



UNIVERSIDAD NACIONAL DE LA PLATA

FACULTAD DE CIENCIAS EXACTAS

DEPARTAMENTO MATEMÁTICA

Trabajo de Tesis Doctoral:

Estimación Bayesiana en el modelo de Riesgos Aditivos

Tesista:

Maximiliano Luis Riddick

Director/a:

Enrique Ernesto Álvarez

Codirector/a:

Nadia Laura Kudraszow

Año: 2020

Agradecimientos

Agradezco a la Universidad Nacional de La Plata, casa de estudios en la que tuve el honor de obtener mi formación de grado y posgrado, en la cual me sentí muy contenido a lo largo de todos estos años. Principalmente, al Departamento de Matemáticas y la Facultad de Ciencias Exactas, donde desarrollé toda mi carrera.

A CONICET, entidad que financió mayoritariamente mis estudios de posgrado, y me conectó con el ambiente de la investigación.

Al Instituto de Cálculo (Universidad de Buenos Aires), donde me acogieron amablemente luego de la obtención de mi título de grado.

A Enrique, por acompañarme no sólo en lo académico, sino a nivel personal en todo este trayecto. Siempre estuviste presente y dispuesto a acompañarme a lo largo de toda esta etapa, aportando generosamente tu consejo y sabiduría.

A Nadia, por haberme incentivado en una primera instancia a continuar mis estudios como estudiante de posgrado, por acompañarme a lo largo de todo el trayecto y por el tiempo dedicado y la minuciosidad con la que abordaste a la corrección de mi tesis.

A Laura, por haber estado siempre presente cuando precisé su ayuda, con absoluta entrega y contribuyendo a solucionar los inconvenientes que surgieran.

A los jurados, quienes se comprometieron con absoluta dedicación y celeridad a la lectura de esta tesis, aportando con amabilidad a su corrección y a concluir de la mejor manera posible esta etapa.

A Leandro, quien siempre me acompañó con su confianza y amistad a lo largo de esta etapa. Has sido un pilar imprescindible en este trayecto.

A Julieta, quien con su buena disposición y voluntad me alentó a seguir adelante.

A Alejandra y Raúl, quienes junto a Nadia y Enrique fueron quienes me incentivaron a continuar mis estudios.

A mi padre, quien siempre me acompañó y me formó como persona, y me apoyó con su orgullo y afecto en cualquier decisión que tomara. Siempre estarás presente.

A mi madre, quien desde muy joven me introdujo en el mundo de las matemáticas, y de quien heredé el respeto y la pasión por la ciencia.

A mis hermanos Lorena, Gerónimo y Erica, y a toda mi familia, quienes siempre me acompañaron de manera incondicional, y me impulsaron a seguir adelante.

A mis amigos, por prestarme su hombro y oído en los momentos difíciles, y compartir los momentos de alegría.

A Marina, quien siempre estuvo presente, y me brindó su apoyo y afecto durante la mayor parte de esta etapa.

A todos, muchas gracias! Lograron hacer placentero este tramo.

Resumen

Desde la década de 1950, el Análisis de Supervivencia se ha convertido en una de las áreas más populares dentro del Análisis Estadístico. De hecho, los dos artículos estadísticos más citados hasta el momento, escritos por Kaplan & Meier (1958) y Cox (1972), pertenecen a esta área. En dicha área, se estudian las variables de Supervivencia, las cuales consisten en variables no negativas que miden el tiempo hasta la ocurrencia de cierto evento de interés, y se definen nuevas funciones buscando expresar distintos aspectos de ellas, tales como la función de Supervivencia y la función de riesgo $\lambda(t)$. Los modelos de Supervivencia son expresados de acuerdo a las expresiones de sus respectivas funciones de riesgo, y el Modelo de Riesgos Aditivos es un modelo semiparamétrico muy utilizado, cuya función de riesgo es de la forma

$$\lambda(t, \boldsymbol{\beta}) := \lambda_0(t) + \boldsymbol{\beta}'z,$$

donde $\lambda_0(\cdot)$ representa a la componente no paramétrica del modelo, $\boldsymbol{\beta}$ es un vector Euclideo de parámetros regresores y z representa un vector de variables asociado a cada observación.

El Análisis Bayesiano es un enfoque que permite, entre muchas otras aplicaciones, introducir información *a priori*, modelizando los parámetros como variables aleatorias.

En esta tesis Doctoral, desarrollamos un análisis Bayesiano extenso del Modelo de Riesgos Aditivos, detallando los trabajos presentados en el área, e incluyendo nuevas propuestas de estimadores Bayesianos. Una ventaja importante de los estimadores hallados es que obtenemos sus expresiones de manera explícita, y evaluamos su eficiencia a través de simulaciones, en donde veremos el impacto obtenido al variar los parámetros de las *prioris* seleccionadas. Además, proponemos distintos procesos *a priori* para modelizar la componente no paramétrica $\lambda_0(\cdot)$. Presentamos métodos de selección de *prioris automáticas*, así como opciones de elucidación, lo que refiere a la traducción del conocimiento experto.

Palabras clave: Modelo de Riesgos Aditivos, Análisis de Supervivencia, Inferencia Bayesiana.

Abstract

Since the decade of 1950, Survival Analysis became one of the most popular areas in Statistical Analysis. In fact, the two most cited statistical articles in history, written by Kaplan & Meier (1958) and Cox (1972) respectively, belong to this area. New functions are defined in order to express specific issues of Survival variables (which are nonnegative variables that focus on the time until the occurrence of a certain event of interest) such as the Survival function and the hazard rate function, $\lambda(t)$. The Survival models are expressed based on their hazard functions, and the Additive Hazards Model is an important semiparametric model, whose hazard expression is

$$\lambda(t, \boldsymbol{\beta}) := \lambda_0(t) + \boldsymbol{\beta}'z,$$

where $\lambda_0(\cdot)$ is the nonparametric component of the model, $\boldsymbol{\beta}$ is an Euclidean vector of regressor parameters and z represents an Euclidean vector of variables associated to each element in the sample.

Bayesian Analysis is an approach that enables, for instance, to introduce prior information as expert knowledge into the model, by modeling the parameters as random variables. In this PhD thesis, we developed a comprehensive Bayesian Analysis of the Additive Hazards Model, giving a literature review of research in the area, and surveying new proposals of Bayesian estimators. An important advantage of our estimators is that they are available in closed form. The accuracy of them is tested through a simulation study, in which the values and weights of the selected priors are chosen in order to measure the impact and flexibility of the model. Moreover, we express in detail several prior processes to model the nonparametric component $\lambda_0(\cdot)$. Further, we discuss *automatic priors* selection methods, and prior elicitation of expert knowledge.

Keywords: Additive Hazards Model, Survival Analysis, Bayesian Inference.

Índice general

1..	<i>Introducción</i>	10
1.1.	Análisis de Supervivencia	13
1.1.1.	Preliminares	13
1.1.2.	Modelos Paramétricos	15
1.1.3.	Datos Censurados	17
1.1.4.	Expresión de la función de Verosimilitud en Modelos de Supervivencia	18
1.1.5.	Incorporación de <i>Covariables</i>	22
1.1.6.	Modelos semi-paramétricos	23
1.2.	Estadística Bayesiana	25
1.2.1.	Un poco de historia	26
1.2.2.	Aspectos técnicos: Enfoque Bayesiano del problema de la estimación puntual	28
2..	<i>El Modelo de Riesgo Aditivo</i>	34
2.1.	Estimación Bayesiana en Modelos de Riesgo Aditivo mediante el modo de la posteriori	37
2.1.1.	Simulación	38
2.2.	Bibliografía en el área	40
2.2.1.	Principales trabajos de inferencia Bayesiana en el modelo de riesgos aditivos	41
2.2.2.	Inclusión de Frailties	48

2.2.3. Otras Propuestas	53
3.. <i>Análisis Bayesiano del Modelo de Riesgo Aditivo</i>	57
3.1. Expresión para la verosimilitud	57
3.2. Selección de prioris	60
3.3. Distribución a Posteriori.	62
3.4. El Método Bayesiano Híbrido	62
3.4.1. Cálculo del estimador Bayesiano de β	64
3.4.2. Cálculo de los estimadores Bayesianos de los parámetros de la baseline hazard	67
3.5. Proceso Gamma en la cumulative baseline $\Lambda_0(\cdot)$	68
3.6. Simulaciones	72
3.7. Simulaciones de los parámetros de regresión β	72
3.8. Simulaciones para los parámetros de la baseline	74
3.9. Datos Reales	77
4.. <i>Prioris no informativas</i>	78
4.1. Priori Uniforme de Laplace	79
4.2. Priori de Jeffreys	80
4.3. Priori de Máxima Entropía	82
5.. <i>Otras opciones de procesos a priori para la baseline</i>	85
5.1. Proceso Gamma sobre la baseline hazard $\lambda_0(\cdot)$	85
5.2. Proceso Beta sobre la cumulative baseline $\Lambda_0(\cdot)$	86
5.3. Proceso Dirichlet bajo conocimiento específico de $S(t)$ para t_F fijo	88
6.. <i>Elucidación de prioris</i>	93
6.1. Elucidación para los parámetros regresores	93

6.2. Elucidación para los parámetros de la baseline	94
6.2.1. Susarla & Van Ryzin	94
7.. <i>Extensiones y Conclusiones</i>	97
7.1. Extensiones	97
7.2. Conclusiones	99

