathcal{S}_n^j|$. Por lo tanto, L_n alcanza su máximo en este conjunto.

Para demostrar que este máximo existe sobre el conjunto $\mathcal{B} \times \mathcal{G} \times [0, \infty)^{s_n}$, se demostrara que existe L finito tal que para todo $(\mathbf{B}^L, \mathbf{G}^L, (\Lambda_j^L\{t_k^j\})_{j,k}) \in (\mathcal{B} \times \mathcal{G} \times [0, \infty)^{s_n}) \backslash (\mathcal{B} \times \mathcal{G} \times [0, L]^{s_n})$, existe $(\mathbf{B}, \mathbf{G}, (\Lambda_j\{t_k^j\})_{j,k}) \in \mathcal{B} \times \mathcal{G} \times [0, L]^{s_n}$ el cual tiene un valor más grande que L_n. Para ver esto, se procedera por contradicción.

Supóngase que no existe L. Entonces para todo $L < \infty$, existe un $(\mathbf{B}^L, \mathbf{G}^L, (\Lambda_j^L\{t_k^j\})_{j,k}) \in (\mathcal{B} \times \mathcal{G} \times [0, \infty)^{s_n}) \backslash (\mathcal{B} \times \mathcal{G} \times [0, L]^{s_n})$ tal que para todo $(\mathbf{B}, \mathbf{G}, (\Lambda_j\{t_k^j\})_{j,k}) \in \mathcal{B} \times \mathcal{G} \times [0, L]^{s_n}$, $L_n(\mathbf{B}, \mathbf{G}, (\Lambda_j\{t_k^j\})_{j,k}) \leq L_n(\mathbf{B}^L, \mathbf{G}^L, (\Lambda_j^L\{t_k^j\})_{j,k})$. Pero esto demuestra que el término

$$L_n(\mathbf{B}^L, \mathbf{G}^L, (\Lambda_j^L\{t_k^j\})_{j,k})$$

puede ser arbitrariamente pequeña cuando L crece, lo cual es una contradicción. Para ver esto, nótese que (5.6) se encuentra acotado por

$$J^{n-s_n} \prod_{i=1}^{n} \prod_{j \in \mathcal{J}} \left\{\Lambda_j\{T_i\} c_3\right\}^{\Delta_i^j} \exp\left(-c_2 \Delta_i^j \sum_{k=1}^{|\mathcal{S}_n^j|} \Lambda_j\{t_k^j\} 1\{t_k^j \leq T_i\}\right).$$

Si $(\mathbf{B}^L, \mathbf{G}^L, (\Lambda_j^L\{t_k^j\})_{j,k}) \in (\mathcal{B} \times \mathcal{G} \times [0, \infty)^{s_n}) \backslash (\mathcal{B} \times \mathcal{G} \times [0, L]^{s_n})$, entonces existe al menos un $j \in \mathcal{J}$ y un $l \in \{1, \ldots, |\mathcal{S}_n^j|\}$ tal que $\Lambda_j^L\{t_l^j\} > L$. Sea i^* el índice del individuo tal que $\Delta_{i^*}^j = 1$ y $T_{i^*} = t_l^j$. Entonces

$$\left\{\Lambda_j^L\{T_{i^*}\} c_3\right\}^{\Delta_{i^*}^j} \exp\left(-c_2 \Delta_{i^*}^j \sum_{k=1}^{|\mathcal{S}_n^j|} \Lambda_j^L\{t_k^j\} 1\{t_k^j \leq T_{i^*}\}\right)$$

tiende a 0 cuando L tiende a $+\infty$. Entonces, la cota superior de $L_n(\mathbf{B}^L, \mathbf{G}^L, (\Lambda_j^L\{t_k^j\})_{j,k})$ puede ser cercana a 0 como se desee mientras L crece, lo cual produce una contradicción. Por lo tanto, para cualquier n fijo, el máximo de L_n es obtenido en el conjunto $\mathcal{B} \times \mathcal{G} \times [0, L]^{s_n}$, para algún $L < \infty$ y en este conjunto, el máximo $\widehat{\theta_n}$ es alcanzado.

□

Para cada n, el problema de maximización de L_n sobre (5.5) se reduce a una dimensión finita, ya que el número total de saltos s_n de N_i $(i = 1, \ldots, n)$ es inferior o igual a n.

El algoritmo esperanza-maximización(EM) (Dempster*et al.*, 1977) puede ser usado para calcular los NPMLEs. Escarela y Bowater (2008) explican e implementan a detalles el algoritmo EM (ver Apéndice).

Denotemos por $g^j(\mathbf{O}; \theta)$, $j \in \mathcal{J}$ la esperanza condicional de Γ^j dado $\mathbf{O}$ y el valor del parámetro θ, es decir,

$$g^j(\mathbf{O}; \theta) = \Delta^j + (1 - \Delta) w^j(\mathbf{O}; \theta),$$

donde

$$w^j(\mathbf{O};\theta) = \frac{\exp\left(-\Lambda_j(T)e^{\beta_j'\mathbf{Z}} + \gamma_j'\mathbf{X}\right)}{\sum_{k\in\mathcal{J}} \exp\left(-\Lambda_k(T)e^{\beta_k'\mathbf{Z}} + \gamma_k'\mathbf{X}\right)}.$$

En el paso M del algoritmo EM, se resuelve la ecuación de score para los datos completos condicionada a los datos observados. En particular, una ecuación útil para $\widehat{\Lambda}_{j,n}$ puede ser obtenida, (véase Lema 5.5.1). Denótemos por $\mathbb{P}_n$ la medida empírica. Entonces se satisface el siguiente resultado:

Lema 5.5.1. *El NPMLE $\widehat{\theta}_n$ satisface la siguiente ecuación,*

$$\widehat{\Lambda}_{j,n}(t) = \int_0^t \frac{1}{H_n^j(s;\widehat{\theta}_n)}\, dG_n^j(s), \tag{5.7}$$

para cada $j \in \mathcal{J}$, donde $H_n^j(s;\theta) = \mathbb{P}_n[h^j(s,\mathbf{O};\theta)]$, $h^j(s,\mathbf{O};\theta) = Y(s)e^{\beta_j'\mathbf{Z}}g^j(\mathbf{O};\theta)$, y $G_n^j(s) = \mathbb{P}_n N^j(s)$.

Demostración: Este resultado es obtenido siguiendo los siguientes pasos:

1. Se toma la derivada con respecto a los tamaños de los saltos $\Lambda_j\{t_k^j\}$ de la esperanza condicional de la función log-verosimílitud de los datos completos y del NPMLE, la cual esta dada por

$$\begin{aligned} l_{\widehat{\theta}_n}(\theta) = \sum_{i=1}^n \sum_{j\in\mathcal{J}} \Bigg\{ & \Delta_i^j \sum_{k=1}^{|\mathcal{S}_n^j|} 1\{T_i = t_k^j\} \log \Lambda_j\{t_k^j\} + \Delta_i^j \beta_j'\mathbf{Z}_i \\ & - g^j(\mathbf{O}_i;\widehat{\theta}_n) e^{\beta_j'\mathbf{Z}_i} \sum_{k=1}^{|\mathcal{S}_n^j|} \Lambda_j\{t_k^j\} 1\{t_k^j \le T_i\} + g^j(\mathbf{O}_i;\widehat{\theta}_n) \log p_{\mathbf{G},i}^{j,\mathbf{X}} \Bigg\}. \end{aligned}$$

2. Se pone $(\partial l_{\widehat{\theta}_n}(\theta)/\partial \Lambda_j\{t_k^j\})|_{\theta=\widehat{\theta}_n} = 0$ y se resuelve para $\Lambda_j\{t_k^j\}$.

3. Se suma sobre $\{k \in \{1,\ldots,|\mathcal{S}_n^j|\} : t_k^j \le t\}$.

□

5.6. Identificabilidad del modelo semiparamétrico para riesgos competitivos

En esta sección se estudia la identificabilidad del modelo de mezclas semiparamétrico para riesgos competitivos propuesto en la sección 5.3. esta sección comienza con la definición de identificabilidad y con la definición de la información de Kullback-Leibler para posteriormente demostrar que el modelo (5.2)-(5.3) es identificable.

5.6.1. Definición de identificabilidad e información en el sentido de Kullback-Leibler

Sea $\mathcal{P} = \{P_\phi : \phi \in \Phi\}$ un modelo estadístico donde el paramétro ϕ puede ser finito o infinito dimensional o tener ambos términos.

Definición 5.6.1. *Un modelo $\mathcal{P} = \{P_\phi : \phi \in \Phi\}$ es identificable si la aplicación P_ϕ es inyectiva.*

Si la familia de probabilidades $\mathcal{P}$ puede ser definida en términos de la familia de densidades $\mathcal{F}$, es decir, $\mathcal{P} = \{P_\phi = (f(y;\phi).\mu) : f \in \mathcal{F}\}$, para una medida μ, entonces la condición de indentificabilidad precedente es expresada de la siguiente manera:

$$\forall \phi_1, \phi_2 \in \Phi, f(y;\phi_1) = f(y;\phi_2) \Rightarrow \phi_1 = \phi_2. \tag{5.8}$$

Definición 5.6.2. *Sean dos distribuciones $P_{\phi_1} = f(y,\phi_1).\mu$ y $P_{\phi_2} = f(y,\phi_2).\mu$. Se llama la infomación en el sentido de Kullback-Leibler de P_{ϕ_2} sobre P_{ϕ_1} a la cantidad,*

$$K(P_{\phi_1}, P_{\phi_2}) = \int_{\mathcal{Y}} \ln \frac{f(y;\phi_2)}{f(y;\phi_1)} f(y;\phi_2)\mu(dy).$$

La información en el sentido de Kullback-Leibler no es una distancia clásica porque las condiciones de simetría y la desigualdad del triángulo no son satisfechas. Sin embargo, ella traduce una idea de aproximación entre las probabilidades P_{ϕ_1} y P_{ϕ_2}, este hecho es garantizado por las siguientes dos propiedades (ver Konishi y Kitagawa, 2008 y Kullback y Leibler, 1951)

Proposición 5.6.1. *Sean dos probabilidades P_{ϕ_1} y P_{ϕ_2}. La información en el sentido de Kullback-Leibler satisface las sigueintes propiedades:*

1. $K(P_{\phi_1}, P_{\phi_2}) \geq 0$
2. $K(P_{\phi_1}, P_{\phi_2}) = 0 \Leftrightarrow P_{\phi_1} = P_{\phi_2}$.

Proposición 5.6.2. *El parámetro ϕ es identificable sí y sólo sí:*

$$\forall \phi_1, \phi_2 \in \Phi, K(P_{\phi_1}, P_{\phi_2}) = 0 \Rightarrow \phi_1 = \phi_2.$$

5.6.2. Identificabilidad para el modelo semiparamétrico para riesgos competitivos

En esta parte se demuestra que el parámetro $\theta = (\mathbf{B}, \mathbf{G}, \Lambda_j; j \in \mathcal{J})$ para el modelo semiparamétrico para riesgos competitivos es identificable. El resultado es obtenido siguiendo las definición de la información de Kullback-Leibler.

Proposición 5.6.3. *El parámetro $\theta = (\mathbf{B}, \mathbf{G}, \Lambda_j; j \in \mathcal{J})$ es identificable; es decir, si $\theta = (\mathbf{B}, \mathbf{G}, \Lambda_j; j \in \mathcal{J})$ y $\theta^* = (\mathbf{B}^*, \mathbf{G}^*, \Lambda_j^*; j \in \mathcal{J})$ son dos elementos de Θ tales que $L(\theta) = L(\theta^*)$ entonces, $\theta = \theta^*$.*

Demostración: Sean $\theta = (\mathbf{B}, \mathbf{G}, \Lambda_j; j \in \mathcal{J})$ y $\theta^* = (\mathbf{B}^*, \mathbf{G}^*, \Lambda_j^*; j \in \mathcal{J})$ dos elementos de Θ. Si $L(\theta) = L(\theta^*)$ c.s., por la condición (**C8**) existe un $l \in \mathcal{J}$ tal que $\Delta^l = 1$, $y \in [0, \tau]$, $||\mathbf{z}|| \leq c_1$ y $||\mathbf{x}|| \leq c_1$, por lo tanto,

$$\lambda_l(y) e^{\beta_l' \mathbf{z}} \exp\left(e^{\beta_l' \mathbf{z}} \Lambda_l(y)\right) p_{\mathbf{G}}^{l,\mathbf{x}} = \lambda_l^*(y) e^{\beta_l^{*\prime} \mathbf{z}} \exp\left(e^{\beta_l^{*\prime} \mathbf{z}} \Lambda_l^*(y)\right) p_{\mathbf{G}^*}^{l,\mathbf{x}},$$

lo cual puede ser reescrito como,

$$p_{\mathbf{G}}^{l,\mathbf{x}} \frac{\partial \exp\left(e^{\beta_l' \mathbf{z}} \Lambda_l(s)\right)}{\partial s} = p_{\mathbf{G}^*}^{l,\mathbf{z}} \frac{\partial \exp\left(e^{\beta_l^{*\prime} \mathbf{z}} \Lambda_l^*(s)\right)}{\partial s}.$$

Sea $t \in [0, \tau]$, integrando ambos lados de la ecuación anterior de 0 a t se obtiene

$$p_{\mathbf{G}}^{l,\mathbf{x}} \exp\left(e^{\beta_l' \mathbf{z}} \Lambda_l(t))\right) = p_{\mathbf{G}^*}^{l,\mathbf{x}} \exp\left(e^{\beta_l^{*\prime} \mathbf{z}} \Lambda_l^*(t)\right)$$

equivalentemente,

$$\frac{\exp\left(e^{\beta_l' \mathbf{z}} \Lambda_l(t))\right)}{\exp\left(e^{\beta_l^{*\prime} \mathbf{z}} \Lambda_l^*(t)\right)} = \frac{p_{\mathbf{G}^*}^{l,\mathbf{x}}}{p_{\mathbf{G}}^{l,\mathbf{x}}}. \tag{5.9}$$

Obsérvese que el lado derecho de la ecuación (5.9) es independiente de t, entonces

$$\frac{\exp\left(e^{\beta_l' \mathbf{z}} \Lambda_l(t))\right)}{\exp\left(e^{\beta_l^{*\prime} \mathbf{z}} \Lambda_l^*(t)\right)} = \frac{p_{\mathbf{G}^*}^{l,\mathbf{x}}}{p_{\mathbf{G}}^{l,\mathbf{x}}} = \kappa_1, \tag{5.10}$$

donde κ_1 es una constante positiva. En particular, tomando $\mathbf{x} = 0$ implica que $\frac{p_{\mathbf{G}^*}^{l,\mathbf{x}}}{p_{\mathbf{G}}^{l,\mathbf{x}}} = 1$, es decir, $\kappa_1 = 1$. De éste resultado y de la ecuación (5.10) se obtiene

$$\frac{\exp\left(e^{\beta_l' \mathbf{z}} \Lambda_l(t)\right)}{\exp\left(e^{\beta_l^{*\prime} \mathbf{z}} \Lambda_l^*(t)\right)} = 1,$$

aplicando la función logaritmo y reordenando términos, la ecuación anterior se reduce a

$$\frac{e^{\beta_l'\mathbf{z}}}{e^{\beta_l^{*\prime}\mathbf{z}}} = \frac{\Lambda_l^*(t)}{\Lambda_l(t)}.$$

Nótese que el lado izquierdo de la ecuación anterior no depende de t, lo que implica que

$$\frac{e^{\beta_l'\mathbf{z}}}{e^{\beta_l^{*\prime}\mathbf{z}}} = \frac{\Lambda_l^*(t)}{\Lambda_l(t)} = \kappa_2 \tag{5.11}$$

con κ_2 una constante positiva. Tomando $\mathbf{z} = 0$, implica que $\kappa_2 = 1$, por lo tanto

$$\frac{e^{\beta_l'\mathbf{z}}}{e^{\beta_l^{*\prime}\mathbf{z}}} = 1,$$

lo que es equivalente a $(\beta_l - \beta_l^*)'\mathbf{z} = 0$ por (**C6**) se obtiene que $\beta_l - \beta_l^* = 0$. Por otro lado, como $\kappa_2 = 1$, por la ecuación (5.11) se sigue que $\Lambda_l(t) = \Lambda_l^*(t)$. Por último $\gamma = \gamma^*$, este resultado es demostrado en el Teorema 1 de Bettina Grün y Friedrich Leisch en *Identifiability of Finite Mixtures of Multinomial Logit Models with Varying and Fixed Effects* no presentamos la demostración de este hecho en ésta tesis debido a que es larga la demostración, pero el lector puede encontrar todos los detalles en esta referencia.

Por lo tanto, $\theta = \theta^*$, lo que concluye la demostración.

□

5.7. Consistencia

El propósito de esta sección es establecer, bajo las condiciones **C1**-**C9** la consistencia de los estimadores semiparamétricos del modelo de riesgos competitivos. La demostración esta basada en las ideas desarrolladas por Murphy (1994) quien estudia los modelos frailty. Véase también, Chang et al. (2005), Kosorok y Song (2007) y Lu (2008) quienes aplican éstas técnicas para diferentes modelos de supervivencia donde los datos son censurados por la derecha.

Para comenzar, se demostraran dos lemas que serán de utilidad en la demostración de la consistencia de los estimadores.

Lema 5.7.1. *Para cada* $j \in \mathcal{J}$, $\limsup_n \widehat{\Lambda}_{j,n}(\tau) < \infty$ *casi seguramente.*

Demostración: Sea $j \in \mathcal{J}$ y $s \in [0, \tau)$, bajo la suposisción **C6**,

$$H_n^j(s; \widehat{\theta}_n) = \mathbb{P}_n[Y(s)e^{\widehat{\beta}_{j,n}'\mathbf{Z}}S^j(\mathbf{O}; \widehat{\theta}_n)] \geq c_2\mathbb{P}_n[Y(s)\Delta^j],$$

utilizando la ley de lo grandes números se deduce que $H_n^j(s;\widehat{\theta}_n) \geq c_2 P_{\theta_0}[Y(s)\Delta^j] + o(1)$ casi seguramente. Bajo las suposisciones establecidas en la sección 5.3, se obtiene que $P_{\theta_0}[Y(s)\Delta^j]$ es mayor que cero. Entonces, para cada $s \in [0,\tau)$, $H_n^j(s;\widehat{\theta}_n)$ es mayor que 0 casi seguramente cuando n tiende a infinito. Más aún, los tamaños de los saltos de $\widehat{\Lambda}_{j,n}$ en τ son acotados por $1/c_2$. Entonces,

$$0 \leq \widehat{\Lambda}_{j,n}(\tau) \leq O(1)\mathbb{P}_n N^j(\tau-) + \frac{1}{c_2}$$

casi seguramente cuando n tiende a infinito, lo cual concluye la prueba.

□

Lema 5.7.2. *Para cada $j \in \mathcal{J}$ y $t \in [0,\tau]$, defínase*

$$\widetilde{\Lambda}_{j,n}(t) = \int_0^t \frac{1}{H_n^j(s;\theta_0)}\, dG_n^j(s).$$

Entonces, $\sup_{t\in[0,\tau]} |\widetilde{\Lambda}_{j,n}(t) - \Lambda_{j,0}(t)|$ *converge a 0 casi seguramente cuando n tiende a infinito.*

Demostración: Primero se demostra que la clase de funciones $\{h^j(s,\mathbf{O};\theta) : s \in [0,\tau], \theta \in \Theta\}$ es Donsker. Para ésto se utilizaran los ejemplos de clases Donsker dados en la sección 4.2. Recuérdese que el espacio de parámetros esta dado por $\Theta = \mathcal{B} \times \mathcal{G} \times \mathcal{L}^{\otimes J}$. Durante la prueba se denotará por $\mathcal{B}_j$ y $\mathcal{G}_j$ al espacio de parámetros para β_j y γ_j respectivamente. Considérese la clase

$$\mathcal{F} = \left\{ w^j(\mathbf{O};\theta) = \frac{\exp\left(-\Lambda_j(T)e^{\beta_j'\mathbf{Z}} + \gamma_j'\mathbf{X}\right)}{\sum_{k\in\mathcal{J}} \exp\left(-\Lambda_k(T)e^{\beta_k'\mathbf{Z}} + \gamma_k'\mathbf{X}\right)} : \theta \in \Theta \right\}. \tag{5.12}$$

Dado que $\mathbf{Z}$ y $\mathbf{X}$ son acotadas, entonces por el ejemplo 2, implica que las clases $\{\beta_j'\mathbf{Z} : \beta_j \in \mathcal{B}_j\}$ y $\{\gamma_j'\mathbf{X} : \gamma_j \in \mathcal{G}_j\}$ son Donsker. Por la diferenciabilidad de $e^{\beta_j'\mathbf{Z}}$ en $\mathbf{Z}$ y dado que la derivada es acotada, implican que $\{e^{\beta_j'\mathbf{Z}} : \beta_j \in \mathcal{B}_j\}$ es Donsker (ver ejemplo 3). Más aún, la clase de funciones que mapean T en $\Lambda_j(T)$ indexadas por $\Lambda_j \in \mathcal{L}$ son también Donsker (ver ejemplo 8). Se sigue del ejemplo 5 que $-\Lambda_j(T)e^{\beta_j'\mathbf{Z}} + \gamma_j'\mathbf{X}$ con θ variando sobre Θ es Donsker, para cada $j \in \mathcal{J}$. Por el ejemplo 4, se concluye que tanto el numerador así como también el denominador en (5.12) con θ variando sobre Θ son clases Donsker. Como el denominador es mayor que cero, implica que $\mathcal{F}$ es Donsker (ver ejemplos 6 y 7). Por el ejemplo 6, $\Delta^j + (1-\Delta)w^j(\mathbf{O};\theta)$ es Donsker cuando θ pertenece a Θ. Finalmente, $\{Y(s) : s \in [0,\tau]\}$ es Donsker, multiplicando clases Donsker se concluye que la clase $\{h^j(s,\mathbf{O};\theta) : s \in [0,\tau], \theta \in \Theta\}$ es Donsker. Siguiendo argumentos similares se puede demostrar que $\{\Delta^j 1\{T \leq t\}/P_{\theta_0}\left[h^j(s,\mathbf{O};\theta_0)\right]|_{s=T} : t \in [0,\tau]\}$ es también una clase Donsker.

Ahora bien, para cada $j \in \mathcal{J}$ y $t \in [0,\tau]$, defínase

$$\Lambda_j(t;\theta_0) = \int_0^t \frac{1}{P_{\theta_0}[h^j(s,\mathbf{O};\theta_0)]}\, P_{\theta_0} dN^j(s).$$

Entonces,

$$
\begin{aligned}
&\sup_{t\in[0,\tau]} \left|\widetilde{\Lambda}_{j,n}(t) - \Lambda_j(t;\theta_0)\right| \\
&\quad = \sup_{t\in[0,\tau]} \left| \frac{1}{n}\sum_{i=1}^{n} \frac{\Delta_i^j 1\{T_i \le t\}}{H_n^j(T_i;\theta_0)} - P_{\theta_0}\left[\frac{\Delta^j 1\{T\le t\}}{P_{\theta_0}\left[h^j(s,\mathbf{O};\theta_0)\right]|_{s=T}}\right]\right| \\
&\quad \le \sup_{t\in[0,\tau]} \left| \frac{1}{n}\sum_{i=1}^{n} \Delta_i^j 1\{T_i \le t\} \left\{\frac{1}{H_n^j(s;\theta_0)} - \frac{1}{P_{\theta_0}\left[h^j(s,\mathbf{O};\theta_0)\right]}\right\}_{|_{s=T_i}} \right| \\
&\qquad\quad + \sup_{t\in[0,\tau]} \left|(\mathbb{P}_n - P_{\theta_0})\left[\frac{\Delta^j 1\{T\le t\}}{P_{\theta_0}\left[h^j(s,\mathbf{O};\theta_0)\right]|_{s=T}}\right]\right| \\
&\quad \le \sup_{s\in[0,\tau]} \left| \frac{1}{H_n^j(s;\theta_0)} - \frac{1}{P_{\theta_0}\left[h^j(s,\mathbf{O};\theta_0)\right]}\right| \\
&\qquad\quad + \sup_{t\in[0,\tau]} \left|(\mathbb{P}_n - P_{\theta_0})\left[\frac{\Delta^j 1\{T\le t\}}{P_{\theta_0}\left[h^j(s,\mathbf{O};\theta_0)\right]|_{s=T}}\right]\right|. \qquad (5.13)
\end{aligned}
$$

Como $\{h^j(s,\mathbf{O};\theta_0) : s \in [0,\tau]\}$ es Donsker y por la definición de la clase de funciones Glivenko-Cantelli, el término $\sup_{s\in[0,\tau]}\left|H_n^j(s;\theta_0) - P_{\theta_0}\left[h^j(s,\mathbf{O};\theta_0)\right]\right|$ converge a 0 casi seguramente. Más aún, para cada $s \in [0,\tau]$, $P_{\theta_0}\left[h^j(s,\mathbf{O};\theta_0)\right] \ge c_2 P_{\theta_0}\left[Y(s)\Gamma^j\right]$ (por **C6**) y bajo la suposición **C7**, $P_{\theta_0}\left[h^j(s,\mathbf{O};\theta_0)\right] > 0$ sobre $[0,\tau]$. Entonces, el primer término del lado derecho de (5.13) converge a 0 casi seguramente, mientras que el segundo converge casi seguramente a 0 por la propiedad de Glivenko-Cantelli de $\{\Delta^j 1\{T \le t\}/P_{\theta_0}\left[h^j(s,\mathbf{O};\theta_0)\right]|_{s=T} : t \in [0,\tau]\}$. Por tanto, se concluye que $\widetilde{\Lambda}_{j,n}$ converge uniformemente a $\Lambda_j(\cdot;\theta_0)$, casi seguramente. Además obsérvese que $\Lambda_j(\cdot;\theta_0)$ es igual a $\Lambda_{j,0}$, lo cual concluye la demostración.

□

Teorema 5.7.1. *Bajo las condiciones **C1**-**C9** y para cada $j \in \mathcal{J}$,*

$$
\|\widehat{\mathbf{B}}_n - \boldsymbol{B}_0\|, \quad \|\widehat{\mathbf{G}}_n - \boldsymbol{G}_0\|, \quad y \quad \sup_{t\in[0,\tau]} |\widehat{\Lambda}_{j,n}(t) - \Lambda_{j,0}(t)|,
$$

convergen a 0 casi seguramente cuando n tiende a infinito.

Demostración: La demostración consiste en dos pasos,

1. Se demostrará que cada sucesión de n contiene otra subsucesión donde el NPMLE $\widehat{\theta}_n$ converge.

2. Se demostrará que el conjunto límite de todas las subsucesiones convergentes de $\widehat{\theta}_n$ se reduce a $\{\theta_0\}$.

Demostración de (1). Como $\mathcal{B} \times \mathcal{G}$ es un conjunto compacto y utilizando el teorema de Bolzanno-Weierstrass se tiene que para cada sucesión de $(\widehat{\mathbf{B}}_n, \widehat{\mathbf{G}}_n)$ existe una subsucesión llamémosla $(\widehat{\mathbf{B}}_{\phi(n)}, \widehat{\mathbf{G}}_{\phi(n)})$ que converge a un límite $(\mathbf{B}^*, \mathbf{G}^*)$ en $\mathcal{B} \times \mathcal{G}$. Utilizando el lema 5.7.1 y el teorema de Helly se puede encontrar con probabilidad 1 una subsucesión $\widehat{\Lambda}_{j,\varphi(n)}$ de $\widehat{\Lambda}_{j,\phi(n)}$ y una función creciente continua por la derecha Λ_j^* tal que $\widehat{\Lambda}_{j,\varphi(n)}(t) \to \Lambda_j^*(t)$ débilmente para todo $t \in [0, \tau]$, donde Λ_j^* es continua. Extrayendo sucesivamente sub-subsuseciones, se puede encontrar una subsucesión $\xi(n)$ de $\varphi(n)$ de tal manera que esta convergencia débil se mantiene a lo largo de $\xi(n)$ para cada $j \in \mathcal{J}$. Ahora se mostrará que Λ_j^* es contínua para cada $j \in \mathcal{J}$ en $[0, \tau]$. Obsérvese que

$$\widehat{\Lambda}_{j,\xi(n)}(t) = \int_0^t \frac{\mathbb{P}_{\xi(n)}[h^j(s, \mathbf{O}; \theta_0)]}{\mathbb{P}_{\xi(n)}[h^j(s, \mathbf{O}; \widehat{\theta}_{\xi(n)})]} \, d\widetilde{\Lambda}_{j,\xi(n)}(s), \tag{5.14}$$

donde $\widetilde{\Lambda}_{j,n}$ es definida en el lema 5.7.2. Por la propiedad de Glivenko-Cantelli del conjunto $\{h^j(s, \mathbf{O}; \theta) : s \in [0, \tau], \theta \in \Theta\}$ se sigue que (ver la demostración del lema 5.7.2)

$$\begin{aligned}
&\sup_{s\in[0,\tau]} \left|\mathbb{P}_{\xi(n)}[h^j(s, \mathbf{O}; \theta_0)] - P_{\theta_0}[h^j(s, \mathbf{O}; \theta_0)]\right| \longrightarrow 0 \quad c.s., \\
&\sup_{s\in[0,\tau]} \left|\mathbb{P}_{\xi(n)}[h^j(s, \mathbf{O}; \widehat{\theta}_{\xi(n)})] - P_{\theta_0}[h^j(s, \mathbf{O}; \widehat{\theta}_{\xi(n)})]\right| \longrightarrow 0 \quad c.s..
\end{aligned} \tag{5.15}$$

Utilizando del teorema de convergencia acotada, el hecho de que, $(\widehat{\mathbf{B}}_{\xi(n)}, \widehat{\mathbf{G}}_{\xi(n)})$ converge a $(\mathbf{B}^*, \mathbf{G}^*)$ y que $\widehat{\Lambda}_{j,\xi(n)}$ converge débilmente a Λ_j^*, se obtiene que $P_{\theta_0}[h^j(s, \mathbf{O}; \widehat{\theta}_{\xi(n)})]$ converge a $P_{\theta_0}[h^j(s, \mathbf{O}; \theta^*)]$ para cada $s \in [0, \tau]$, donde $\theta^* = (\mathbf{B}^*, \mathbf{G}^*, \Lambda_j^*; j \in \mathcal{J})$. Más aún, por **C3**, la derivada de $P_{\theta_0}[h^j(s, \mathbf{O}; \widehat{\theta}_{\xi(n)})]$ con respecto a s es uniformemente acotada; por lo tanto, la sucesión de funciones $P_{\theta_0}[h^j(\cdot, \mathbf{O}; \widehat{\theta}_{\xi(n)})]$ es equicontínua. Por el teorema de Arzela-Ascoli, existe una subsucesión $\psi(n)$ de $\xi(n)$, tal que $P_{\theta_0}[h^j(\cdot, \mathbf{O}; \widehat{\theta}_n)]$ converge uniformemente a $P_{\theta_0}[h^j(\cdot, \mathbf{O}; \theta^*)]$ en $[0, \tau]$ a lo largo de esta subsucesión. Se puede suponer que esta subsucesión es la misma para todos los $j \in \mathcal{J}$. Utilizando éste resultado, la ecuación (5.15) y la desigualdad del triángulo, se obtiene que

$$\frac{d\widehat{\Lambda}_{j,\psi(n)}(t)}{d\widetilde{\Lambda}_{j,\psi(n)}(t)} = \frac{\mathbb{P}_{\psi(n)}[h^j(t, \mathbf{O}; \theta_0)]}{\mathbb{P}_{\psi(n)}[h^j(s, \mathbf{O}; \widehat{\theta}_{\psi(n)})]} \longrightarrow \frac{P_{\theta_0}[h^j(t, \mathbf{O}; \theta_0)]}{P_{\theta_0}[h^j(t, \mathbf{O}; \theta^*)]}$$

uniformemente en $t \in [0, \tau]$. Tomando límites en ambos lados de $\widehat{\Lambda}_{j,\psi(n)}(t)$ en (5.14), se obtiene que

$$\Lambda_j^*(t) = \int_0^t \frac{P_{\theta_0}[h^j(s, \mathbf{O}; \theta_0)]}{P_{\theta_0}[h^j(s, \mathbf{O}; \theta^*)]} \, d\Lambda_{j,0}(s).$$

Se concluye que Λ_j^* es absolutamente continua con respecto a $\Lambda_{j,0}$, y por lo tanto $\Lambda_j^*(t)$ es diferenciable con respecto t, y por consecuencia continua. Una segunda conclusión se obtiene utilizando el teorema de Dini, el que garantiza que $\widehat{\Lambda}_{j,\psi(n)}$ converge uniformemente a Λ_j^* con probabilidad 1. Más aún, $d\widehat{\Lambda}_{j,\psi(n)}(t)/d\widetilde{\Lambda}_{j,\psi(n)}(t)$ converge a $d\Lambda_j^*(t)/d\Lambda_{j,0}(t) := \lambda_j^*(t)/\lambda_{j,0}(t)$ uniformemente en t.

En resumen, para cada sucesión de n se puede encontrar una subsucesión $\psi(n)$ y un elemento $(\mathbf{B}^*, \mathbf{G}^*, \Lambda_j^*; j \in \mathcal{J})$ tal que $\|\widehat{\mathbf{B}}_{\psi(n)} - \mathbf{B}^*\|$, $\|\widehat{\mathbf{G}}_{\psi(n)} - \mathbf{G}^*\|$ y $\sup_{t\in[0,\tau]} |\widehat{\Lambda}_{j,\psi(n)}(t) - \Lambda_j^*(t)|$ convergen a 0 casi seguramente para cada $j \in \mathcal{J}$.