1. INTRODUCCIÓN

El estudio de procesos de supervivencia (o sobrevida) juega un papel muy importante en campos como la medicina, ciencias biomédicas y epidemiología, entre otros. Así, por ejemplo, la comparación entre la supervivencia observada en dos grupos de pacientes puede llevar a validar un determinado tratamiento o, alternativamente, a identificar un factor de riesgo importante. En general, los estudios que intentan evaluar la supervivencia en una determinada situación presentan características que determinan que deba utilizarse una modelización adecuada a cada caso en particular como, por ejemplo, a partir de qué momento se mide la variable. La situación más común corresponde a un estudio en el cual la supervivencia del paciente se estudia a partir de un determinado instante de tiempo en el que se interviene sobre dicho paciente (administración de un tratamiento, intervención quirúrgica, etc.). El estudio de procesos de supervivencia implica el seguimiento de los individuos a lo largo del tiempo, pudiéndose producir una serie de situaciones que complican la caracterización de los mismos. Así, la situación más favorable la tendremos cuando podemos observar de manera exacta el tiempo T de aparición del suceso de interés (muerte, aparición de complicaciones post-operatorias, rechazo de un órgano transplantado, etc.). En esta situación, hablaremos de datos *no censurados*. Por otra parte, es habitual que algunos de los pacientes se pierdan a lo largo del seguimiento. Por ejemplo, puede pasar que un paciente transplantado deje de acudir a la consulta por cambio de domicilio, perdiéndose su rastro a efectos de observar el suceso de interés. En este caso, sabremos que el suceso no se ha producido durante el tiempo que hemos seguido al paciente, pero no sabemos, pasado este tiempo, cuándo se ha producido el suceso de interés. En esta situación, hablaremos de datos *censurados*. Por otra parte,

los estudios clínicos suelen tener un inicio y un final. Para los pacientes que no presentan el suceso cuando el estudio se acaba, sabremos que al menos han sobrevivido a ese tiempo, pero no sabremos cuando han presentado el suceso de interés.

El objetivo de la tesis es realizar un análisis de un modelo particular dentro del Análisis de Supervivencia (el Modelo de Riesgos Aditivos) desde un enfoque Bayesiano, obteniendo estimadores Bayesianos de forma explícita para la totalidad de los parámetros del modelo, objetivo que no se encuentra presente hasta el momento en la literatura. Además, presentaremos diferentes alternativas para la selección de prioris, y técnicas de elucidación utilizadas para traducir la información a priori disponible como información sobre los parámetros. La tesis se encuentra dividida por capítulos, sobre los cuales detallaremos brevemente a continuación.

En lo que resta del Capítulo 1, trataremos las definiciones básicas y motivacionales del Análisis de Supervivencia, incluyendo diversos ejemplos y presentando los conceptos que se utilizarán a lo largo de los siguientes capítulos. Además, presentaremos el enfoque Bayesiano, contando brevemente el contexto histórico en el que se desarrolló y su evolución, para luego dar una descripción técnica de su implementación.

En el Capítulo 2, presentaremos el modelo sobre el que se desarrollan los contenidos de esta tesis, explicitando sus propiedades y motivaciones, y una aproximación usual a la componente no paramétrica del modelo, llamada “*constante a trozos*”. Además, se adjunta una breve sección de simulaciones, en la cual desarrollamos una primera propuesta de estimadores Bayesianos para el Modelo de Riesgo Aditivo, proponiendo una función de riesgo distinta a la más usada en la literatura (función de pérdida cuadrática). Luego, presentamos una extensa revisión de los trabajos publicados en el área, detallando las técnicas que fueron utilizadas y los inconvenientes con los que se encontraron los distintos autores al abordar al Modelo de Riesgos Aditivos desde un enfoque Bayesiano.

En el Capítulo 3, calculamos estimadores Bayesianos para la totalidad de los parámetros del modelo mediante un método híbrido (denominado así ya que los estimadores de los parámetros regresores se obtienen mediante una “*ecuación de estimación*”), utilizando un algoritmo que permite obtener una nueva expresión para la función de verosimilitud. Mediante esta nueva expresión, se obtienen estimadores Bayesianos explícitos para la totalidad de los parámetros del modelo. Finalizaremos la Sección con diversas simulaciones, y una aplicación a datos reales.

En el Capítulo 4 detallaremos la implementación de *prioris no informativas*, las cuales buscan minimizar el impacto de la distribución *a priori* seleccionada acorde a la información previa.

En el Capítulo 5 ofreceremos otras alternativas de procesos *a priori* para la componente no paramétrica del modelo, explicitando la manera de implementarlas.

En el Capítulo 6 presentaremos opciones de elucidación del conocimiento experto tanto para las componentes paramétrica como no paramétrica del modelo, el cual en general se encuentra asociado a la Función de Supervivencia. Finalmente, en el Capítulo 7 realizaremos un compendio de los resultados obtenidos, y presentaremos extensiones a realizar como investigación futura.

1.1. Análisis de Supervivencia

A continuación, presentaremos las funciones básicas del análisis de Supervivencia, e introduciremos las nociones de datos censurados y covariables, ejemplificando con modelos clásicos.

1.1.1. Preliminares

Definición. Llamaremos **función de Supervivencia** a la función definida como:

$$S(t) := P(T \geq t),$$

donde P denota a la función de Probabilidad de la variable T . Cuando T es continua, $S(t) = 1 - F(t)$, siendo F la función de distribución acumulada.

Definición. La **función de Riesgo** de una variable aleatoria T en un momento específico t es la tasa instantánea de ocurrencia de eventos entre la población que aún está en riesgo en el instante t . Si T es continua, tenemos que:

$$\begin{aligned} \lambda(t) &= \lim_{\Delta \rightarrow 0} \frac{P(t < T \leq t + \Delta | T > t)}{\Delta} \\ &= \lim_{\Delta \rightarrow 0} \frac{P(T \in (t, t + \Delta] \cap T > t) / P(T > t)}{\Delta} \\ &= \lim_{\Delta \rightarrow 0} \frac{P(T \in (t, t + \Delta])}{\Delta} / P(T > t) \\ &= \lim_{\Delta \rightarrow 0} \frac{\int_t^{t+\Delta} f(u) du}{\Delta} / P(T > t) \\ &= \frac{f(t)}{S(t)}. \end{aligned}$$

Definición. Si T es continua, la **función de Riesgo Acumulado** se define como:

$$\Lambda(t) = \int_0^t \lambda(s) ds.$$

Esto implica que:

$$\Lambda(t) = -\log[S(t)],$$

$$S(t) = \exp[-\Lambda(t)].$$

De esta manera, advertimos que

$$f(t) \longleftrightarrow F(t) \longleftrightarrow \lambda(t) \longleftrightarrow \Lambda(t) \longleftrightarrow S(t),$$

donde \longleftrightarrow representa la relación “uno a uno” que existe entre ellas. Es decir, conocer a una de estas funciones nos permite conocer a las otras, lo que implica que tener conocimiento de alguna de ellas se puede traducir en conocimiento de las demás.

La motivación de definir estas nuevas funciones es que describen distintos aspectos de una misma variable aleatoria, los cuales en ocasiones se presentan con mayor naturalidad en la práctica, facilitando su modelización. En medicina, por ejemplo, suele ser más fácil modelizar la supervivencia o el riesgo de una variable específica, que su función de densidad.

1.1.2. Modelos Paramétricos

Los Modelos Paramétricos consisten en suponer que la distribución $F(x)$ de la variable aleatoria pertenece a una familia de distribuciones que depende de un número finito de parámetros reales. Veamos, a continuación, algunos modelos clásicos:

Exponencial: Es la función más simple y una de las distribuciones más importantes dentro del análisis de supervivencia. Algunas de las aplicaciones de esta función se encuentran al describir patrones de vida en sistemas electrónicos, como por ejemplo sistemas de generación de estados de cuenta bancarios, cajeros automáticos y componentes de radares. Es usualmente asociada a patrones de falla puramente aleatorios. También es famosa por la propiedad de *falta de memoria* con el único detalle que esta propiedad requiere que la edad de la persona (componente, animal, etc.) no afecte a su supervivencia futura. Está caracterizada por una tasa de riesgo constante ν , como su único parámetro. Valores grandes de ν se asocian a riesgo alto y por lo tanto a baja supervivencia y viceversa. A continuación, se presentan sus respectivas funciones de densidad, supervivencia, riesgo y riesgo acumulado correspondientes a este modelo:

$$\begin{aligned} f(t) &= \nu e^{-\nu t}, \quad t \geq 0, \\ S(t) &= 1 - F(t) = 1 - (1 - e^{-\nu t}) = e^{-\nu t}, \\ \lambda(t) &= \frac{f(t)}{S(t)} = \frac{\nu e^{-\nu t}}{e^{-\nu t}} = \nu, \quad t \geq 0 \\ \Lambda(t) &= -\log[S(t)] = -\log[e^{-\nu t}] = \nu t. \end{aligned}$$

Weibull: Es una generalización de la distribución *Exponencial*; se diferencia de esta en que no supone una tasa de riesgo constante, logrando con esto aumentar sus posibles aplicaciones. Sus primeras aplicaciones se dieron en rentabilidad, mortalidad humana, en pacientes con cáncer de pulmón y tiempos de mejora en cirugías mayores. Está caracterizada por dos parámetros: α que determina la forma de la distribución y ν que

determina su escala. Cuando $\alpha = 1$ se obtiene el caso exponencial. Cuando t crece, la tasa de riesgo aumenta cuando $\alpha > 1$, y disminuye cuando $\alpha < 1$. Sus respectivas funciones de densidad, supervivencia, riesgo y riesgo acumulado son:

$$f(t) = \alpha \nu t^{\alpha-1} e^{-\nu t^\alpha}, \quad t \geq 0 \implies F(t) = 1 - e^{-\nu t^\alpha},$$

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

$$\lambda(t) = \frac{\alpha \nu t^{\alpha-1} e^{-\nu t^\alpha}}{e^{-\nu t^\alpha}} = \alpha \nu t^{\alpha-1}, \quad t \geq 0,$$

$$\Lambda(t) = -\log[S(t)] = -\log[e^{-\nu t^\alpha}] = \nu t^\alpha.$$

La condición de monotonicidad de su función de riesgo hace que esta variable no resulte adecuada para modelizar, por ejemplo, el tiempo de vida de una persona.

Gamma: Esta distribución incluye, como casos particulares, a las distribuciones *Exponencial* y *Ji-cuadrado*. Inicialmente fue utilizada para describir el tiempo útil de los vasos de cristal de cafeteras y para la duración de materiales, de igual forma se ha utilizado en la industria y supervivencia humana.

Esta distribución surge de suponer que la falla o muerte ocurre en n periodos, o cuando n subfallas han sucedido, es decir, al tiempo T_1 , la primera subfalla ocurre, al tiempo T_2 ocurre la segunda subfalla, y así sucesivamente. La falla total o muerte ocurre al final del n -ésimo periodo con la n -ésima subfalla. El tiempo de supervivencia T es entonces $T_1 + T_2 + \dots + T_n$, se supone que dichos tiempos son independientes y distribuidos exponencialmente con función de densidad $\beta \exp(-\beta t_i)$, $t \geq 0, i = 1, \dots, n$. Lo cual quiere decir que las subfallas ocurren de forma independiente con tasa constante β . La distribución de T es llamada distribución de *Erlang* y ha sido ampliamente usada en teoría de colas y procesos de vida. Si se generaliza la distribución de *Erlang* reemplazando el parámetro n por un parámetro real $\alpha > 0$ se obtiene la distribución *Gamma*.

La distribución *Gamma* está caracterizada por dos parámetros α y β , los cuales determinan la forma y la escala respectivamente. Si $0 < \alpha < 1$ la tasa de riesgo disminuye monótonamente de infinito a β conforme avanza el tiempo; si $\alpha > 1$ la tasa de riesgo crece monótonamente de 0 a β cuando t avanza y si $\alpha = 1$ la tasa de riesgo se vuelve constante con tasa β como en el caso exponencial. En este caso, dado que no podemos hallar una primitiva para la función *Gamma*, las respectivas funciones de densidad, distribución, supervivencia y riesgo están dadas por la forma:

$$\begin{aligned} f(t) &= \frac{\beta^\alpha t^{\alpha-1} e^{-\beta t}}{\Gamma(\alpha)}, \\ F(t) &= \int_0^t \frac{\beta^\alpha x^{\alpha-1} e^{-x\beta}}{\Gamma(\alpha)} dx, \\ S(t) &= \int_t^\infty \frac{\beta^\alpha x^{\alpha-1} e^{-x\beta}}{\Gamma(\alpha)} dx, \\ \lambda(t) &= \frac{t^{\alpha-1} e^{-t\beta}}{\int_t^\infty x^{\alpha-1} e^{-x\beta} dx}. \end{aligned}$$

1.1.3. Datos Censurados

La presencia de *datos censurados por la derecha* es bastante común en los estudios de supervivencia. Los mecanismos de censura por derecha más usuales son:

- Tipo I: el mecanismo de censura Tipo I se presenta cuando cada individuo/elemento de la muestra tiene una potencial censura fija C_i , de manera que T_i es observado sólo si $T_i \leq C_i$. Las censuras de Tipo I usualmente se presentan en estudios que son llevados a cabo durante un período de tiempo específico.
- Censura progresiva (o aleatoria independiente): Cuando C es una variable aleatoria independiente de T^* . Usualmente, C es el último tiempo observado.
- Tipo II: Cuando el estudio concluye al llegar a cierto número fijo de observaciones y se considera a C como una variable aleatoria.

El modelo de muestreo censurado es muy útil para analizar muestras que contienen este tipo de datos. Sean T_1^*, \dots, T_n^* variables aleatorias independientes e idénticamente distribuidas (i.i.d.) con función de distribución F , y sean C_1, \dots, C_n variables aleatorias i.i.d. (e independientes de las anteriores), con función de distribución G . En nuestro modelo, las T_i^* representan los tiempos reales de “vida” (algunos de ellos posiblemente no observables), y las C_i los tiempos de censura. Con muestreo censurado, observamos los pares:

$$(T_1, \Delta_1), (T_2, \Delta_2), \dots, (T_n, \Delta_n)$$

donde:

$$T_j = \min(T_j^*, C_j),$$

$$\Delta_j = \begin{cases} 1, & \text{si } T_j^* \leq C_j \text{ (dato no censurado),} \\ 0, & \text{si } T_j^* > C_j \text{ (dato censurado).} \end{cases}$$

1.1.4. Expresión de la función de Verosimilitud en Modelos de Supervivencia

La función de verosimilitud tiene como argumento el/los parámetros θ del modelo (dada la muestra). En ella, cada factor representa la información que aporta cada observación como función del parámetro. Considerando censura por la derecha, la información que brinda cada elemento de la muestra será $f(t)$ (en el caso de observaciones no censuradas), o $S(t)$ (en el caso de observaciones censuradas). En base a esto, la expresión para la función de verosimilitud en modelos de Supervivencia está dada por:

$$\mathcal{L}_n(\theta | (t_1, \delta_1), \dots, (t_n, \delta_n)) \propto \prod_{i=1}^n [f(t_i | \theta)]^{\delta_i} [S(t_i | \theta)]^{1-\delta_i} = \prod_{i=1}^n [\lambda(t_i | \theta)]^{\delta_i} S(t_i | \theta), \quad (1.1)$$

donde el símbolo \propto denota que las expresiones son proporcionales. Esta expresión es la misma para una gran variedad de modelos de censura por la derecha, incluyendo los tres más usuales (Tipo I, Censura progresiva y Tipo II). Un mayor detalle de estos aspectos se puede

apreciar en Lawless (2011).

A continuación, presentamos algunos ejemplos importantes aplicando censura por la derecha, y obteniendo sus funciones y estimadores de máxima verosimilitud.

Exponencial con censura: Consideramos el caso de un modelo de supervivencia con distribución *Exponencial* con censura por la derecha. Tal sería el caso si midiéramos, por ejemplo, el correcto funcionamiento de una componente electrónica durante 3 días. Si funciona correctamente al cabo de dicho tiempo, la consideramos apta y cesamos la medición. En casos como éste, el aporte a la función de verosimilitud de cada elemento t_i de la muestra es de la forma:

$$(\nu e^{-\nu t_i})^{\delta_i} (e^{-\nu t_i})^{1-\delta_i} = \nu^{\delta_i} e^{-\nu t_i}$$

donde δ_i indica si lo que se observó fue un evento o una censura. Luego, la expresión de la función de verosimilitud es:

$$\mathcal{L}_n(\nu | (t_1, \delta_1), \dots, (t_n, \delta_n)) = \prod_{l=1}^n \lambda(t_l)^{\delta_l} S(t_l) = \prod_{l=1}^n \nu^{\delta_l} e^{-\nu t_l} = \nu^{\sum_{l=1}^n \delta_l} e^{-\nu \sum_{l=1}^n t_l}.$$

Luego, hallamos su logaritmo

$$\log \mathcal{L}_n = \left(\sum_{l=1}^n \delta_l \right) \log(\nu) - \nu \sum_{l=1}^n t_l,$$

y derivamos para hallar la *función de Score*

$$\frac{\partial \log \mathcal{L}_n}{\partial \nu} = \frac{\sum_{l=1}^n \delta_l}{\nu} - \sum_{l=1}^n t_l.$$

Para hallar el valor donde se maximiza la función de verosimilitud, en este caso alcanza con encontrar el valor de ν que anula la función de Score. Luego, el EMV (*Estimador de Máxima Verosimilitud*) de ν es:

$$\hat{\nu}^{MV} = \frac{\sum_{l=1}^n \delta_l}{\sum_{l=1}^n t_l}.$$

Advertimos que, en ausencia de datos censurados (es decir, $\sum_{l=1}^n \delta_l = n$), este estimador coincide con el EMV clásico de la distribución exponencial.

Exponencial a trozos: Este caso será ampliamente usado a lo largo de esta tesis, dado que lo utilizaremos como aproximación a una curva continua mediante funciones constantes a trozos (de la misma manera que en *Teoría de la Medida* se utilizan funciones simples para aproximar funciones medibles).