Demostración de (2). Considérese la siguiente diferencia,

$$0 \leq \frac{1}{\psi(n)} \log L_{\psi(n)}(\widehat{\mathbf{B}}_{\psi(n)}, \widehat{\mathbf{G}}_{\psi(n)}, \widehat{\Lambda}_{j,\psi(n)}; j \in \mathcal{J}) - \frac{1}{\psi(n)} \log L_{\psi(n)}(\mathbf{B}_0, \mathbf{G}_0, \widetilde{\Lambda}_{j,\psi(n)}; j \in \mathcal{J}).$$

Haciendo tender n hacia infinito, se obtiene que

$$\begin{aligned} 0 \leq P_{\theta_0} \Bigg[\sum_{j\in\mathcal{J}} \log & \left(\frac{\lambda_j^*(T) \exp\left(\beta_j^{*\prime}\mathbf{Z} - e^{\beta_j^{*\prime}\mathbf{Z}} \Lambda_j^*(T)\right) p_{\mathbf{G}^*}^{j,\mathbf{X}}}{\lambda_{j,0}(T) \exp\left(\beta_{j,0}^{\prime}\mathbf{Z} - e^{\beta_{j,0}^{\prime}\mathbf{Z}} \Lambda_{j,0}(T)\right) p_{\mathbf{G}_0}^{j,\mathbf{X}}} \right)^{\Delta^j} \\ & + (1 - \Delta) \log \left(\frac{\sum_{j\in\mathcal{J}} \exp\left(-e^{\beta_j^{*\prime}\mathbf{Z}} \Lambda_j^*(T)\right) p_{\mathbf{G}^*}^{j,\mathbf{X}}}{\sum_{j\in\mathcal{J}} \exp\left(-e^{\beta_{j,0}^{\prime}\mathbf{Z}} \Lambda_{j,0}(T)\right) p_{\mathbf{G}_0}^{j,\mathbf{X}}} \right) \Bigg], \end{aligned}$$

este resultado es obtenido aplicando los resultados del inciso anterior. Obsérvese que el lado derecho de la desigualdad es el negativo de la información de Kulback-Leibler, entonces, $P_{\theta_0}\left[log \frac{L(\theta^*)}{L(\theta_0)}\right] = 0$ y por lo tanto, se sigue de la identificabilidad del parámetro θ (ver proposicion, 5.6.3) que $\theta^* = \theta$.

Combinando los resultados de los pasos (1) y (2), se concluye que la sucesión $\|\widehat{\mathbf{B}}_n - \mathbf{B}_0\|$, $\|\widehat{\mathbf{G}}_n - \mathbf{G}_0\|$ y $\sup_{t\in[0,\tau]} |\widehat{\Lambda}_{j,n}(t) - \Lambda_{j,0}(t)|$ para cada $j \in \mathcal{J}$, convergen a 0 casi seguramente cuando n tiende a infinito.

□

5.8. Normalidad asintótica

5.8.1. Score e información

Una vez establecida la consistencia, se puede estudiar la distribución asintótica de los NPMLEs propuestos por Escarela y Bowater (2008) para el modelo de riesgos competitivos. Para derivar la normalidad asintótica, se adopta una función analítica desarrollada por Murphy (1995) para el estudio de modelos frailty (ver también Fang *et al.*, 2005; Kosorok y Song, 2007 y Lu, 2008; quienes recientemente han adoptado este acercamiento para varios modelos semi-paramétricos de regresión para el análisis de datos de supervivencia).

Para calcular las funciones de score, se trabaja con submodelos unidimensionales $\widehat{\theta}_{n,\epsilon}$ los cuales pasan por el estimador $\widehat{\theta}_n$. Espacíficamente se considera el submodelo

$$\epsilon \longmapsto \widehat{\theta}_{n,\epsilon} = \left(\widehat{\mathbf{B}}_n + \epsilon\mathbf{h}_{\mathbf{B}}, \widehat{\mathbf{G}}_n + \epsilon\mathbf{h}_{\mathbf{G}}, \int_0^{\cdot} \left(1 + \epsilon h_{\Lambda_j}(s)\right) d\widehat{\Lambda}_{j,n}(s); j \in \mathcal{J}\right),$$

donde $\mathbf{h}_{\mathbf{B}} = (h'_{\beta_1} \dots h'_{\beta_J})'$, $\mathbf{h}_{\mathbf{G}} = (h'_{\gamma_1} \dots h'_{\gamma_{J-1}})'$, h_{β_j} es un vector de dimensión p ($j \in \mathcal{J}$), h_{γ_j} es un vector de dimensión q ($j = 1, \dots, J-1$) y h_{Λ_j} es una función no negativa sobre $[0, \tau]$ ($j \in \mathcal{J}$). Denotemos por $\mathbf{h}$ la colección $(\mathbf{h}_{\mathbf{B}}, \mathbf{h}_{\mathbf{G}}, h_{\Lambda_j}; j \in \mathcal{J})$.

Las ecuaciones de score son obtenidas derivando $l_{\widehat{\theta}_n}(\widehat{\theta}_{n,\epsilon})$ con respecto a ϵ y evaluándolas en $\epsilon = 0$. Como $\widehat{\theta}_n$ maximiza $l_{\widehat{\theta}_n}(\theta)$ entonces se satisface,

$$\left.\frac{\partial l_{\widehat{\theta}_n}(\widehat{\theta}_{n,\epsilon})}{\partial \epsilon}\right|_{\epsilon=0} = 0 \tag{5.16}$$

para cada $\mathbf{h}$. Defínase $\Psi_{\mathbf{B}}(\theta) = (\Psi_{\beta_1}(\theta)' \dots \Psi_{\beta_J}(\theta)')'$ y $\Psi_{\mathbf{G}}(\theta) = (\Psi_{\gamma_1}(\theta)' \dots \Psi_{\gamma_{J-1}}(\theta)')'$, donde

$$\Psi_{\beta_j}(\theta) = \Delta^j \mathbf{Z} - g^j(\mathbf{O}, \theta)\mathbf{Z}e^{\beta_j'\mathbf{Z}}\Lambda_j(T),$$

para cada $j \in \mathcal{J}$, y

$$\Psi_{\gamma_j}(\theta) = \mathbf{X}\left(g^j(\mathbf{O}, \theta) - p_{\mathbf{G}}^{j,\mathbf{X}}\right).$$

para cada $j = 1, \dots, J-1$. Para cada $j \in \mathcal{J}$, definase también,

$$\Psi_{\Lambda_j}(\theta)(h_{\Lambda_j}) = \Delta^j h_{\Lambda_j}(T) - g^j(\mathbf{O}, \theta)e^{\beta_j'\mathbf{Z}} \int_0^T h_{\Lambda_j}(s)\, d\Lambda_j(s).$$

Después de hacer los cálculos correspondientes, la ecuación de score (5.16) puede ser reex-presada como $\Psi_n(\widehat{\theta}_n)(\mathbf{h}) = 0$, donde $\Psi_n(\widehat{\theta}_n)(\mathbf{h})$ tiene la forma,

$$\Psi_n(\widehat{\theta}_n)(\mathbf{h}) = \mathbb{P}_n\left[\mathbf{h}_{\mathbf{B}}'\Psi_{\mathbf{B}}(\widehat{\theta}_n) + \mathbf{h}_{\mathbf{G}}'\Psi_{\mathbf{G}}(\widehat{\theta}_n) + \sum_{j\in\mathcal{J}} \Psi_{\Lambda_j}(\widehat{\theta}_n)(h_{\Lambda_j})\right]. \tag{5.17}$$

El espacio de elementos de $\mathbf{h}$ es definido como,

$$\mathcal{H} = \Big\{\mathbf{h} = (\mathbf{h}_{\mathbf{B}}, \mathbf{h}_{\mathbf{G}}, h_{\Lambda_j}; j \in \mathcal{J}) : \mathbf{h}_{\mathbf{B}} \in \mathbb{R}^P, ||\mathbf{h}_{\mathbf{B}}|| < \infty; \mathbf{h}_{\mathbf{G}} \in \mathbb{R}^Q, ||\mathbf{h}_{\mathbf{G}}|| < \infty;$$
$$h_{\Lambda_j} : [0, \tau] \to \mathbb{R}, ||h_{\Lambda_j}||_v < \infty, j \in \mathcal{J}\Big\},$$

donde $\|h_{\Lambda_j}\|_v$ denota la variación total de h_{Λ_j} sobre $[0, \tau]$. Más aún, se toma a las funciones h_{Λ_j} como continuas por la derecha en 0.

Defínase,

$$\theta(\mathbf{h}) = \mathbf{h}'_{\mathbf{B}}\mathbf{B} + \mathbf{h}'_{\mathbf{G}}\mathbf{G} + \sum_{j \in \mathcal{J}} \int_0^{\tau} h_{\Lambda_j}(s)\, d\Lambda_j(s),$$

donde $\mathbf{h} \in \mathcal{H}$. A partir de esto, se puede reconsiderar al parámetro θ como un funcional lineal sobre $\mathcal{H}$ y al espacio de parámetros Θ como un subconjunto de $l^{\infty}(\mathcal{H})$, el cual representa el espacio de todas las funciones real-valuadas sobre $\mathcal{H}$ cuya representación esta dada por la norma uniforme. Por otra parte, el operador de score Ψ_n es un mapeo de Θ sobre el espacio $l^{\infty}(\mathcal{H})$.

Observación. La selección apropiada de $\mathbf{h}$ permite obtener los valores originales de θ; denotaremos por $\mathbf{0}_r$ $(r \geq 2)$ el vector columna de dimensión r cuyas entradas son todas iguales a 0.

Por ejemplo, si se toma $\mathbf{h}_{\mathbf{G}} = \mathbf{0}_Q$, $h_{\Lambda_j}(\cdot) = 0$ para cada $j \in \mathcal{J}$ y $\mathbf{h}_{\mathbf{B}} = (h'_{\beta_1} \ldots h'_{\beta_J})'$ tal que $h_{\beta_j} = \mathbf{0}_p$ para cada $j \in \mathcal{J}$ excepto para algún $j = l$, con h_{β_l} un vector de dimensión p que es igual a 1 en la i-ésima posición y cero en cualquier otro lado, dando así el i-ésimo componente de β_l. Otro ejemplo, es tomar $\mathbf{h}_{\mathbf{B}} = \mathbf{0}_P$, $\mathbf{h}_{\mathbf{G}} = \mathbf{0}_Q$, $h_{\Lambda_j}(\cdot) = 0$ para cada $j \in \mathcal{J}$ excepto para $h_{\Lambda_l}(\cdot) = 1\{\cdot \leq t\}$, para algún $t \in (0, \tau)$. En este caso, $\theta(\mathbf{h})$ se reduce a $\Lambda_l(t)$.

Defínase el *operador información* $\sigma = (\sigma_{\mathbf{B}}, \sigma_{\mathbf{G}}, \sigma_{\Lambda_j}; j \in \mathcal{J}) : \mathcal{H} \to \mathcal{H}$ por

$$\sigma_{\mathbf{B}}(\mathbf{h}) = P_{\theta_0}\left[2\Psi_{\mathbf{B}}(\theta_0)\sum_{j \in \mathcal{J}} \Delta^j h_{\Lambda_j}(T)\right] + P_{\theta_0}\left[\Psi_{\mathbf{B}}(\theta_0)^{\otimes 2}\right]\mathbf{h}_{\mathbf{B}} + P_{\theta_0}\left[\Psi_{\mathbf{B}}(\theta_0)\Psi_{\mathbf{G}}(\theta_0)'\right]\mathbf{h}_{\mathbf{G}}$$

$$\sigma_{\mathbf{G}}(\mathbf{h}) = P_{\theta_0}\left[2\Psi_{\mathbf{G}}(\theta_0)\sum_{j \in \mathcal{J}} \Delta^j h_{\Lambda_j}(T)\right] + P_{\theta_0}\left[\Psi_{\mathbf{G}}(\theta_0)^{\otimes 2}\right]\mathbf{h}_{\mathbf{G}} + P_{\theta_0}\left[\Psi_{\mathbf{G}}(\theta_0)\Psi_{\mathbf{B}}(\theta_0)'\right]\mathbf{h}_{\mathbf{B}}$$

$$\begin{aligned}
\sigma_{\Lambda_j}(\mathbf{h})(s) &= h_{\Lambda_j}(s)P_{\theta_0}\left[W^j(s, \mathbf{O}, \theta_0)\right] \\
&\quad - P_{\theta_0}\left[2\Delta^j h_{\Lambda_j}(T)W^j(s, \mathbf{O}, \theta_0) - \left\{W^j(s, \mathbf{O}, \theta_0)\right\}^2 \int_0^T h_{\Lambda_j}(u)\, d\Lambda_{j,0}(u)\right] \\
&\quad + P_{\theta_0}\Bigg[2W^j(s, \mathbf{O}, \theta_0)\sum_{k > j}\Bigg\{W^k(s, \mathbf{O}, \theta_0)\int_0^T h_{\Lambda_k}(u)\, d\Lambda_{k,0}(u) \\
&\qquad\qquad - W^k(s, \mathbf{O}, \theta_0)\int_0^s h_{\Lambda_k}(u)\, d\Lambda_{k,0}(u) - \Delta^k h_{\Lambda_k}(T)\Bigg\}\Bigg] \\
&\quad - \mathbf{h}'_{\mathbf{B}}P_{\theta_0}\left[2\Psi_{\mathbf{B}}(\theta_0)g^j(\mathbf{O}; \theta_0)e^{\beta'_{j,0}\mathbf{Z}}Y(s)\right] - \mathbf{h}'_{\mathbf{G}}P_{\theta_0}\left[2\Psi_{\mathbf{G}}(\theta_0)g^j(\mathbf{O}; \theta_0)e^{\beta'_{j,0}\mathbf{Z}}Y(s)\right],
\end{aligned}$$

donde $W^j(s, \mathbf{O}, \theta_0) = Y(s)e^{\beta'_{j,0}\mathbf{Z}}g^j(\mathbf{O}, \theta_0),\ j \in \mathcal{J},\ s \in [0, \tau]$.

Observación. Algunos de los términos en σ puede ser simplifacados mediante el uso de las propiedades de la esperanza condicional. Por ejemplo, $P_{\theta_0}[W^j(s, \mathbf{O}, \theta_0)]$ en el término $\sigma_{\Lambda_j}(\mathbf{h})$ se simplifica a $P_{\theta_0}[Y(s)e^{\beta'_{j,0}\mathbf{Z}}\Gamma^j]$. Sin embargo, para fines de estimación de la varianza, se construirá más adelante una versión empírica de σ reemplazando θ_0 y P_{θ_0} por $\widehat{\theta}_n$ y $\mathbb{P}_n$ respectivamente en $\sigma_{\mathbf{B}}$,$\sigma_{\mathbf{G}}$ y σ_{Λ_j}.

Los siguientes lemas establecen algunas propiedades útiles para el operador de score y el operador información.

Lema 5.8.1. *Sea* $\mathbf{h} \in \mathcal{H}$. *Entonces* $P_{\theta_0}[\Psi_1(\theta_0)(\mathbf{h})] = 0$ *y*

$$P_{\theta_0}\left[\Psi_1(\theta_0)(\mathbf{h})^2\right] = \mathbf{h}'_{\boldsymbol{B}}\sigma_{\boldsymbol{B}}(\mathbf{h}) + \mathbf{h}'_{\boldsymbol{G}}\sigma_{\boldsymbol{G}}(\mathbf{h}) + \sum_{j\in\mathcal{J}}\int_0^\tau \sigma_{\Lambda_j}(\mathbf{h})(s)h_{\Lambda_j}(s)\,d\Lambda_{j,0}(s).$$

donde $\sigma_{\boldsymbol{B}}$, $\sigma_{\boldsymbol{G}}$, *y* σ_{Λ_j} $(j \in \mathcal{J})$ *siguen las expresiones mencionadas anteriormente.*

Demostración: Sea $j \in \mathcal{J}$. Entonces

$$\begin{aligned} P_{\theta_0}\left[\Psi_{\beta_j}(\theta_0)\right] &= P_{\theta_0}\left[\Delta^j\mathbf{Z} - g^j(\mathbf{O}, \theta_0)\mathbf{Z}e^{\beta'_{j,0}\mathbf{Z}}\Lambda_{j,0}(T)\right] \\ &= P_{\theta_0}\left[\Delta^j\mathbf{Z} - \Gamma^j\mathbf{Z}e^{\beta'_{j,0}\mathbf{Z}}\Lambda_{j,0}(T)\right] \\ &= P_{\theta_0}\left[\mathbf{Z}\Gamma^jM(\tau)\right], \end{aligned}$$

donde la segunda línea es obtenida por las propiedades de la esperanza condicional, definiendo $M(t) = N(t) - \sum_{l\in\mathcal{J}}\int_0^t \Gamma^l e^{\beta'_{l,0}\mathbf{Z}}Y(s)\,d\Lambda_{l,0}(s)$ la martingala del proceso contable con respecto a la fltaración $\sigma\{N(s), 1\{T \le s, \Delta = 0\}, \mathbf{Z}, \mathbf{X}, H : 0 \le s \le t\}$. Como $\mathbf{Z}$ y Γ^j son acotadas y medibles con respecto a la filtarción M, entonces $P_{\theta_0}\left[\mathbf{Z}\Gamma^jM(\tau)\right] = 0$. Argumentos similares implican que $P_{\theta_0}\left[\Psi_{\gamma_j}(\theta_0)\right] = 0$ y $P_{\theta_0}\left[\Psi_{\Lambda_j}(\theta_0)(h_{\Lambda_j})\right] = 0$ para cada $j \in \mathcal{J}$. Esto conluye la primera parte del lema.

Desarrollando $\Psi_1(\theta_0)(\mathbf{h})^2$ se obtiene,

$$\begin{aligned} \Psi_1(\theta_0)(\mathbf{h})^2 &= \left[\mathbf{h}'_{\mathbf{B}}\Psi_{\mathbf{B}}(\theta_0) + \mathbf{h}'_{\mathbf{G}}\Psi_{\mathbf{G}}(\theta_0) + \sum_{j\in\mathcal{J}}\Psi_{\Lambda_j}(\theta_0)(h_{\Lambda_j})\right]^2 \\ &= \mathbf{h}'_{\mathbf{B}}\Psi^2_{\mathbf{B}}(\theta_0)\mathbf{h}_{\mathbf{B}} + \mathbf{h}'_{\mathbf{G}}\Psi^2_{\mathbf{G}}(\theta_0)\mathbf{h}_{\mathbf{G}} + \left(\sum_{j\in\mathcal{J}}\Psi_{\Lambda_j}(\theta_0)(h_{\Lambda_j})\right)^2 + \mathbf{h}'_{\mathbf{B}}\Psi_{\mathbf{B}}(\theta_0)\Psi_{\mathbf{G}}(\theta_0)'\mathbf{h}_{\mathbf{G}} \\ &\quad + \mathbf{h}'_{\mathbf{G}}\Psi_{\mathbf{G}}(\theta_0)\Psi_{\mathbf{B}}(\theta_0)'\mathbf{h}_{\mathbf{B}} + 2\mathbf{h}'_{\mathbf{B}}\Psi_{\mathbf{B}}(\theta_0)\sum_{j\in\mathcal{J}}\Psi_{\Lambda_j}(\theta_0)(h_{\Lambda_j}) \\ &\quad + 2\mathbf{h}'_{\mathbf{G}}\Psi_{\mathbf{G}}(\theta_0)\sum_{j\in\mathcal{J}}\Psi_{\Lambda_j}(\theta_0)(h_{\Lambda_j}) \end{aligned} \tag{5.18}$$

$$\begin{aligned}
&= \mathbf{h}_{\mathbf{B}}'\Psi_{\mathbf{B}}^2(\theta_0)h_{\mathbf{B}} + \mathbf{h}_{\mathbf{G}}'\Psi_{\mathbf{G}}^2(\theta_0)\mathbf{h}_{\mathbf{G}} + \left(\sum_{j\in\mathcal{J}}\Psi_{\Lambda_j}(\theta_0)(h_{\Lambda_j})\right)^2 \\
&+\mathbf{h}_{\mathbf{B}}'\Psi_{\mathbf{B}}(\theta_0)\Psi_{\mathbf{G}}(\theta_0)'\mathbf{h}_{\mathbf{G}} + \mathbf{h}_{\mathbf{G}}'\Psi_{\mathbf{G}}(\theta_0)\Psi_{\mathbf{B}}(\theta_0)'\mathbf{h}_{\mathbf{B}} \\
&+2\mathbf{h}_{\mathbf{B}}'\Psi_{\mathbf{B}}(\theta_0)\sum_{j\in\mathcal{J}}\left[\Delta^j h_{\Lambda_j}(T) - g^j(\mathbf{O},\theta)e^{\beta_j'\mathbf{Z}}\int_0^T h_{\Lambda_j}(s)\,d\Lambda_j(s)\right] \\
&+2\mathbf{h}_{\mathbf{G}}'\Psi_{\mathbf{G}}(\theta_0)\sum_{j\in\mathcal{J}}\left[\Delta^j h_{\Lambda_j}(T) - g^j(\mathbf{O},\theta)e^{\beta_j'\mathbf{Z}}\int_0^T h_{\Lambda_j}(s)\,d\Lambda_j(s)\right]
\end{aligned} \tag{5.19}$$

Obsérvese que el término

$$\begin{aligned}
\left(\sum_{j\in\mathcal{J}}\Psi_{\Lambda_j}(\theta_0)(h_{\Lambda_j})\right)^2 &= \left(\sum_{j\in\mathcal{J}}\Delta^j h_{\Lambda_j}(T) - g^j(\mathbf{O},\theta)e^{\beta_j'\mathbf{Z}}\int_0^T h_{\Lambda_j}(s)\,d\Lambda_j(s)\right)^2 \\
&= \sum_{j\in\mathcal{J}}\left(\left(\Delta^j h_{\Lambda_j}(T)\right)^2 + \left(g^j(\mathbf{O},\theta)e^{\beta_j'\mathbf{Z}}\int_0^T h_{\Lambda_j}(s)\,d\Lambda_j(s)\right)^2\right. \\
&\quad -2\Delta^j h_{\Lambda_j}(T)g^j(\mathbf{O},\theta)e^{\beta_j'\mathbf{Z}}\int_0^T h_{\Lambda_j}(s)\,d\Lambda_j(s) \\
&\quad +P_{\theta_0}\left[2W^j(s,\mathbf{O},\theta_0)\sum_{k>j}\left\{W^k(s,\mathbf{O},\theta_0)\int_0^T h_{\Lambda_k}(u)\,d\Lambda_{k,0}(u)\right.\right. \\
&\quad \left.\left.\left. -W^k(s,\mathbf{O},\theta_0)\int_0^s h_{\Lambda_k}(u)\,d\Lambda_{k,0}(u) - \Delta^k h_{\Lambda_k}(T)\right\}\right]\right)
\end{aligned} \tag{5.20}$$

Por otro lado, como $P_{\theta_0}\left[\Psi_{\Lambda_j}(\theta_0)(h_{\Lambda_j})\right] = 0$, implica que

$$P_{\theta_0}\left[\Delta^j h_{\Lambda_j}(T)\right] = P_{\theta_0}\left[g^j(\mathbf{O},\theta)e^{\beta_j'\mathbf{Z}}\int_0^T h_{\Lambda_j}(s)\,d\Lambda_j(s)\right]$$

para todo $\mathbf{h} \in \mathcal{H}$. En paricular, tomando $\mathbf{h}^2$ se cumple la igualdad de arriba , es decir,

$$P_{\theta_0}\left[\Delta^j h_{\Lambda_j}^2(T)\right] = P_{\theta_0}\left[g^j(\mathbf{O},\theta)e^{\beta_j'\mathbf{Z}}\int_0^T h_{\Lambda_j}^2(s)\,d\Lambda_j(s)\right] \tag{5.21}$$

Finalmente, sustituyendo los resultados de (5.20) y (5.21) en (5.18) y tomando la esperanza se obtiene el resultado deseado.

□

Lema 5.8.2. *El operador σ es inyectivo.*

Demostración: Supóngase que $\sigma(\mathbf{h}) = 0$. Por el lema 5.8.1, $P_{\theta_0}\left[\Psi_1(\theta_0)(\mathbf{h})^2\right] = 0$ y por lo tanto $\Psi_1(\theta_0)(\mathbf{h}) = 0$ casi seguramente.

Sea $j \in \mathcal{J}$, por la suposición **C8**, para cada $t \in [0,\tau]$, $\|\mathbf{z}\| < c_1$ y $\|\mathbf{x}\| < c_1$, existe un subconjunto no despreciable de Ω (llamémoslo Ω') tal que $T(\omega) = t$, $\mathbf{Z}(\omega) = \mathbf{z}$, $\mathbf{X}(\omega) = \mathbf{x}$, $\Delta(\omega) = 1$ y $H(\omega) = j$ cuando $\omega \in \Omega'$. Si la igualdad $\Psi_1(\theta_0)(\mathbf{h}) = 0$ se cumple casi seguramente, entonces, en particular eso se cumple para algún $\omega \in \Omega'$. Entonces para algún ω, $\Psi_1(\theta_0)(\mathbf{h}) = 0$ se reduce a,

$$h_{\Lambda_j}(t) + h'_{\beta_j}\mathbf{z} + h'_{\gamma_j}x - \sum_{l=1}^{J-1} h'_{\gamma_l}\mathbf{x}p_{\mathbf{G}_0}^{l,x}$$
$$-e^{\beta'_{j,0}\mathbf{z}}\left[\int_0^t h_{\Lambda_j}(s)\,d\Lambda_{j,0}(s) + h'_{\beta_j}\mathbf{z}\Lambda_{j,0}(t)\right] = 0, \tag{5.22}$$

con la convención de que $h_{\gamma_J} = 0$. Escogiendo t arbitrariamente cercano a 0, la ecuación (5.22) se reduce a

$$h_{\Lambda_j}(0) + h'_{\beta_j}\mathbf{z} + h'_{\gamma_j}\mathbf{x} - \sum_{l=1}^{J-1} h'_{\gamma_l}\mathbf{x}p_{\mathbf{G}_0}^{l,\mathbf{x}} = 0, \tag{5.23}$$

donde h_{Λ_j} y $\Lambda_{j,0}$ son continuas por la derecha en 0 y $\Lambda_{j,0}(0) = 0$ (por **C5**). Tomando la diferencia de (5.22)-(5.23) se obtiene la siguiente ecuación,

$$h_{\Lambda_j}(t) - h_{\Lambda_j}(0) - e^{\beta'_{j,0}\mathbf{z}}\left[\int_0^t h_{\Lambda_j}(s)\,d\Lambda_{j,0}(s) + h'_{\beta_j}\mathbf{z}\Lambda_{j,0}(t)\right] = 0, \tag{5.24}$$

para todo $t \in [0,\tau]$, $\|\mathbf{z}\| < c_1$ y $\|\mathbf{x}\| < c_1$.

Sea $t > 0$, entonces $\Lambda_{j,0}(t) > 0$ (por **C5**) y la ecuación (5.24) puede ser reescrita como

$$\frac{h_{\Lambda_j}(t) - h_{\Lambda_j}(0)}{\Lambda_{j,0}(t)} = e^{\beta'_{j,0}\mathbf{z}}\left[r_j(t) + h'_{\beta_j}\mathbf{z}\right], \tag{5.25}$$

donde $r_j(t) = \int_0^t h_{\Lambda_j}(s)\,d\Lambda_{j,0}(s)/\Lambda_{j,0}(t)$.

Considérese primero el caso cuando $\beta_{j,0} = 0$. Entonces el lado izquierdo de (5.25) no depende de $\mathbf{z}$ y por tanto h_{β_j} debría ser igual a 0. Ahora, considérese el caso cuando $\beta_{j,0} \neq 0$. Sean $t_1, t_2 > 0$, entonces $e^{\beta'_{j,0}\mathbf{z}}[r_j(t_1) - r_j(t_2)]$ debría no depender de $\mathbf{z}$. Por la suposisción **C6**, la matriz de covarianza de $\mathbf{Z}$ es positiva definida, entonces se pueden encontrar dos distintos valores $\mathbf{z}_1$ y $\mathbf{z}_2$ de $\mathbf{Z}$ tales que,

$$e^{\beta'_{j,0}\mathbf{z}_1}\left[r_j(t_1) - r_j(t_2)\right] = e^{\beta'_{j,0}\mathbf{z}_2}\left[r_j(t_1) - r_j(t_2)\right].$$

Esto implica que $r_j(t_1) = r_j(t_2)$ y por lo tanto, $h_{\Lambda_j}(t)$ tiene que ser una constante c_6 para cada $t \in (0,\tau]$. De (5.25), se deduce que $h_{\Lambda_j}(0) = c_6$, lo cual implica que $h_{\beta_j} = 0$, $c_6 = 0$ y que

$h_{\Lambda_j}(t) = 0$ para cada $t \in [0,\tau]$. Lo anterior junto con (5.23) implican que $h_{\gamma_j} = 0$ para cada $j = 1,\ldots,J-1$.

Se concluye que $\mathbf{h_B} = 0$, $\mathbf{h_G} = 0$ y que para cada $j \in \mathcal{J}$, $h_{\Lambda_j}(t) = 0$ con $t \in [0,\tau]$.

Poniendo estos resultados en $\sigma_{\Lambda_j}(\mathbf{h})(s) = 0$, se obitiene que $h_{\Lambda_j}(s)P_{\theta_0}\left[W^j(s,\mathbf{O},\theta_0)\right] = 0$ para cada $s \in [0,\tau]$ y $j \in \mathcal{J}$. Por las suposiciones **C2**, **C5** y **C6**, $P_{\theta_0}\left[W^j(\cdot,\mathbf{O},\theta_0)\right]$ es uniformemente acotada por 0 en $[0,\tau]$. Entonces, h_{Λ_j} es identicamente igual a 0 sobre $[0,\tau]$ para cada $j \in \mathcal{J}$. Por lo tanto, se concluye que σ es inyectiva.

□

Lema 5.8.3. *El operador σ es continuamente invertible.*

Demostración: Sea $\mathcal{H}$ un espacio de Banach, para probar que σ es continuamente invertible, es suficiente probar que σ es inyectiva y que puede ser escrita como la suma $A + (\sigma - A)$ de un operador acotado A con inversa acotada y un operador compacto $\sigma - A$ (ver lema 25.93 de van der Vaart, 1998).

Por el lema 5.8.2 σ es inyectiva. Defínase el operador lineal $A : \mathcal{H} \to \mathcal{H}$ por $A(\mathbf{h}) = (\mathbf{h_B}, \mathbf{h_G}, h_{\Lambda_j}(\cdot)P_{\theta_0}\left[W^j(\cdot,\mathbf{O},\theta_0)\right]; j \in \mathcal{J})$. Obsérvese que A es acotado (por las condiciones **C4** y **C6**) y continuamente invertible con inversa $A^{-1}(\mathbf{h}) = (\mathbf{h_B}, \mathbf{h_G}, h_{\Lambda_j}(\cdot)P_{\theta_0}\left[W^j(\cdot,\mathbf{O},\theta_0)\right]^{-1}; j \in \mathcal{J})$ porque el término $P_{\theta_0}\left[W^j(\cdot,\mathbf{O},\theta_0)\right]$ es uniformemente acotado por 0 en $[0,\tau]$ (por **C2**, **C5**, **C6**).

Como un operador lineal acotado con dimensión finita es compacto, entonces solo es necesario mostrar que el operador $K_{\Lambda_j} : VB(0,\tau) \to VB(0,\tau)$ con $j \in \mathcal{J}$, dado por

$$\begin{aligned}
K_{\Lambda_j}(h_{\Lambda_j})(s) \quad = \quad & h_{\Lambda_j}(s)P_{\theta_0}\left[W^j(s,\mathbf{O},\theta_0)\right] \\
& -P_{\theta_0}\left[2\Delta^j h_{\Lambda_j}(T)W^j(s,\mathbf{O},\theta_0) - \left\{W^j(s,\mathbf{O},\theta_0)\right\}^2 \int_0^T h_{\Lambda_j}(u)\,d\Lambda_{j,0}(u)\right] \\
& +P_{\theta_0}\left[2W^j(s,\mathbf{O},\theta_0)\sum_{k>j}\left\{W^k(s,\mathbf{O},\theta_0)\int_0^T h_{\Lambda_k}(u)\,d\Lambda_{k,0}(u)\right.\right. \\
& \left.\left. -W^k(s,\mathbf{O},\theta_0)\int_0^s h_{\Lambda_k}(u)\,d\Lambda_{k,0}(u) - \Delta^k h_{\Lambda_k}(T)\right\}\right]
\end{aligned}$$

es compacto. Es decir, dada una sucesión de funciones $h_{\Lambda_j,n}$ con $||h_{\Lambda_j,n}||_v \leq 1$, se debe demostrar que existe una subsucesión y un elemento $g \in VB(0,\tau)$ tal que $||K_{\Lambda_j}h_{\Lambda_j,\eta(n)} - g||_v \to 0$.

Como K_{Λ_j} es un operador lineal acotado entonces, $||K_{\Lambda_j}h_{\Lambda_j}||_v \leq M\int |h_{\Lambda_j}(u)|d\Lambda_{j,0}(u)$ para cada h_{Λ_j} y una constante positiva M. Entonces es suficiente mostrar que existe una subsucesión $h_{\Lambda_j,\eta(n)}$ de $h_{\Lambda_j,n}$ que converge. Como h_{Λ_j} es de variación acotada, se puede escribir $h_{\Lambda_j,n}$ como una diferencia de funciones acotadas crecientes $h^{(1)}_{\Lambda_j,n}$ y $h^{(2)}_{\Lambda_j,n}$. Por el teorema de Helly, existe una subsucesión $h^{(1)}_{\Lambda_j,\eta(n)}$ de $h^{(1)}_{\Lambda_j,n}$ la cual converge puntualmente a $h^{(1)*}_{\Lambda_j}$, de la misma manera existe una subsucesión $h^{(2)}_{\Lambda_j,\eta(n)}$ de $h^{(2)}_{\Lambda_j,n}$ la cual converge puntualmente a $h^{(2)*}_{\Lambda_j}$. Entonces $h_{\Lambda_j,n}$ converge a la diferencia de los límites por el teorema de convergencia dominada se sigue $\sigma - A$ es un operador compacto.

□

Se denotará la inversa de σ por $\widetilde{\sigma} = (\widetilde{\sigma}_{\mathbf{B}}, \widetilde{\sigma}_{\mathbf{G}}, \widetilde{\sigma}_{\Lambda_j}; j \in \mathcal{J}) : \mathcal{H} \to \mathcal{H}$.

5.8.2. Normalidad asintótica de los NPMLEs

Para establecer la normalidad asintótica se comienza estableciendo algunas notaciones que serán utilizadas más adelante. Sea $\{\mathbf{e}_1, \ldots, \mathbf{e}_P\}$ la base canónica de $\mathbb{R}^P$, donde $\mathbf{e}_m$ es el vector columna de dimensión P con 1 en la m-ésima entrada y 0 en otro lado, $m = 1, \ldots, P$. Denotemos por $(\mathbf{u}, \mathbf{0}_Q, 0; j \in \mathcal{J})$ el vector conformado por $\mathbf{h}$ tal que $\mathbf{h}_{\mathbf{B}} = \mathbf{u}$, $\mathbf{h}_{\mathbf{G}} = \mathbf{0}_Q$, y h_{Λ_j} es identicamente igual a 0 para cada $j \in \mathcal{J}$. Defínase el operador lineal $\varpi : \mathbb{R}^P \to \mathbb{R}^P$ de la siguiente manera, $\mathbf{u} \mapsto \varpi(\mathbf{u}) = \widetilde{\sigma}_{\mathbf{B}}((\mathbf{u}, \mathbf{0}_Q, 0; j \in \mathcal{J}))$. Sea Σ la matriz de dimensión $(P \times P)$ definida por

$$\Sigma = (\varpi(\mathbf{e}_1) \ldots \varpi(\mathbf{e}_P)) .$$

Teorema 5.8.1. *Bajo las condiciones **C1-C9**, $\sqrt{n}(\widehat{\boldsymbol{B}}_n - \boldsymbol{B}_0)$ converge en distribución a una distribución normal P-variada con media cero y varianza eficiente* Σ.

Demostración del Teorema 5.8.1: La demostración se basa en las ideas desarrolladas en la demostración del Teorema 3 de Fang *et al.* (2005). Como

$$\Psi_1^2(\theta_0)\,\sqrt{n}(\hat{\theta}_n - \theta) = \sqrt{n}\,(\Psi_n(\theta_0) - \Psi(\theta_0))\,(g) + o_p(1). \tag{5.26}$$

Por el lema 5.8.1

$$\begin{aligned}\Psi_1^2(\theta_0)(\mathbf{h})\,\sqrt{n}(\hat{\theta}_n - \theta) \quad = \quad & \sqrt{n}(\hat{\mathbf{B}}_n - \mathbf{B}_0)'\sigma_{\mathbf{B}}(g) + \sqrt{n}(\hat{\mathbf{G}}_n - \mathbf{G}_0)'\sigma_{\mathbf{G}}(g) \\ & + \sum_{j\in\mathcal{J}} \int_0^\tau \sigma_{\Lambda_j}(g)(s) h_{\Lambda_j}(s)\,\sqrt{n}\,d(\hat{\Lambda}_n - \Lambda_{j,0})(s).\end{aligned} \tag{5.27}$$

Combinando las ecuaciones (5.26) y (5.27) y el hecho de que σ es continuamente invertible entonces se puede escribir $g = \widetilde{\sigma}(\mathbf{h})$, permitiendo obtener,

$$\begin{aligned}&\sqrt{n}\left(\mathbf{h}'_{\mathbf{B}}\left(\widehat{\mathbf{B}}_n - \mathbf{B}_0\right) + \mathbf{h}'_{\mathbf{G}}\left(\widehat{\mathbf{G}}_n - \mathbf{G}_0\right) + \sum_{j\in\mathcal{J}} \int_0^\tau h_{\Lambda_j}(s)\, d\left(\widehat{\Lambda}_{j,n} - \Lambda_{j,0}\right)(s)\right) = \\ &\qquad \sqrt{n}\,(\Psi_n(\theta_0)(\widetilde{\sigma}(\mathbf{h})) - P_{\theta_0}\left[\Psi_1(\theta_0)(\widetilde{\sigma}(\mathbf{h}))\right]) + o_p(1).\end{aligned}$$

Tomando $\mathbf{h_G} = \mathbf{0}_Q$ y h_{Λ_j} la función identicamente igual a 0 para cada $j \in \mathcal{J}$, entonces la ecuación anterior se reduce a

$$\sqrt{n}\mathbf{h}'_{\mathbf{B}}\left(\widehat{\mathbf{B}}_n - \mathbf{B}_0\right) = \sqrt{n}\left(\Psi_n(\theta_0)(\widetilde{\sigma}(\check{\mathbf{h}})) - P_{\theta_0}\left[\Psi_1(\theta_0)(\widetilde{\sigma}(\check{\mathbf{h}}))\right]\right) + o_p(1), \tag{5.28}$$

donde $\check{\mathbf{h}} = (\mathbf{h_B}, \mathbf{0}_Q, 0; j \in \mathcal{J})$. Por el teorema del Límite Central y el lema 5.8.1 el término

$$\sqrt{n}\mathbf{h}'_{\mathbf{B}}(\widehat{\mathbf{B}}_n - \mathbf{B}_0)$$

converge en distribución a una distribución normal con media cero y varianza

$$P_{\theta_0}[\Psi_1(\theta_0)(\widetilde{\sigma}(\check{\mathbf{h}}))^2],$$

para cada $\mathbf{h_B} \in \mathbb{R}^P$. Obsérvese que $\check{\mathbf{h}} = \sigma(\widetilde{\sigma}(\check{\mathbf{h}})) = (\sigma_{\mathbf{B}}(\widetilde{\sigma}(\check{\mathbf{h}})), \sigma_{\mathbf{G}}(\widetilde{\sigma}(\check{\mathbf{h}})), \sigma_{\Lambda_j}(\widetilde{\sigma}(\check{\mathbf{h}})); j \in \mathcal{J})$ y por el lema 5.8.1 se sigue que,

$$P_{\theta_0}\left[\Psi_1(\theta_0)(\widetilde{\sigma}(\check{\mathbf{h}}))^2\right] = \mathbf{h}'_{\mathbf{B}}\widetilde{\sigma}_{\mathbf{B}}(\check{\mathbf{h}}) = \mathbf{h}'_{\mathbf{B}}\varpi(\mathbf{h_B}),$$

y entonces, $P_{\theta_0}[\Psi_1(\theta_0)(\widetilde{\sigma}(\check{\mathbf{h}}))^2] = \mathbf{h}'_{\mathbf{B}}\Sigma\mathbf{h_B}$. Por lo tanto, por la *Cramer-Wold device* (definida en van der Vaart , 1998), $\sqrt{n}(\widehat{\mathbf{B}}_n - \mathbf{B}_0)$ converge en distribución a una distribución normal con media cero y matriz de varianza-covarianza Σ. Ahora bien, sea $\check{\mathbf{h}}$ igual a $\check{\mathbf{h}}_m = (\mathbf{e}_m, \mathbf{0}_Q, 0; j \in \mathcal{J})$ en (5.28), para cada $m = 1, \ldots, P$, lo que permite establecer un sistema de P ecuaciones:

$$\sqrt{n}\left(\widehat{\mathbf{B}}_n - \mathbf{B}_0\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} l\left(\mathbf{O}_i; \theta_0\right) + o_p(1),$$

donde

$$l\left(\mathbf{O}; \theta_0\right) = \Sigma\Psi_{\mathbf{B}}(\theta_0) + \Sigma^*\Psi_{\mathbf{G}}(\theta_0) + \sum_{j\in\mathcal{J}}\Psi_{\Lambda_j}(\theta_0)\left(\Sigma^{**}\right),$$

$$\Sigma^* = \begin{pmatrix} \widetilde{\sigma}_{\mathbf{G}}(\check{\mathbf{h}}_1)' \\ \vdots \\ \widetilde{\sigma}_{\mathbf{G}}(\check{\mathbf{h}}_P)' \end{pmatrix}, \qquad \Sigma^{**} = \begin{pmatrix} \widetilde{\sigma}_{\Lambda_j}(\check{\mathbf{h}}_1) \\ \vdots \\ \widetilde{\sigma}_{\Lambda_j}(\check{\mathbf{h}}_P) \end{pmatrix},$$

aquí $\Psi_{\Lambda_j}(\theta_0)$ es aplicado componente a componente a Σ^{**}. Así, $\widehat{\mathbf{B}}_n$ es un estimador lineal asintótico de $\mathbf{B}_0$, y su función de influencia pertenece al espacio tangente generado por las funciones de score. Por lo tanto $\widehat{\mathbf{B}}_n$ es semiparamétricamente eficiente (ver Tsiatis, 2006).

□

Observación. En el análisis de riesgos competitivos el parámetro de regresión $\mathbf{B}$ es usualmente el parámetro de interés. Sin embargo, también se puede establecer el resultado de la normalidad asintótica para los NPMLEs $\mathbf{G}$ y Λ_j. Denotemos por $\varsigma : \mathbb{R}^Q \to \mathbb{R}^Q$ y defínase $\varsigma(\mathbf{u}) = \widetilde{\sigma}_{\mathbf{G}}((\mathbf{0}_P, \mathbf{u}, 0; j \in \mathcal{J}))$, sea $\{\mathbf{f}_1, \ldots, \mathbf{f}_Q\}$ una base canónica de $\mathbb{R}^Q$ y sea $\Upsilon = (\varsigma(\mathbf{f}_1)\ldots\varsigma(\mathbf{f}_Q))$ la matriz de dimensión $(Q \times Q)$ de ς con respecto a esta base. Sea también $\mathbf{h}_{j,t} = (\mathbf{h_B}, \mathbf{h_G}, h_{\Lambda_j}; j \in \mathcal{J})$ tal que $\mathbf{h_B} = \mathbf{0}_P$, $\mathbf{h_G} = \mathbf{0}_Q$,$h_{\Lambda_j}(\cdot) = 1\{\cdot \leq t\}$ para algún $t \in (0, \tau)$, $j \in \mathcal{J}$ y h_{Λ_l} es identicamente igual a cero para cada $l \in \mathcal{J}$, $l \neq j$.

Teorema 5.8.2. *Bajo las condciones **C1-C9**, $\sqrt{n}(\widehat{\mathbf{G}}_n - \mathbf{G}_0)$ converge en distribución a una distribución normal Q-variada con media cero y matriz de varianza Υ. Además, para cada $t \in (0, \tau)$ y $j \in \mathcal{J}$, $\sqrt{n}(\widehat{\Lambda}_{j,n}(t) - \Lambda_{j,0}(t))$ converge en distribución normal con media cero y varianza $\sigma^2_{j,t} = \int_0^t \widetilde{\sigma}_{\Lambda_j}(\mathbf{h}_{j,t})(s)\, d\Lambda_{j,0}(s)$.*

Demostración del Teorema 5.8.2: La demostración es similar a la del Teorema 5.8.1.

□

5.9. Estimación de la varianza

Esta sección se enfoca en la estimación de la varianza asintótica de $\widehat{\mathbf{B}}_n$, $\widehat{\mathbf{G}}_n$ y $\widehat{\Lambda}_{j,n}(t)$. La estimación de $\widehat{\Lambda}_{j,n}(t)$ puede ser útil para obtener intervalos de confianza las funciones de riesgo acumulado $\widehat{\Lambda}_{j,n}(t)$.

Sean $\mathbb{A}_n^{\mathbf{B}}$, $\mathbb{A}_n^{\mathbf{G}}$, $\mathbb{B}_n^{\mathbf{B}}$ y $\mathbb{B}_n^{\mathbf{G}}$ matrices de dimensiones $(P \times P)$, $(P \times Q)$, $(Q \times P)$ $(Q \times Q)$ respectivamente definidas por,

$$\begin{aligned}
\mathbb{A}_n^{\mathbf{B}} &= \mathbb{P}_n\left[\Psi_{\mathbf{B}}(\widehat{\theta}_n)^{\otimes 2}\right],\\
\mathbb{B}_n^{\mathbf{G}} &= \mathbb{P}_n\left[\Psi_{\mathbf{G}}(\widehat{\theta}_n)^{\otimes 2}\right],\\
\mathbb{A}_n^{\mathbf{G}} &= \mathbb{P}_n\left[\Psi_{\mathbf{B}}(\widehat{\theta}_n)\Psi_{\mathbf{G}}(\widehat{\theta}_n)'\right] = \left(\mathbb{B}_n^{\mathbf{B}}\right)'.
\end{aligned}$$

Sea A_n^{Λ} una matriz particionada de dimensión $(P \times s_n)$ definada como,

$$\mathbb{A}_n^{\Lambda} = (\mathbb{A}_n^{\Lambda_1} \ldots \mathbb{A}_n^{\Lambda_J}),$$

donde $\mathbb{A}_n^{\Lambda_j}$ es una matriz de $(P \times |\mathcal{S}_n^j|)$, donde la l-ésima columna $(l = 1, \ldots, |\mathcal{S}_n^j|)$ es expresada por

$$\frac{2}{n}\Psi_{\mathbf{B},(j,l)}(\widehat{\theta}_n),$$

para cada $j \in \mathcal{J}$ y donde $\Psi_{\mathbf{B},(j,l)}(\widehat{\theta}_n)$ denota el valor de $\Psi_{\mathbf{B}}(\widehat{\theta}_n)$ pero evaluado en el individuo i, tal que $\Delta_i = 1$ y $T_i = t_l^j$ $(j \in \mathcal{J}$ y $l = 1, \ldots, |\mathcal{S}_n^j|)$. De la misma manera, defínase la matriz particionada $\mathbb{B}_n^{\Lambda}$ de dimensión $(Q \times s_n)$ dada por,

$$\mathbb{B}_n^{\Lambda} = (\mathbb{B}_n^{\Lambda_1} \ldots \mathbb{B}_n^{\Lambda_J}),$$

para cada $j \in \mathcal{J}$, donde $\mathbb{B}_n^{\Lambda_j}$ es una matriz de dimensión $(Q \times |\mathcal{S}_n^j|)$ con la l-ésima columna ($l = 1, \ldots, |\mathcal{S}_n^j|$) esta dada por $(2/n)\Psi_{\mathbf{G},(j,l)}(\widehat{\theta}_n)$, donde el término $\Psi_{\mathbf{G},(j,l)}(\widehat{\theta}_n)$ es definido similarmente como en $\Psi_{\mathbf{B},(j,l)}(\widehat{\theta}_n)$. Defínase también las matrices particionadas $C_n^{\mathbf{B}}$ y $C_n^{\mathbf{G}}$ de tamaño $(s_n \times P)$ y $(s_n \times Q)$ respectivamente, de la manera siguiente,

$$\mathbb{C}_n^{\mathbf{B}} = \begin{pmatrix} \mathbb{C}_{n,1}^{\mathbf{B}} \\ \vdots \\ \mathbb{C}_{n,J}^{\mathbf{B}} \end{pmatrix} \text{ y } \mathbb{C}_n^{\mathbf{G}} = \begin{pmatrix} \mathbb{C}_{n,1}^{\mathbf{G}} \\ \vdots \\ \mathbb{C}_{n,J}^{\mathbf{G}} \end{pmatrix},$$

para cada $j \in \mathcal{J}$, $\mathbb{C}_{n,j}^{\mathbf{B}}$ es una matriz de $(|\mathcal{S}_n^j| \times P)$ donde el l-ésimo renglón ($l = 1, \ldots, |\mathcal{S}_n^j|$) esta dado por,

$$-\mathbb{P}_n\left[2\Psi_{\mathbf{B}}(\widehat{\theta}_n)' g^j(\mathbf{O}; \widehat{\theta}_n) e^{\widehat{\beta}'_{j,n}\mathbf{Z}} Y(t_l^j)\right],$$

mientras que $\mathbb{C}_{n,j}^{\mathbf{G}}$ es una matriz de $(|\mathcal{S}_n^j| \times Q)$ donde el l-ésimo renglón esta dado por,

$$-\mathbb{P}_n\left[2\Psi_{\mathbf{G}}(\widehat{\theta}_n)' g^j(\mathbf{O}; \widehat{\theta}_n) e^{\widehat{\beta}'_{j,n}\mathbf{Z}} Y(t_l^j)\right].$$

Defínase $\mathbb{C}_n^{\Lambda}$ como una matriz particionada de $(s_n \times s_n)$ donde la entrada (j,k)-ésima (para $j \in \mathcal{J}, k \in \mathcal{J}$) es la submatriz $\mathbb{C}_{n,j}^{\Lambda_k}$ de dimensión $(|\mathcal{S}_n^j| \times |\mathcal{S}_n^k|)$ donde el elemento (l,m)-ésimo es definido como sigue,

$$\begin{aligned} \mathbb{C}_{n,j}^{\Lambda_k}(l,m) \quad = \quad & 1\{j=k\}\Big\{1\{l=m\}\mathbb{P}_n\left[W^k(t_m^k, \mathbf{O}, \widehat{\theta}_n)\right] - \frac{2}{n}W^k(t_l^k, \mathbf{O}_{(k,m)}, \widehat{\theta}_n) \\ & \qquad + \mathbb{P}_n\left[\left\{W^k(t_l^k, \mathbf{O}, \widehat{\theta}_n)\right\}^2 \widehat{\Delta\Lambda_{k,n}}(t_m^k) 1\{t_m^k \le T\}\right]\Big\} \\ & + 1\{j<k\}\Big\{\mathbb{P}_n\Big[2W^j(t_l^j, \mathbf{O}, \widehat{\theta}_n) W^k(t_l^j, \mathbf{O}, \widehat{\theta}_n) \widehat{\Delta\Lambda_{k,n}}(t_m^k)\Big\{1\{t_m^k \le T\} \\ & \qquad - 1\{t_m^k \le t_l^j\}\Big\}\Big] - \frac{2}{n}W^j(t_l^j, \mathbf{O}_{(k,m)}, \widehat{\theta}_n)\Big\} \end{aligned}$$

para $l = 1, \ldots, |\mathcal{S}_n^j|$ y $m = 1, \ldots, |\mathcal{S}_n^k|$. En la expresión de $\mathbb{C}_{n,j}^{\Lambda_k}(l,m)$, el término $\widehat{\Delta\Lambda_{k,n}}(t)$ denota los tamaños de los saltos de $\widehat{\Lambda}_{k,n}$ al tiempo t; esto es, $\widehat{\Delta\Lambda_{k,n}}(t) = \widehat{\Lambda}_{k,n}(t) - \widehat{\Lambda}_{k,n}(t-)$. Más aún, $\mathbf{O}_{(k,m)}$ denota el valor de $\mathbf{O}$ para el individuo i tal que $\Delta_i = 1$ y $T_i = t_m^k$.

Defínanse D_n la matriz particionada definida por,

$$\mathbb{D}_n = \begin{pmatrix} \mathbb{A}_n^{\mathbf{B}} & \mathbb{A}_n^{\mathbf{G}} & \mathbb{A}_n^{\Lambda} \\ \mathbb{B}_n^{\mathbf{B}} & \mathbb{B}_n^{\mathbf{G}} & \mathbb{B}_n^{\Lambda} \\ \mathbb{C}_n^{\mathbf{B}} & \mathbb{C}_n^{\mathbf{G}} & \mathbb{C}_n^{\Lambda} \end{pmatrix}$$

y las matrices

$$
\begin{aligned}
\Sigma_n = \Big\{ & \mathbb{A}_n^{\mathbf{B}} - \mathbb{A}_n^{\mathbf{G}}(\mathbb{B}_n^{\mathbf{G}})^{-1}\mathbb{B}_n^{\mathbf{B}} - (\mathbb{A}_n^{\Lambda} - \mathbb{A}_n^{\mathbf{G}}(\mathbb{B}_n^{\mathbf{G}})^{-1}\mathbb{B}_n^{\Lambda}) \\
& \times(\mathbb{C}_n^{\Lambda} - \mathbb{C}_n^{\mathbf{G}}(\mathbb{B}_n^{\mathbf{G}})^{-1}\mathbb{B}_n^{\Lambda})^{-1}(\mathbb{C}_n^{\mathbf{B}}\mathbb{C}_n^{\mathbf{G}}(\mathbb{B}_n^{\mathbf{G}})^{-1}\mathbb{B}_n^{\mathbf{B}})\Big\}^{-1}, \\
\Upsilon_n = \Big\{ & \mathbb{B}_n^{\mathbf{G}} - \mathbb{B}_n^{\mathbf{B}}(\mathbb{A}_n^{\mathbf{B}})^{-1}\mathbb{A}_n^{\mathbf{G}} - (\mathbb{B}_n^{\Lambda} - \mathbb{B}_n^{\mathbf{B}}(\mathbb{A}_n^{\mathbf{B}})^{-1}\mathbb{A}_n^{\Lambda}) \\
& \times(\mathbb{C}_n^{\Lambda} - \mathbb{C}_n^{\mathbf{B}}(\mathbb{A}_n^{\mathbf{B}})^{-1}\mathbb{A}_n^{\Lambda})^{-1}(\mathbb{C}_n^{\mathbf{G}} - \mathbb{C}_n^{\mathbf{B}}(\mathbb{A}_n^{\mathbf{B}})^{-1}\mathbb{A}_n^{\mathbf{G}})\Big\}^{-1}, \\
\Xi_n = \Big\{ & \mathbb{C}_n^{\Lambda} - \mathbb{C}_n^{\mathbf{B}}(\mathbb{A}_n^{\mathbf{B}})^{-1}\mathbb{A}_n^{\Lambda} - (\mathbb{C}_n^{\mathbf{G}} - \mathbb{C}_n^{\mathbf{B}}(\mathbb{A}_n^{\mathbf{B}})^{-1}\mathbb{A}_n^{\mathbf{G}}) \\
& \times(\mathbb{B}_n^{\mathbf{G}} - \mathbb{B}_n^{\mathbf{B}}(\mathbb{A}_n^{\mathbf{B}})^{-1}\mathbb{A}_n^{\mathbf{G}})^{-1}(\mathbb{B}_n^{\Lambda} - \mathbb{B}_n^{\mathbf{B}}(\mathbb{A}_n^{\mathbf{B}})^{-1}\mathbb{A}_n^{\Lambda})\Big\}^{-1}.
\end{aligned}
$$

Y defínanse también los vectores $\Phi_{j,t,n}$ y $U_{j,t,n}$ de dimensiones s_n, con $t \in (0,\tau)$ y $j \in \mathcal{J}$ tales que,

$$\Phi_{j,t,n} = \left(\mathbf{0}'_{l_n^j} \quad \widehat{\Delta\Lambda_{j,n}}(t_1^j)1\{t_1^j \le t\} \;\dots\; \widehat{\Delta\Lambda_{j,n}}(t^j_{|\mathcal{S}_n^j|})1\{t^j_{|\mathcal{S}_n^j|} \le t\} \quad \mathbf{0}'_{u_n^j}\right)'$$

y

$$U_{j,t,n} = \left(\mathbf{0}'_{l_n^j} \quad 1\{t_1^j \le t\} \;\dots\; 1\{t^j_{|\mathcal{S}_n^j|} \le t\} \quad \mathbf{0}'_{u_n^j}\right)',$$

donde $l_n^j = \sum_{k=1}^{j-1} |\mathcal{S}_n^k|$ y $u_n^j = \sum_{k=j+1}^{J} |\mathcal{S}_n^k|$ con $l_n^1 = u_n^J = 0$. Con todas estas especificaciones se satisface lo siguiente:

Teorema 5.9.1. *Bajo las condiciones **C1-C9**, los estimadores de la varianza Σ_n, Υ_n y $\sigma^2_{j,t,n} = \Phi'_{j,t,n}\Xi_n U_{j,t,n}$ converge en probabilidad a Σ, Υ y $\sigma^2_{j,t}$ ($t \in (0,\tau)$, $j \in \mathcal{J}$) respectivamente.*

Demostración del Teorema 5.9.1: La demostración se basa en las ideas desarrolladas por Parner (1998), Dupuy y Mesbah (2004) y Fang et al. (2005).

En primer lugar, se estima σ a través de una versión empírica $\sigma_n = (\sigma_{\mathbf{B},n}, \sigma_{\mathbf{G},n}, \sigma_{\Lambda_j,n}; j \in \mathcal{J})$ obtenida al remplazar θ_0 y P_{θ_0} por $\widehat{\theta}_n$ y $\mathbb{P}_n$ respectivamente en $\sigma_{\mathbf{B}}$, $\sigma_{\mathbf{G}}$ y σ_{Λ_j}. Siguiendo los mismos argumentos del lema 5.7.2, las funciones $\sigma_{\mathbf{B},n}, \sigma_{\mathbf{G},n}, \sigma_{\Lambda_j,n}$ con $j \in \mathcal{J}$ forman clases Donsker de donde $\|\sigma_n(\mathbf{h}) - \sigma(\mathbf{h})\|_H \to 0$. Como σ_n es invertible y con inversa continua, entonces se puede expresar $h = \widetilde{\sigma}_n(g)$. Entonces,

$$
\begin{aligned}
\|\widetilde{\sigma}_n(g) - \widetilde{\sigma}(g)\|_{\mathcal{H}} &= \|\widetilde{\sigma}(\sigma(\mathbf{h})) - \widetilde{\sigma}_n(\sigma_n(\mathbf{h}))\|_{\mathcal{H}} \\
&\le \sup_{\mathbf{h}\in\mathcal{H}} \frac{\|\widetilde{\sigma}(\mathbf{h})\|_{\mathcal{H}}}{\|\mathbf{h}\|_{\mathcal{H}}} \|\sigma(\mathbf{h}) - \sigma_n(\mathbf{h})\|_{\mathcal{H}}.
\end{aligned}
$$

Como $\|\sigma_n(\mathbf{h}) - \sigma(\mathbf{h})\|_H \to 0$, entonces $\widetilde{\sigma}_n = (\widetilde{\sigma}_{\mathbf{B},n}, \widetilde{\sigma}_{\mathbf{G},n}, \widetilde{\sigma}_{\Lambda_j,n}; j \in \mathcal{J})$ converge a $\widetilde{\sigma}(\mathbf{h})$ en probabilidad (ver Dupuy y Mesbah, 2004).

Para cada $\mathbf{h}_{\mathbf{B}}$, la varianza asintótica de $\sqrt{n}\mathbf{h}'_{\mathbf{B}}(\widehat{\mathbf{B}}_n - \mathbf{B}_0)$ es $\mathbf{h}'_{\mathbf{B}}\varpi(\mathbf{h}_{\mathbf{B}})$, la cual es calculada sistemáticamente por $\mathbf{h}'_{\mathbf{B}}\widetilde{\sigma}_{\mathbf{B},n}(\check{\mathbf{h}})$, donde $\check{\mathbf{h}} = (\mathbf{h}_{\mathbf{B}}, \mathbf{0}_Q, 0; j \in \mathcal{J})$. Defínase $\check{\mathbf{h}}_n$ tal que, $\check{\mathbf{h}}_n =$

$(\breve{\mathbf{h}}_{\mathbf{B},n}, \breve{\mathbf{h}}_{\mathbf{G},n}, \breve{h}_{\Lambda_j,n}; j \in \mathcal{J}) = \widetilde{\sigma}_n(\breve{\mathbf{h}})$. Entonces $\sigma_n(\breve{\mathbf{h}}_n) = \breve{\mathbf{h}}$, lo cual se puede reescribir de la manera siguiente,

$$\begin{cases} \sigma_{\mathbf{B},n}(\breve{\mathbf{h}}_n) = \mathbf{h}_{\mathbf{B}} \\ \sigma_{\mathbf{G},n}(\breve{\mathbf{h}}_n) = \mathbf{0}_Q \\ \sigma_{\Lambda_1,n}(\breve{\mathbf{h}}_n)(s) = 0, \quad s \in [0,\tau] \\ \vdots \\ \sigma_{\Lambda_J,n}(\breve{\mathbf{h}}_n)(s) = 0, \quad s \in [0,\tau]. \end{cases}$$

En particular, sea $s = t_1^j, \ldots, t^j_{|\mathcal{S}_n^j|}$ para cada $j \in \mathcal{J}$, en el sistema anterior. Esto produce un sistema de $(P + Q + s_n)$ ecuaciones las cuales pueden ser escritas en forma matricial, cuya expresión es la siguiente,

$$\mathbb{D}_n \begin{pmatrix} \breve{\mathbf{h}}_{\mathbf{B},n} \\ \breve{\mathbf{h}}_{\mathbf{G},n} \\ \breve{\mathbf{h}}_{\Lambda,n} \end{pmatrix} = \begin{pmatrix} \mathbf{h}_{\mathbf{B}} \\ 0_Q \\ 0_{s_n} \end{pmatrix} \tag{5.29}$$

donde $\breve{\mathbf{h}}_{\Lambda,n} = (\breve{h}_{\Lambda_1,n}(t_1^1) \ldots \breve{h}_{\Lambda_1,n}(t^1_{|\mathcal{S}_n^1|}) \ldots \breve{h}_{\Lambda_J,n}(t_1^J) \ldots \breve{h}_{\Lambda_J,n}(t^J_{|\mathcal{S}_n^J|}))'$. Haciendo algunos calculos en (6.32) se puede demostrar que $\breve{\mathbf{h}}_{\mathbf{B},n} = \Sigma_n \mathbf{h}_{\mathbf{B}}$, donde Σ_n fue definida anteriormente entonces, $\mathbf{h}'_{\mathbf{B}}\Sigma_n\mathbf{h}_{\mathbf{B}}$ es un estimador consistente de la varianza asintótica de $\sqrt{n}\mathbf{h}'_{\mathbf{B}}(\widehat{\mathbf{B}}_n - \mathbf{B}_0)$ para cada $\mathbf{h}_{\mathbf{B}}$. A partir de esto se sigue que Σ_n es un estimador consistente de Σ. Para obtener la consistencia de Υ_n se procede de la misma manera.

Sea $t \in (0,\tau)$ y $j \in \mathcal{J}$. Utilizando el teorema de convergencia dominada y la consistencia de $\widetilde{\sigma}_n$ podemos concluir que $\sigma^2_{j,t,n} = \int_0^t \widetilde{\sigma}_{\Lambda_j,n}(\mathbf{h}_{j,t})(s)\, d\widehat{\Lambda}_{j,n}(s)$ converge en probabilidad a $\sigma^2_{j,t}$. Similarmente como el procedimiento desarrollado arriba, sea $\mathbf{h}_n = (\mathbf{h}_{\mathbf{B},n}, \mathbf{h}_{\mathbf{G},n}, h_{\Lambda_j,n}; j \in \mathcal{J}) = \widetilde{\sigma}_n(\mathbf{h}_{j,t})$. Entonces $\sigma_n(\mathbf{h}_n) = \mathbf{h}_{j,t}$, lo cual puede ser escrito

$$\begin{cases} \sigma_{\mathbf{B},n}(\mathbf{h}_n) = 0_P \\ \sigma_{\mathbf{G},n}(\mathbf{h}_n) = 0_Q \\ \sigma_{\Lambda_j,n}(\mathbf{h}_n)(s) = 1\{s \leq t\}, \quad s \in [0,\tau] \\ \sigma_{\Lambda_l,n}(\mathbf{h}_n)(s) = 0, \quad l \in \mathcal{J}, \l \neq j, s \in [0,\tau]. \end{cases} \tag{5.30}$$

En particular, tomando $s = t_1^j, \ldots, t^j_{|\mathcal{S}_n^j|}$ para cada $j \in \mathcal{J}$ en (6.33) se obtiene el sistema de ecuaciones

$$\mathbb{D}_n \begin{pmatrix} \mathbf{h}_{\mathbf{B},n} \\ \mathbf{h}_{\mathbf{G},n} \\ \mathbf{h}_{\Lambda,n} \end{pmatrix} = \begin{pmatrix} 0_P \\ 0_Q \\ U_{j,t,n} \end{pmatrix}$$

donde $\mathbf{h}_{\Lambda,n} = (h_{\Lambda_1,n}(t_1^1) \ldots h_{\Lambda_1,n}(t^1_{|\mathcal{S}_n^1|}) \ldots h_{\Lambda_J,n}(t_1^J) \ldots h_{\Lambda_J,n}(t^J_{|\mathcal{S}_n^J|}))'$ y $U_{j,t,n}$ ya han sido establecidas anteriormente. Similarmente a lo anterior se puede demostrar que $\mathbf{h}_{\Lambda,n} = \Xi_n U_{j,t,n}$, y también que,

$$\begin{aligned}
\sigma^2_{j,t,n} &= \int_0^t \widetilde{\sigma}_{\Lambda_j,n}(\mathbf{h}_{j,t})(s)\, d\widehat{\Lambda}_{j,n}(s) \\
&= \sum_{l=1}^{|\mathcal{S}_n^j|} \widetilde{\sigma}_{\Lambda_j,n}(\mathbf{h}_{j,t})(t_l^j)\widehat{\Delta\Lambda_{j,n}}(t_l^j) 1\{t_l^j \leq t\} \\
&= \Phi'_{j,t,n}\mathbf{h}_{\Lambda,n}
\end{aligned}$$

y por lo tanto, $\Phi'_{j,t,n}\Xi_n U_{j,t,n}$ es un estimador consistente para $\sigma^2_{j,t}$.