Dada una grilla $[s_0 = 0, s_1, \dots, s_{J-1}, s_J = \infty)$, la cual define una partición de \mathbb{R}_0^+ , consideramos una variable aleatoria cuya función de riesgo sea:

$$\lambda_0(t) = \sum_{i=1}^J \nu_i 1_{[s_{i-1}, s_i)}(t). \quad (1.2)$$

La función de verosimilitud para este modelo adopta la siguiente expresión

$$\mathcal{L}_n(\nu_1, \dots, \nu_J | (t_1, \delta_1), \dots, (t_n, \delta_n)) = \prod_{l=1}^n \left(\sum_{i=1}^J \nu_i 1_{[s_{i-1}, s_i)}(t_l) \right)^{\delta_l} \exp\{-\Lambda_0(t_l)\},$$

de lo que se deduce que la función de verosimilitud para cada ν_i fijo resulta en:

$$\mathcal{L}_n(\nu_i | (t_1, \delta_1), \dots, (t_n, \delta_n)) \propto \exp\left\{ \sum_{l=1}^n -\Lambda_0(t_l) \right\} \prod_{k_i=1}^{n_i} \nu_i^{\delta_{k_i}}.$$

donde el conjunto $\{t_{k_i}\}_{k_i=\{1, \dots, n_i\}}$ representa a la cantidad de observaciones en la muestra que se encuentran en el intervalo $[s_{i-1}, s_i)$, y $\{\delta_{k_i}\}_{k_i=\{1, \dots, n_i\}}$ son sus respectivos indicadores de censura. Dado un ν_i fijo, podemos expresar la función de riesgo acumulado evaluada en cada observación t como

$$\Lambda_0(t) := \begin{cases} \Lambda_0(t) & t \leq s_{i-1}, \\ \Lambda_0(s_{i-1}) + \nu_i(t - s_{i-1}) & s_{i-1} < t \leq s_i, \\ \Lambda_0(s_{i-1}) + \nu_i(s_i - s_{i-1}) + \int_{s_i}^t \lambda_0(u) du & s_i < t, \end{cases}$$

logrando diferenciar la parte que depende de ν_i de la que no.

Llamando $m_i = \#\{t_l > s_i\}_{l=1, \dots, n}$, y dado ν_i fijo, obtenemos

$$\begin{aligned} \exp \left\{ - \sum_{l=1}^n \Lambda_0(t_l) \right\} &= \exp \left\{ - \left(\sum_{t_l < s_{i-1}} \Lambda_0(t_l) + \sum_{t_l \in [s_{i-1}, s_i)} \Lambda_0(t_l) + \sum_{t_l \geq s_i} \Lambda_0(t_l) \right) \right\} \\ &\propto \exp \left\{ - \left(\sum_{t_l \in [s_{i-1}, s_i)} \nu_i(t_l - s_{i-1}) + m_i \nu_i(s_i - s_{i-1}) \right) \right\} \\ &= \exp \left\{ -\nu_i \left(\sum_{k_i=1}^{n_i} (t_{k_i} - s_{i-1}) + m_i(s_i - s_{i-1}) \right) \right\}. \end{aligned} \quad (1.3)$$

Luego, la función de verosimilitud se puede reescribir como

$$\mathcal{L}_n(\nu_i | (t_1, \delta_1), \dots, (t_n, \delta_n)) = e^{-\nu_i \left(\sum_{k_i=1}^{n_i} (t_{k_i} - s_{i-1}) + m_i(s_i - s_{i-1}) \right)} \prod_{k_i=1}^{n_i} \nu_i^{\delta_{k_i}}.$$

Calculamos su logaritmo:

$$\begin{aligned} \log \mathcal{L}_n(\nu_i | (t_1, \delta_1), \dots, (t_n, \delta_n)) &= -\nu_i \left(\sum_{k_i=1}^{n_i} (t_{k_i} - s_{i-1}) + m_i(s_i - s_{i-1}) \right) + \log \left(\prod_{k_i=1}^{n_i} \nu_i^{\delta_{k_i}} \right) \\ &= -\nu_i \left(\sum_{k_i=1}^{n_i} (t_{k_i} - s_{i-1}) + m_i(s_i - s_{i-1}) \right) + \sum_{k_i=1}^{n_i} \log \left(\nu_i^{\delta_{k_i}} \right) \\ &= -\nu_i \left(\sum_{k_i=1}^{n_i} (t_{k_i} - s_{i-1}) + m_i(s_i - s_{i-1}) \right) + \log \nu_i \sum_{k_i=1}^{n_i} \delta_{k_i}, \end{aligned}$$

para luego hallar la función de Score:

$$\frac{\partial \log \mathcal{L}_n(\nu_i | (t_1, \delta_1), \dots, (t_n, \delta_n))}{\partial \nu_i} = - \left(\sum_{k_i=1}^{n_i} (t_{k_i} - s_{i-1}) + m_i(s_i - s_{i-1}) \right) + \frac{\sum_{k_i=1}^{n_i} \delta_{k_i}}{\nu_i}.$$

Luego, despejando, hallamos una expresión explícita para el estimador de máxima verosimilitud de cada ν_i :

$$\hat{\nu}_i^{MV} = \frac{\sum_{k_i=1}^{n_i} \delta_{k_i}}{\sum_{k_i=1}^{n_i} (t_{k_i} - s_{i-1}) + m_i(s_i - s_{i-1})},$$

donde $\sum_{k_i=1}^{n_i} \delta_{k_i}$ resulta ser la cantidad de observaciones no censuradas que caen en el intervalo $[s_{i-1}, s_i)$.

Cuando tratamos sobre \mathbb{R}_0^+ (es decir, no consideramos una partición), nos encontramos en el caso particular de la *Exponencial*, y el estimador hallado coincide con el de la *Exponencial con censura*, ya que en el numerador se encuentra la cantidad de observaciones no censuradas, y en el denominador advertimos que $s_{i-1} = 0$ y $m_i = 0$.

1.1.5. Incorporación de Covariables

Es usual considerar a la expresión de la función de supervivencia desconocida, en cuyo caso procedemos a estimarla mediante la muestra. Podemos asumir que dicha distribución pertenece a una familia paramétrica (por ejemplo: *Log-normal*, *Exponencial*, *Weibull*, *Gamma*), o podemos modelizarla de manera no paramétrica. En cualquier caso, asumiremos independencia entre las observaciones (t_i, δ_i) , para $i \in \{1, \dots, n\}$. Eventualmente, podemos asumir la existencia de ciertas variables asociadas a cada unidad de la muestra, a las que llamaremos *covariables* y las denotaremos $z(t_i) = z_i$, que permitirán explicar ciertas particularidades de cada unidad de muestreo. Dichas *covariables* pueden ser datos como la edad del paciente al inicio del estudio, indicaciones particulares de un tratamiento en cada muestra, etc. En ocasiones, podemos considerar a dichas *covariables* en función del tiempo, y las llamaremos *covariables dependientes del tiempo*.

Se presenta a continuación un modelo donde incluimos una *covariable*.

Ejemplo: Consideremos una variable aleatoria con función de riesgo:

$$\lambda(t) = \nu e^{\beta z},$$

donde z es una *covariable* que no depende del tiempo, y ν y β son los parámetros de interés.

La función de riesgo acumulado en este modelo resulta en

$$\Lambda(t) = \nu e^{\beta z} t,$$

por lo que su función de Supervivencia adopta la expresión

$$S(t) = \exp\{-\nu e^{\beta z} t\}.$$

Luego, la función de Verosimilitud con esta función de riesgo es:

$$\mathcal{L}_n(\nu, \beta | (t_1, \delta_1), \dots, (t_n, \delta_n)) = \prod_{l=1}^n (\nu e^{\beta z_l})^{\delta_l} \exp\{-\nu e^{\beta z_l} t_l\},$$

y su logaritmo es

$$\log \mathcal{L}_n(\nu, \beta | (t_1, \delta_1), \dots, (t_n, \delta_n)) = \sum_{l=1}^n [\delta_l (\log \nu + \beta z_l) - \nu e^{\beta z_l} t_l].$$

Así, llegamos a la función de Score para ν :

$$\frac{\partial \log \mathcal{L}_n(\nu, \beta | (t_1, \delta_1), \dots, (t_n, \delta_n))}{\partial \nu} = \sum_{l=1}^n \left[\frac{\delta_l}{\nu} - e^{\beta z_l} t_l \right],$$

y la función de Score para β :

$$\frac{\partial \log \mathcal{L}_n(\nu, \beta | (t_1, \delta_1), \dots, (t_n, \delta_n))}{\partial \beta} = \sum_{l=1}^n z_l (\delta_l - \nu t_l e^{\beta z_l}).$$

Igualando ambas ecuaciones a 0, nos queda un sistema del cual no se puede despejar $\hat{\nu}^{MV}$ y $\hat{\beta}^{MV}$. No obstante, se puede recurrir al uso de *algoritmos computacionales* y hallar el EMV de manera numérica, mediante la maximización de la función de Verosimilitud.

1.1.6. Modelos semi-paramétricos

Se denominan *modelos semiparamétricos* a los modelos que tienen tanto una parte *paramétrica* como una *no-paramétrica*.

Algunos de estos modelos, los cuales son presentados mediante una expresión de su función de riesgo, y sobre los que hay una extensa bibliografía dedicada a su tratamiento, son:

Modelo de riesgos proporcionales de Cox (1972): Sin dudas el más popular de los modelos semiparamétricos, fue propuesto por David Cox, siendo entre los trabajos de estadística el segundo más citado en la historia. También llamado PHM por las siglas en inglés *Proportional Hazards Model*. Sean $(T_1, \Delta_1), \dots, (T_n, \Delta_n)$ independientes, y sea Z el vector de covariables asociado a esta muestra, donde la *función de riesgo* de $T^*|Z = z$ está dada por:

$$\lambda(t, \boldsymbol{\beta}) = \lambda_0(t)e^{z'\boldsymbol{\beta}}$$

y λ_0 es una función de riesgo arbitraria, la que llamaremos en adelante *baseline hazard*, y que determina una *distribución base* con densidad f_0 .