□

5.10. Simulaciones

La realización de simulaciones representa una importante herramienta estadística para investigar el rendimiento, las propiedades y la adecuación de los modelos estadísticos en situaciones pre-especificadas. En esta sección se ilustrará el funcionamiento del modelo de mezclas semi-paramétrico para el análisis de riesgos competitivos estudiado en las secciones anteriores.

Dos estudios de simulación fueron realizados; en el primer estudio se consideró un tamaño de muestra $n = 30$ mientras que para el segundo estudio el tamaño de muestra fue $n = 200$. En ambas situaciones se consideraron, dos tipos de causa ($J = 2$) y una variable explicativa $\mathbf{Z}$ que fue generada a través de una distribución normal estándar. El modelo de mezclas para riesgos competitivos representado por las ecuaciones (5.2) y (5.3) fue ajustado mediante dos métodos. El primer método utilizado fue el estimador producto límite (PLE) descrito por Escarela y Bowater (2008). Ellos especifican la distribución de supervivencia condicional para el j-ésimo riesgo como

$$S_j(t) = \prod_{m: t_{j,(m)} \leq t} \alpha_{j,m} \qquad j \in \mathcal{J}, \quad m = 1, ..., k_j,$$

donde k_j, denota el número de muertes por causa j, $\alpha_{j,m}$ son parámetros no negativos con $\alpha_{j0} = 1$ y

$$\alpha_{jm} = \frac{S_j(t_{j,(m+1)})}{S_j(t_{j,(m)})}.$$

Por lo tanto la estimación de la función de supervivencia base condicional es expresada como

$$\hat{S}_j(t) = \exp\left\{ - \sum_{m: t_{j,(m)} < t} \frac{d_{jm}}{\sum_{m \in R_{jm}} g_m^j(\theta) \exp(\boldsymbol{\beta}_j \mathbf{Z}_m)} \right\},$$

donde, para cada tipo de fallo j, $t_{j,(1)} < ... < t_{j,(k_j)}$ denota los distintos tiempos de fallo, R_{jl} es el conjunto de individuos que se encuentran en riesgo antes de $t_{j,(l)}$ y d_{jl} es el número de muerte por causa j al tiempo $t_{j,(l)}$.

El segundo método utilizado fue el modelo exponencial donde la distribución condicional es especificada por $\lambda_j(t) = \exp\{\kappa_j\}$, donde $-\infty < \kappa_j < \infty$, $j = 1, 2$.

Los tiempos de supervivencia en cada estudio fueron obtenidos por medio del método de transformación inversa, dos casos fueron considerados. Primero, se consideró una función de mortalidad para cada riesgo con distribución exponencial, es decir,

$$\lambda_j(t|\mathbf{Z}) = h_j \exp(\beta_j'\mathbf{Z}) \qquad j = 1, 2.$$

Por lo tanto, si un elemento pertenece al primer componente, un tiempo de supervivencia para la causa 1, es generado de acuerdo a $\lambda_1(t|\mathbf{Z})$, el cual es obtenido por el método de transformación propuesto por Bender *et al.* (2005) (quienes presentan técnicas para generar tiempos de supervivencia aplicando el modelo de riesgos proporcionales de Cox). Por tanto el tiempo de supervivencia sigue la siguiente formula

$$T = -\frac{-\log(U)}{h_1 \exp(\beta_1'\mathbf{Z})},$$

donde U es una variable que sigue una distribución uniforme en el intervalo [0,1]. De la misma manera, para un elemento que pertenece al segundo componente, el tiempo de supervivencia para la causa 2 es generado de acuerdo a $\lambda_2(t|\mathbf{Z})$ y aplicando el método de transformación inversa este tiempo puede ser generado por,

$$T = -\frac{-\log(U)}{h_2 \exp(\beta_2'\mathbf{Z})},$$

donde U es una variable que sigue una distribución uniforme sobre el intervalo [0,1].

Los valores verdaderos considerados para generar los tiempos de supervivencia en este caso fueron,

$$(h_1, \beta_1, h_2, \beta_2) = (0.5, -0.5, 1.0, -1.0).$$

Para el segundo caso se considero que ambas componentes de la función de mortalidad siguen una distribución Weibull, es decir,

$$\lambda_j(t|\mathbf{Z}) = h_j \exp(\beta_j'\mathbf{Z})\nu_j t^{\nu_j - 1} \qquad j = 1, 2,$$

donde $\nu > 0$ y por consecuencia, el tiempo de supervivencia puede ser generado siguiendo la formula (Bender *et al.*, 2005)

$$T = \left(-\frac{-\log(U)}{h_i \exp(\beta_i'\mathbf{Z})}\right)^{1/\nu_i} \qquad j = 1, 2,$$

donde U es una variable que sigue una distribución uniforme en el intervalo [0,1].

Para este caso los verdaderos valores de los parámetros considerados fueron,

$$(h_1, \beta_1, \nu_1, h_2, \beta_2, \nu_2) = (0.5, -0.5, 0.5, 1.0, -1.0, 1.5).$$

Obsérvese que al tomar estos valores de parámetros, la primera fuerza de mortalidad corresponde a una función decreciente, mientras que la segunda función de mortalidad corresponde a una función creciente.

En los dos estudios el modelo logístico propuesto en la ecuación (5.3), se considero el vector de covariables $\mathbf{X} = (1, \mathbf{Z})$, donde los verdaderos valores considerados para los coeficientes de regresión fueron, $\gamma = (-1.0, 0.5)$. Para cada elemento el tiempo de censura fue generado por una distribución uniforme $U(d_1, d_2)$ donde d_1 y d_2 son cualesquiera contantes. Si el k-ésimo tiempo de fallo fuese mayor que el k-ésimo tiempo de censura, entonces es considerado el k-ésimo elemento como censurado. En este estudio se consideran 3 conjuntos diferentes de d_1 y d_2, lo que permitirá investigar el comportamiento del modelo para diferentes niveles de censura. Para cada estudio, se generaron 100 muestras independientes para ajustar los datos mediante los dos métodos.

En las Tablas 5.1 y 5.2 se presenta un promedio de las estimaciones de los coeficientes con sus respectivos errores estándares para ambos métodos (estimador producto límite y modelo exponencial) cuando el tamaño de la muestra es igual a 30. La Tabla 5.1 muestra los resultados obtenidos cuando los tiempos de supervivencia siguen el modelo exponencial, obsérvese que cuando el nivel de censura es bajo tanto el modelo semiparamétrico y el enfoque paramétrico dan buenas estimaciones, pero conforme el nivel de censura aumenta en este caso mostramos niveles de censura al 23.3 % y al 39.8 % ambos modelos dan estimadores malos. La Tabla 5.2 muestra el promedio de las estimaciones de los coeficientes cuando los tiempos de supervivencia siguen un modelo weibull, para este caso se obtuvo malas estimaciones para ambos modelos y para los diferentes niveles de censura. En particular, en el caso del enfoque semiparamétrico cuando el nivel de censura es alto no fue posible obtener las estimaciones, se dejo correr el programa por varias horas (más de 8) sin tener respuesta alguna, el problema es que son pocos datos con muchas observaciones censuradas y tal vez la distribución de los datos también ocasiona más problemas.

En las Tablas 5.3 y 5.4 se presenta un promedio de los estimaciones de los coeficientes con sus respectivos errores estándares para ambos métodos (estimador producto límite y modelo exponencial) cuando el tamaño de la muestra es igual a 200. Los resultados presentados en la Tabla 5.3 fueron obtenidos cuando los tiempos de supervivencia fueron considerados provenientes de una distribución exponencial. A partir de estos resultados se puede observar que el método semiparamétrico y el enfoque paramétrico son comparables cuando el nivel de censura es bajo. Mientras que para niveles medios y amplios de censura el enfoque paramétrico proporciona mejores estimadores de los coeficientes del modelo a comparación del enfoque semiparamétrico. Obsérvese que mientras el nivel de censura se incrementa las estimaciones del modelo semiparamétrico se van alejando ligeramente de los verdaderos valores de los parámetros. Por otro lado, la Tabla 5.4 presenta los resultados obtenidos cuando los tiempos de supervivencia siguen un modelo weibull. En este caso, se puede observar que el enfoque semiparamétrico estudiado en este capítulo proporciona mejores estimaciones respecto al modelo exponencial. Y esto se debe a que en este caso se considero que la distribución de los tiempos de supervivencia fuera diferente a una distribución exponencial. Obsérvese también que para niveles bajos de censura el modelo exponencial da malas estimaciones para los coeficientes de probabilidad de tipo de riesgo, mientras que para altos niveles de censura el problema es con la estimación de los coeficientes de la función de riesgo.

En conclusión, con los resultados de las simulaciones realizadas podemos afirmar que cuando el tamaño de la muestra es 200 el método semiparamétrico estudiado en este capítulo es teóricamente satisfactorio pero también es computacionalmente intensivo.

Cuadro 5.1: Promedio de las estimaciones de los coeficientes y sus correspondientes errores estándares usando el estimador producto límite (PLE) y el modelo exponencial, cuando n=30 y los tiempos de supervivencia siguen un modelo exponencial.

Distribución de censura	**Censura**				**PLE**		**Exponencial**	
					Coeficientes de probabilidad de tipo de riesgo			
		Parámetro	Covariable	Valor verdadero	Estimación	S.E.	Estimación	S.E
U(2.0,9.0)	8.5 %	γ_1	Intercepto	-1.0	-1.074	0.181	-1.083	0.206
		γ_1	**X**	0.5	0.539	0.048	0.621	0.074
					Coeficientes en función de riesgo			
		β_1	**Z**	-0.5	-0.494	0.055	-0.480	0.320
		β_2	**Z**	-1.0	-0.984	0.105	-0.985	0.054
					Coeficientes de probabilidad de tipo de riesgo			
		Parámetro	Covariable	Valor verdadero	Estimación	S.E.	Estimación	S.E.
U(0.5,5.0)	23.3 %	γ_1	Intercepto	-1.0	-1.249	0.085	-1.076	0.131
		γ_1	**X**	0.5	0.750	0.180	0.702	0.179
					Coeficientes en función de riesgo			
		β_1	**Z**	-0.5	-0.073	0.069	-1.700	0.095
		β_2	**Z**	-1.0	-1.068	0.023	-0.986	0.0613
					Coeficientes de probabilidad de tipo de riesgo			
		Parámetro	Covariable	Valor verdadero	Estimación	S.E.	Estimación	S.E.
U(0.5,1.8)	39.8 %	γ_1	Intercepto	-1.0	-2.330	0.423	-2.154	0.130
		γ_1	**X**	0.5	2.718	2.5	3.164	1.957
					Coeficientes en función de riesgo			
		β_1	**Z**	-0.5	-5.214	2.15	-7.583	2.338
		β_2	**Z**	-1.0	-0.993	0.204	-0.909	0.056

Cuadro 5.2: Promedio de las estimaciones de los coeficientes y sus correspondientes errores estándares usando el estimador producto límite (PLE) y el modelo exponencial, cuando n=30 y los tiempos de supervivencia siguen un modelo weibull.

Distribución de censura	**Censura**				**PLE**		**Exponencial**	
					Coeficientes de probabilidad de tipo de riesgo			
		Parámetro	Covariable	Valor verdadero	Estimación	S.E.	Estimación	S.E
U(2.0,19.0)	8.7 %	γ_1	Intercepto	-1.0	-1.241	0.123	-1.167	0.017
		γ_1	**X**	0.5	0.259	0.124	0.562	0.032
					Coeficientes en función de riesgo			
		β_1	**Z**	-0.5	-1.375	0.204	-0.834	0.043
		β_2	**Z**	-1.0	-1.110	0.077	-0.738	0.262
					Coeficientes de probabilidad de tipo de riesgo			
		Parámetro	Covariable	Valor verdadero	Estimación	S.E.	Estimación	S.E.
U(0.5,5.7)	23.7 %	γ_1	Intercepto	-1.0	-6.131	2.516	-1.401	0.052
		γ_1	**X**	0.5	2.613	0.925	0.363	0.325
					Coeficientes en función de riesgo			
		β_1	**Z**	-0.5	-5.052	2.103	-3.959	1.612
		β_2	**Z**	-1.0	-1.138	0.402	-0.820	0.136
					Coeficientes de probabilidad de tipo de riesgo			
		Parámetro	Covariable	Valor verdadero	Estimación	S.E.	Estimación	S.E.
U(0.5,2)	39.0 %	γ_1	Intercepto	-1.0	-	-	-9.411	0.688
		γ_1	**X**	0.5	-	-	3.692	0.608
					Coeficientes en función de riesgo			
		β_1	**Z**	-0.5	-	-	-6.842	1.45
		β_2	**Z**	-1.0	-	-	-0.864	0.155

Cuadro 5.3: Promedio de las estimaciones de los coeficientes y sus correspondientes errores estándares usando el estimador producto límite (PLE) y el modelo exponencial, cuando n=200 y los tiempos de supervivencia siguen un modelo exponencial.

Distribución de censura	**Censura**				**PLE**		**Exponencial**	
					Coeficientes de probabilidad de tipo de riesgo			
		Parámetro	Covariable	Valor verdadero	Estimación	S.E.	Estimación	S.E
U(2.0,9.0)	8.9 %	γ_1	Intercepto	-1.0	-1.051	0.078	-1.040	0.041
		γ_1	**X**	0.5	0.502	0.050	0.508	0.027
					Coeficientes en función de riesgo			
		β_1	**X**	-0.5	-0.518	0.035	-0.492	0.019
		β_2	**X**	-1.0	-1.004	0.079	-1.00	0.010
					Coeficientes de probabilidad de tipo de riesgo			
		Parámetro	Covariable	Valor verdadero	Estimación	S.E.	Estimación	S.E.
U(0.5,5.0)	23.0 %	γ_1	Intercepto	-1.0	-1.032	0.042	-0.991	0.031
		γ_1	**X**	0.5	0.511	0.015	0.537	0.042
					Coeficientes en función de riesgo			
		β_1	**X**	-0.5	-0.545	0.048	-0.530	0.046
		β_2	**X**	-1.0	-0.996	0.023	-0.987	0.014
					Coeficientes de probabilidad de tipo de riesgo			
		Parámetro	Covariable	Valor verdadero	Estimación	S.E.	Estimación	S.E.
U(0.5,1.8)	40.8 %	γ_1	Intercepto	-1.0	-1.019	0.022	-1.039	0.045
		γ_1	**X**	0.5	0.611	0.122	0.579	0.011
					Coeficientes en función de riesgo			
		β_1	**X**	0.5	-0.593	0.098	-0.509	0.030
		β_2	**X**	-1.0	-1.016	0.034	-0.993	0.007

Cuadro 5.4: Promedio de las estimaciones de los coeficientes y sus correspondientes errores estándares usando el estimador producto límite (PLE) y el modelo exponencial, cuando n=200 y los tiempos de supervivencia siguen un modelo weibull.

Distribución de censura	**Censura**				**PLE**		**Exponencial**	
					Coeficientes de probabilidad de tipo de riesgo			
		Parámetro	Covariable	Valor verdadero	Estimación	S.E.	Estimación	S.E
U(2.0,19.0)	9 %	γ_1	Intercepto	-1	-1.086	0.095	-1.053	0.063
		γ_1	**X**	0.5	0.415	0.110	0.495	0.018
					Coeficientes en función de riesgo			
		β_1	**X**	-0.5	-0.450	0.051	-0.700	0.036
		β_2	**X**	-1.0	-1.032	0.037	-0.694	0.033
					Coeficientes de probabilidad de tipo de riesgo			
		Parámetro	Covariable	Valor verdadero	Estimación	S.E.	Estimación	S.E.
U(0.5,5.7)	22.3 %	γ_1	Intercepto	-1	-1.125	0.140	-1.354	0.155
		γ_1	**X**	0.5	0.365	0.145	0.248	0.103
					Coeficientes en función de riesgo			
		β_1	**X**	-0.5	-0.428	0.091	-0.376	0.131
		β_2	**X**	-1.0	-1.042	0.072	-0.834	0.38
					Coeficientes de probabilidad de tipo de riesgo			
		Parámetro	Covariable	Valor verdadero	Estimación	S.E.	Estimación	S.E.
U(0.5,2)	38.7 %	γ_1	Intercepto	-1.0	-1.155	0.158	-1.844	0.140
		γ_1	**X**	0.5	0.523	0.026	0.248	0.200
					Coeficientes en función de riesgo			
		β_1	**X**	-0.5	-0.560	0.066	-0.414	0.012
		β_2	**X**	-1	-1.039	0.042	-0.900	0.087

Capítulo 6

Modelo de cura semiparamétrico utilizando modelos de transformación

6.1. Introduction

Los modelos de cura permiten estudiar datos en donde existe una proporción de individuos que nunca experimentan el evento de interés. Estos individuos son conocidos como *curados*, mientras que los individuos susceptibles al evento son considerados como *no curados*.

Un ejemplo típico de este tipo de datos es cuando en estudios clínicos, una proporción de individuos responden favorablemente a algún tratamiento, siendo libres de cualquier signo o síntoma de la enfermedad y son considerados como individuos completamente curados, mientras que el resto de los individuos pueden recaer al tratamiento.

Farewell (1986) y Taylor (1995) dan algunos ejemplos en el tratamiento de cáncer y en el efecto de radiaciones. El objetivo de estos estudios es estimar la tasa de cura, (proporción de individuos curados en la población) y la distribución de los tiempos de fallo para los individuos no curados ajustando covariables. Aplicaciones de modelos de cura pueden ser encontrados en muchas disciplinas incluyendo las ciencias biomédicas, la economía, la sociología y la ingeniería, entre muchas otras. Maller y Zhou (1996) dan una amplía lista de aplicaciones de este tema.

Gran literatura puede ser encontrada para el estudio de modelos de cura. En particular, una variedad de modelos paramétricos han sido considerados, los cuales utilizan un modelo logístico para la proporción curada y una distribución particular para los tiempos de fallo de los individuos no curados. Discusiones de modelos paramétricos pueden ser encontradas en: Berkson y Gage (1952), quienes utilizan una distribución exponencial para los tiempos de supervivencia y una constante para la fracción de cura para ajustar datos en el análisis de cáncer de seno y de estómago. Farewell (1982, 1986) proponen una regresión Weibull para la supervivencia y una regesión logística para la fracción de cura. Otras referencias importantes son: Boag (1949), Jones *et al.* (1981), Pack y Morgan (1990), Cantor y Shuster (1992), Maller y Zhou (1992), Sposto *et al.* (1992), Lo *et al.* (1993) y Ghitany *et al.* (1994). Distribuciones más generales tales como la gamma generalizada y la F generalizada son propuestas por Yamaguchi (1992) y Peng *et al.* (1998), respectivamente.

Los métodos paramétricos son parsimoniosos y fácil de interpretar, sin embargo pueden ser sensibles a errores de especificación del modelo, más aún, existe poca evidencia para sugerir y justificar el empleo de un modelo paramétrico.

Una alternativa a los modelos paramétricos son los modelos semiparamétricos. Entre los autores que estudian este tipo de modelos se pueden mencionar a: Kuk y Chen (1992), quienes consideran la estimación de los parámetros de regresión utilizando el método de máxima verosimilitud marginal, proponen el modelo llamado *modelo de cura de riesgos proporcionales*, en el cual el modelo de riesgos proporcionales de Cox (1972) es especificado en el tiempo de supervivencia de los individuos susceptibles y el modelo de regresión logística para la fracción curable. El modelo tiene la flexibilidad de ser semiparamétrico pero su método de estimación depende de aproximaciones de Monte Carlo, lo cual lo hace computacionalmente inconveniente para ser empleado.

Por otro lado, Taylor (1995) emplea el estimador de Kaplan-Meier para estimar la distribución de los tiempos de supervivencia para los individuos no curados y el algoritmo EM para estimar los coeficientes del modelo logístico. El problema con esta especificación es que no tiene la flexibilidad de incluir variables explicativas en la distibución de los tiempo de fallo para los individuos no curados.

Peng y Dear (2000) y Sy y Taylor (2000) estudian un modelo de mezclas noparamétrico. Ellos extienden el modelo de riesgos proporcionales de Cox permitiendo que existan pacientes curados en la población e investigan efectos de covariables sobre la distribución del tiempo de fallo t. El algoritmo EM y la verosimilitud marginal son empleados para estimar los parámetros de interés. Sus modelos extienden los modelos y los métodos de estimación existentes.

Fang *et al.* (2005) consideran la inferencia de un modelo de mezclas de riesgos logístico/proporcional semiparamétrico. Estudian la consistencia y la normalidad asintótica de sus estimadores semiparamétricos de máxima verosimilitud. Derivan la estimación consistente de la varianza para los componentes paramétricos y noparamétricos. Lu (2008) propone un acercamiento de verosimilitud noparamétrico para estimar la función de mortalidad integrada y los parámetros de regresión. Finalmente, establece la propiedades asintóticas de sus estimadores usando la teoría moderna de procesos empíricos.

Recientemente, Lu y Ying (2004) proponen una estimación de ecuaciones para modelos de cura semiparamétricos empleando modelos de transformación, donde la clase de modelos de transformación lineal son usados para el tiempo de fallo de los individuos susceptibles y un modelo logístico es usado para modelar la fracción curable. Emplean la teoría de martingalas para construir las ecuaciones de estimación. Las propiedades de muestras grandes de los estimadores resultantes son estudiadas. Sin embargo, el algoritmo para resolver las ecuaciones de estimación puede no converger y los estimadores resultantes de los parámetros de regresión no son eficientes aun cuando el modelo es especificada la propiedad de riesgos proporcionales.

Modelos semiparamétricos de transformación han sido estudiados en otros contextos como por ejemplo, para datos agrupados (Zeng y Lin, 2008), para eventos recurrentes (Zeng y Lin, 2007) y para situaciones de punto de cambio (Kosorok y Song, 2007).

El propósito de este capítulo es generalizar el modelo propuesto por Lu y Ying (2004) utilizando la clase de transformación lineal (ver Clayton y Cuzick ,1985; Cuzick, 1988; Bickel *et al.* 1993; Cheng *et al.*, 1995) para analizar datos con fracción curable. El modelo propuesto tiene la flexibilidad de incluir variables explicativas dependientes del tiempo, además combina una regresión logística para la probabilidad de ocurrencia y la clase de modelo de transformación para el tiempo de ocurrencia. Incluye como casos particulares, el modelo de cura de riesgos proporcionales (Farewell 1982; Kuk y Chen, 1992; Sy y Taylor, 2000; Peng y Dear, 2000) y el modelo momios-proporcionales (Bennet, 1983; Murphy *et al.* 1997; Zeng *et al.* 2005; Martinussen y Scheike 2006). Una vez especificado el modelo, se establecen las propiedades asintóticas de los estimadores resultantes del modelo empleando la teoría moderna de los procesos empíricos. Se muestra que los estimadores de los parámetros son semiparamétricamente eficientes y finalmente se deriva la estimación de la varianza consistente para los componentes finito e infinito dimensionales.

6.2. Modelos de cura

El objetivo de esta sección es presentar el marco teórico de los modelos de cura, los cuales suponen que la población subyacente es una mezcla de individuos susceptibles y no susceptibles. Se analizaran datos de supervivencia con censura a la derecha con la posibilidad de que existan individuos curados en la población, por lo tanto todos los individuos susceptibles deberán de experimentar el evento de interés siempre y cuando no exista censura, mientras que los individuos que son no susceptibles son considerados como inmunes al evento de interés. Bajo el modelado de mezclas se puede descomponer el tiempo del evento de la siguiente manera,

$$T = \eta T^* + (1 - \eta)\infty$$

donde $T^* < \infty$, y denota el tiempo de fallo de un individuo susceptible, mientras que η es una variable indicadora, la cual toma el valor de 1 si el individuo es susceptible y 0 si no. Así, se puede modelar por separado la distribución de supervivencia de los individuos susceptibles y la fracción de los susceptibles. El modelo para datos de supervivencia considerando una facción curable, es representado por las dos fórmulas siguientes

$$\lambda(t|\mathbf{Z}) = \lim_{dt\to 0} \frac{\mathbb{P}(t \le T^* < t + dt|T^* \ge t, \mathbf{Z})}{dt} \tag{6.1}$$

$$\pi(\gamma'\mathbf{X}) = \mathbb{P}(\eta = 1|\mathbf{X}, \gamma) = \frac{\exp(\gamma'\mathbf{X})}{1 + \exp(\gamma'\mathbf{X})} \tag{6.2}$$

donde el primer término representa la función de mortalidad para un individuo susceptible y el segundo representa la tasa de cura, la cual sigue un modelo logístico. $\mathbf{Z}$ y $\mathbf{X}$ son vectores de covariables en $\mathbb{R}^q$ y $\mathbb{R}^p$ respectivamente (las componentes del vector $\mathbf{Z}$ pueden ser variables dependientes del tiempo). $\mathbf{Z}$ y $\mathbf{X}$ pueden tener componentes independientes del tiempo en común, γ es el vector de coeficientes, el cual contiene el componente correspondiente a la ordenada. A partir de esta notación, la función de supervivencia de T es expresada como,

$$\begin{aligned} S_T(t|\mathbf{X}, \mathbf{Z}) &= \mathbb{P}(T > t|\mathbf{X}, \mathbf{Z}) \\ &= \mathbb{P}(T > t, \eta = 1|\mathbf{X}, \mathbf{Z}) + \mathbb{P}(T > t, \eta = 0|\mathbf{X}, \mathbf{Z}) \\ &= \mathbb{P}(T > t|\eta = 1, \mathbf{X}, \mathbf{Z})\mathbb{P}(\eta = 1|\mathbf{X}) + \mathbb{P}(T > t|\eta = 0, \mathbf{X}, \mathbf{Z})\mathbb{P}(\eta = 0|\mathbf{X}) \\ &= \mathbb{P}(T^* > t|\mathbf{Z})\pi(\gamma'\mathbf{X}) + \mathbb{P}(\eta = 0|\mathbf{X}) \\ &= \pi(\gamma'\mathbf{X})S_{T^*}(t|\mathbf{Z}) + 1 - \pi(\gamma'\mathbf{X}), \end{aligned}$$

donde $S_{T^*}(t|\mathbf{Z})$ es la función de supervivencia del tiempo de fallo para los individuos susceptibles. La función de densidad correspondiente es definida por

$$\begin{aligned} f_T(t|\mathbf{X}, \mathbf{Z}) &= \frac{d}{dt}\{F_T(t|\mathbf{X}, \mathbf{Z})\} \\ &= \frac{d}{dt}\{1 - S_T(t|\mathbf{X}, \mathbf{Z})\} \\ &= \frac{d}{dt}\{1 - \pi(\gamma'\mathbf{X})S_{T^*}(t|\mathbf{Z}) - 1 + \pi(\gamma'\mathbf{X})\} \\ &= -\pi(\gamma'\mathbf{X})\frac{d}{dt}S_{T^*}(t|\mathbf{Z}) \\ &= \pi(\gamma'\mathbf{X})f_{T^*}(t|\mathbf{Z}), \end{aligned}$$

donde $f_{T^*}(t|\mathbf{Z})$ es la función de densidad del tiempo de fallo para los individuos susceptibles.

Obsérvese que la función de supervivencia y la función de densidad para los individuos curados toman el valor de 0 y 1 respectivamente, para cada valor finito t, esto es porque los individuos curados no experimentan el evento de interé por lo tanto, sus tiempos de fallo pueden ser definidos como infinitos.

Supóngase que la variable T puede ser censurada a la derecha por una variable aleatoria positiva C. Denótese por τ el tiempo final de seguimiento en el estudio. Definase $Y = \min(T, \min(\tau, C))$ y $\Delta = 1\{T \le \min(\tau, C)\}$, donde $1\{\cdot\}$ denota la función indicadora. Más aún, se asume que el tiempo de censura C es independiente de T y η dado $\mathbf{Z}$ y $\mathbf{X}$. Los datos consisten en n vectores independientes $(Y_i, \Delta_i, \mathbf{Z}_i,\ \mathbf{X}_i)$ $(i, \ldots, n)$.

Supóngase que la distribución marginal de las covariables $\mathbf{Z}$ y $\mathbf{X}$ no dependen de los parámetro del tiempo de fallo y de la distribución de la tasa de cura, entonces la función de verosimilitud para el modelo de cura dadas n replicas *i.i.d* $(Y_i, \Delta_i, \mathbf{Z}_i,\ \mathbf{X}_i)$, $i, \ldots, n$, es representada por

$$\prod_{i=1}^{n} \{f_T(Y_i|\mathbf{X}_i, \mathbf{Z}_i)\}^{\Delta_i} \{S_T(Y_i|\mathbf{X}_i, \mathbf{Z}_i)\}^{(1-\Delta_i)}$$
$$= \prod_{i=1}^{n} \{\pi(\gamma'\mathbf{X}_i) f_{T^*}(Y_i|\mathbf{Z}_i)\}^{\Delta_i} \{1 - \pi(\gamma'\mathbf{X}_i) + \pi(\gamma'\mathbf{X}_i) S_{T^*}(Y_i|\mathbf{X}_i, \mathbf{Z}_i)\}^{(1-\Delta_i)}.$$

La construcción de esta función de verosimiltud para los modelos de cura fue también derivada por Fang *et al.* (2005) y Lu (2008).

6.3. Modelos de transformación

Emplear modelos de transformación permite obtener una gran variedad de modelos para el análisis de tiempos de supervivencia censurados. Esta clase de modelos incluyen el modelo de riesgos proporcionales y el modelo de odds-proporcionales como casos particulares, así como tambíen muchas otras alternativas. Estos modelos han sido de principal interés para muchos autores, véase por ejemplo: Cheng *et al.* (1995), Bagdonavičius y Nikulin (1999), Bagdonavičius y Nikulin (2002), Chen *et al.*, (2002), Slud y Vonta (2004), Kosorok y Song (2007), Martinussen y Scheike (2006), Zeng y Lin (2007) y Zeng *et al.* (2008).

6.3.1. Modelos de transformación semiparamétricos

Sea T^* un tiempo de fallo aleatorio y $\mathbf{Z}$ un vector de covariables de dimensión q, el cual es considerado por el momento, independiente del tiempo. La clase de transformación lineal es expresada por la escuación:

$$h(T^*) = -\beta'\mathbf{Z} + \varepsilon, \tag{6.3}$$

donde h es una función de transformación creciente desconocida, $\beta = (\beta_1, \ldots, \beta_q)'$ es el vector de dimensión q de los parámetros de regresión de interes y ε es una variable de error aleatoria con función de distribución conocida F_ε (ε es independiente de $\mathbf{Z}$).

De esta manera, el modelo (6.3) puede ser reparametrizado como,

$$\Lambda(T^*) = e^{-\beta'\mathbf{Z}} e^{\varepsilon},$$

donde $\Lambda(u) = \exp(h(u))$ es una función desconocida, la cual es estrictamente creciente y positiva tal que $\Lambda(0) = 0$ y $\lim_{u\to\infty} \Lambda(u) = \infty$.

Sea $F_{T^*|\mathbf{Z}}$ la función de distribución de T^* dado $\mathbf{Z}$. Entonces

$$\begin{aligned} F_{T^*|\mathbf{Z}}(t) &= \mathbb{P}(T^* \leq t|\mathbf{Z}) \\ &= \mathbb{P}(\Lambda(T^*) \leq \Lambda(t)|\mathbf{Z}) \\ &= \mathbb{P}(e^{-\beta'\mathbf{Z}} e^{\varepsilon} \leq \Lambda(t)|\mathbf{Z}) \\ &= \mathbb{P}(e^{\varepsilon} \leq \Lambda(t) e^{\beta'\mathbf{Z}}|\mathbf{Z}) \\ &= F_{e^\varepsilon}(\Lambda(t) e^{\beta'\mathbf{Z}}) \end{aligned}$$

donde F_{e^ε} denota la función de distribución condicional de e^{ε}. Por lo tanto, la función de riesgo para T^* dado $\mathbf{Z}$ es representada por

$$\lambda_{T^*|\mathbf{Z}}(t|\mathbf{Z}) = \frac{dF_{T^*|\mathbf{Z}}(t)/dt}{1 - F_{T^*|\mathbf{Z}}(t)} = \frac{dF_{e^\varepsilon}(\Lambda(t)e^{\beta'\mathbf{Z}})/dt}{1 - F_{e^\varepsilon}(\Lambda(t)e^{\beta'\mathbf{Z}})} = \frac{f_{e^\varepsilon}(\Lambda(t)e^{\beta'\mathbf{Z}})\lambda(t)e^{\beta'\mathbf{Z}}}{1 - F_{e^\varepsilon}(\Lambda(t)e^{\beta'\mathbf{Z}})},$$

donde $\lambda(\cdot)$ es la derivada de $\Lambda(\cdot)$ y f_{e^ε} denota la función de densidad de e^{ε}. Denotando λ_{e^ε} la función de mortalidad de $\exp(\varepsilon)$, la expresion anterior se reduce a

$$\lambda_{T^*|\mathbf{Z}}(t|\mathbf{Z}) = \lambda_{e^\varepsilon}(e^{\beta'\mathbf{Z}}\Lambda(t))e^{\beta'\mathbf{Z}}\lambda(t). \tag{6.4}$$

A partir de la ecuación (6.4) se pueden deducir algunos ejemplos de modelos de transformación importantes:

Ejemplo 1. Supóngase que ε tiene distribución valor extremo, es decir, $F_\varepsilon(u) = 1-\exp(-e^u)$. Entonces, $\exp(\varepsilon)$ es distribuida como una variable aleatoria exponencial, por lo cual $\lambda_{e^\varepsilon}(u) = 1$, para cada $u \geq 0$. De la ecuación (6.4) se sigue que

$$\lambda_{T^*|\mathbf{Z}}(t|\mathbf{Z}) = e^{\beta'\mathbf{Z}}\lambda(t),$$

obteniendo asi, una función de riesgo la cual sigue el modelo de riesgos proporcionales de Cox con función de riesgo base $\lambda(\cdot)$ y función de riego acumulativa $\Lambda(\cdot)$.

Ejemplo 2. Supóngase que ε tiene una distribución logística estándar, es decir, $F_\varepsilon(u) = \exp(u)/(1 + \exp(u))$. Entonces $\lambda_{e^\varepsilon}(u) = (1 + u)^{-1}$, por lo cual (6.4) se reduce a

$$\lambda_{T^*|\mathbf{Z}}(t|\mathbf{Z}) = \frac{\lambda(t)}{\Lambda(t) + e^{-\beta'\mathbf{Z}}}.$$

Para este caso la función de supervivencia condicional de T^* dado $\mathbf{Z}$ puede ser expresada como

$$S_{T^*|\mathbf{Z}}(t) = \frac{e^{-\beta'\mathbf{Z}}}{\Lambda(t) + e^{-\beta'\mathbf{Z}}},$$

o de manera equivalente por

$$logit(1 - S_{T^*|\mathbf{Z}}(t)) = \beta'\mathbf{Z} + h(t),$$

lo que proporciona el modelo de odd-proporcionales. Algunos aplicaciones de estos modelos pueden ser encontrados en Kosorok y Song (2007) y Ma y Kosorok (2005).

A continuación se dara una generalización del modelo de transformación, explicado anteriormente. Reescribiendo F_ε como

$$F_\varepsilon(u) = 1 - G(e^u),$$

donde G es una función decreciente conocida, tal que $G(0) = 1$ y $G(\infty) = 0$. Entonces,

$$\begin{aligned} \mathbb{P}(T^* > t|\mathbf{Z}) &= \mathbb{P}(h(T^*) > h(t)|\mathbf{Z}) = \mathbb{P}(-\beta'\mathbf{Z} + \varepsilon > h(t)|\mathbf{Z}) \\ &= \mathbb{P}(\varepsilon > h(t) + \beta'\mathbf{Z}|\mathbf{Z}) = 1 - F_\varepsilon(h(t) + \beta'\mathbf{Z}) \\ &= G\left(e^{h(t)}e^{\beta'\mathbf{Z}}\right) = G\left(e^{\beta'\mathbf{Z}}\Lambda(t)\right) \\ &= G\left(\int_0^t e^{\beta'\mathbf{Z}} d\Lambda(s)\right). \end{aligned}$$

Esta expresión puede ser extendida para incluir covariables dependientes del tiempo. Sea $\tilde{\mathbf{Z}}(t) = \{\mathbf{Z}(s) : 0 \leq s \leq t\}$, la cual denota la historia de $\mathbf{Z}(\cdot)$ en el intervalo de tiempo $[0, t]$, algunos autores consideran el modelo de transformación definido por la siguiente función de supervivencia,

$$\mathbb{P}(T^* > t|\tilde{\mathbf{Z}}(t)) = G\left(\int_0^t e^{\beta'\mathbf{Z}(s)} d\Lambda(s)\right).$$

o equivalentemente, por la siguiente función de riesgo acumulativa $\Lambda(t|\tilde{\mathbf{Z}}(t))$

$$\Lambda(t|\tilde{\mathbf{Z}}(t)) = -\log G\left(\int_0^t e^{\beta'\mathbf{Z}(s)} d\Lambda(s)\right) \equiv H\left(\int_0^t e^{\beta'\mathbf{Z}(s)} d\Lambda(s)\right). \tag{6.5}$$

Obsérve que H es una función creciente conocida con $H(0) = 0$ y $H(\infty) = \infty$. En particular, esta formulación es adoptada por: Kosorok y Song (2007) quienes estudian modelos de transformación lineal aplicados a datos de supervivencia con censura a la derecha y punto de cambio en los coeficientes de regresion; Zeng y Lin (2007) estudian un modelo de transformación con efectos aleatorios para eventos recurrentes; Zeng *et al.* (2008) estudian modelos de transformación lineal con efectos aleatorios para tiempos de fallo agrupados.

Observación. Especificando la función H dejando al mismo tiempo sin especificar la función Λ en (6.5) es equivalente a especificar la distribución de ε dejando al mismo tiempo sin especificar la función h en (6.3).

6.4. Notaciones y suposiciones del modelo

Se considerara un modelo de transformación lineal semiparamétrico con covariables dependientes del tiempo y fracción inmune, el cual es especificado por la ecuación (6.2) para la fracción curada y por la clase de modelos de ransformación lineal (6.5) para los tiempos de fallo para los individuos susceptibles.

Para comenzar con el estudio del modelo se daran algunas notaciones y suposiciones que seran utilizadas a través del capítulo. Para comenzar, se supondra que todas la variables son definidas sobre un espacio de probabilidad $(\Omega, \mathcal{C}, \mathbb{P})$.

Se asume que el vector de covariables $\mathbf{Z}$ es dependiente del tiempo, es decir, $\tilde{\mathbf{Z}}(t) = \{\mathbf{Z}(s) : 0 \leq s \leq t\}$; junto con las notaciones dadas en la sección 1, los datos consisten de n copias *i.i.d* $\mathbf{O}_i = (Y_i, \Delta_i, \tilde{\mathbf{Z}}_i(Y_i), \mathbf{X}_i)$, $i = 1, \ldots, n$. Denótese por $\theta = (\beta, \gamma, \Lambda)$ el vector de parámetros en el modelo (6.2)-(6.5) y por $\theta_0 = (\beta_0, \gamma_0, \Lambda_0)$ los verdaderos valores de los parámetros.

Para establecer las propiedades de los estimadores se necesita establecer algunas suposiciones de regularidad:

(C1) Los verdaderos valores β_0 y γ_0 pertenecen al interior de conjuntos compactos conocidos $\mathcal{B} \subset \mathbb{R}^q$ y $\mathcal{G} \subset \mathbb{R}^p$ respectivamente.

(C2) El vector de covariables $\mathbf{X}$ es acotado con matriz de covarianza positiva definida. $\mathbf{Z}(\cdot)$ es un proceso càglàd con variación uniformemente acotada en $[0, \tau]$. Sean

$$M_1 = \max_{t \in [0,\tau], \beta \in \mathcal{B}} e^{\beta' \mathbf{Z}(t)} \qquad M_2 = \min_{t \in [0,\tau], \beta \in \mathcal{B}} e^{\beta' \mathbf{Z}(t)}.$$

(C3) $\Lambda(\cdot)$ es una función estrictamente creciente sobre $[0, \tau]$. $\Lambda(\cdot)$ es continuamente diferenciable.

(C4) $H(\cdot)$ es tres veces difereciable en $[0, \infty)$, con $H^{(1)}(u) > 0$ y $\sup_{u \geq 0}\{|H^{(k)}(u)|\} < \infty$ $(k = 1, 2, 3)$, donde $H^{(k)}(\cdot)$ denota la k-ésima derivada de $H(\cdot)$.

(C5) Sea $\Psi : [0, \infty) \to [0, 1]$ una función definida por, $\Psi(u) = 1 - \exp\{-H(u)\}$. Existe una constante $\rho_0 > 0$ tal que

$$\limsup_{x \to \infty} (1 + x)^{\rho_0} (1 - \Psi(x)) < \infty, \quad \limsup_{x \to \infty} (1 + x)^{1+\rho_0} (\Psi^{(1)}(x)) < \infty$$

Bajo el valor verdadero θ_0, la esperanza de las variables aleatorias sera denotada por P_{θ_0}.

(C6) Con probabilidad 1, existe una constante positiva y finita M_3 tal que

$$P_{\theta_0}[\Delta | Y, X, \tilde{\mathbf{Z}}(Y)] > M_3, \quad t \in [0, \tau].$$

(C7) La siguiente condición de identificabilidad se cumple para cada $t \in [0,\tau]$: si existe un vector $\mu \in \mathbb{R}^q$ y una función determinista $\alpha_0(t)$, tal que con probabilidad 1, $\alpha_0(t)+\mu'\mathbf{Z}(t) = 0$, entonces $\mu = 0$ y $\alpha_0(t) = 0$.

Bajo el modelo (6.2)-(6.5) y las condiciones **C1**-**C7**, la función de verosimilitud para el parámetro θ de las observaciones $\mathbf{O}_i$ $(i,\ldots,n)$ es proporcional a

$$\begin{aligned} L_n(\theta) &= \prod_{i=1}^{n}\left\{\pi(\gamma'\mathbf{X}_i)e^{\beta'\mathbf{Z}_i(Y_i)}\lambda(Y_i)H^{(1)}\left(\int_0^{Y_i} e^{\beta'\mathbf{Z}_i(s)}d\Lambda(s)\right)\right. \\ &\quad \left.\times\exp\left\{-H\left(\int_0^{Y_i} e^{\beta'\mathbf{Z}_i(s)}d\Lambda(s)\right)\right\}\right\}^{\Delta_i} \\ &\quad \times\left\{1-\pi(\gamma'\mathbf{X}_i)+\pi(\gamma'\mathbf{X}_i)\exp\left\{-H\left(\int_0^{Y_i} e^{\beta'\mathbf{Z}_i(s)}d\Lambda(s)\right)\right\}\right\}^{(1-\Delta_i)} \end{aligned} \tag{6.6}$$

donde, para cualquier función $f(\cdot)$, $f^{(1)}(\cdot)$ denota la derivada de $f(\cdot)$ y $\lambda(\cdot)$ es la derivada de $\Lambda(\cdot)$.

6.5. Identificabilidad

En esta sección se estudiara la identificabilidad de los parámetros del modelo de cura semiparamétrico utlizando modelos de transformación. Los resultados son obtenidos siguiendo las definiciones de la información de Kullback-Leibler.

Proposición 6.5.1. *El modelo es identificable, es decir, si $L_1(\theta) = L_1(\theta^*)$ c.s. entonces $\theta = \theta^*$.*

Demostración: Supóngase que $L_1(\theta) = L_1(\theta^*)$ c.s. Por **C6**, existe un $\omega \in \Omega$ fuera de un conjunto despreciable donde $L_1(\theta)$ puede diferir de $L_1(\theta^*)$ cuando $\Delta(\omega) = 1$. Entonces, con probabilidad 1, se obtiene

$$\begin{aligned} &\left\{\pi(\gamma'\mathbf{x})e^{\beta'\mathbf{z}(y)}\lambda(y)\Psi^{(1)}\left(\int_0^{y} e^{\beta'\mathbf{z}(s)}d\Lambda(s)\right)\right\} \\ &\qquad = \left\{\pi(\gamma^{*\prime}\mathbf{x})e^{\beta^{*\prime}\mathbf{z}(y)}\lambda^*(y)\Psi^{(1)}\left(\int_0^{y} e^{\beta^{*\prime}\mathbf{z}(s)}d\Lambda^*(s)\right)\right\}. \end{aligned}$$

Esta igualdad puede ser reexpresada como:

$$\pi(\gamma'\mathbf{x})\frac{\partial\Psi\left(\int_0^y e^{\beta'\mathbf{z}(s)}d\Lambda(s)\right)}{\partial y} = \pi(\gamma^{*\prime}\mathbf{x})\frac{\partial\Psi\left(\int_0^y e^{\beta^{*\prime}\mathbf{z}(s)}d\Lambda^*(s)\right)}{\partial y}.$$

Sea $t \in [0,\tau]$, integrando ambos lados de esta ecuación de 0 a t se obtiene

$$\pi(\gamma'\mathbf{x})\Psi\left(\int_0^t e^{\beta'\mathbf{z}(s)}d\Lambda(s)\right) = \pi(\gamma^{*\prime}\mathbf{x})\Psi\left(\int_0^t e^{\beta^{*\prime}\mathbf{z}(s)}d\Lambda^*(s)\right),$$

de esta manera,

$$\frac{\Psi\left(\int_0^t e^{\beta'\mathbf{z}(s)}d\Lambda(s)\right)}{\Psi\left(\int_0^t e^{\beta^{*\prime}\mathbf{z}(s)}d\Lambda^*(s)\right)} = \frac{\pi(\gamma^{*'}\mathbf{x})}{\pi(\gamma'\mathbf{x})}.$$

Obsérvese que el lado derecho de esta ultima igualdad es independiente de t entonces,

$$\frac{\Psi\left(\int_0^t e^{\beta'\mathbf{z}(s)}d\Lambda(s)\right)}{\Psi\left(\int_0^t e^{\beta^{*\prime}\mathbf{z}(s)}d\Lambda^*(s)\right)} = \frac{\pi(\gamma^{*'}\mathbf{x})}{\pi(\gamma'\mathbf{x})} = \kappa,$$

con κ una constante positiva. Ahora, se demostrara que si $\frac{\pi(\gamma^{*'}\mathbf{x})}{\pi(\gamma'\mathbf{x})} = \kappa$ para todo $\mathbf{x}$, entonces $\gamma^* = \gamma$. En efecto, tomando $\mathbf{x} = 0$, entonces $\kappa = \frac{\pi(0)}{\pi(0)}$, y por lo tanto $\kappa = 1$.

Ahora se necesita demostrar que si $\frac{\pi(\gamma^{*'}\mathbf{x})}{\pi(\gamma'\mathbf{x})} = 1$ para todo $\mathbf{x}$, entonces $\gamma^* = \gamma$. En efecto, $\frac{\pi(\gamma^{*'}\mathbf{x})}{\pi(\gamma'\mathbf{x})} = 1$ implica que

$$\pi(\gamma^{*'}\mathbf{x}) = \pi(\gamma'\mathbf{x}). \tag{6.7}$$

Por definición de $\pi(\cdot)$, la ecuación (6.7) puede ser expresada como

$$\mathbb{P}(\eta = 1|\mathbf{x}, \gamma^*) = \mathbb{P}(\eta = 1|\mathbf{x}, \gamma),$$

o equivalentemente,

$$1 - \mathbb{P}(\eta = 0|\mathbf{x}, \gamma^*) = 1 - \mathbb{P}(\eta = 0|\mathbf{x}, \gamma),$$

lo que da como resultado,

$$\frac{1}{1 + e^{\gamma^{*'}\mathbf{x}}} = \frac{1}{1 + e^{\gamma'\mathbf{x}}}.$$

Por la condición **C2**, se sigue que $\gamma = \gamma^*$.

Como $\kappa = 1$, entonces,

$$\frac{\Psi\left(\int_0^t e^{\beta'\mathbf{z}(s)}d\Lambda(s)\right)}{\Psi\left(\int_0^t e^{\beta^{*\prime}\mathbf{z}(s)}d\Lambda^*(s)\right)} = 1,$$

para todo $t \in [0, \tau]$. Entonces,

$$\Psi\left(\int_0^t e^{\beta'\mathbf{z}(s)}d\Lambda(s)\right) = \Psi\left(\int_0^t e^{\beta^{*'}\mathbf{z}(s)}d\Lambda^*(s)\right).$$

Por definición de $\Psi(\cdot)$ esta expresión es equivalente a,

$$1 - \exp\left\{-H\left(\int_0^t e^{\beta'\mathbf{z}(s)}d\Lambda(s)\right)\right\} = 1 - \exp\left\{-H\left(\int_0^t e^{\beta^{*'}\mathbf{z}(s)}d\Lambda^*(s)\right)\right\},$$

y por lo tanto,

$$H\left(\int_0^t e^{\beta'\mathbf{z}(s)}d\Lambda(s)\right) = H\left(\int_0^t e^{\beta^{*\prime}\mathbf{z}(s)}d\Lambda^*(s)\right).$$

Se sigue por la condición **C4**

$$\int_0^t e^{\beta'\mathbf{z}(s)}d\Lambda(s) = \int_0^t e^{\beta^{*\prime}\mathbf{z}(s)}d\Lambda^*(s).$$

Tomando la derivada de Radon-Nikodym respecto a Λ^* en ambos lados de la igualdad anterior y aplicando logaritmos se obtiene

$$\beta'\mathbf{z}(t) + \log(\lambda(t)) = \beta^{*\prime}\mathbf{z}(t) + \log(\lambda^*(t)),$$

o equivalentemente,

$$(\beta - \beta^*)'\mathbf{z}(t) + \log\left(\frac{\lambda(t)}{\lambda^*(t)}\right) = 0.$$

Por la condición **C7**, se sigue que $\beta = \beta^*$ y $\lambda(t) = \lambda^*(t)$. Por lo tanto el modelo es identificable.

□

Pasamos ahora a la estimación en el modelo (6.2)-(6.5).

6.6. Máxima verosimilitud

Parecería lógico que para calcular el estimador de máxima verosimilitud (MLE) de θ_0 es a través de la maximización de la función de verosimilitud (6.6). Sin embargo, el máximo de ésta función es infinito cuando la función $\Lambda(\cdot)$ pertenece a la clase de funciones continuas. Zeng *et al.* (2008) proponen utilizar la estimación noparamétrica la cual consiste en remplazar la función continua $\Lambda(\cdot)$ por una función creciente de saltos definida en el intervalo $[0, \tau]$, con saltos en los tiempos de fallo. Siguiendo con este enfoque, supóngase que los datos contienen k ($k \leq n$) distintos tiempos de fallo, los cuales son denotados y ordenados como $s_1 < \ldots < s_k$. $\Lambda\{t\}$ denota el tamaño del salto en el instante t. Entonces se busca maximizar la función $L_n(\beta, \gamma, \Lambda) =$

$$\prod_{i=1}^{n}\left\{\pi(\gamma'\mathbf{X}_i)e^{\beta'\mathbf{Z}_i(Y_i)}\Lambda\{Y_i\}H^{(1)}\left(\sum_{j=1}^{k}e^{\beta'\mathbf{Z}_i(s_j)}\Lambda\{s_j\}1\{s_j \leq Y_i\}\right)\right.$$
$$\left.\times\exp\left\{-H\left(\sum_{j=1}^{k}e^{\beta'\mathbf{Z}_i(s_j)}\Lambda\{s_j\}1\{s_j \leq Y_i\}\right)\right\}\right\}^{\Delta_i}$$
$$\times\left\{1 - \pi(\gamma'\mathbf{X}_i) + \pi(\gamma'\mathbf{X}_i)\exp\left\{-H\left(\sum_{j=1}^{k}e^{\beta'\mathbf{Z}_i(s_j)}\Lambda\{s_j\}1\{s_j \leq Y_i\}\right)\right\}\right\}^{(1-\Delta_i)} \quad (6.8)$$

sobre el espacio

$$\Theta_n = \left\{(\beta,\gamma,\Lambda) : \beta \in \mathcal{B}, \gamma \in \mathcal{G}, \Lambda\{s_j\} \in [0,\infty), j = 1,\ldots,k\right\}.$$

Para cada n fijo, se define el estimador de máxima verosimilitud (MLE) $\hat{\theta}_n$ de θ_0, como el valor (si existe) que maximiza L_n sobre Θ_n. El estimador de máxima verosimilitud obtenido por este procedimiento es nombrado a veces como un MLE no paramétrico (NPMLE). A continuación se demuestra la existencia del estimador de máxima verosimilitud.

Proposición 6.6.1. *El estimador de máxima verosimilitud* $\hat{\theta}_n = (\hat{\beta}_n, \hat{\gamma}_n, \hat{\Lambda}_n(\cdot))$ *existe y es alcanzado.*

Demostración: La demostración es obtenida por contradicción y sigue el enfoque utilizado por Fang *et al.* (2005) y Hernández Quintero *et al.* (2009) para el estudio de la existencia de los estimadores de máxima verosimilitud de modelos de supervivencia en otros contextos.

Es suficiente mostrar que la función Λ tiene saltos finitos. Para esto, supóngase primero que $\Lambda\{s_j\} \leq U < \infty$ para cada $j = 1,\ldots,k$. La función L_n es una función contínua de β, γ y $\Lambda\{s_j\}$ sobre el conjunto compacto, $\mathcal{B}\times\mathcal{G}\times[0,U]^k$. Entonces L_n alcanza su máximo sobre este conjunto.

Para mostra que un máximo existe sobre el conjunto $\mathcal{B}\times\mathcal{G}\times[0,\infty)^k$, se mostrará que existe una U finita tal que para todo $(\beta^U,\gamma^U,\Lambda^U\{s_j\}; j = 1,\ldots,k) \in (\mathcal{B}\times\mathcal{G}\times[0,\infty)^k)\backslash(\mathcal{B}\times\mathcal{G}\times[0,U]^k)$, existe un $(\beta,\gamma,\Lambda\{s_j\}; j = 1,\ldots,k) \in \mathcal{B}\times\mathcal{G}\times[0,U]^k$ el cual tiene un valor mas grande que L_n.

El resultado es obtenido por contradicción. Esto es, supóngase que no existe tal U. Entonces, para todo $U < \infty$, existe un $(\beta^U,\gamma^U,\Lambda^U\{s_j\}; j = 1,\ldots,k) \in (\mathcal{B}\times\mathcal{G}\times[0,\infty)^k)\backslash(\mathcal{B}\times\mathcal{G}\times[0,U]^k)$ tal que para todo $(\beta,\gamma,\Lambda\{s_j\}; j = 1,\ldots,k) \in \mathcal{B}\times\mathcal{G}\times[0,U]^k$, $L_n(\beta,\gamma,\Lambda\{s_j\}; j = 1,\ldots,k) \leq L_n(\beta^U,\gamma^U,\Lambda^U\{s_j\}; j = 1,\ldots,k)$.

Se demostrará que $L_n(\beta^U,\gamma^U,\Lambda^U\{s_j\}; j = 1,\ldots,k)$ pude ser arbitrariamente pequeño mientras U crece. Para ver esto, observe que la ecuación (6.8) puede ser acotada superiormente por

$$\prod_{i=1}^{n}\left\{M_1\Lambda\{Y_i\}H^{(1)}\left(\sum_{j=1}^{k} e^{\beta'\mathbf{Z}_i(s_j)}\Lambda\{s_j\}1\{s_j \leq Y_i\}\right)\right.$$
$$\left.\times\exp\left\{-H\left(\sum_{j=1}^{k} M_2\Lambda\{s_j\}1\{s_j \leq Y_i\}\right)\right\}\right\}^{\Delta_i}$$

Si $(\beta^U,\gamma^U,\Lambda^U\{s_j\}; j = 1,\ldots,k) \in (\mathcal{B}\times\mathcal{G}\times[0,\infty)^k)\backslash(\mathcal{B}\times\mathcal{G}\times[0,U]^k)$ entonces existe al menos un $l \in \{1,\ldots,k\}$ tal que $\Lambda^U\{s_l\} > U$. También existe un $i^* \in \{1,\ldots,n\}$ tal que $\Delta_{i^*} = 1$ y

$Y_{i^*} = s_l$. Para este individuo se tiene que el término,

$$\Lambda^U\{Y_{i^*}\}H^{(1)}\left(\sum_{j=1}^{k} e^{\beta'\mathbf{Z}_i(s_j)}\Lambda^U\{s_j\}1\{s_j \leq Y_{i^*}\}\right)$$
$$\times \exp\left\{-H\left(\sum_{j=1}^{k} M_2\Lambda^U\{s_j\}1\{s_j \leq Y_{i^*}\}\right)\right\}$$

tiende a 0 cuando U tiende a $+\infty$, esto es obtenido utilizando la condición **C4** y el hecho de que $\Lambda^U\{Y_{i^*}\}$ tiende a $+\infty$. Entonces la cota superior de $L_n(\beta^U, \gamma^U, \Lambda^U\{s_j\}; j = 1, \ldots, k)$ puede ser arbitrariamente pequeña cuando U crece, lo cual es una contradicción. Por lo tanto, se sigue que para cualquier n fijo, el máximo de L_n es obtenido en el conjunto $\mathcal{B} \times \mathcal{G} \times [0, U]^k$, para algún $U < \infty$ y sobre este conjunto el máximo $\widehat{\theta_n}$ es alcanzado.

□

Denotemos por $\mathbb{P}_n$ la distribución empirica de los datos y por P_{θ_0} la esperanza con respecto al valor verdadero.

Lema 6.6.1. *El NPMLE $\widehat{\theta_n}$ satisface la siguiente ecuación, para cada $t \in [0, \tau]$*

$$\hat{\Lambda}_n(t) = \int_0^t \frac{dG_n(u)}{W_n(u; \hat{\theta}_n)} \tag{6.9}$$

donde $(1/W_n)(u; \hat{\theta}_n)$ y $G_n(u)$ son funciones no decrecientes en u, definidas por

$$W_n(u; \theta) = \mathbb{P}_n[w(u, \mathbf{O}; \theta)] \qquad y \qquad G_n(u) = \mathbb{P}_n\left[\Delta 1\{Y \leq u\}\right]$$

respectivamente, con $w(u, \mathbf{O}; \theta) = e^{\beta'\mathbf{Z}(u)}1\{u \leq Y\}\phi(\mathbf{O}; \theta)$ y

$$\begin{aligned}\phi(\mathbf{O}; \theta) \quad = \quad & \frac{(1-\Delta)\pi(\gamma'\boldsymbol{X})\Psi^{(1)}\left(\int_0^Y e^{\beta'\mathbf{Z}(u)}d\Lambda(u)\right)}{1-\pi(\gamma'\boldsymbol{X}) + \pi(\gamma'\boldsymbol{X})\left(1-\Psi\left(\int_0^Y e^{\beta'\mathbf{Z}(u)}d\Lambda(u)\right)\right)} \\ & -\frac{\Delta\Psi^{(2)}\left(\int_0^Y e^{\beta'\mathbf{Z}(u)}d\Lambda(u)\right)}{\Psi^{(1)}\left(\int_0^Y e^{\beta'\mathbf{Z}(u)}d\Lambda(u)\right)}.\end{aligned} \tag{6.10}$$

Demostración: Esta demostración consiste en de dos pasos:

1. Se toma la derivada respecto a los tamaños de los saltos $\Lambda\{s_j\}$ $(j = 1, \ldots, k)$ de la función-log-verosimilitud, es decir, la derivada del logarítmo de la ecuación (6.8)), la cual se re-

duce a la siguiente forma

$$
\begin{aligned}
l_n(\theta) = \sum_{i=1}^{n} \Delta_i \ln \pi(\gamma' \mathbf{X}_i) + \Delta_i \beta' \mathbf{Z}_i(Y_i) + \Delta_i \sum_{j=1}^{k} 1\{Y_i = s_j\} \ln(\Lambda\{s_j\}) \\
+\Delta_i \ln\left(\Psi^{(1)}\left(\sum_{j=1}^{k} e^{\beta' \mathbf{Z}_i(s_j)} \Lambda\{s_j\} 1\{s_j \leq Y_i\}\right)\right) \\
+(1-\Delta_i) \ln\left(1 - \pi(\gamma' \mathbf{X}_i) + \pi(\gamma' \mathbf{X}_i)\left(1 - \Psi\left(\sum_{j=1}^{k} e^{\beta' \mathbf{Z}_i(s_j)} \Lambda\{s_j\} 1\{s_j \leq Y_i\}\right)\right)\right).
\end{aligned}
$$

2. Poniendo $(\partial l_n(\theta)/\partial \Lambda\{s_j\})|_{\theta=\widehat{\theta}_n} = 0$ y resolviendo para $\Lambda\{s_j\}$.

Realizando estos dos pasos el resultado es obtenido, concluyendo así la prueba.

□

6.7. Consistencia

Ahora se esta interesado en el estudio de la consitencia de los estimadores noparamétricos del modelo de cura semiparamétrico utilizando modelos de transformación.

Teorema 6.7.1. *Bajo las condiciones **C1**-**C7**, el NPMLE es consistente, es decir,*

$$
\sup_{t \in [0,\tau]} |\hat{\Lambda}_n(t) - \Lambda_0(t)|, \quad ||\hat{\gamma}_n - \hat{\gamma}_0|| \quad and \quad ||\hat{\beta}_n - \hat{\beta}_0||
$$

convergen a 0 casi seguramente cuando n tiende a ∞.

La prueba de la consistencia sigue las ideas desarrolladas por Murphy (1994) para modelos frailty, véase también Zeng y Lin (2007) y Zeng *et al.* (2008), quienes usan técnicas similares en modelos de transformación con eventos recurrentes y para tiempos de fallo agrupados respectivamente. Tres lemas son necesarios antes de presentar la demostracion del Teorema 6.7.1.

Los siguientes lemas se satisfacen bajo las condiciones **C4** y **C5**.

Lema 6.7.1. *Con probabilidad 1, se satisface:*

$$
\left\{\pi(\gamma' \boldsymbol{X}_i) e^{\beta' \mathbf{Z}_i(Y_i)} \Psi^{(1)}\left(\int_0^{Y_i} e^{\beta' \mathbf{Z}_i(s)} d\Lambda(s)\right)\right\}^{\Delta_i}
$$

$$\times\left\{1-\pi(\gamma'\boldsymbol{X}_i)+\pi(\gamma'\boldsymbol{X}_i)\left(1-\Psi\left(\int_0^{Y_i} e^{\beta'\mathbf{Z}_i(s)}d\Lambda(s)\right)\right)\right\}^{(1-\Delta_i)}$$
$$\leq M_4\,(1+\Lambda(Y_i))^{-(1+\rho_0)\Delta_i}+M_5\,(1+\Lambda(Y_i))^{-(\rho_0+\Delta_i)}\,, \tag{6.11}$$

donde M_4 y M_5 son constantes positivas independientes de β, γ y Λ.

Demostración. Por la condición **C5**, existe una constante $\rho_0 > 0$ tal que,

$$\limsup_{x\to\infty}(1+x)^{\rho_0}(1-\Psi(x)) < \infty \quad \text{y} \quad \limsup_{x\to\infty}(1+x)^{1+\rho_0}(\Psi^{(1)}(x)) < \infty.$$

Por la segunda desigualdad, existe una constante $m_0 < \infty$ tal que

$$\pi(\gamma'\mathbf{X}_i)\Psi^{(1)}\left(\int_0^{Y_i} e^{\beta'\mathbf{Z}_i(s)}d\Lambda(s)\right) \leq m_0\left(1+\int_0^{Y_i} e^{\beta'\mathbf{Z}_i(s)}d\Lambda(s)\right)^{-(1+\rho_0)},$$

entonces,

$$\left\{\pi(\gamma'\mathbf{X}_i)\Psi^{(1)}\left(\int_0^{Y_i} e^{\beta'\mathbf{Z}_i(s)}d\Lambda(s)\right)\right\}^{\Delta_i} \leq m_0^{\Delta_i}\left(1+\int_0^{Y_i} e^{\beta'\mathbf{Z}_i(s)}d\Lambda(s)\right)^{-(1+\rho_0)\Delta_i} \tag{6.12}$$

Por la primera desigualdad de la condición **C5**, existe una constante $m_1 < \infty$ tal que

$$\begin{aligned}
1-\pi(\gamma'\mathbf{X}_i)+\pi(\gamma'\mathbf{X}_i)&\left(1-\Psi\left(\textstyle\int_0^{Y_i} e^{\beta'\mathbf{Z}_i(s)}d\Lambda(s)\right)\right)\\
&\leq 1+\left(1-\Psi\left(\textstyle\int_0^{Y_i} e^{\beta'\mathbf{Z}_i(s)}d\Lambda(s)\right)\right)\\
&\leq 1+m_1\left(1+\textstyle\int_0^{Y_i} e^{\beta'\mathbf{Z}_i(s)}d\Lambda(s)\right)^{-\rho_0},
\end{aligned}$$

entonces,

$$\begin{aligned}
\left\{1-\pi(\gamma'\mathbf{X}_i)+\pi(\gamma'\mathbf{X}_i)\left(1-\Psi\left(\int_0^{Y_i} e^{\beta'\mathbf{Z}_i(s)}d\Lambda(s)\right)\right)\right\}^{(1-\Delta_i)}&\\
\leq \quad \left\{1+m_1\left(1+\int_0^{Y_i} e^{\beta'\mathbf{Z}_i(s)}d\Lambda(s)\right)^{-\rho_0}\right\}^{(1-\Delta_i)}&\\
\leq \quad 1+m_1\left(1+\int_0^{Y_i} e^{\beta'\mathbf{Z}_i(s)}d\Lambda(s)\right)^{-\rho_0(1-\Delta_i)}&.
\end{aligned} \tag{6.13}$$

Sea $m_2 = \text{mín}\{M_2, 1\}$, donde M_2 es definida en la condición **C2**. Entonces,

$$1+\int_0^{Y_i} e^{\beta'\mathbf{Z}_i(s)}d\Lambda(s) \geq 1+M_2\Lambda(Y_i) \geq m_2\,(1+\Lambda(Y_i))\,. \tag{6.14}$$

De las desigualdades (6.12), (6.13), (6.14) y por la condición (**C2**), se obtiene

$$\begin{aligned}
&\left\{\pi(\gamma'\mathbf{X}_i)e^{\beta'\mathbf{Z}_i(Y_i)}\Psi^{(1)}\left(\int_0^{Y_i} e^{\beta'\mathbf{Z}_i(s)}d\Lambda(s)\right)\right\}^{\Delta_i}\\
&\qquad\times\left\{1-\pi(\gamma'\mathbf{X}_i)+\pi(\gamma'\mathbf{X}_i)\left(1-\Psi\left(\int_0^{Y_i} e^{\beta'\mathbf{Z}_i(s)}d\Lambda(s)\right)\right)\right\}^{(1-\Delta_i)}\\
&\leq M_1^{\Delta_i}m_0^{\Delta_i}\left[m_2\left(1+\Lambda(Y_i)\right)\right]^{-(1+\rho_0)\Delta_i}\left[1+m_1\left(m_2\left(1+\Lambda(Y_i)\right)\right)^{-\rho_0(1-\Delta_i)}\right]\\
&= M_4\left(1+\Lambda(Y_i)\right)^{-(1+\rho_0)\Delta_i}+M_5\left(1+\Lambda(Y_i)\right)^{-(\rho_0+\Delta_i)}
\end{aligned}$$

donde $M_4 = M_1^{\Delta_i}m_0^{\Delta_i}m_2^{-(1+\rho_0)\Delta_i}$ y $M_5 = M_1^{\Delta_i}m_0^{\Delta_i}m_2^{-(\rho_0+\Delta_i)}m_1$. Esto concluye la prueba.