Dados dos vectores de covariables z_1 y z_2 que difieren sólo en la i -ésima covariable, definimos el *HR* (del inglés *hazard ratio*) como:

$$HR = \frac{\lambda_0(t) \exp\{z_1'\boldsymbol{\beta}\}}{\lambda_0(t) \exp\{z_2'\boldsymbol{\beta}\}} = \exp[(z_{1i} - z_{2i})\beta_i],$$

donde z_{1i} , z_{2i} y β_i representan las i -ésimas componentes de z_1 , z_2 y $\boldsymbol{\beta}$ respectivamente. Advertimos que *HR* no depende del tiempo t . El nombre *modelo de riesgos proporcionales* se deduce de la última ecuación, donde se advierte dicha propiedad.

Considerando datos de supervivencia con censura estándar (por la derecha) $\{(t_l, \delta_l)\}_{l=1}^n$, la verosimilitud para este modelo está dada por:

$$\mathcal{L}_n(\boldsymbol{\beta}, \lambda_0 | (t_1, \delta_1), \dots, (t_n, \delta_n)) = \prod_{l=1}^n \{e^{z_l'\boldsymbol{\beta}} \lambda_0(t_l)\}^{\delta_l} [S_0(t_l)]^{e^{z_l'\boldsymbol{\beta}}}.$$

Este modelo es *semiparamétrico* ya que posee una parte *paramétrica*, que es la que se corresponde con $e^{z'\boldsymbol{\beta}}$, y una no paramétrica, que se corresponde con la *baseline* $\lambda_0(t)$.

Modelo de riesgos aditivos (Aalen, 1980): En adelante, nos referiremos a este modelo como AHM (del inglés *Additive Hazards Model*). Este modelo propone la siguiente función

de riesgo

$$\lambda(t, \boldsymbol{\beta}) = \lambda_0(t) + z'\boldsymbol{\beta},$$

donde, nuevamente, $\lambda_0(\cdot)$ representa a la componente no paramétrica del modelo, $\boldsymbol{\beta}$ representa al vector de parámetros regresores y z al vector de covariables de la muestra. En los próximos capítulos trataremos en detalle este modelo.

Modelo de tiempo de falla acelerada: El modelo de *tiempo de falla acelerada* o AFTM (*Accelerated Failure Time Model*), presenta un enfoque ligeramente distinto a los modelos previos, proponiendo un cambio de escala para los tiempos de sobrevivida. Destacamos que, a diferencia del PHM y AHM, la expresión del AFTM es de acuerdo a los tiempos de sobrevivida, y no a la función de riesgo. Dicha expresión es:

$$T^* = T_0^* \exp(z'\boldsymbol{\beta}),$$

donde $T_0^* \sim F_0(\cdot)$ es la componente no paramétrica del modelo, z es el vector de covariables y $\boldsymbol{\beta}$ el vector de parámetros regresores. Advertimos que, en este tipo de modelos, la función de riesgo es $\lambda(t, \boldsymbol{\beta}) = \lambda_0 [t \exp(z'\boldsymbol{\beta})] \exp(z'\boldsymbol{\beta})$, la cual no resulta ni multiplicativa ni aditiva.

Estos modelos se hicieron muy populares debido a su amplio rango de aplicación a distintos procesos. El desarrollo de propuestas de estimación clásica en cada uno de ellos es muy amplio. Una recopilación extensa de estos puede ser hallada en algunos libros clásicos, como Kalbfleisch & Prentice (1980), Klein & Moeschberger (2005), o Lawless (2003), entre otros.

1.2. Estadística Bayesiana

A continuación, presentaremos una breve introducción a la Estadística Bayesiana. Comenzaremos retratando en orden cronológico las principales figuras que contribuyeron al

desarrollo y validez de este enfoque de la estadística, para luego dar una definición más formal, con ejemplos ilustrativos.

1.2.1. Un poco de historia

Los orígenes y consolidación de la Estadística Bayesiana fueron impulsados mayormente por las siguientes reconocidas figuras históricas dentro de la historia de la ciencia. Se presentan sus aportes y desarrollos de manera muy breve y en orden cronológico.

Bernoulli: En 1713, James Bernoulli presenta “Ars Conjectandi” (El Arte de la Conjetura), en el cual desarrolla una primera aproximación a la Teoría de la Probabilidad. En dicho trabajo, el concepto de probabilidad está asociado al límite de la frecuencia relativa, logrando probar la primer conexión matemática entre *probabilidad* y *frecuencia*, hoy en día conocida como *Ley débil de grandes números*.

Bayes: Thomas Bayes fue un hombre del clero y matemático amateur (uno muy bueno, aparentemente fue el primero en comprender la naturaleza de las expansiones asintóticas). Luego de su muerte, en 1761, entre sus trabajos fue hallado un curioso manuscrito sin publicar. No se sabe a ciencia cierta cual era la intención de Bayes con este manuscrito, o cuánto se editó el trabajo mientras pasaba de mano en mano hasta llegar a publicarse finalmente en 1763, pero con él se dio inicio a la “Estadística Bayesiana”. Ésta brindó soluciones completamente diferentes al problema inconcluso de Bernoulli.

Laplace: En uno de sus primeros trabajos (en 1774), Laplace “redescubrió” el principio de Bayes, y lo desarrolló e implementó con mayor claridad y generalidad. Durante los próximos

40 años utilizó la Estadística Bayesiana en problemas de Astronomía, Geodésicas, Meteorología, Estadísticas de Población e incluso Jurisprudencia.

El teorema básico aparece hoy en día de manera trivialmente simple, aún siendo por mucho el principio más importante sobre el que se desarrolla la inferencia científica.

Denotando varias proposiciones A , B y C ,

$$P(A|B \cap C) = \frac{P(A|C)P(B|A \cap C)}{P(B|C)}. \quad (1.4)$$

Lo importante de la ecuación (1.4) es que representa el “proceso de aprendizaje”: $P(A|C)$ representa la probabilidad “a priori” de A , cuando sólo conocemos C . $P(A|B \cap C)$ representa la probabilidad “a posteriori”, que representa la modificación del resultado cuando se incorpora nueva información B . En general, A representará alguna hipótesis o teoría, B representa la nueva información adquirida mediante observaciones, y la información “a priori” C representa la totalidad de lo que sabemos de A antes de la obtención de los datos B .

Un ejemplo famoso (un problema real que Laplace resolvió) es el del cálculo de la masa M_S de Saturno. En este caso, A representaría la hipótesis de que M_S pertenece a determinado intervalo, B consistiría en los datos obtenidos de los observatorios sobre las perturbaciones mutuas entre Júpiter y Saturno, y C la información obtenida del sentido común acerca del tamaño de Saturno (su masa no puede ser demasiado pequeña ya que sinó perdería sus anillos, ni demasiado grande ya que perturbaría el equilibrio del sistema solar). Laplace reportó, a finales del siglo XVIII, una estimación mediante el Teorema de Bayes, en la cual M_S era igual a $1/3512$ de la masa solar, y dio una probabilidad de 0.99991 de que el verdadero valor M_S se encuentra dentro del 1% de ese valor. Otros 150 años de acumulación de datos han elevado la estimación de Laplace en un 0.63% (es decir, dentro del intervalo hallado).

Luego de su muerte, todo el desarrollo analítico y los resultados obtenidos por Laplace fueron desestimados por la ausencia de explicaciones suficientemente claras de algunos conceptos importantes que consoliden su teoría.

Jeffreys: Temprano en el siglo XX, Harold Jeffreys redescubrió la lógica de Laplace y, en la década de 1930, explicó con mucha mayor solidez y claridad los conceptos subyacentes en el trabajo de Laplace. Pero no le fue fácil: durante los próximos treinta años estuvo bajo constante ataque de la comunidad académica de la época. Luego de 1960, la escuela Bayesiana comenzó a crecer, lento al principio, e incrementando su ritmo hasta llegar al ritmo vertiginoso de estos últimos años. Esta extraña historia es uno de los motivos por los cuales, quizás, hoy en día quienes realizan estadística Bayesiana sienten temor al defender su lógica.

Cox: En la década de 1930, el trabajo de Jeffreys por validar la estadística Bayesiana desató mordaces debates sobre este tema. Los frequentistas parecieron no tomar nota de las grandes masas de evidencia aportadas por Laplace y Jeffreys, en las cuales demostraron de manera pragmática su éxito cuando se aplican sin ser necesaria alguna conexión con la *frecuencia*. En 1946, un modesto y breve artículo, escrito por R. T. Cox, dejaba finalmente sentadas bases sólidas sobre las que se erige la estadística Bayesiana. Luego de 1946, aquellos que se seguían oponiendo a los métodos Bayesianos, estaban obligados a ignorar no sólo el éxito pragmático, sino también el teorema demostrado en este artículo.

1.2.2. Aspectos técnicos: Enfoque Bayesiano del problema de la estimación puntual

El enfoque Bayesiano supone que se tiene alguna información previa sobre el parámetro θ sobre el que se hacen inferencias. Esta información está expresada por medio de una distribución sobre θ , denominada distribución *a priori*. La distribución *a priori* puede tener distintas interpretaciones. Se pueden dar, por ejemplo, las siguientes alternativas:

- la distribución *a priori* está basada en experiencias previas similares,
- la distribución *a priori* expresa una creencia subjetiva.