□

Lema 6.7.2. *El* lím $\sup_n \hat{\Lambda}_n(\tau) < \infty$ *casi seguramente.*

Demostración: Los principales argumentos para esta prueba se basan en Murphy (1994), Zeng and Lin (2007) and Zeng *et al.* (2008).

Considérese la función de saltos $\tilde{\Lambda}_n(\cdot)$ definida como:

$$\tilde{\Lambda}_n(t) = \int_0^t \frac{dG_n(u)}{W_n(u;\theta_0)}.$$

Obsérvese que $\tilde{\Lambda}_n(t)$ respecto a G_n es similar a $\hat{\Lambda}_n(t)$, pero con la diferencia de que W_n es evaluada en θ_0. Sea $\tilde{\theta}_n = (\beta_0, \gamma_0, \tilde{\Lambda}_n)$. La prueba es hecha por contradicción.

Claramente, $0 \leq n^{-1}[\ln L_n(\hat{\theta}_n) - \ln L_n(\tilde{\theta}_n)]$. El lado derecho de esta desigualdad puede ser acotado superiormente utilizando el Lema 6.7.1, lo cual conduce a

$$\begin{aligned}
0 \leq M_6 &+ \tfrac{1}{n}\textstyle\sum_{i=1}^n\Big\{\Delta_i\ln(\hat{\Lambda}_n\{Y_i\})\\
&\qquad -(1+\rho_0)\Delta_i\ln\left(1+\hat{\Lambda}_n(Y_i)\right)-(\rho_0+\Delta_i)\ln\left(1+\hat{\Lambda}_n(Y_i)\right)\Big\}\\
&-\tfrac{1}{n}\textstyle\sum_{i=1}^n\Big\{\Delta_i\ln(\tilde{\Lambda}_n\{Y_i\})+\Delta_i\ln\left(\Psi^{(1)}\left(\sum_{j=1}^k e^{\beta_0'\mathbf{Z}_i(s_j)}\tilde{\Lambda}_n\{s_j\}1\{s_j\leq Y_i\}\right)\right)\\
&\qquad +(1-\Delta_i)\ln\Big(1-\pi(\gamma_0'\mathbf{X}_i)\\
&\qquad\qquad +\pi(\gamma_0'\mathbf{X}_i)\left(1-\Psi\left(\textstyle\sum_{j=1}^k e^{\beta_0'\mathbf{Z}_i(s_j)}\tilde{\Lambda}_n\{s_j\}1\{s_j\leq Y_i\}\right)\right)\Big)\Big\},
\end{aligned}$$

donde M_6 es una constante positiva. De la construcción de $\tilde{\Lambda}_n(t)$, se sigue que

$$\begin{aligned}
0 &\leq M_6+\frac{1}{n}\sum_{i=1}^n\left\{\Delta_i\ln(\hat{\Lambda}_n\{Y_i\})-(2\Delta_i+\rho_0+\rho_0\Delta_i)\ln\left(1+\hat{\Lambda}_n(Y_i)\right)\right\}\\
&\leq M_6+\frac{1}{n}\sum_{i=1}^n\Big\{\Delta_i\ln(1+\hat{\Lambda}_n(Y_i))-(2\Delta_i+\rho_0+\rho_0\Delta_i)1\{Y_i=\tau\}\ln\left(1+\hat{\Lambda}_n(Y_i)\right)\\
&\qquad -(2\Delta_i+\rho_0+\rho_0\Delta_i)1\{Y_i<\tau\}\ln\left(1+\hat{\Lambda}_n(Y_i)\right)\Big\}. \qquad (6.15)
\end{aligned}$$

Ahora siguiendo los argumentos de Murphy (1994), se puede demostrar que el lado derecho de (6.15) es negativo si $\hat{\Lambda}_n(\tau)$ diverge. Para esto, considérese una partición $\tau = s_0 > \ldots > s_N = 0$ de $[0, \tau]$. Entonces el lado derecho de (6.15) puede ser acotado superiormente por

$$\begin{aligned} M_6 &- \frac{1}{2n}\sum_{i=1}^{n}(2\Delta_i + \rho_0 + \rho_0\Delta_i)1\{Y_i = \tau\}\ln\left(1 + \hat{\Lambda}_n(\tau)\right) \\ &- \left\{\frac{1}{2n}\sum_{i=1}^{n}(2\Delta_i + \rho_0 + \rho_0\Delta_i)1\{Y_i = \tau\}\ln\left(1 + \hat{\Lambda}_n(\tau)\right)\right. \\ &\qquad\qquad \left. -\frac{1}{n}\sum_{i=1}^{n}\Delta_i 1\{Y_i \in [s_1, s_0)\}\ln\left(1 + \hat{\Lambda}_n(\tau)\right)\right\} \\ &- \left\{\sum_{q=1}^{N}\left\{\frac{1}{n}\sum_{i=1}^{n}(2\Delta_i + \rho_0 + \rho_0\Delta_i)1\{Y_i \in [s_q, s_{q-1})\}\ln\left(1 + \hat{\Lambda}_n(s_q)\right)\right.\right. \\ &\qquad\qquad \left.\left. -\frac{1}{n}\sum_{i=1}^{n}\Delta_i 1\{Y_i \in [s_{q+1}, s_q)\}\ln\left(1 + \hat{\Lambda}_n(s_q)\right)\right\}\right\}. \end{aligned} \tag{6.16}$$

Usando la ideas de Murphy (1994) para la construcción de la partición, la sucesión $s_0 > s_1 > \ldots > s_N$ puede ser escogida de tal manera que el primer térmno de (6.16) diverge a $-\infty$ cuando $\hat{\Lambda}_n(\tau) \to \infty$, mientras que el segundo y el tercero son negativos cuando n es grande. Esto contradice al hecho de que (6.16) debería ser no negativa. Por lo tanto $\limsup \hat{\Lambda}_n(\tau) < \infty$.

□

Lema 6.7.3. $\tilde{\Lambda}_n(t)$ *converge uniformemente a* $\Lambda_0(t)$ *casi seguramente.*

Demostración. Primero se demostrará que la clase de funciones

$$\{w(u, \mathbf{O}; \theta); u \in [0, \tau], \theta \in \Theta\}$$

es una clase Donsker. Por el ejemplo 8 de la sección 4.2 establce que si $\mathbf{Z}$ es un proceso caglad sobre $[0, \tau]$ uniformemente acotado entonces, $\mathbf{Z}(\cdot)$ es Donsker, multiplicar dos clases Donsker se obtiene que $\{\beta'\mathbf{Z}(u) : u \in [0, \tau], \beta \in \mathcal{B}\}$ es Donsker. La función exponencial es Lipschitziana sobre un conjunto compacto de la recta real. (Esto se sigue de la expansión de Taylor de primer orden). Dado que $\mathbf{Z}(\cdot)$ es uniformemente acotada, se sigue de los ejemplos 3 y 4 de la sección 4.2, que la clase $\{e^{\beta'\mathbf{Z}(u)} : u \in [0, \tau], \beta \in \mathcal{B}\}$ es Donsker.

Como $\mathbf{X}$ es acotada, por el ejemplo 2 de la sección 4.2 implica que la clase $\{\gamma'\mathbf{X} : \gamma \in \mathcal{G}\}$ es Donsker. Diferenciabilidad de $e^{\gamma'\mathbf{X}}$ en $\mathbf{X}$ y que la derivada es acotada implican que $\{e^{\gamma'\mathbf{X}} : \gamma \in \mathcal{G}\}$ es Donsker (ver ejemplo 3 de la sección 4.2). Como $1 + e^{\gamma'\mathbf{X}} > 0$, entonces $\{e^{\gamma'\mathbf{X}}/(1 + e^{\gamma'\mathbf{X}})\}$ es Donsker (ver ejemplo, 7 de la sección 4.2). Por la condición **C4**, la función $\Psi(\cdot)$ es Donsker (porque su derivada es acotada y Lipschitziana), entonces $1-\Psi(\cdot)$ es Donsker. Por el ejemplo 5 y 6 de la sección 4.2 (multiplicando y sumando clases Donsker uniformemente acotadas, preservan la propiedad de Donsker) entonces las clases, $\{1-\pi(\gamma' X)+\pi(\gamma'\mathbf{X})\Psi(\cdot)\}$ y $\{(1-\Delta_i)\pi(\gamma'\mathbf{X})\Psi^{(1)}(\cdot)\}$ son Donsker.

Como $1-\pi(\gamma'\mathbf{X})+\pi(\gamma'\mathbf{X})\Psi(\cdot)>0$ y por los ejemplos 6 y 7 de la sección 4.2, la clase

$$\left\{\frac{(1-\Delta)\pi(\gamma'\mathbf{X})\Psi^{(1)}(\cdot)}{1-\pi(\gamma'\mathbf{X})+\pi(\gamma'\mathbf{X})\Psi(\cdot)}\right\},$$

es Donsker. Por la condición **C4** las funciones $\Psi^{(1)}(\cdot)$ y $\Psi^{(2)}(\cdot)$ son Lipschitzianas con $\Psi^{(1)}(\cdot)>0$, entonces,

$$\left\{\frac{\Delta_i\Psi^{(2)}(\cdot)}{\Psi^{(1)}(\cdot)}\right\}$$

es Donsker. De este análisis se obtiene que $\phi(Y,\mathbf{O};\theta)$ es una clase Donsker. Finalmente, la función indicadora es Donsker.

Entonces, se concluye que la clase

$$\{w(u,\mathbf{O};\theta);u\in[0,\tau],\theta\in\Theta\}$$

es Donsker. Argumentos similares permiten demostrar que el término

$$\{\Delta 1\{u\leq t\}/P_{\theta_0}[w(u,\mathbf{O};\theta_0)]\}$$

es tambien una clase Donsker.

Entonces,

$$\begin{aligned}
\sup_{t\in[0,\tau]}\left|\tilde{\Lambda}_n(t)-\Lambda_0(t)\right| &= \sup_{t\in[0,\tau]}\left|\frac{1}{n}\sum_{i=1}^{n}\frac{\Delta_i 1\{Y_i\leq t\}}{W_n(Y_i;\theta_0)}-P_{\theta_0}\left[\frac{\Delta 1\{Y\leq t\}}{P_{\theta_0}[w(u,\mathbf{O};\theta_0)]|_{s=Y}}\right]\right| \\
&\leq \sup_{s\in[0,\tau]}\left|\frac{1}{W_n(s;\theta_0)}-\frac{1}{P_{\theta_0}[w(s,\mathbf{O};\theta_0)]}\right| \\
&\quad+\sup_{t\in[0,\tau]}\left|(\mathbb{P}_n-P_{\theta_0})\left[\frac{\Delta 1\{Y\leq t\}}{P_{\theta_0}[w(u,\mathbf{O};\theta_0)]|_{s=Y}}\right]\right| \qquad (6.17)
\end{aligned}$$

Del resultado obtenido arriba, $\{w(u,\mathbf{O};\theta_0);u\in[0,\tau],\theta\in\Theta\}$ es una clase Donsker y por lo tanto una clase Glivenko Cantelli, entonces

$$\sup_{u\in[0,\tau]}|W_n(u;\theta_0)-P_{\theta_0}[w(u,\mathbf{O};\theta_0))]|$$

converge a 0 casi seguramente. Más aún, para $u\in[0,\tau]$, $P_{\theta_0}[w(u,\mathbf{O};\theta_0)]>0$ en $[0,\tau]$. Entonces el primer término de lado derecho de (6.17) converge a 0 casi seguramente y el segundo término converge a 0 casi seguramente por la propiedad de Glivenko-Cantelli del término

$$\{\Delta 1\{u\leq t\}/P_{\theta_0}[w(u,\mathbf{O};\theta_0)]|_{u=Y}\}.$$

Se concluye que $\widetilde{\Lambda}_n$ converge uniformemente a Λ_0, lo que finaliza la prueba.

□

Demostración del Teorema 6.7.1. La demostración consiste en dos pasos:

1. Se demostrará que cada sucesión de n contiene una subsucesión donde el NPMLE $\widehat{\theta}_n$ converge.

2. Se demostrará que el conjunto de límite de todas las subsucesiones convergentes de $\widehat{\theta}_n$ se reduce a $\{\theta_0\}$.

Demostración 1. Como $\mathcal{B} \times \mathcal{G}$ es un conjunto compacto y por el Teorema de Bolzanno-Weierstrass, cada sucesión de $(\widehat{\beta}_n, \widehat{\gamma}_n)$ tiene una subsucesión $(\widehat{\beta}_{\phi(n)}, \widehat{\gamma}_{\phi(n)})$, la cual converge a algún (β^*, γ^*) en $\mathcal{B} \times \mathcal{G}$. Por el Lema 6.7.2 y por el Teorema de Helly, se puede encontrar con probabilidad 1 una subsucesión $\widehat{\Lambda}_{\varphi(n)}$ de $\widehat{\Lambda}_{\phi(n)}$ y una función no decreciente continua por la derecha Λ^* tal que $\widehat{\Lambda}_{\varphi(n)}(t) \to \Lambda^*(t)$ para todo $t \in [0, \tau]$, donde Λ^* es continua; se dice que $\widehat{\Lambda}_{\varphi(n)}$ converge débilemnte a Λ^*. Entonces, se puede encontrar una subsucesión $\xi(n)$ de $\varphi(n)$ de tal manera que esta convergencia débil se cumple a lo largo de $\xi(n)$. Ahora se demostará Λ^* es continua en $[0, \tau]$. Nótese que

$$\begin{aligned} \hat{\Lambda}_{\xi(n)}(t) &= \int_0^t \frac{dG_{\xi(n)}(u)}{\mathbb{P}_{\xi(n)}[w(u, \mathbf{O}; \hat{\theta}_{\xi(n)})]} \\ &= \int_0^t \frac{\mathbb{P}_{\xi(n)}[w(u, \mathbf{O}; \theta_0)]}{\mathbb{P}_{\xi(n)}[w(u, \mathbf{O}; \hat{\theta}_{\xi(n)})]} \frac{dG_{\xi(n)}}{\mathbb{P}_{\xi(n)}[w(u, \mathbf{O}; \theta_0)]} \\ &= \int_0^t \frac{\mathbb{P}_{\xi(n)}[w(u, \mathbf{O}; \theta_0)]}{\mathbb{P}_{\xi(n)}[w(u, \mathbf{O}; \theta_{\xi(n)})]} d\tilde{\Lambda}_{\xi(n)}(u), \end{aligned} \tag{6.18}$$

donde $\widetilde{\Lambda}_n$ es definida en el lema 6.7.2.

Se sigue de la propiedad de Glivenko-Cantelli de $\{w(u, \mathbf{O}; \theta)\}$ que

$$\begin{aligned} &\sup_{u \in [0,\tau]} \left| \mathbb{P}_{\xi(n)}[w(u, \mathbf{O}; \theta_0)] - P_{\theta_0}[w(s, \mathbf{O}; \theta_0)] \right| \longrightarrow 0 \quad c.s., \\ &\sup_{u \in [0,\tau]} \left| \mathbb{P}_{\xi(n)}[w(s, \mathbf{O}; \widehat{\theta}_{\xi(n)})] - P_{\theta_0}[w(s, \mathbf{O}; \widehat{\theta}_{\xi(n)})] \right| \longrightarrow 0 \quad c.s.. \end{aligned} \tag{6.19}$$

Entonces, $P_0[w(u, \mathbf{O}; \hat{\theta}_{\xi(n)})]$ converge uniformemente a $P_0[w(u, \mathbf{O}; \theta^*)]$. Usando este resultado, la ecuación 6.19 y la desigualdad del triángulo, se obiene que,

$$\frac{d\hat{\Lambda}_{\xi(n)}(t)}{d\tilde{\Lambda}_{\xi(n)}(t)} \to \frac{P_{\theta_0}\left[w(s, \mathbf{0}; \theta_0)\right]}{P_{\theta_0}\left[w(s, \mathbf{0}; \theta^*)\right]}$$

uniformemente en $t \in [0, \tau]$. Tomando límites en ambos lados de (6.18), se obtiene

$$\Lambda^*(t) = \int_0^t \frac{P_0 \left[w(u, \mathbf{O}; \theta_0) \right]}{P_0 \left[w(u, \mathbf{O}; \theta^*) \right]} d\Lambda_0(u).$$

Entonces $\Lambda^*(t)$ es absolutamente continua con respecto a $\Lambda_0(t)$, por lo tanto, $\Lambda^*(t)$ es diferenciable con respecto a t.

Demostración de 2. Finalmente, se quiere demostrar que $\beta^* = \beta_0$, $\gamma^* = \gamma_0$ y $\Lambda^* = \Lambda_0$. Considérese la diferencia

$$0 \leq \frac{1}{\xi(n)} l_{\xi(n)} \left(\hat{\gamma}_{\xi(n)}, \hat{\beta}_{\xi(n)}, \hat{\Lambda}_{\xi(n)} \right) - \frac{1}{\xi(n)} l_{\xi(n)} \left(\gamma_0, \beta_0, \tilde{\Lambda}_{\xi(n)} \right).$$

Al permitir que n vaya a infinito, se obtiene que $P_{\theta_0} \left[l\left(\gamma^*, \beta^*, \Lambda^*\right) - l\left(\gamma_0, \beta_0, \Lambda_0\right) \right] \geq 0$. El lado izquierdo de esta desigualdad es el negativo de la información Kullback-Leibler entre la densidad indexada por θ^* y la densidad verdadera, lo cual implica que $\theta^* = \theta_0$. Esto prueba que $\hat{\beta}_n, \hat{\gamma}_n$ y $\hat{\Lambda}(t)$ ($t \in [0, \tau]$) convergen casi seguramente a β_0, γ_0 y $\Lambda_0(t)$.

□

6.8. Normalidad Asintótica

6.8.1. Score e Información

Para obtener la normalidad asintótica de los estimadores de máxima verosimilitud, se adapta el enfoque de una función analítica desarrollada por Murphy(1995) para modelos frailty; véase también Fang *et al.* (2005), Kosorok y Song (2007), Lu (2008), y Hernández Quintero *et al.* (2010), quienes recientemente han adpotado es enfoque para otros modelos de regresión semiparamétricos en el análisis de supervivencia. Para obtener la distribución asintótica de los estimadores, debemos verificar que un análogo de la matriz de información (ahora un operador) es continuamente invertible y que las ecuaciones de score son asintóticamente normales. En esta última verificación se usan resultados de la teoría de procesos empíricos.

Considérese el submodelo

$$\epsilon \to \hat{\theta}_{n,\epsilon} = \left(\hat{\beta}_n + \epsilon \mathbf{h}_\beta, \hat{\gamma}_n + \epsilon \mathbf{h}_\gamma, \int_0^{\cdot} (1 + \epsilon h_\Lambda(s)) \, d\hat{\Lambda}_n(s) \right)$$

donde h_Λ es una función no negativa en $[0, \tau]$, $\mathbf{h}_\beta$ y $\mathbf{h}_\gamma$ son vectores en $\mathbb{R}^q$ y $\mathbb{R}^p$ respectivamente. Sea $\mathbf{h} = (\mathbf{h}_\beta, \mathbf{h}_\gamma, h_\Lambda)$.

Porque el estimador de máxima verosimilitud $\hat{\theta}_n = (\hat{\beta}_n, \hat{\gamma}_n, \hat{\Lambda}_n)$ para el modelo completo tambien maximiza la verosimilitud bajo cualquier submodelo paramétrico que pasa a través de $\hat{\theta}_n$, este debería satisfacer la función de score la cual es obtenida diferenciando $l_n(\widehat{\theta_{n,\epsilon}})$ respecto a ϵ y evaluandola en $\epsilon = 0$. Esto es,

$$S_n(\hat{\theta}_n)(\mathbf{h}) = \left.\frac{\partial l_n(\widehat{\theta_{n,\epsilon}})}{\partial \epsilon}\right|_{\epsilon=0} = 0 \tag{6.20}$$

para cada $\mathbf{h}$. Defínase

$$\begin{aligned}
S_\gamma(\theta) &= \Delta \mathbf{X}\,(1 - \pi(\gamma'\mathbf{X})) \\
&\quad - \frac{(1-\Delta)\mathbf{X}\pi(\gamma'\mathbf{X})(1-\pi(\gamma'\mathbf{X}))\Psi\left(\int_0^Y e^{\beta'\mathbf{Z}(s)} d\Lambda(s)\right)}{1 - \pi(\gamma'\mathbf{X}) + \pi(\gamma'\mathbf{X})\left(1 - \Psi\left(\int_0^Y e^{\beta'\mathbf{Z}(s)} d\Lambda(s)\right)\right)}, \\
S_\beta(\theta) &= \Delta \mathbf{Z}(Y) - \phi(\mathbf{O};\theta) \int_0^Y e^{\beta'\mathbf{Z}(u)} \mathbf{Z}(u) d\Lambda(u), \\
S_\Lambda(\theta)(h_\Lambda) &= \Delta h_\Lambda(Y) - \phi(\mathbf{O};\theta) \int_0^Y e^{\beta'\mathbf{Z}(u)} h_\Lambda(u) d\Lambda(u).
\end{aligned}$$

Entonces, el operador de score $S_n(\hat{\theta}_n)(\mathbf{h})$ tiene la forma

$$S_n(\hat{\theta}_n)(\mathbf{h}) = \mathbb{P}_n\left[\mathbf{h}'_\beta S_\beta(\hat{\theta}_n) + \mathbf{h}'_\gamma S_\gamma(\hat{\theta}_n) + S_\Lambda(\hat{\theta}_n)(h_\Lambda)\right]. \tag{6.21}$$

Considérese el espacio de elementos de $\mathbf{h}$ como

$$\mathcal{H} = \left\{\mathbf{h} = (\mathbf{h}_\beta, \mathbf{h}_\gamma, h_\Lambda) : \mathbf{h}_\beta \in \mathbb{R}^q; \mathbf{h}_\gamma \in \mathbb{R}^p; h_{\Lambda_j} \in VB([0,\tau])\right\},$$

donde VB denota la variación acotada sobre $[0,\tau]$. Más aún, se toma la función h_Λ continua por la derecha en 0. Se define la siguiente norma sobre $\mathcal{H}$: si $\mathbf{h} \in \mathcal{H}$,

$$||\mathbf{h}||_H = ||\mathbf{h}_\beta|| + ||\mathbf{h}_\gamma|| + ||h_\Lambda||_v,$$

donde $||\cdot||$ es la norma Euclideana y $||h_\Lambda||_v$ denota la variación total de h_Λ sobre $[0,\tau]$. Defínase tambien $\mathcal{H}_r = \{\mathbf{h} \in \mathcal{H}, ||\mathbf{h}||_{\mathcal{H}} \leq r\}$ y $\mathcal{H}_\infty = \{\mathbf{h} \in \mathcal{H}, ||\mathbf{h}||_{\mathcal{H}} < \infty\}$.

Obsérvese que el parámetro puede ser escrito de la siguiente manera,

$$\theta(\mathbf{h}) = \mathbf{h}'_\beta \beta + \mathbf{h}'_\gamma \gamma + \int_0^\tau h_\Lambda(s) d\Lambda(s),$$

donde $\mathbf{h} \in \mathcal{H}$. A partir de esta notación, se puede reconsiderar el parámetro θ como un funcional lineal en $\mathcal{H}_r$ y al espacio de parámetros Θ como un subconjunto de $l^\infty(\mathcal{H}_r)$ el cual es el espacio de funciones real-valuadas en $\mathcal{H}_r$.

Más aún, el operador de score S_n es un mapeo aleatorio de Θ al espacio $l^\infty(\mathcal{H}_r)$.

Observación. Escogiendo apropiadamente $\mathbf{h}$ se pueden obtener los componentes originales del parámetro θ; se denotará por $\mathbf{0}_r$ $(r \geq 2)$ al vector columna de dimensión r que tiene todas sus entadas iguales a 0. Por ejemplo, considerando $\mathbf{h}_\gamma = \mathbf{0}_p$, $h_\Lambda(\cdot) = 0$, y $\mathbf{h}_\beta$ como el vector de dimensión q cuya entrada i-ésima es igual a 1 y cero en las otras. Entonces se puede obtener el componente i- ésimo de β. Otro ejemplo es, que si se considera $\mathbf{h}_\beta = \mathbf{0}_q$, $\mathbf{h}_\gamma = \mathbf{0}_p$ y $h_\Lambda(\cdot) = 1\{\cdot \leq t\}$, para algún $t \in (0, \tau)$. En este caso, $\theta(\mathbf{h})$ se reduce a $\Lambda(t)$.

Similar a los modelos paramétricos, se define la información de Fisher por

$$I(\theta_0)(\mathbf{h}) = P_{\theta_0}[S_1(\theta_0)(\mathbf{h})^2]$$

donde S_1 es el operador de score (6.21) basado en una observación. Una expresión explícita de $I(\theta_0)(\mathbf{h})$ es dada por el siguiente lema.

Lema 6.8.1. *Sea* $\mathbf{h} \in \mathcal{H}_r$*. Entonces* $P_{\theta_0}[S_1(\theta_0)(\mathbf{h})] = 0$ *y el operador de Fisher es representado por*

$$I(\theta_0)(\mathbf{h}) = \mathbf{h}_\beta'\sigma_\beta(\mathbf{h}) + \mathbf{h}_\gamma'\sigma_\gamma(\mathbf{h}) + \int_0^\tau \sigma_\Lambda(\mathbf{h})(s)h_\Lambda(s)d\Lambda_0(s), \tag{6.22}$$

donde

$$\begin{aligned}
\sigma_\beta(\mathbf{h}) &= P_{\theta_0}\left[2S_\beta(\theta_0)\Delta h_\Lambda(Y)\right] + P_{\theta_0}\left[S_\beta(\theta_0)^{\otimes 2}\right]\mathbf{h}_\beta + P_{\theta_0}\left[S_\beta(\theta_0)S_\gamma(\theta_0)'\right]\mathbf{h}_\gamma\\
\sigma_\gamma(\mathbf{h}) &= P_{\theta_0}\left[2S_\gamma(\theta_0)\Delta h_\Lambda(Y)\right] + P_{\theta_0}\left[S_\gamma(\theta_0)^{\otimes 2}\right]\mathbf{h}_\gamma + P_{\theta_0}\left[S_\gamma(\theta_0)S_\beta(\theta_0)'\right]\mathbf{h}_\beta\\
\sigma_\Lambda(\mathbf{h})(s) &= -P_{\theta_0}\left[\Delta h_\Lambda(Y)1\{s \leq Y\}\phi(\mathbf{O};\theta_0)e^{\beta_0'\mathbf{Z}(s)}\right]\\
&\quad +P_{\theta_0}\left[\{\phi(\mathbf{O};\theta_0)\}^2\left\{\int_0^Y e^{\beta_0'\mathbf{Z}(u)}h_\Lambda(u)d\Lambda_0(u)\right\}1\{s \leq Y\}e^{\beta_0'\mathbf{Z}(s)}\right]\\
&\quad -\mathbf{h}_\beta' P_{\theta_0}\left[2S_\beta(\theta_0)1\{s \leq Y\}\phi(\mathbf{O};\theta_0)e^{\beta_0'\mathbf{Z}(s)}\right]\\
&\quad -\mathbf{h}_\mathbf{G}' P_{\theta_0}\left[2S_\gamma(\theta_0)1\{s \leq Y\}\phi(\mathbf{O};\theta_0)e^{\beta_0'\mathbf{Z}(s)}\right].
\end{aligned}$$

Demostración: Se demostrará primero que $P_{\theta_0}[S_1(\theta_0)(\mathbf{h})] = 0$. Obsévese que

$$\begin{aligned}
P_{\theta_0}\left[S_\beta(\theta_0)\right] &= P_{\theta_0}\left[\Delta\mathbf{Z}(Y) - \phi(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}\mathbf{Z}(u)d\Lambda(u)\right]\\
&= P_{\theta_0}\left[\int_0^\tau \mathbf{Z}(u)dN(u) - \phi(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}\mathbf{Z}(u)d\Lambda(u)\right]\\
&= P_{\theta_0}\left[\int_0^\tau \mathbf{Z}(u)dM(u)\right]
\end{aligned}$$

donde,

$$M(t) = N(t) - \int_0^t 1\{Y \geq u\}\phi(\mathbf{O};\theta)e^{\beta'\mathbf{Z}(u)}d\Lambda(u)$$

la cual es una martingala del proceso contable respecto a la filtración $\sigma\{N(s), 1\{Y \leq s, \Delta = 1\}, \mathbf{X}, \mathbf{Z}(s) : 0 \leq s \leq t\}$. Obsérvese que se ha obtenido un proceso en Y, el cual es una integral estocástica martingala previsto la covariable independiente del tiempo, este proceso es previsible (para verificar esto es suficiente verificar que sea acotado, lo cual es obtenido por la condición **C2**). Entonces $P_{\theta_0}\left[S_\beta(\theta_0)\right] = 0$.

Argumentos similares implican que $P_{\theta_0}[S_\Lambda(\theta_0)] = 0$ y $P_{\theta_0}\left[S_\gamma(\theta_0)\right] = 0$. Concluyendo así la primera parte de la demostración.

Para probar el segundo resultado, es suficiente desarrollar

$$S_1(\theta_0)(\mathbf{h})^2 = \left[\mathbf{h}'_\beta S_\beta(\theta_0) + \mathbf{h}'_\gamma S_\gamma(\theta_0) + S_\Lambda(\theta_0)(h_\Lambda)\right]^2,$$

tomando en cuenta la siguiente relación,

$$P_{\theta_0}[\Delta h_\Lambda(Y)] = P_{\theta_0}\left[\phi(\mathbf{O};\theta)\int_0^Y e^{\beta_0'\mathbf{Z}(u)} h_\Lambda(u) d\Lambda_0(u)\right],$$

la cual es obtenida utilizando el hecho de que $P_{\theta_0}[S_\Lambda(\theta_0)(h_\Lambda)] = 0$ para cualquier función acotada h_Λ.

Algunas manipualciones algebraicas, agrupación de términos, y la aplicación de la esperanza de la expresión resultante, se obtiene el resultado (ver demostración de lema 5.8.1)

□

Lema 6.8.2. *El operador $\sigma = (\sigma_\beta, \sigma_\gamma, \sigma_\Lambda) : \mathcal{H}_r \to \mathcal{H}_r$ es uno a uno.*

Demostración: Supóngase que $\sigma(\mathbf{h}) = 0$. Por el lema 6.8.1, $P_{\theta_0}\left[S_1(\theta_0)(\mathbf{h})^2\right] = 0$. Se sigue que $S_1(\theta_0)(\mathbf{h}) = 0$ casi seguramente.

Tomando $\Delta = 1$ y $\Delta = 0$, se obtienen 2 ecuaciones. Considerando la diferencia entre estas dos ecuaciones y realizando algunas simplificaciones algebraicas y reordenando términos se obtiene la siguiente ecuación:

$$\mathbf{h}'_\gamma\left[\mathbf{X}\left(1-\pi(\gamma_0'\mathbf{X})\right) + \frac{\mathbf{X}\pi(\gamma_0'\mathbf{X})(1-\pi(\gamma_0'\mathbf{X}))\Psi\left(\int_0^Y e^{\beta_0'\mathbf{Z}(s)} d\Lambda_0(s)\right)}{1-\pi(\gamma_0'\mathbf{X}) + \pi(\gamma_0'\mathbf{X})\left(1-\Psi\left(\int_0^Y e^{\beta_0'\mathbf{Z}(s)} d\Lambda_0(s)\right)\right)}\right]$$
$$+\mathbf{h}'_\beta\mathbf{Z}(Y) + h_\Lambda(Y) + \Gamma(\theta_0)\int_0^Y e^{\beta_0'\mathbf{Z}(s)}\left\{\mathbf{h}'_\beta\mathbf{Z}(s) + h_\Lambda(s)\right\} d\Lambda_0(s) = 0, \tag{6.23}$$

donde

$$\Gamma(\theta_0) = \left(\frac{\pi(\gamma_0'\mathbf{X})\Psi^{(1)}\left(\int_0^Y e^{\beta_0'\mathbf{Z}(s)} d\Lambda_0(s)\right)}{1-\pi(\gamma_0'\mathbf{X}) + \pi(\gamma_0'\mathbf{X})\left(1-\Psi\left(\int_0^Y e^{\beta_0'\mathbf{Z}(s)} d\Lambda_0(s)\right)\right)} + \frac{\Psi^{(2)}\left(\int_0^Y e^{\beta_0'\mathbf{Z}(s)} d\Lambda_0(s)\right)}{\Psi^{(1)}\left(\int_0^Y e^{\beta_0'\mathbf{Z}(s)} d\Lambda_0(s)\right)}\right).$$

Escogiendo Y arbitrariamente cercano a 0, la ecuación (6.23) se reduce a

$$\mathbf{h}_\gamma'\mathbf{X}\,(1-\pi(\gamma_0'\mathbf{X}))+\mathbf{h}_\beta'\mathbf{Z}(0)+h_\Lambda(0)=0,$$

donde h_Λ y Λ_0 son continuas por la derecha en 0 y $\Lambda_0(0)=0$. De manera equivalente, la igualdad anterior puede ser reescrita como,

$$\mathbf{h}_\gamma'\mathbf{X}\,(1-\pi(\gamma_0'\mathbf{X}))=-\mathbf{h}_\beta'\mathbf{Z}(0)-h_\Lambda(0),$$

obsérvese que el lado derecho de esta ecuación no depende de $\mathbf{X}$, entonces $\mathbf{h}_\gamma$ debería ser igual 0. Por lo tanto la ecuación (6.23) se reduce a

$$\mathbf{h}_\beta'\mathbf{Z}(Y)+h_\Lambda(Y)+\Gamma(\theta_0)\int_0^Y e^{\beta_0'\mathbf{Z}(s)}\left\{\mathbf{h}_\beta'\mathbf{Z}(s)+h_\Lambda(s)\right\}d\Lambda_0(s)=0. \tag{6.24}$$

Obteniendo así una ecuación homogénea para $\mathbf{h}_\beta'\mathbf{Z}(t)+h_\Lambda(t)$ (es decir, no existe término constante aislado en la ecuación) la cual tiene como soluciones sólo la solución trivial (véase Zeng y Lin, 2007 y el reporte técnico de Zeng y Lin, 2006). Entonces, $\mathbf{h}_\beta'\mathbf{Z}(t)+h_\Lambda(t)=0$ para todo $t\in[0,\tau]$. Por la condición **C7**, se sigue que $\mathbf{h}_\beta=0$ y $h_\Lambda=0$. Esto concluye la demostración.

□

Lema 6.8.3. *El operador σ es continuamente invertible con inversa σ^{-1} denotada como $\sigma^{-1}=(\sigma_\beta^{-1},\sigma_\gamma^{-1},\sigma_\Lambda^{-1})$.*

Demostración: Dado que $\mathcal{H}_r$ es un espacio de Banach, para mostrar que σ es continuamente invertible, es suficiente probar que σ es uno a uno y que puede ser reescrito como una suma $A+(\sigma-A)$ de un operador lineal acotado A con inversa acotada y un operador compracto $\sigma-A$ (lema 25.93 de van der Vaart, 1998).

σ es uno a uno por el lema 6.8.2. Defínase el operador lineal, $A:\mathcal{H}_r\to\mathcal{H}_r$ por $A(\mathbf{h})=(\mathbf{h}_\beta,\mathbf{h}_\gamma,h_\Lambda(\cdot)P_{\theta_0}[W(\cdot,\mathbf{O};\theta_0)])$. A es acotado (por la condiciones **C1** y **C2**). Además, el término $P_{\theta_0}[W(\cdot,\mathbf{O},\theta_0)]$ es mayor que 0 en el intervalo $[0,\tau]$. Entonces A es invertible con inversa acotada $A^{-1}(\mathbf{h})=(\mathbf{h}_\beta,\mathbf{h}_\gamma,h_\Lambda(\cdot)P_{\theta_0}[W(\cdot,\mathbf{O},\theta_0)]^{-1})$.

Demostrar que el operador $\sigma-A$ es compacto es obtenido usando el enfoque de Lu (2008). Como un operador lineal acotado con dimensión finita es compacto, entonces solo es necesario mostrar que el operador $K_\Lambda: VB(0,\tau)\to VB(0,\tau)$, dado por

$$K_\Lambda(h_\Lambda)(s)=-P_{\theta_0}\left[\Delta h_\Lambda(Y)1\{s\le Y\}\phi(\mathbf{O};\theta_0)e^{\beta_0'\mathbf{Z}(s)}\right]$$
$$+P_{\theta_0}\left[\{\phi(\mathbf{O};\theta_0)\}^2\left\{\int_0^Y e^{\beta_0'\mathbf{Z}(u)}h_{\Lambda(u)}d\Lambda_0(u)\right\}1\{s\le Y\}e^{\beta_0'\mathbf{Z}(s)}\right]$$

es compacto. Es decir, dada una sucesión de funciones $h_{\Lambda,n}$ con $||h_{\Lambda,n}||_v \leq 1$, se debe demostrar que existe una subsucesión y un elemento $g \in VB(0,\tau)$ tal que $||K_\Lambda h_{\Lambda,\eta(n)} - g||_v \to 0$.

Como K_Λ es un operador lineal acotado entonces, $||K_\Lambda h_\Lambda||_v \leq M_7 \int |h_\Lambda(u)| d\Lambda_0(u)$ para cada h_Λ y una constante fija M_7. Entonces es suficiente mostrar que existe una subsucesión $h_{\Lambda,\eta(n)}$ de $h_{\Lambda,n}$ que converge. Como h_Λ es de variación acotada, se puede escribir $h_{\Lambda,n}$ como una diferencia de funciones acotadas crecientes $h^{(1)}_{\Lambda,n}$ y $h^{(2)}_{\Lambda,n}$. Por el teorema de Helly, existe una subsucesión $h^{(1)}_{\Lambda,\eta(n)}$ de $h^{(1)}_{\Lambda,n}$ la cual converge puntualmente a $h^{(1)*}_{\Lambda}$, de la misma manera existe una subsucesión $h^{(2)}_{\Lambda,\eta(n)}$ de $h^{(2)}_{\Lambda,n}$ la cual converge puntualmente a $h^{(2)*}_{\Lambda}$. Entonces $h_{\Lambda,n}$ converge a la diferencia de los límites por el teorema de convergencia dominada se sigue $\sigma - A$ es un operador compacto.

□

6.8.2. Resultados de normalidad asintótica

Teorema 6.8.1. *Bajo las condiciones **C1-C7**,*

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{d} N(0, \Sigma_\beta) \qquad y \qquad \sqrt{n}(\hat{\gamma}_n - \gamma_0) \xrightarrow{d} N(0, \Sigma_\gamma),$$

donde

$$\Sigma_\beta = \left(\sigma_\beta^{-1}(\mathbf{e}_1, \mathbf{0}_p, 0), \ldots, \sigma_\beta^{-1}(\mathbf{e}_q, \mathbf{0}_p, 0)\right)$$

y

$$\Sigma_\gamma = \left(\sigma_\gamma^{-1}(\mathbf{0}_q, \mathbf{d}_1, 0), \ldots, \sigma_\gamma^{-1}(\mathbf{0}_q, \mathbf{d}_p, 0)\right)$$

son las varianzas eficientes para la estimación de β_0 y γ_0. $\mathbf{e}_i$ y $\mathbf{d}_i$ son vectores de dimensión q y p respectivamente, donde la i-ésima componente es igual a 1 y cero en las demás. Mas aún, para cualquier $t \in [0,\tau]$,

$$\sqrt{n}\left(\hat{\Lambda}_n(t) - \Lambda_0(t)\right) \xrightarrow{d} N(0, \upsilon^2(t)),$$

donde $\upsilon^2(t) = \int_0^t \sigma_\Lambda^{-1}\left(\mathbf{0}_q, \mathbf{0}_p, 1\{u \leq t\}\right) d\Lambda_0(u)$.

Para realizar la demostración del Teorema 6.8.1 es necesario el siguiennte lema.

Lema 6.8.4. *Las funciones de score para la estimación de β_0 y γ_0 son*

$$l_\beta = S_\beta(\theta_0) - S_\Lambda(\theta_0)(g^*_\beta) - \Sigma_{23} S_\gamma(\theta_0) \tag{6.25}$$

y

$$l_\gamma = S_\gamma(\theta_0) - S_\Lambda(\theta_0)(g^*_\gamma) - \Sigma_{32} S_\beta(\theta_0), \tag{6.26}$$

respectivamente, donde

$$\Sigma_{23} = -\Sigma_{\beta}^{-1}\begin{pmatrix} \sigma_{\gamma}^{-1}(\mathbf{e}_1, \mathbf{0}_p, 0)' \\ \vdots \\ \sigma_{\gamma}^{-1}(\mathbf{e}_q, \mathbf{0}_p, 0)' \end{pmatrix} \qquad y \qquad \Sigma_{32} = -\Sigma_{\gamma}^{-1}\begin{pmatrix} \sigma_{\beta}^{-1}(\mathbf{0}_q, \mathbf{d}_1, 0)' \\ \vdots \\ \sigma_{\beta}^{-1}(\mathbf{0}_q, \mathbf{d}_p, 0)' \end{pmatrix},$$

y S_Λ es aplicado a las componentes de los vectores

$$g_{\beta}^{*} = -\Sigma_{\beta}^{-1}\begin{pmatrix} \sigma_{\Lambda}^{-1}(\mathbf{e}_1, \mathbf{0}_p, 0)' \\ \vdots \\ \sigma_{\Lambda}^{-1}(\mathbf{e}_q, \mathbf{0}_p, 0)' \end{pmatrix} \qquad y \qquad g_{\gamma}^{*} = -\Sigma_{\gamma}^{-1}\begin{pmatrix} \sigma_{\Lambda}^{-1}(\mathbf{0}_q, \mathbf{d}_1, 0)' \\ \vdots \\ \sigma_{\Lambda}^{-1}(\mathbf{0}_q, \mathbf{d}_p, 0)' \end{pmatrix}.$$

Más aún, las matrices de varianza asintóticas eficientes de $\hat{\beta}_n$ y $\hat{\gamma}_n$ son

$$\left(P_{\theta_0}[l_\beta l_\beta]\right)^{-1} = \Sigma_\beta \qquad y \qquad \left(P_{\theta_0}[l_\gamma l_\gamma]\right)^{-1} = \Sigma_\gamma,$$

respectivamente.

Demostración: Se demostrará que l_β es ortogonal a la funcion de score $S_\Lambda(\theta_0)(g)$ para cualquier función acotada g. Considérese el término $\mathbf{e}_i'\Sigma_\beta\mathbb{P}_{\theta_0}[l_\beta S_\Lambda(\theta_0)(g)]$, el cual es igual a

$$P_{\theta_0}\left[\left(e_i'\Sigma_\beta S_\beta(\theta_0) - S_\Lambda(\theta_0)(e_i'\Sigma_\beta g_\beta^*) - e_i'\Sigma_\beta\Sigma_{23}S_\gamma(\theta_0)\right)S_\Lambda(\theta_0)(g)\right] \tag{6.27}$$

Como $e_i'\Sigma_\beta = \sigma_\beta^{-1}(\mathbf{e}_i, \mathbf{0}_p, 0)'$, entonces la ecuación (6.27) es igual a

$$P_{\theta_0}\left(\left[\sigma_\beta^{-1}(\mathbf{e}_i, \mathbf{0}_p, 0)'S_{1\beta}(\theta_0) + S_\Lambda(\theta_0)(\sigma_\Lambda^{-1}(\mathbf{e}_i, \mathbf{0}_p, 0)) + \sigma_\gamma^{-1}(\mathbf{e}_i, \mathbf{0}_p, 0)'S_\gamma(\theta_0)\right)S_\Lambda(\theta_0)(g)\right]. \tag{6.28}$$

Como

$$P_{\theta_0}[\Delta h_\Lambda(Y)] = P_{\theta_0}\left[\phi(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}h_\Lambda(u)d\Lambda(u)\right]. \tag{6.29}$$

Desarrollando los términos en la ecuación (6.28), se obtiene que $S_\beta(\theta_0)S_\Lambda(\theta_0)(g)$ es igual a

$$\Delta\mathbf{Z}(Y)\Delta g_\Lambda(Y) - \Delta\mathbf{Z}(Y)\phi(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}g_\Lambda(u)d\Lambda(u) - \Delta g_\Lambda(Y)\phi(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}\mathbf{Z}(u)d\Lambda(u)$$
$$+\phi^2(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}\mathbf{Z}(u)d\Lambda(u)\int_0^Y e^{\beta'\mathbf{Z}(u)}g_\Lambda(u)d\Lambda(u).$$

Utilizando la ecuación (6.29) lo anterior es equivalente a

$$\Delta\mathbf{Z}(Y)\phi(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}g_\Lambda(u)d\Lambda(u) - \Delta\mathbf{Z}(Y)\phi(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}g_\Lambda(u)d\Lambda(u)$$
$$-\Delta g_\Lambda(Y)\phi(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}\mathbf{Z}(u)d\Lambda(u) + \phi^2(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}\mathbf{Z}(u)d\Lambda(u)\int_0^Y e^{\beta'\mathbf{Z}(u)}g_\Lambda(u)d\Lambda(u)$$

reduciendo términos esto es igual a $\int g\sigma_\Lambda(\mathbf{e}_i, \mathbf{0}_p, 0)$. De manera similar, el término

$$S_\Lambda(\theta_0)(\sigma_\Lambda^{-1}(\mathbf{e}_i, \mathbf{0}_p, 0))S_\Lambda(\theta_0)(g)$$

es igual a

$$\begin{aligned}&\Delta\sigma_\Lambda^{-1}(\mathbf{e}_i, \mathbf{0}_p, 0)\Delta g(Y) - \Delta\sigma_\Lambda^{-1}(\mathbf{e}_i, \mathbf{0}_p, 0)\phi(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}g(u)d\Lambda(u)\\ -&\Delta g(Y)\phi(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}\sigma_\Lambda^{-1}(\mathbf{e}_i, \mathbf{0}_p, 0)d\Lambda(u) + \phi^2(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}\sigma_\Lambda^{-1}(\mathbf{e}_i, \mathbf{0}_p, 0)d\Lambda(u).\end{aligned}$$

Utilizando la ecuación (6.29) lo anterior es equivalente a

$$\begin{aligned}&\Delta\sigma_\Lambda^{-1}(\mathbf{e}_i, \mathbf{0}_p, 0)\phi(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}g(u)d\Lambda(u) - \Delta\sigma_\Lambda^{-1}(\mathbf{e}_i, \mathbf{0}_p, 0)\phi(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}g(u)d\Lambda(u)\\ -&\Delta g(Y)\phi(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}\sigma_\Lambda^{-1}(\mathbf{e}_i, \mathbf{0}_p, 0)d\Lambda(u) + \phi^2(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}\sigma_\Lambda^{-1}(\mathbf{e}_i, \mathbf{0}_p, 0)d\Lambda(u)\end{aligned}$$

reduciendo términos esto es igual a $\int g\sigma_\Lambda(\sigma^{-1}(\mathbf{0}_q, \mathbf{e}_i, 0))$. Finalmente, el término

$$S_\gamma(\theta_0)S_\Lambda(\theta_0)(g)$$

es igual

$$\begin{aligned}&\Delta\mathbf{X}\left(1-\pi(\gamma'\mathbf{X})\right)\Delta g(Y) - \Delta\mathbf{X}\left(1-\pi(\gamma'\mathbf{X})\right)\phi(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}g d\Lambda(u)\\ &-\frac{\Delta g(Y)(1-\Delta)\mathbf{X}\pi(\gamma'\mathbf{X})(1-\pi(\gamma'\mathbf{X}))\Psi\left(\int_0^Y e^{\beta'\mathbf{Z}(s)}d\Lambda(s)\right)}{1-\pi(\gamma'\mathbf{X})+\pi(\gamma'\mathbf{X})\left(1-\Psi\left(\int_0^Y e^{\beta'\mathbf{Z}(s)}d\Lambda(s)\right)\right)}\\ &+\phi(\mathbf{O};\theta)\int_0^Y e^{\beta'\mathbf{Z}(u)}g(u)d\Lambda(u)\frac{\Delta g(Y)(1-\Delta)\mathbf{X}\pi(\gamma'\mathbf{X})(1-\pi(\gamma'\mathbf{X}))\Psi\left(\int_0^Y e^{\beta'\mathbf{Z}(s)}d\Lambda(s)\right)}{1-\pi(\gamma'\mathbf{X})+\pi(\gamma'\mathbf{X})\left(1-\Psi\left(\int_0^Y e^{\beta'\mathbf{Z}(s)}d\Lambda(s)\right)\right)}.\end{aligned}$$

Utilizando la ecuación (6.29) y reduciendo términos lo anterior es equivalente a $\int g\sigma_\Lambda(\mathbf{0}_q, \mathbf{e}_i, 0)$. De los resultados obtenidos se reduce que $\mathbf{e}_i'\Sigma_\beta\mathbb{P}_{\theta_0}[l_\beta S_\Lambda(\theta_0)(g)] = \int g\sigma_\Lambda(\sigma^{-1}(\mathbf{e}_i, \mathbf{0}_p, 0))d\Lambda_0 = 0$. Por lo tanto, por el teorema 3.4.1 de Bickel *et al.* (1993) se sigue que l_β es una función de score eficiente para la estimación de β. Que l_γ es una función de score eficiente para la estimación de γ es probado de la misma manera.

Por lo tanto,

$$\mathbf{e}_i\mathbb{P}_{\theta_0}[\Sigma_\beta l_\beta l_\beta'\Sigma_\beta]\mathbf{e}_j = \mathbb{P}_{\theta_0}[\sigma_\beta^{-1}(\mathbf{e}_j, \mathbf{0}_p, 0)'S_\beta(\theta_0)S_\beta'(\theta_0)\sigma_\beta^{-1}(\mathbf{e}_i, \mathbf{0}_p, 0)'] = \mathbf{e}_i'\Sigma_\beta\mathbf{e}_j$$

pata todo $i, j = 1, \ldots, q$, donde la segunda igualdad es obtenida aplicando el lema 6.8.1. Entonces, $\mathbb{P}_{\theta_0}[\Sigma_\beta l_\beta l'_\beta \Sigma_\beta] = \Sigma_\beta$, lo cual implica que $(\mathbb{P}_{\theta_0}[l_\beta l'_\beta])^{-1} = \Sigma_\beta$. Similarmente se puede obtener que $(\mathbb{P}_{\theta_0}[l_\gamma l'_\gamma])^{-1} = \Sigma_\gamma$.

□

Demostración del Teorema 6.8.1: La demostración sigue las ideas desarrolladas en el Teorema 5.8.1. De la misma manera se demuestra que

$$\sqrt{n}\left(\mathbf{h}'_\beta\left(\widehat{\beta}_n - \beta_0\right) + \mathbf{h}'_\gamma\left(\widehat{\gamma}_n - \gamma_0\right) + \int_0^\tau h_\Lambda(s) d\left(\widehat{\Lambda}_n - \Lambda_0\right)(s)\right) =$$
$$\sqrt{n}\left(S_n(\theta_0)(\sigma^{-1}(\mathbf{h})) - P_{\theta_0}\left[S_1(\theta_0)(\sigma^{-1}(\mathbf{h}))\right]\right) + o_p(1) \tag{6.30}$$

uniformemente en $\mathbf{h}$ cuando $n \to \infty$.

Poniendo $\mathbf{h}_\gamma = \mathbf{0}_p$ y h_Λ identicamente igual a 0. La ecuación (6.30) se reduce a

$$\sqrt{n}\mathbf{h}'_\beta\left(\widehat{\beta}_n - \beta_0\right) = \sqrt{n}\left(S_n(\theta_0)(\sigma^{-1}(\mathbf{h}_\beta, \mathbf{0}_p, 0)) - P_{\theta_0}\left[S_1(\theta_0)(\sigma^{-1}(\mathbf{h}_\beta, \mathbf{0}_p, 0))\right]\right) + o_p(1),$$

Por el teorema del límite central, $\sqrt{n}\mathbf{h}'_\beta(\hat{\beta}_n - \beta_0)$ es asintóticamente normal con media 0 y varianza

$$P_{\theta_0}[S(\sigma^{-1}(\mathbf{h}_\beta, \mathbf{0}_p, 0))]^2 = [(\sigma_\beta^{-1}(\mathbf{h}_\beta, \mathbf{0}_p, 0))]'\mathbf{h}_\beta = \mathbf{h}'_\beta \Sigma_\beta \mathbf{h}_\beta, \tag{6.31}$$

para cualquier $\mathbf{h}_\beta$, donde la primera igualdad se sigue de la ecuación 6.22. Entonces, por la Cramer-Wold device (van der Vaart, 1998), $\sqrt{n}(\hat{\beta}_n - \beta_0)$ es asintóticamente normal con varianza Σ_β. Más aún, por el lema 6.8.4, Σ_β es la varianza eficiente.

Que $\sqrt{n}(\hat{\gamma}_n - \gamma_0)$ es asintóticamente normal con media 0 y varianza Σ_γ es probada de la misma forma poniendo $h_\Lambda = 0$ y $\mathbf{h}_\beta = \mathbf{0}_q$ en 6.30.

Por último, escogiendo $\mathbf{h} = (\mathbf{0}_q, \mathbf{0}_p, 1\{u \leq t\})$, se tiene que

$$\sqrt{n}(\widehat{\Lambda}_n(t) - \Lambda_0(t)) = \sqrt{n}\Big(S_n(\theta_0)(\sigma^{-1}(\mathbf{0}_q, \mathbf{0}_p, 1\{u \leq t\}))$$
$$- P_{\theta_0}\left[S_1(\theta_0)(\sigma^{-1}(\mathbf{0}_q, \mathbf{0}_p, 1\{u \leq t\}))\right]\Big) + o_p(1).$$

Lo cual tiene una distribución asntóticamente normal con media 0 y varianza

$$\begin{aligned} \upsilon^2(t) &= P_{\theta_0}[S_1(\sigma^{-1}(\mathbf{0}_q, \mathbf{0}_p, 1\{u \leq t\}))]^2 \\ &= I(\theta_0)(\sigma^{-1}(\mathbf{0}_q, \mathbf{0}_p, 1\{u \leq t\})) \\ &= \int_0^t \sigma_\Lambda^{-1}\left(\mathbf{0}_q, \mathbf{0}_p, 1\{u \leq t\}\right) d\Lambda_0(u). \end{aligned}$$

Esto completa la prueba.

□

6.9. Estimación de la varianza

La varianza asintótica de $\hat{\beta}_n$, $\hat{\gamma}_n$ y $\hat{\Lambda}_n$ involucra invertir el operador lineal σ en un espacio funcional. Como la inversa σ^{-1} no tiene forma cerrada, la estimación de las varianzas asintóticas no es sencillo. Un posible método para la estimación de las varianzas asintóticas de las estimaciones de los parámetros de regresión euclidianos y de $\hat{\Lambda}_n$ es invertiendo una matriz de información observada discreta, lo cual es sugerido por Sy y Taylor (2000). Otro posible enfoque es derivando una variación de la verosimilitud profile (véase Nielsen *et al.*, 1992 y Murphy *et al.*, 1997), lo cual requiere diferenciaciones de la verosimilitud profile. Sin embargo, no es claro como estos métodos pueden dar estimaciones consistentes de las varianzas asintóticas. Aqui se presenta estimaciones consistentes de las varianzas asintóticas tanto para la estimación de los parámetros euclidianos $\hat{\beta}_n$ y $\hat{\gamma}_n$, así como tambien para la estimación del parámetro infinito-dimensional $\hat{\Lambda}_n$. La demostración sigue el enfoque desarrollado por Fang *et al.* (2005).

Sean las matrices $\mathbb{A}_n^\beta$, $\mathbb{A}_n^\gamma$, $\mathbb{B}_n^\beta$ y $\mathbb{B}_n^\gamma$ de dimensión $(q\times q)$, $(q\times p)$, $(p\times q)$ y$(q\times q)$ respectivamente definidas por

$$\begin{aligned}
\mathbb{A}_n^\beta &= \mathbb{P}_n\left[S_\beta(\widehat{\theta_n})^{\otimes 2}\right],\\
\mathbb{B}_n^\gamma &= \mathbb{P}_n\left[S_\gamma(\widehat{\theta_n})^{\otimes 2}\right],\\
\mathbb{A}_n^\gamma &= \mathbb{P}_n\left[S_\beta(\widehat{\theta_n})S_\gamma(\widehat{\theta_n})'\right] = \left(\mathbb{B}_n^\beta\right)'.
\end{aligned}$$

Defínase

$$\mathbb{A}_n^\Lambda = \frac{2}{n}S_\beta(\widehat{\theta_n}).$$

como una matriz de dimensión $(q\times k)$. Similarmente, defínase la matriz particionada

$$\mathbb{B}_n^\Lambda = \frac{2}{n}S_\gamma(\widehat{\theta_n}),$$

de dimensión $(Q\times s_n)$. Sean

$$\begin{aligned}
\mathbb{C}_n^\beta &= -\mathbb{P}_n\left[2S_\beta(\widehat{\theta_n})'\phi(Y,\mathbf{O};\hat{\theta}_n)e^{\hat{\beta}_n'\mathbf{Z}(s)}1\{s\le Y\}\right],\\
\mathbb{C}_n^\gamma &= -\mathbb{P}_n\left[2S_\gamma(\widehat{\theta_n})'\phi(Y,\mathbf{O};\hat{\theta}_n)e^{\hat{\beta}_n'\mathbf{Z}(s)}1\{s\le Y\}\right],
\end{aligned}$$

matrices de dimensiones $(k\times q)$ y $(k\times p)$ respectivamente.

Defínase la matriz $\mathbb{C}_n^\Lambda$ de dimensión $(k\times k)$ donde el (l,m)-ésimo elemento es definido por:

$$\begin{aligned}
\mathbb{C}_n^\Lambda(l,m) = \; & \mathbb{P}_n\left[\left\{\phi(Y,\mathbf{O};\hat{\theta}_n)\right\}^2\widehat{\Delta\Lambda}(Y_m)e^{\hat{\beta}_n'\mathbf{Z}(Y_m)}1\{Y_m\le Y\}\right]\\
& -1\{l=m\}\mathbb{P}_n\left[\phi(Y,\mathbf{O};\hat{\theta}_n)e^{\hat{\beta}_n'\mathbf{Z}(s)}1\{Y_m\le Y\}\right]
\end{aligned}$$

donde, $\widehat{\Delta\Lambda}(t)$ denota los tamaños de los saltos de $\widehat{\Lambda}$ en el tiempo t; que es, $\widehat{\Delta\Lambda}(t) = \widehat{\Lambda}(t) - \widehat{\Lambda}(t-)$. Defínase la matriz particionada

$$\mathbb{D}_n = \begin{pmatrix} \mathbb{A}_n^\beta & \mathbb{A}_n^\gamma & \mathbb{A}_n^\Lambda \\ \mathbb{B}_n^\beta & \mathbb{B}_n^\gamma & \mathbb{B}_n^\Lambda \\ \mathbb{C}_n^\beta & \mathbb{C}_n^\gamma & \mathbb{C}_n^\Lambda \end{pmatrix}$$

y las matrices

$$\begin{aligned}
\Sigma_{\beta,n} &= \Big\{\mathbb{A}_n^\beta - \mathbb{A}_n^\gamma(\mathbb{B}_n^\gamma)^{-1}\mathbb{B}_n^\beta - (\mathbb{A}_n^\Lambda - \mathbb{A}_n^\gamma(\mathbb{B}_n^\gamma)^{-1}\mathbb{B}_n^\Lambda) \\
&\qquad \times(\mathbb{C}_n^\Lambda - \mathbb{C}_n^\gamma(\mathbb{B}_n^\gamma)^{-1}\mathbb{B}_n^\Lambda)^{-1}(\mathbb{C}_n^\beta - \mathbb{C}_n^\gamma(\mathbb{B}_n^\gamma)^{-1}\mathbb{B}_n^\beta)\Big\}^{-1}, \\
\Sigma_{\gamma,n} &= \Big\{\mathbb{B}_n^\gamma - \mathbb{B}_n^\beta(\mathbb{A}_n^\beta)^{-1}\mathbb{A}_n^\gamma - (\mathbb{B}_n^\Lambda - \mathbb{B}_n^\beta(\mathbb{A}_n^\beta)^{-1}\mathbb{A}_n^\Lambda) \\
&\qquad \times(\mathbb{C}_n^\Lambda - \mathbb{C}_n^\beta(\mathbb{A}_n^\beta)^{-1}\mathbb{A}_n^\Lambda)^{-1}(\mathbb{C}_n^\gamma - \mathbb{C}_n^\beta(\mathbb{A}_n^\beta)^{-1}\mathbb{A}_n^\gamma)\Big\}^{-1}, \\
\Xi_n &= \Big\{\mathbb{C}_n^\Lambda - \mathbb{C}_n^\beta(\mathbb{A}_n^\beta)^{-1}\mathbb{A}_n^\Lambda - (\mathbb{C}_n^\gamma - \mathbb{C}_n^\beta(\mathbb{A}_n^\beta)^{-1}\mathbb{A}_n^\gamma) \\
&\qquad \times(\mathbb{B}_n^\gamma - \mathbb{B}_n^\beta(\mathbb{A}_n^\beta)^{-1}\mathbb{A}_n^\gamma)^{-1}(\mathbb{B}_n^\Lambda - \mathbb{B}_n^\beta(\mathbb{A}_n^\beta)^{-1}\mathbb{A}_n^\Lambda)\Big\}^{-1}.
\end{aligned}$$

También, para cualquier $t \in (0,\tau)$ defínanse los vectores de dimensión k, por

$$\Phi_{t,n} = \left(\widehat{\Delta\Lambda_n}(t_1)1\{t_1 \le t\} \ \ldots \ \widehat{\Delta\Lambda_n}(t_k)1\{t_k \le t\}\right)'$$

y

$$U_{j,t,n} = (1\{t_1 \le t\} \ \ldots \ 1\{t_k \le t\})',$$

Con estas notaciones, se cumple lo siguiente:

Teorema 6.9.1. *Bajo las condiciones **C1-C7**, los estimadores de la varianza $\Sigma_{\beta,n}$, $\Sigma_{\gamma,n}$ y $\upsilon_n^2(t) = \Phi_{t,n}'\Xi_n U_{t,n}$ convergen en probability a Σ_β, Σ_γ y $\upsilon^2(t)$ ($t \in (0,\tau)$) respectivamente cuando $n \to \infty$.*

Demostración del Teorema 6.9.1: La demostración se basa en los argumentos dados en Parner (1998) y Fang *et al.* (2005).

Primero, se estima σ por medio de una versión empírica $\sigma_n = (\sigma_{\beta,n}, \sigma_{\gamma,n}, \sigma_{\Lambda_n})$ obtenida por reemplazar θ_0 y P_{θ_0} por $\widehat{\theta_n}$ y $\mathbb{P}_n$ respectivamente en σ_β, σ_γ y σ_{Λ_j}. Similar a la prueba del Teorema (5.9.1), se puede demostrar que σ_n converge en probabilidad a σ en $\mathcal{H}$ y que su inversa $\sigma_n^{-1} = (\sigma_{\beta,n}^{-1}, \sigma_{\gamma,n}^{-1}, \sigma_{\Lambda_n}^{-1})$ es tal que $\sigma_n^{-1}(\mathbf{h})$ converge a $\sigma^{-1}(\mathbf{h})$ en probabilidad.

De 6.31 para cada $\mathbf{h}_\beta \in \mathbb{R}^q$, la varianza asintótica de $\sqrt{n}\mathbf{h}_\beta'(\widehat{\beta}_n - \beta_0)$ es $[\sigma_\beta^{-1}(\mathbf{h}_\beta, \mathbf{0}_p, 0)]'\mathbf{h}_\beta$. Por la consistencia de $\hat{\theta}_n$, el teorema de convergencia dominada y el Teorema 2.10.6 de van der Vaart y Wellner (1996), se puede mostrar que $\hat{\sigma}_n(\mathbf{h}_\beta, \mathbf{0}_p, 0)$ es un estimador consistente de $\sigma(\mathbf{h}_\beta, \mathbf{0}_p, 0)$. Por lo tanto, $[\hat{\sigma}_{\beta,n}^{-1}(\mathbf{h}_\beta, \mathbf{0}_p, 0)]'\mathbf{h}_\beta$ da un estimador consistente de la varianza asintótica de $\sqrt{n}\mathbf{h}_\beta'(\widehat{\beta}_n - \beta_0)$.

Denótese $\hat{\mathbf{h}}_n = (\hat{\mathbf{h}}_{\beta,n}, \hat{\mathbf{h}}_{\gamma,n}, \hat{h}_{\Lambda,n}) = \sigma_n^{-1}(\mathbf{h}_\beta, \mathbf{0}_p, 0)$. Entonces $\sigma_n(\hat{\mathbf{h}}_n) = (\mathbf{h}_\beta, \mathbf{0}_p, 0)$, lo cual puede ser escrito como,

$$\begin{cases} \sigma_{\beta,n}(\hat{\mathbf{h}}_n) = \mathbf{h}_\beta \\ \sigma_{\gamma,n}(\hat{\mathbf{h}}_n) = \mathbf{0}_p \\ \sigma_{\Lambda_n}(\hat{\mathbf{h}}_n)(u) = 0, \quad \text{for all} \quad u \in [0, \tau]. \end{cases}$$

En particular, sea $u = t_1, \ldots, t_k$, en el sistema anterior. Lo que da un sistema de $(q + p + k)$ ecuaciones:

$$\mathbb{D}_n \begin{pmatrix} \hat{\mathbf{h}}_{\beta,n} \\ \hat{\mathbf{h}}_{\gamma,n} \\ \breve{\mathbf{h}}_{\Lambda,n} \end{pmatrix} = \begin{pmatrix} \mathbf{h}_\beta \\ \mathbf{0}_p \\ 0_{s_n} \end{pmatrix} \tag{6.32}$$

donde $\breve{\mathbf{h}}_{\Lambda,n} = (\hat{h}_{\Lambda_n}(t_1) \ldots \hat{h}_{\Lambda_n}(t_k))'$. Se deduce directamente de algunos cálculos que $\hat{\mathbf{h}}_{\beta,n} = \Sigma_n \mathbf{h}_\beta$, con Σ_n dada arriba y entonces, $\mathbf{h}'_\beta \Sigma_n \mathbf{h}_\beta$ es un estimador consistente de la varianza asintótica de $\sqrt{n}\mathbf{h}'_\beta(\widehat{\beta}_n - \beta_0)$ para cada $\mathbf{h}_\beta$. Se sigue que Σ_n es un estimador consistente de Σ.