El hecho de que el enfoque bayesiano considere una distribución de probabilidades sobre θ , supone tratar a θ como una variable aleatoria. En este caso, llamaremos Θ a la variable aleatoria sobre el parámetro, cuya distribución es $\tau(\theta)$. Una vez observada la muestra, se puede calcular la distribución condicional de la variable aleatoria Θ dada la muestra \mathcal{X} . Esta distribución se denomina distribución *a posteriori* y está definida a través de la función:

$$f_{\theta|\mathcal{X}}(\theta, \mathcal{X}) = \frac{f_{\mathcal{X}|\theta}(\mathcal{X}, \theta)\tau(\theta)}{\int f_{\mathcal{X}|\theta}(\mathcal{X}, s)\tau(s)ds}.$$

Definición. Una **función de pérdida** es una función $l(\theta, d)$ no negativa, con $l(\theta, \theta) = 0$, que nos indica cual es la pérdida cuando el valor del estimador es d y el valor del parámetro es θ . Luego, si usamos el estimador $\delta(\mathcal{X})$, la pérdida será

$$l(\theta, \delta(\mathcal{X}))$$

y esta pérdida será una variable aleatoria ya que depende de \mathcal{X} .

Definimos el *Riesgo Frecuentista* como

$$R(\theta, \delta) := \int_{\mathcal{X}} l(\theta, \delta(x))f(x|\theta)dx.$$

Notemos que la integral se realiza sobre el espacio \mathcal{X} .

Desde el enfoque Bayesiano, uno considera la *pérdida esperada a posteriori*:

$$\varrho(\tau, d|\mathcal{X}) := \int_{\Theta} l(\theta, d)\tau(\theta|\mathcal{X})d\theta,$$

el cual promedia el error (o la pérdida) de las distribuciones a posteriori para el parámetro θ , condicionadas a la muestra observada. Notemos que, en este caso, la integral se realiza sobre el espacio paramétrico.

Luego, definimos el *Riesgo Integrado*, el cual es el Riesgo Frecuentista promediado sobre los valores de θ con respecto a su distribución a posteriori τ :

$$r(\tau, \delta) := \int_{\Theta} \int_{\mathcal{X}} l(\theta, \delta(x))f(x|\theta)dx \tau(\theta)d\theta,$$

el cual podemos reescribir utilizando el Teorema de Fubini (ya que $l(\theta, \delta) \geq 0$):

$$\begin{aligned}
 r(\tau, \delta) &= \int_{\Theta} \int_{\mathcal{X}} l(\theta, \delta(x)) f(x|\theta) dx \tau(\theta) d\theta \\
 &= \int_{\mathcal{X}} \int_{\Theta} l(\theta, \delta(x)) f(x|\theta) \tau(\theta) d\theta dx \\
 &= \int_{\mathcal{X}} \int_{\Theta} l(\theta, \delta(x)) \tau(\theta|x) d\theta m(x) dx \\
 &= \int_{\mathcal{X}} \varrho(\tau, \delta(x)|x) m(x) dx
 \end{aligned}$$

donde $m(x)$ proviene de la completación de la distribución a posteriori. En vista de esta última expresión, resulta intuitivo proponer un estimador como aquel que minimice $r(\tau, \delta)$.

Definición. *Un estimador Bayesiano asociado a una distribución a priori τ y a una función de pérdida l es cualquier estimador δ^τ que minimice $r(\tau, \delta)$. Al valor $r(\tau) = r(\tau, \delta^\tau)$ se lo denomina riesgo de Bayes.*

Es importante destacar que, desde un punto de vista estrictamente Bayesiano, sólo la *pérdida esperada a posteriori* $\varrho(\tau, d|\mathcal{X})$ es importante, ya que el paradigma Bayesiano se basa en la distribución condicional. Notemos que, bajo funciones de pérdida estrictamente convexas, los estimadores Bayesianos son únicos.

El estimador Bayesiano utilizado con mayor frecuencia es el asociado a la función de pérdida cuadrática $l(\theta, \delta(\mathcal{X})) = (\theta - \delta(\mathcal{X}))^2$. Esta función de pérdida fue propuesta por Legendre (1805) y Gauss (1810), y presenta como principal cuestionamiento el hecho de que sobrepenaliza observaciones muy alejadas de la media. Una de sus ventajas es su forma convexa, y la simplicidad de trabajar con ella (argumento que utilizó Gauss en su trabajo de 1810 defendiendo su utilización, siendo consciente de la arbitrariedad de la propuesta). Bajo esta función de pérdida, el estimador Bayesiano resultante es la esperanza a posteriori. Es importante destacar que esta no es la única función de pérdida que conduce a este estimador Bayesiano.

Otro estimador Bayesiano que utilizaremos en el presente trabajo es el denominado *MAP*

Maximum a posteriori, definido como el modo de la posteriori. Este estimador es el resultante de la utilización de la función de pérdida 0 – 1:

$$l_{\epsilon}(\theta, \delta(\mathcal{X})) := \begin{cases} 0, & |\theta - d| \leq \epsilon, \\ 1, & |\theta - d| > \epsilon, \end{cases}$$

y tomando $\epsilon \rightarrow 0$ se obtiene dicho estimador. Este estimador puede expresarse como un *estimador máximo verosímil penalizado* en el sentido clásico, conservando sus propiedades de optimalidad asintóticas (suficiencia, consistencia), bajo algunas condiciones de regularidad de $f_{\mathcal{X}|\theta}(\mathcal{X}, \theta)$ y $\tau(\theta)$. Estas propiedades resultan intuitivas ya que, conforme el tamaño de la muestra aumenta, la información contenida en la muestra se vuelve predominante con respecto a la información a priori.

Información adicional sobre Estadística Bayesiana se puede consultar, por ejemplo, en Robert (2007).

A modo de ejemplo, consideremos los siguientes casos:

Exponencial: Consideremos una variable aleatoria con distribución $\mathcal{E}(\Psi)$, donde suponemos que $\Psi \sim \text{Gamma}(\alpha, \beta)$. Recordemos las expresiones obtenidas en la sección 1.1.2 para las distintas funciones en el caso de la distribución Exponencial:

$$f(t|\Psi = \psi) = \psi e^{-\psi t},$$

$$S(t|\Psi = \psi) = 1 - F(t) = 1 - (1 - e^{-\psi t}) = e^{-\psi t},$$

$$\lambda(t|\Psi = \psi) = \frac{f(t)}{S(t)} = \frac{\psi e^{-\psi t}}{e^{-\psi t}} = \psi,$$

$$\Lambda(t|\Psi = \psi) = -\log[S(t)] = -\log[e^{-\psi t}] = \psi t.$$

Luego, considerando una muestra $\mathcal{X} = \{t_i\}_{i \in \{1, \dots, n\}}$:

$$\begin{aligned}
 f_{\psi|\mathcal{X}} &= \frac{f(t_1, \dots, t_n, \psi)\Psi(\psi)}{\int f(t_1, \dots, t_n, y)\Psi(y)dy} \\
 &= \frac{\prod_{i=1}^n \psi e^{-\psi t_i} \frac{\beta^\alpha \psi^{\alpha-1} e^{-\beta\psi}}{\Gamma(\alpha)}}{\int \prod_{i=1}^n y e^{-y t_i} \frac{\beta^\alpha y^{\alpha-1} e^{-\beta y}}{\Gamma(\alpha)} dy} \\
 &= \frac{\psi^n e^{-\psi \sum_{i=1}^n t_i} \psi^{\alpha-1} e^{-\beta\psi}}{\int y^n e^{-y \sum_{i=1}^n t_i} y^{\alpha-1} e^{-\beta y} dy} \\
 &= \frac{[(\sum_{i=1}^n t_i) + \beta]^{n+\alpha} \psi^{(n+\alpha)-1} e^{-[(\sum_{i=1}^n t_i) + \beta]\psi}}{\Gamma(n + \alpha)}.
 \end{aligned}$$

Es decir $f_{\psi|\mathcal{X}}$ es la función de densidad de una v.a. con distribución $Gamma(n+\alpha, (\sum_{i=1}^n t_i) + \beta)$. En este caso, advertimos que partimos de una priori $Gamma$, y llegamos a una posteriori que se encuentra en la misma familia, ya que también es $Gamma$. Se dice entonces que la familia de distribuciones $Gamma$ es conjugada de la familia de distribuciones $Exponencial$.

Luego, el estimador Bayesiano de ψ correspondiente a la función de pérdida $l(\Theta, d) = (q(\Theta) - d)^2$ es la esperanza de la posteriori. Es decir:

$$\widehat{\psi}^B = \frac{n + \alpha}{(\sum_{i=1}^n t_i) + \beta}.$$

Consideremos, ahora, un caso más general.

Exponencial con censura: Sea $\{(t_1, \delta_1), \dots, (t_n, \delta_n)\}$ el conjunto de valores observados de una muestra i.i.d. $\{(T_1, \Delta_1), \dots, (T_n, \Delta_n)\}$, donde los tiempos reales $T_i^* \sim \mathcal{E}(\psi)$. Nuevamente, mediante las expresiones halladas en 1.1.2, expresamos la función de verosimilitud de la forma:

$$\mathcal{L}_n(\psi|(t_1, \delta_1), \dots, (t_n, \delta_n)) = \prod_{l=1}^n \psi^{\delta_l} e^{-\psi t_l},$$

donde consideramos a ψ como una realización de la variable aleatoria Ψ . Consideramos como distribución a priori para $\Psi \sim Gamma(\alpha, \beta)$:

$$f_{\Psi}(\psi) = \frac{\psi^{\alpha-1} \beta^{\alpha} e^{-\psi\beta}}{\Gamma(\alpha)}, \text{ con } \alpha, \beta > 0.$$

De manera análoga al caso de la distribución exponencial sin censura, se obtiene que la densidad a posteriori no es más que la función de densidad de una variable aleatoria con distribución *Gamma* $(\alpha + \sum_{l=1}^n \delta_l, \beta + \sum_{l=1}^n t_l)$.