La consistencia de Υ_n es probada de la misma manera.

Sea $t \in (0, \tau)$, aplicando el teorema de convergencia dominada y la consistencia de σ_n^{-1}, se sigue que $\upsilon_n^2(t) = \int_0^t \sigma_{\Lambda,n}^{-1}(\mathbf{0}_q, \mathbf{0}_p, 1\{s \leq t\})\, d\widehat{\Lambda}_n(s)$ converge en probabilidad a $\upsilon^2(t)$. Sea $\mathbf{h}_n = (\mathbf{h}_{\beta,n}, \mathbf{h}_{\gamma,n}, h_{\Lambda_n}) = \sigma_n^{-1}(\mathbf{0}_q, \mathbf{0}_p, 1\{s \leq t\})$. Entonces $\sigma_n(\mathbf{h}_n) = (\mathbf{0}_q, \mathbf{0}_p, 1\{s \leq t\})$, lo cual puede ser escrito como

$$\begin{cases} \sigma_{\beta,n}(\mathbf{h}_n) = \mathbf{0}_q \\ \sigma_{\gamma,n}(\mathbf{h}_n) = \mathbf{0}_p \\ \sigma_{\Lambda_n}(\mathbf{h}_n)(u) = 1\{s \leq t\}, \quad \text{for all} \quad u \in [0, \tau] \end{cases} \tag{6.33}$$

En particular, tomando $u = t_1, \ldots, t_k$ en (6.33) se obtiene el siguiente sistema de ecuaciones

$$\mathbb{D}_n \begin{pmatrix} \mathbf{h}_{\beta,n} \\ \mathbf{h}_{\gamma,n} \\ \mathbf{h}_{\Lambda,n} \end{pmatrix} = \begin{pmatrix} \mathbf{0}_q \\ \mathbf{0}_p \\ U_{t,n} \end{pmatrix}$$

donde $\mathbf{h}_{\Lambda,n} = (h_{\Lambda,n}(t_1) \ldots h_{\Lambda,n}(t_k))'$ y $U_{t,n}$ son definidos anteriormente. Resolviendo el sistema de ecuaciones (6.33) se obtiene que $\mathbf{h}_{\Lambda,n} = \Xi_n U_{t,n}$ y entonces

$$\begin{aligned} \upsilon_n^2(t) &= \int_0^t \sigma_{\Lambda,n}^{-1}(\mathbf{0}_q, \mathbf{0}_p, 1\{s \leq t\})\, d\widehat{\Lambda}_n(s) \\ &= \sum_{l=1}^k \sigma_{\Lambda,n}^{-1}(\mathbf{0}_q, \mathbf{0}_p, 1\{t_l \leq t\}) \widehat{\Delta\Lambda_n}(t_l) 1\{t_l \leq t\} \\ &= \Phi'_{t,n} \mathbf{h}_{\Lambda,n} \end{aligned}$$

por lo tanto $\Phi'_{t,n} \Xi_n U_{t,n}$ es un estimador consistente para $\upsilon^2(t)$.

□

Capítulo 7

Conclusiones

Actualmente los modelos de regresión semiparamétricos con aplicaciones a datos de supervivencia con censura a la derecha han ganado popularidad y una abundante literatura ha sido desarrollada para esta clase de modelos. Estas nuevas formulaciones presentan desafíos teóricos por la presencia de parámetros infinito-dimensionales. En este trabajo se han estudiado dos modelos semiparamétricos del análisis de supervivencia: un modelo de mezclas semiparamétrico para riesgos competitivos y un modelo de cura semiparamétrico basado en modelos de transformación. Para el primer modelo se adoptó la especificación propuesta por Escarela y Bowater (2008), mientras que para el segundo modelo se tomó una especificación que generaliza los modelos bien establecidos al utilizar modelos de transformación cuyos estimadores son semparamétricamente eficientes.

La contribución principal de este trabajo es el desarrollo de una teoría general para los NPMLEs de los dos modelos. La teoría aquí presentada puede ser fácilmente empleada para derivar resultados asintóticos para otros modelos semiparamétricos del análisis de supervivencia. Para las dos clases de modelos tratadas aquí se identificó un conjunto de condiciones de regularidad bajo las cuales los estimadores son consistentes, asintóticamente normales y eficientes. Para cada modelo una representación del estimador NPML de la fuerza de mortalidad integrada fue mostrada (ecuaciones 5.7 y 6.9, respectivamente) la cual facilita y simplifica mucho las derivaciones matemáticas de los resultados posteriores. Extendiendo las técnicas desarrolladas por Murphy(1994), empleando clases Donsker y la propiedad Glivenko-Cantelli fueron parte fundamental para obtener la consistencia de los NPMLEs (Teorema 5.7.1 y Teorema 6.7.1, respectivamente). La normalidad asintótica fue demostrada aplicando la metodología desarrollada por Murphy (1995) para modelos frailty, la cual recomienda adoptar una función analítica que permite trabajar con modelos unidimensionales que pasan a través del estimador del modelo, y que a su vez permite ver al espacio de parámetros como un espacio de funciones; esta nueva función analítica permite una manera fácil de definir el operador información del modelo.

A partir de estos resultados es posible obtener la normalidad asintótica para cada modelo semiparamétrico (Teorema 5.8.1, Teorema, 5.8.2 y Teorema 6.8.1), y siguiendo la teoría establecida en Bickel *et al.* (1993) y Tsiatis (2006) se pudo demostrar que la función de eficiencia pertenece al espacio tangente generado por las funciones de score, dando como resultado que los estimadores de los parámetros de regresión sean eficientes. Finalmente, las varianzas son obtenidas invirtiendo un operador lineal en un espacio funcional. Las varianzas consistentes para los parámetros finito e infinito dimensionales para cada modelo fueron obtenidas siguiendo las ideas desarrolladas por Parner (1998),Dupuy y Mesbah (2004) y Fang *et al.* (2005) (Teorema 5.9.1 y Teorema 6.9.1). Como las inferencias expuestas en este trabajo tienen características deseables, las cuales son equivalentes a las del modelo original de Cox, los bioestadísticos y estadísticos en general encontrarán en estas formulaciones una forma conveniente, concisa y precisa para obtener inferencias para riesgos competitivos y modelos de cura.

Aún siendo una disciplina ya madura, el análisis de supervivencia sigue presentando brechas. Algunas cuestiones relacionadas con el modelo semiparamétrico de mezclas para riesgos competitivos y con el modelo semiparamétrico de cura utilizando modelos de transformación siguen abiertas. Aquí se señalan algunas sobresalientes:

- En el caso del modelo de mezclas semiparamétrico para riesgos competitivos es posible notar que la covariable $\mathbf{Z}$ en el modelo (5.2) ha sido tomada como una variable independiente del tiempo, como lo hace Ng y McLachlan (2003) y Escarela y Bowater (2008). Sin embargo, esta suposición puede ser ampliada; es decir, se puede permitir que el modelo (5.2) incluya variables dependientes del tiempo, lo que implicaría que ciertas condiciones de regularidad deben de ser establecidas.

- Sería conveniente ampliar el modelo de tiempo de fallo condicional (5.2) a una clase flexible de modelos, tales como lo modelos de transformación lineal (ver por ejemplo Slud y Vonta, 2004 y el capítulo 6 de este trabajo).

- También sería interesante incorporar estudios de diseños más complejos en la inferencia estadística para los modelos de mezclas semiparamétricos para riesgos competitivos, como son, la incorporación de censura por intervalos o tiempos de datos agrupados.

- En el caso del modelo de cura semiparamétrico basado en modelos de transformación, es importante desarrollar un proceso de estimación empleando el algoritmo EM para calcular los NPMLEs como lo desarrollan Peng y Dear (2000) y Sy y Taylor (2000) para el caso del modelo de cura semiparamétrico de riesgos proporcionales. Esto permitirá realizar estudios de simulación para demostrar la eficiencia del modelo en situaciones prácticas. La extensión del modelo para que permita observaciones censuradas por intervalos es también un problema abierto.

- Establecer pruebas de bondad de ajuste para ambas clases de modelos debe de constituir una dirección importante para trabajos futuros.

Capítulo 8

Apéndice

El algoritmo de Esperanza-Maximización (EM) introducido por Dempster *et al.* (1977) puede ser utilizado para calcular los estimadores noparamétricos. Esta sección explica el algoritmo EM desarrollado por Escarela y Bowater (2008).

8.1. Algoritmo EM

Con el fin de estimar **B** y **G**, es posible modificar el algoritmo EM propuesto por Larson y Dinse (1985) para el caso en el que el modelo de mezclas toma una forma más restringida. Defínase la variable Γ_i^j la cual toma el valor de 1 si el individuo i muere de la causa j y el valor 0 en cualquier otro caso. Obsérvese que si $\Delta_i = 1$, entonces $\Gamma_i^j = \Delta_i^j$, $j \in \mathcal{J}$; sin embargo, si $\Delta_i = 0$, entonces Γ_i^j es indefinida para toda $j \in \mathcal{J}$. Por lo tanto $p_{\mathbf{G},i}^{j,\mathbf{X}}$ y la matriz $\mathcal{N} = [\Gamma_i^j]$, son parcialmente observadas lo que permite usar el algoritmo EM, como lo proponen Dempster *et al.* (1977).

Bajo la suposición de que la matriz $\mathcal{N}$ es completamente observada, la función de verosimilitud dada en la ecuación (5.4) puede ser ahora referida como la función de verosimilitud completa para n individuos, la cual puede ser reescrita como

$$\begin{aligned} L_C &= \prod_{i=1}^{n} \left\{ \left(\prod_{j \in \mathcal{J}} \left[p_{\mathbf{G},i}^{j,\mathbf{X}} f_j(t_i) \right]^{\Gamma_i^j \Delta_i^j} \right) \times \prod_{j=1}^{J} [p_{\mathbf{G},i}^{j,\mathbf{X}} S_j(t_i)]^{(1-\Delta_i)\Gamma_i^j} \right\} \\ &= \prod_{i=1}^{n} \prod_{j=1}^{J} \left(p_{\mathbf{G},i}^{j,\mathbf{X}} \right)^{\Gamma_i^j} \times \prod_{i=1}^{n} \prod_{j=1}^{J} [\lambda_j(t_i)]^{\Delta_i^j \Gamma_i^j} [S_j(t_i)]^{\Gamma_i^j} \qquad (8.1) \\ &= L_p \times L_S . \end{aligned}$$

Paso E

El paso E en el algoritmo EM calcula la esperanza del logaritmo de la ecuación (8.1) dado las estimaciones actuales de $f_j(t_i)$, $S_j(t_i)$ y $p^{j,\mathbf{X}}_{\mathbf{G},i}$, la cual es la suma de las siguientes funciones

$$l_p = \sum_{i=1}^{n} \sum_{j=1}^{J} g_i^j(\theta) \log p^{j,\mathbf{X}}_{\mathbf{G},i}, \tag{8.2}$$

y

$$\begin{aligned} ls &= \sum_{i=1}^{n} \sum_{j=1}^{J} \Delta_i^j \log \lambda_j(t_i; \mathbf{Z}_i) + g_i^j(\theta) \log S_j(t_i; \mathbf{Z}_i) \qquad (8.3)\\ &= \sum_{i=1}^{n} \sum_{j=1}^{J} \Delta_i^j [\log \lambda_j(t_i) + \beta_j' \mathbf{Z}_i] - g_i^j(\theta) \Lambda_j(t) \exp\{\beta_j' \mathbf{Z}_i\} \end{aligned}$$

donde, $g_i^j(\theta)$ es la esperanza de Γ_i^j dadas las estimaciones de $p^{j,\mathbf{X}}_{\mathbf{G},i}$ y $S_j(t_i)$ y cuyo valor es

$$g_i^j(\theta) = \Delta_i^j + (1 - \Delta_i)\omega_i^j(\theta), \tag{8.4}$$

donde $\omega_i^j(\theta) = \mathbb{P}\{\Gamma_i^j = 1 | T > t_i\}$, y tiene la siguiente fórmula:

$$\omega_i^j(\theta) = \frac{p^{j,\mathbf{X}}_{\mathbf{G},i} S_j(t_i)}{\sum_{l=1}^{J} p^{l,\mathbf{X}}_{\mathbf{G},i} S_l(t_i)}.$$

Para el j-ésimo tipo de fallo, sean $t_{j,(1)} < ... < t_{j,(k_j)}$ el ordenamiento de los distintos tiempos de fallo no censurados por causa j. Sea D_{il} es el conjunto de muertes no censuradas por causa j que ocurren al tiempo $t_{j,(l)}$ y R_{jl} el conjunto de individuos en riesgo antes de $t_{j,(l)}$. Finalemente, sea $t_{j,(0)} = 0$ y $t_{j,(k_j+1)} = \infty$. Siguiendo Kalbfleisch y Prentice (1973, 1978) y Holt (1978), la ecuación (8.3) puede ser aproximada por,

$$\log \prod_{j=1}^{J} \prod_{l=1}^{k_j} \frac{\exp\{\beta_j' \mathbf{Z}_{j,(l)}\}}{\sum_{m \in R_{jl}} g_m^j(\theta) \exp\{\beta_j' \mathbf{Z}_m\}} \tag{8.5}$$

donde $\mathbf{Z}_{j,(l)}$ es el vector de covariables correspondientes a la observación, cuyo tiempo de supervivencia observada es $t_{j,(l)}$. Esta aproximación es similar a la verosimilitud parcial de Cox, excepto por la inclusión de los pesos g_{mj}.

La verosimilitud parcial en (8.5) se adapta cuando existen tiempos de fallo empatados cuando los tipos de causa son diferentes, sin embargo, una modificación es requerida si los tiempos de fallo empatados ocurren en la misma causa. Una aproximación es propuesta por Breslow (1974) lo que implica que el ajuste de la función de verosimilitud se convierte en:

$$\log \prod_{j=1}^{J} \prod_{l=1}^{k_j} \frac{\exp\{\beta_j' \mathbf{s}_{j,(l)}\}}{\left[\sum_{m \in R_{jl}} g_m^j(\theta) exp\{\beta_j' \mathbf{Z}_m\}\right]^{d_{jl}}} \tag{8.6}$$

donde $\mathbf{s}_{jl} = \sum_{i \in D_{jl}} \mathbf{Z}_i$ representa la suma de los vectores de covariables para los individuos en D_{jl}. El valor de d_{jl} denota el número de fallos empatados por la causa j al tiempo $t_{j,(l)}$.

Paso M utilizando el estimador producto límite

El paso M involucra la maximización de la ecuación (8.6) con respecto a $\mathbf{B}$ dado el valor de los pesos g_m^j. Para esto, es necesario que las funciones de supervivencia condicional $S_j(t; \mathbf{Z})$ sean estimadas. Con el fin de obtener un estimador adecuado para las funciones de supervivencia condicionales base se puede usar los enfoques desarrollados por Johansen (1983) y Klein (1992) quienes utilizan el enfoque de verosimilitud parcial. Escarela y Bowater (2008) adoptan el estimador noparamétrico producto límite (PME) descrito por Kalflbfleisch y Prentice (1980, p. 85) el cual específica a la función de supervencia base para el j-ésimo riego como,

$$S_j(t) = \prod_{m: t_{j,(m)} \leq t} \alpha_{j,m} \qquad j \in \mathcal{J}, \quad m = 1, ..., k_j, \tag{8.7}$$

donde los $\alpha_{j,m}$ son parámetros no negativos, con $\alpha_{j0} = 1$ y

$$\alpha_{jm} = \frac{S_j(t_{j,(m+1)})}{S_j(t_{j,(m)})}.$$

El modelo de mezclas que resulta de la estructura PLE descrita arriba, es esencialmente el modelo de Kuk (1992) que generaliza el modelo de Larson y Disnse (1985). Sin embargo, el procedimiento de estimación descrito por Kuk, involucra simulaciones de Monte Carlo, lo cual hace computacionalmente caro el modelo. Escarela y Bowater (2008) proponen un método de estimación alternativo que debe de producir inferencias más precisas.

Dada la distribución de supervivencia base expresada en la Eq.(8.7), la correspondiente función de mortalidad para el j-ésimo riesgo es $\lambda_j(t; \mathbf{x}) = 1 - \alpha_{jm}^{\exp(\boldsymbol{\beta}_j' \mathbf{Z})}$ para todo $t_{j,m} \leq t < t_{j,(m+1)}$. Aplicando los valores obtenidos en el paso E y reordenando términos, la ecuación (8.3) es expresada como (Kalbfleisch y Prentice, 1980 pp. 85)

$$log \prod_{j=1}^{J} \prod_{l=1}^{k_j} \left\{ \prod_{i \in D_{jl}} \left(1 - \alpha_{jl}^{\exp(\boldsymbol{\beta}_j' \mathbf{Z}_i)}\right) \prod_{m \in R_{jl} - \bigcup_j D_{jl}} \alpha_{jm}^{g_m^j(\theta) \exp(\boldsymbol{\beta}_j' \mathbf{Z}_m)} \right\}.$$

Entonces, si hay empates, la estimación de máxima verosimilitud de α_{jl} dados los valores de β_j y γ_j es (ver Collet, 1994, pp. 98)

$$\hat{\alpha}_{jl} \approx \exp\left\{ \frac{-d_{jl}}{\sum_{m \in R_{jl}} g_m^j \exp(\boldsymbol{\beta}_j' \mathbf{Z}_m)} \right\}.$$

Sustituyendo este resultado en la Eq.(8.7), se obtiene entonces que el estimador de la función base de supervivencia condicional es expresado como

$$\hat{S}_{j,0}(t) = \exp\left\{ - \sum_{m:t_{j,(m)}<t} \frac{d_{jm}}{\sum_{m \in R_{jm}} g_{mj} \exp(\beta_j' \mathbf{Z}_m)} \right\},$$

donde $d_{j,l}$ denota el número muertes no censuradas por la causa j al tiempo $t_{j,l}$.

Bibliografía

[1] P.K. Andersen and R.D. Gill. Cox's regression model for counting processes: A large sample study. *Ann. Statist.*, 10(4):1100–1120, 1982.

[2] V. Bagdonavicius and M.S. Nikulin. Generalized proportional hazards model based on modified partial likelihood. *Lifetime Data Analysis*, 5(4):329–350, 1999.

[3] V. Bagdonavicius and M.S. Nikulin. *Accelerated Life Models : Modeling and Statistical Analysis*. Chapman & Hall, Boca Ratón (Florida), 2002.

[4] R. Bender, T. Augustin, and M. Blettner. Generating survival times to simulate cox proportional hazards models. *Statist. in Medicine*, 24:1713–1723, 2005.

[5] J. Berkson and R.P. Gage. Survival curve for cancer patients following treatment. *J. Amer. Statis. Assoc.*, 47(259):501–515, 1952.

[6] P. Bickel, C. Klaassen, Y. Ritov, and J. Wellner. *Efficient and adaptive estimation for semiparametric models*. Johns Hopkins University Press, Baltimore, 1993.

[7] J.M. Boag. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *J. R. Statis. Soc. Ser. B*, 11(1):15–53, 1949.

[8] N. Breslow. Covariance analysis of censored survival data. *Biometrics*, 30(1):89–99, 1974.

[9] A.B. Cantor and J.J. Shuster. Parametric versus non-parametric methods for estimating cure rates based on censored survival data. *Statist. in Medicine*, 11(7):931–937, 1992.

[10] J.F. Carrière. Removing cancer when it is correlated with other causes of death. *Biometrical Journal*, 37(3):339–350, 1995.

[11] I.S. Chang, C.A. Hsuing, M.C. Wang, and C.C. Wen. An asymptotic theory for the nonparametric maximum likelihood estimator in the cox gene model. *Bernoulli*, 11(5):863–892, 2005.

[12] K. Chen, Z. Jin, and Z. Ying. Semiparametric analysis of transformation models with censored data. *Biometrika*, 89(3):659–668, 2002.

[13] S.C. Cheng, F.P. Fine, and L.J. Wei. Prediction of cumulative incidence function under the proportional hazards model. *Biometrika*, 82(4):835–845, 1995.

[14] S.C. Cheng, L.J. Wei, and Z. Ying. Analysis of transformation models with censored data. *Biometrics*, 54(1):219–228, 1998.

[15] K.C. Choi and X. Zhou. Large sample properties of mixture models with covariates for competing risks. *J. Multivar. Anal*, 82(2):331–366, 2002.

[16] D. Clayton and J. Cuzick. Multivariate generalizations of the proportional hazards model (with discussion). *J. R. Statis. Soc.. Ser. A*, 148(2):82–117, 1985.

[17] D. Collet. *Modelling survival data in medical research*. Chapman & Hall, London, 1994.

[18] D. R. Cox. Regression models and life-tables (with discussion). *J. Roy. Statist. Soc. Ser. B*, 34:187–220, 1972.

[19] D. R. Cox and D. Oakes. *Analysis of Survival Data*. Chapman & Hall, London, 1984.

[20] D. R. Cox and E.J. Snell. *Analysis of Binary Data*. Chapman & Hall, London, 1989.

[21] J. Cuzick. Rank regression. *Ann. Statist.*, 16(4):1369–1389, 1988.

[22] A.P. Dempster, N.M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *J. R. Stat. Soc. Ser. B*, 39(1):1–38, 1977.

[23] J.-F. Dupuy and G. Escarela. Modélisation de risques concurrents par un modèle de mélange semi-paramétrique. *Comptes Rendus Mathematique*, 34(10):641–644., 2007.

[24] J.-F. Dupuy, I. Grama, and M. Mesbah. Asymptotic theory for the Cox model with missing time-dependent covariate. *Ann. Statist.*, 34(2):903–924, 2006.

[25] J.-F. Dupuy and M. Mesbah. Estimation of the asymptotic variance of semiparametric maximum likelihood estimators in the Cox model with a missing time-dependent covariate. *Comm. Statist. Theory Methods*, 33(6):1385–1401, 2004.

[26] B. Efron. The efficiency of cox's likelihood function for censored data. *J. Amer. Stat. Assoc.*, 72(359):557–565, 1977.

[27] R.C. Eldant-Johnson and N.L. Johnson. *Survival models and data analysis*. Wiley, New York, 1980.

[28] G. Escarela and R. Bowater. Fitting a semi-parametric mixture model for competing risks in survival data. *Comm. Statist. Theory Methods*, 37(2), 2008.

[29] H.-B. Fang, G. Li, and J. Sun. Maximum likelihood estimation in a semiparametric logistic/proportional-hazards mixture model. *Scand. J. Statist.*, 32(1):59–75, 2005.

[30] V.T. Farewell. The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, 38(4):1041–1046, 1982.

[31] V.T. Farewell. Mixture models in survival analysis: are they worth the risk? *Can. J. Statist.*, 14(3):257–262, 1986.

[32] J.P Fine. Analysing competing risks data with transformation models. *J. R. Stat. Soc. Ser. B*, 61(4):817–830, 1999.

[33] T. R. Fleming and D. P. Harrington. *Counting processes and survival analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons, 1991.

[34] J.J. Gaynor, E.J. Feuer, C.C Tan, D.H. Wu, C.R. Little, D.J. Straus, B.D. Clarkson, and M.F. Brennan. On the use of cause-specific failure and conditional failure probabilities: Examples from clinical oncology data. *J. Amer. Statist. Assoc.*, 88(2):218–241, 1993.

[35] M.E. Ghitany, R. A. Maller, and S. Zhou. Exponential mixture models with long-term survivors and covariates. *J. Multivar. Anal*, 49(422):400–409, 1994.

[36] B. Grun and F. Leisch. Identifiability of finite mixtures of multinomial logit models with varying and fixed effects. *Technical Report, Department of Statistics,University of Munich*, (024):329–350, 2008.

[37] A. Hernández Quintero, J.-F. Dupuy, and G. Escarela. Analysis of a semiparametric mixture model for competing risks. *Ann. Inst. Statis. Math.*, 2009.

[38] J. D. Holt. Competing risk analyses with special reference to matched pair experiments. *Biometrika*, 65(1):159–165, 1978.

[39] C. Huber and J.P. Lecoutre. Estimation fonctionnelle dans les modèles de durèe. *Analyse statistique des durèes de vie*, pages 59–120, 1989.

[40] S. Johansen. An extension of cox's regression model. *Int. Stat. Rev.*, 51(2):165–174, 1983.

[41] N.L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions.Volume 1*. Wiley Series in Probability and Statistics, New York, 1994.

[42] D.R. Jones, R.L. Powles, D. Machin, and Sylvester R.J. On estimating the proportion of cured patients in clinical studies. *Biometrie-Praximetrie*, 21:1–11, 1981.

[43] J. D. Kalbfleisch and R. L. Prentice. Marginal likelihoods based on cox's regression and life model. *Biometrika*, 60(2):267–278, 1973.

[44] J.D. Kalbfleisch and R.L. Prentice. *The Statistical Analysis of Failure Time Data*. Willey, New York, 1980.

[45] J. P. Klein and R. Bajorunaite. Inference for competing risks. In *Advances in survival analysis*, volume 23 of *Handbook of Statist.*, pages 291–311. Elsevier, Amsterdam, 2004.

[46] J.P. Klein. Semiparametric estimation of random effects using the cox model based on the em algorithm. *Biometrics*, 48(3):795–806, 1992.

[47] J.P. Klein and M.L. Moeschberger. *Survival Analysis. Techniques for Censored and Truncated Data*. Statistics for Biology and Healt. Springer, New York, 1997.

[48] S. Konishi and G. Kitagawa. *Information Criteria and Statistical Modeling*. Springer, New York, 2008.

[49] M. R. Kosorok and R. Song. Inference under right censoring for transformation models with a change-point based on a covariate threshold. *Ann. Statist.*, 35(3):957–989, 2007.

[50] A.Y.C. Kuk. A semiparametric mixture model for the analysis of competing risks data. *Australian & New Zealand Journal of Statistics*, 34(2):169–180, 1992.

[51] A.Y.C. Kuk and C.H. Chen. A mixture model combining logistic regression with proportional hazards regression. *Biometrika*, 79(3):531–541, 1992.

[52] S. Kullback and R. A. . Leibler. On information and sufficiency. *Ann. Math. Statist.*, 22(1):79–86, 1951.

[53] M.G. Lagakos, C.J Sommer, and M. Zelen. Semi-markov models for partially censored data. *Biometrika*, 65(2):311–317, 1978.

[54] M. G. Larson and G. E. Dinse. A mixture model for the regression analysis of competing risks data. *J. Roy. Statist. Soc. Ser. C*, 34(3):201–211, 1985.

[55] J.F. Lawlees. *Statistical Models and Methods for Lifetime Data*. Wiley Series in Probability and Mathematical Statistics. Wiley, New York, 1982.

[56] Y. Lo, J.M. Taylor, W.H. McBride, and Withers H.R. The effect of fractionated doses of radiation on mouse spinal cord. *Int. J. Radiat. Oncol. Biol. Phys.*, 27(2):309–317, 1993.

[57] W. Lu. Maximum likelihood estimation in the proportional hazards cure model. *Ann. Inst. Statist. Math.*, 60(3):545–574, 2008.

[58] W. Lu and Z. Ying. On semiparametric transformation cure models. *Biometrika*, 91(2):331–343, 2004.

[59] S. Ma and Kosorok M.R. Penalized log-likelihood estimation for partly linear transformation models with current status data. *Ann. Statist.*, 33(5):2256–2290, 2005.

[60] R. A. Maller and S. Zhou. Estimating the proportion of immunes in a censored sample. *Biometrika*, 79(4):731–739, 1992.

[61] R.A Maller and S. Zhou. *Survival Analysis with Long-Term Survivors*. Wiley, New York, 1996.

[62] R.A. Maller and X. Zhou. Analysis of parametric models for competing risks. *Statistica Sinica*, 12:725–750, 2002.

[63] T. Martinussen and T. H. Scheike. *Dynamic regression models for survival data*. Statistics for Biology and Health. Springer, New York, 2006.

[64] W.Q. Meeker and L.A. Escobar. *Statistical Methods for Reliability Data*. Wiley Series in Probability and Statistics, New York, 1998.

[65] S. A. Murphy. Consistency in a proportional hazards model incorporating a random effect. *Ann. Statist.*, 22(2):712–731, 1994.

[66] S. A. Murphy. Asymptotic theory for the frailty model. *Ann. Statist.*, 23(1):182–198, 1995.

[67] S. A. Murphy, A. J. Rossini, and A. W. Van der Vaart. Maximum likelihood estimation in the proportional odds model. *J. Amer. Statist. Assoc.*, 92(439):968–976, 1997.

[68] M. Naskar, K. Das, and J.G. Ibrahim. A semiparametric mixture model for analyzing clustered competing risks data. *Biometrics*, 61(3):729–737, 2005.

[69] S. K. Ng and G. J. McLachlan. An em-based semi-parametric mixture model approach to the regression analysis of competing-risks data. *Stat. Med.*, 22:1097–1111, 2003.

[70] G.G. Nielsen, R.D. Gill, P.K. Andersen, and T.I.A. Sørensen. A counting process approach to maximum likelihood estimation in frailty models. *Scand. J. Statist*, 19(1):25–43, 1992.

[71] S.F. Pack and Morgan B.J. T. A mixture model for interval-censored time-to-response quantal assay data. *Biometrics*, 46(3):749–757, 1990.

[72] E. Parner. Asymptotic theory for the correlated gamma-frailty model. *Ann. Statist.*, 26(1):183–214, 1998.

[73] Y. Peng. Fitting semiparametric cure models. *Comput. Statist. Data Anal.*, 41(3-4):481–490, 2003.

[74] Y. Peng and K.B.G. Dear. A nonparametric mixture model for cure rate estimation. *Biometrics*, 56(1):237–243, 2000.

[75] Y. Peng, K.B.G Dear, and J.W Denham. A generalized f mixture model for cure rate estimation. *Statist. in Med.*, 17(8):813–830, 1998.

[76] R.L. Prentice, J.D. Kalbfleisch, A.V. Peterson, N. Flournoy, V.T. Farewell, and Breslow N.E. The analysis of failure times in the presence of competing risks. *Biometrics.*, 34(4):41–54, 1978.

[77] E. V. Slud and F. Vonta. Consistency of the NPML estimator in the right-censored transformation model. *Scand. J. Statist.*, 31(1):21–41, 2004.

[78] L.B Spivey and A.J Gross. Concomitant information in competing risk analysis. *Biometrical Journal*, 33(4):419 – 427, 2007.

[79] R. Sposto, H.N Sather, and S.A. Baker. A comparison of tests of the difference in the proportion of patients who are cured. *Biometrics*, 48(1):87–99, 1992.

[80] J. P. Sy and J. M. G. Taylor. Estimation in a Cox proportional hazards cure model. *Biometrics*, 56(1):227–236, 2000.

[81] J. M. G. Taylor. Semi-parametric estimation in failure time mixture models. *Biometrics*, 51(3):899–907, 1995.

[82] A. Tsiatis. A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences of the United States of America*, 72(1):20–22, 1975.

[83] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, Cambridge, 1998.

[84] A. W. van der Vaart and J. A. Wellner. *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York, 1996.

[85] V.G. Voinov and M.S. Nikulin. *Unbiased estimators ann their applications. Volume 1: Univariate Case*. Mathematics and Its Applications, Kluwer Academic Publisher, Dordrecht, 1993.

[86] G.A. Whitmore. First-passage-time models for duration data: Regression structures and competing risks. *J. R. Stat. Soc. Ser. D*, 35(2):207–219, 1986.

[87] K. Yamaguchi. Accelerated failure-time regression models with a regression model of surviving fraction: an application to the analysis of "permanent employmentïn japan. *J. Amer. Statist. Assoc.*, 87(418):284–292, 1992.

[88] D. Zeng and D. Y. Lin. Maximum likelihood estimation in semiparametric regression models with censored data. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 69(4):507–564, 2007.

[89] D. Zeng and D. Y. Lin. Semiparametric transformation models with random effects for recurrent events. *J. Amer. Statist. Assoc.*, 102(477):167–180, 2007.

[90] D. Zeng, D. Y. Lin, and G. Yin. Maximum likelihood estimation for the proportional odds model with random effects. *J. Amer. Statist. Assoc.*, 100(470):470–483, 2005.

[91] D. Zeng and D.Y. Lin. Efficient estimation of semiparametric transformation models for counting processes. *Biometrika*, 93(3):627–640, 2006.

[92] D. Zeng and D.Y. Lin. A general asymptotic theory for maximum likelihood estimation in semiparametric regression models with censored data. *Technical Report. University of North Carolina, Chapell Hill*, 2007.

[93] D. Zeng, D.Y. Lin, and X. Lin. Semiparametric transformation models with random effects for clustered failure time data. *Statistica Sinica*, 18(1):355–377, 2008.

UNIVERSIDAD AUTÓNOMA METROPOLITANA

Tesis para obtener el grado de Doctor en Ciencias Matemáticas

Presentada por

Angélica Hernández Quintero

11 de junio de 2010

INFERENCIA ESTADÍSTICA BASADA EN PROCESOS EMPÍRICOS PARA MODELOS SEMIPARAMÉTRICOS DEL ANÁLISIS DE SUPERVIVENCIA

ASESORES Gabriel Escarela Pérez Jean-François Dupuy

Dr. Gabriel Escarela Pérez, Universidad Autónoma Metropolitana, México.
Dr. Jean François Dupuy, Université de La Rochelle, Francia.

JURADO

Dr. Dauxois Jean-Yves, Université de Besançon, Francia.
Dr. Dupuy Jean-François, Université de La Rochelle, Francia.
Dr. Nieto Barajas Enrique, Instituto Tecnológico Autónomo, México.
Dr. Gordienko Illich Evgueni, Universidad Autónoma Metropolitana, México.
Dr. Montes de Oca Machorro José Raúl, Universidad Autónoma Metropolitana, México.