Por consiguiente, el estimador Bayesiano correspondiente a la función de pérdida cuadrática es:

$$\hat{\psi}^B = \frac{\alpha + \sum_{l=1}^n \delta_l}{\beta + \sum_{l=1}^n t_l}.$$

2. EL MODELO DE RIESGO ADITIVO

De ahora en adelante, nos dedicaremos al estudio de muestras del tipo $(T_1, \Delta_1), \dots, (T_n, \Delta_n)$ independientes, donde la *función de riesgo* de T^* asociada a un vector de covariables $Z = z$ está dada por:

$$\lambda(t, \boldsymbol{\beta}) = \lambda_0(t) + \boldsymbol{\beta}'z,$$

con λ_0 función de riesgo arbitraria, que determina una *baseline hazard* con densidad f_0 . En adelante, denotaremos por $\lambda(t)$ en vez de $\lambda(t, \boldsymbol{\beta})$ a la función de riesgo del modelo aditivo, salvo que sea necesario especificar el conjunto de parámetros.

A diferencia del modelo de Cox (donde los riesgos son proporcionales), advertimos que en este caso los riesgos son *paralelos*, ya que la diferencia de riesgos entre dos elementos distintos de la muestra con vectores de covariables z_1 y z_2 que difieren sólo en la i -ésima componente, definimos $((\lambda_0(t) + \boldsymbol{\beta}'z_1) - (\lambda_0(t) + \boldsymbol{\beta}'z_2)) = \boldsymbol{\beta}'z_1 - \boldsymbol{\beta}'z_2 = \beta_i(z_{1i} - z_{2i}))$ es una función constante en t .

Consideraremos una aproximación paramétrica a la componente no paramétrica del modelo λ_0 , que se presenta a continuación:

$$\lambda_0(t) = \begin{cases} a_1, & \text{si } 0 \leq t < s_1, \\ a_2, & \text{si } s_1 \leq t < s_2, \\ \vdots & \\ a_J, & \text{si } s_{J-1} \leq t, \end{cases}$$

con s_1, \dots, s_{J-1} conocidas (J fijo). Una restricción necesaria para este modelo es que $a_J > 0$. Esto es debido a que $S(t) \xrightarrow{t \rightarrow \infty} 0$, por lo que $\Lambda(t) \xrightarrow{t \rightarrow \infty} \infty$ (lo que implica que $a_J > 0$). Advertimos que esta expresión, llamada *constante a trozos*, representa una aproximación paramétrica mediante la función de riesgo de una variable aleatoria *Exponencial a trozos* (lo que motivó el hecho de asignarle el subíndice 0 a la función de riesgo de la Exponencial a trozos definida en la ecuación (1.2)).

Analicemos la forma que toman las distintas funciones bajo este modelo. Integrando la función de riesgo, obtenemos una expresión para la Función de Riesgo Acumulado, la cual resulta en:

$$\Lambda(t) = \int_0^t \lambda(u) du = \int_0^t [\lambda_0(u) + \beta' z] du = \Lambda_0(t) + \beta' z t - \Lambda_0(0) = \Lambda_0(t) + \beta' z t.$$

En base a esta, calculamos la Función de Supervivencia:

$$S(t) = e^{-\Lambda(t)} = e^{-\Lambda_0(t) - \beta' z t} = S_0(t) e^{-\beta' z t}.$$

Hallamos su Función de Densidad calculando el producto entre las funciones de riesgo y supervivencia:

$$f(t) = \lambda(t)S(t) = [\lambda_0(t) + \beta' z] S_0(t) e^{-\beta' z t} = [\lambda_0(t) S_0(t) + S_0(t) \beta' z] e^{-\beta' z t}.$$

Por último, calculamos la función de Verosimilitud para el AHM (recordando la ecuación (1.1)):

$$\begin{aligned} \mathcal{L}_n(a_1, \dots, a_J, \beta | (t_1, \delta_1), \dots, (t_n, \delta_n)) &\propto \prod_{l=1}^n \lambda(t_l)^{\delta_l} S(t_l) \\ &= \prod_{l=1}^n [\lambda_0(t_l) + \beta' z_l]^{\delta_l} S_0(t_l) e^{-\beta' z_l t_l} \\ &= \left\{ \prod_{l=1}^n S_0(t_l) \right\} \left\{ e^{-\beta' \sum_{l=1}^n z_l t_l} \prod_{l=1}^n [\lambda_0(t_l) + \beta' z_l]^{\delta_l} \right\}. \end{aligned}$$

Cuando no haya lugar a confusión, utilizaremos sólo la expresión \mathcal{L}_n , sin especificar el argumento.

A continuación introduciremos la *función de Intensidad* la cual es habitualmente utilizada para caracterizar modelos de supervivencia.

Definición. *Dado un proceso de conteo*

$$N(t) = \text{“número de eventos en } [0, t]\text{”},$$

la **función de Intensidad** definida acorde al proceso es:

$$I(t) = Y(t)\lambda(t), \tag{2.1}$$

donde $Y(t)$ es llamada función “en riesgo” y está definida como

$$Y(t) = 1_{(T^* > t, C > t)}(t) = \begin{cases} 1, & \text{si el evento no se presentó ni se censuró al instante } t, \\ 0, & \text{en caso contrario.} \end{cases}$$

Un tratamiento detallado acerca de procesos de conteo en modelos de supervivencia se puede encontrar, por ejemplo, en Andersen *et al.* (1993) o Fleming & Harrington (2011). Esta función será ampliamente utilizada cuando presentemos los trabajos realizados por diferentes autores sobre este modelo.

La expresión de la *función de Intensidad* definida en base al proceso de conteo $N(t)$ para el AHM es la siguiente:

$$I(t) = Y(t) [\lambda_0(t) + z'\beta].$$

Los problemas usuales al tratar con el modelo aditivo son:

- la restricción impuesta a la componente paramétrica del modelo ($z'\beta$) para garantizar que la función de riesgo sea no negativa,
- la expresión poco tratable que adquiere la verosimilitud.

2.1. Estimación Bayesiana en Modelos de Riesgo Aditivo mediante el modo de la posteriori

A continuación, presentaremos un primer abordaje Bayesiano al modelo, a través de un ejemplo sencillo.

En el caso de la *baseline hazard*, consideraremos el modelo *constante a trozos*, llamando A_j al valor de la *baseline hazard* en el intervalo $[s_{j-1}, s_j)$.

Consideraremos A_1, \dots, A_J i.i.d., con distribución a priori $A_i \sim \mathcal{E}(\alpha), \forall i \in \{1, \dots, J\}$.

En el caso de la componente paramétrica, consideraremos una sola covariable: $B \sim \mathcal{E}(b)$, y llamaremos a_1, \dots, a_j y β a los respectivos valores tomados por las variables A_1, \dots, A_j y B .

Suponiendo independencia entre ambas componentes, es decir:

$$A_1, \dots, A_J \perp B,$$

la posteriori resultante es:

$$\begin{aligned} f_{A_1, \dots, A_J, B | \mathcal{X}} &\propto \left\{ \prod_{l=1}^n S_0(t_l) \right\} e^{-\beta \sum_{i=1}^n z_i t_i} \prod_{l=1}^n (\lambda_0(t_l) + \beta z_l)^{\delta_l} \alpha^J e^{-\alpha \sum_{j=1}^J a_j} b e^{-b\beta} \\ &= \left\{ \prod_{l=1}^n S_0(t_l) \right\} e^{-\beta(\sum_{i=1}^n z_i t_i + b)} \prod_{l=1}^n (\lambda_0(t_l) + \beta z_l)^{\delta_l} \alpha^J e^{-\alpha \sum_{j=1}^J a_j} b \\ &= f_0((t_1, \delta_1), \dots, (t_n, \delta_n), a_1, \dots, a_J, \beta), \end{aligned}$$

donde la expresión f_0 expresa a la posteriori como función de la muestra y los valores a priori. La propuesta en este caso será hallar los estimadores Bayesianos correspondientes a la función de pérdida 0 – 1, es decir, el estimador *MAP* para la totalidad de los parámetros del modelo.

En el caso particular¹ en el que el parámetro $b = 1$ y el parámetro que genera las alturas

¹ Valores elegidos como referencia para describir el método.

de las funciones de riesgo es $\alpha = 1$, la función arriba enunciada adquiere la forma:

$$f_0((t_1, \delta_1), \dots, (t_n, \delta_n), a_1, \dots, a_J, \beta) = \left\{ \prod_{l=1}^n S_0(t_l) \right\} e^{-\beta(\sum_{l=1}^n z_l t_l + 1)} \prod_{l=1}^n (\lambda_0(t_l) + \beta z_l)^{\delta_l} e^{-\sum_{j=1}^J a_j}.$$

Llamaremos:

$$\begin{aligned} g(a_1, \dots, a_J, \beta) &= \log[f_0((t_1, \delta_1), \dots, (t_n, \delta_n), a_1, \dots, a_J, \beta)] \\ &= \sum_{l=1}^n \log S_0(t_l) - \beta \left(\sum_{l=1}^n z_l t_l + 1 \right) - \sum_{j=1}^J a_j + \sum_{l=1}^n \delta_l \log(\lambda_0(t_l) + \beta z_l). \end{aligned}$$

Nuestro objetivo se ha convertido ahora en hallar los valores de a_1, \dots, a_J, β que maximicen esta función g . A continuación, dichos parámetros serán hallados de manera numérica mediante el algoritmo de optimización denominado *nlnmb* del programa de código abierto R.

2.1.1. Simulación

Para probar la eficacia del proceso, consideremos como caso particular la partición de la recta real positiva dada por los 7 puntos de corte 0.25, 0.5, 0.75, 1, 1.5, 2 y 3. Estos son datos que, en el modelo actual, suponemos conocidos. Luego, generamos (en función de un n que haremos variar acorde al tamaño de la muestra que deseemos simular) las *covariables* (que denotaremos por el vector n -dimensional z) con distribución prefijada χ_1^2 , y los tiempos reales (que denotaremos por el vector n -dimensional TS) cuya distribución será $\mathcal{E}(\lambda_0(t) + \beta z)$ (ya que consideramos $\lambda_0(t)$ constante e igual a 1), donde el β será fijado en el valor 1. Estos valores fueron elegidos como referencia para presentar las simulaciones. Luego, generamos las censuras (denotadas por el vector n -dimensional c), con distribución fijada en $\mathcal{E}(1/2)$ (siendo susceptible de cambiarse por cualquier distribución que se ajuste al modelo que se desee estimar). Todos estos serán considerados datos desconocidos, pero a partir de los cuales serán generados los vectores n -dimensionales T (los tiempos observados) y δ (indicador de las censuras), que serán los datos observados en la muestra. De esta forma, por la distribución que

les asignamos a los tiempos reales TS , la *baseline hazard* en cada intervalo (es decir, los a_j , con $j = 1, \dots, 8$) de la partición considerada en este caso particular será la función constante 1.

Para testear la eficacia del procedimiento, es decir, que tan bien estiman los estimadores Bayesianos obtenidos de manera numérica mediante el algoritmo de maximización, presentamos una tabla (2.1), donde varía el tamaño de la muestra n . La cantidad de veces que se repite el procedimiento (réplicas) es 1000.

Tab. 2.1: Cantidad de réplicas 1000

	$n = 100$		$n = 500$		$n = 1000$	
	Media	Desvío estándar	Media	Desvío estándar	Media	Desvío estándar
a_1	0.88935	0.2526	0.9867	0.1373	0.9906	0.0956
a_2	0.8896	0.3426	0.9788	0.1675	0.9859	0.1187
a_3	0.8025	0.3518	0.9604	0.1997	0.9803	0.1406
a_4	0.7678	0.4388	0.9498	0.2392	0.9648	0.1705
a_5	0.7874	0.4288	0.9484	0.2309	0.9702	0.1676
a_6	0.7338	0.4390	0.8695	0.32918	0.9339	0.2510
a_7	0.7505	0.4675	0.8531	0.3805	0.9317	0.2919
a_8	0.5172	0.2152	0.6762	0.3808	0.8092	0.4244
β	1.1474	0.3166	1.0119	0.1331	1.0129	0.0920

En la tabla se observa como mejora la eficacia de los estimadores conforme aumenta el tamaño de la muestra.

Los resultados hallados exhiben el aumento de la eficiencia de los estimadores Bayesianos conforme aumenta el tamaño de la muestra. Además, evidencia una disminución significativa en el desvío estándar, lo que muestra de manera empírica la consistencia de los estimadores.

Es importante notar que los estimadores funcionan mejor en los primeros intervalos. Esto se debe a que la cantidad de observaciones con las que se estiman dichos parámetros es mayor, disminuyendo el desvío estándar. Dado que es sólo una introducción, no presentamos un análisis de como influye la selección de prioris en los estimadores (detallaremos esto en el Capítulo 3).

2.2. Bibliografía en el área

Si bien se define el “estimador Bayesiano” asociado a una función de pérdida, en la literatura lo más usual es utilizar la función de pérdida cuadrática. A continuación, presentamos una revisión minuciosa de la bibliografía en el área, donde apreciaremos que en todos los casos se calcula dicho estimador mediante la esperanza de la posteriori.

Una revisión del análisis del modelo aditivo desde una perspectiva frecuentista fue incluida en el trabajo de Martinussen & Scheike (2006), y desde el enfoque frecuentista robusto, una revisión de los tres modelos fue recopilada por Álvarez & Ferrario (2012).

Desde una perspectiva Bayesiana, el PHM de Cox ha sido el más tratado. Una revisión de esto se encuentra en el trabajo de Sinha & Dey (1997). Con respecto al AFTM, una recopilación de su tratamiento Bayesiano fue incluido en Zhang & Lawson (2011). Sin embargo, el AHM ha ganado atención como una de las principales alternativas al tradicional PHM, especialmente cuando la hipótesis de riesgos proporcionales no se satisface. Los diferentes modelos representan distintos aspectos en la relación entre la variable respuesta y las covariables. En el caso del modelo aditivo, los riesgos son paralelos, mientras que en el multiplicativo son proporcionales.

En esta revisión del tratamiento Bayesiano en el AHM, detallaremos las principales difi-

cultades con las que se encuentran los autores al tratar con este modelo, y las diferentes alternativas por las que optan para sortear dichos problemas.

2.2.1. Principales trabajos de inferencia Bayesiana en el modelo de riesgos aditivos

El primer trabajo en el área fue presentado por Beamonte & Bermúdez (2003).

Una modelización particular del AHM fue considerada por los autores, denominada “modelo gamma poligonal”, en la cual la función de riesgo adquiere la expresión:

$$\lambda(t|Z = z) = \lambda_0(t) + \lambda_1(t|Z = z).$$

Como en el modelo de función de riesgo *constante a trozos*, el eje real se particiona acorde a ciertos puntos específicos. En cada partición de la grilla, la expresión que adquiere la baseline es la de una función lineal, las cuales se unen en los puntos de corte, dando así una forma poligonal. Dicha expresión es la siguiente:

$$\lambda_0(t) = \begin{cases} A_j + \frac{(A_{j+1}-A_j)(t-s_{j-1})}{s_j-s_{j-1}}, & \text{si } s_{j-1} \leq t < s_j, j = 1, \dots, J-1, \\ A_J, & \text{si } t \geq s_{J-1}. \end{cases}$$

La componente paramétrica del modelo $\lambda_1(t)$ es la función de riesgo de una variable con distribución *Gamma* y parámetros α y β , i.e.,

$$\lambda_1(t|Z = z) = \frac{t^{\alpha-1} \exp(-\beta t)}{\int_t^\infty u^{\alpha-1} \exp(-\beta u) du}, \text{ si } t > 0.$$

Una suposición impuesta es que tanto α como β son específicos para cada individuo de la población, y están relacionados con el vector de covariables Z a través de un modelo probabilístico. Luego, una estructura jerárquica es introducida al modelo, cuyo segundo nivel está dado por:

$$\begin{aligned} \frac{\alpha}{\beta} | \beta, z &\sim \mathcal{LN}(b'z, \sigma_\alpha^2), \\ \beta &\sim \mathcal{LN}(\mu_\beta, \sigma_\beta^2), \end{aligned}$$

donde \mathcal{LN} denota la distribución *log Normal*. Los hiperparámetros $b, \sigma_\alpha^2, \mu_\beta$ y σ_β^2 son considerados constantes desconocidas comunes a la población.

Para realizar el análisis Bayesiano, los autores seleccionaron como prioris para el siguiente nivel jerárquico de los parámetros del modelo (donde \mathcal{IG} denota a la distribución *Gamma Inversa*):

$$\begin{aligned}\mu_\beta | \sigma_\beta^2 &\sim \mathcal{N}(m_\beta, v_\beta^2 \sigma_\beta^2), \\ \sigma_\beta^2 &\sim \mathcal{IG}(a_\beta, b_\beta), \\ b | \sigma_\alpha^2 &\sim \mathcal{N}_p(m_\alpha, V_\alpha \sigma_\alpha^2), \\ \sigma_\alpha^2 &\sim \mathcal{IG}(a_\alpha, b_\alpha).\end{aligned}$$

Para los parámetros de la baseline, se seleccionó un proceso autocorrelacionado de primer orden:

$$A_j = A_{j-1} \exp(\epsilon_j), j = 2, \dots, J,$$

donde $(\epsilon_2, \dots, \epsilon_J)$ se asumen independientes, con distribución *Normal*, de media 0 y varianza σ_ϵ^2 , donde:

$$\begin{aligned}A_1 &\sim \text{Gamma}(a_A, b_A), \\ \sigma_\epsilon^2 &\sim \mathcal{IG}(a_\epsilon, b_\epsilon).\end{aligned}$$

La posteriori obtenida resulta intratable (de manera analítica). Sin embargo, las condicionales de la posteriori para los parámetros $(\mu_\beta, \sigma_\beta^2)$, $(\mu_\alpha, \sigma_\alpha^2)$ y σ_ϵ son conocidas, por lo que un análisis mediante simulaciones usando muestreo de Gibbs (Geman & Geman, 1984) o alguna otra técnica del estilo MCMC (*Markov Chain Monte Carlo*²) es posible. De las otras condicionales es posible generar muestras utilizando un algoritmo Metropolis-Hastings (Metrópolis *et al*,

² Son Métodos Monte Carlo basados en Cadenas de Markov. Estas técnicas de simulación poseen una amplia aplicación en la generación de estimadores Bayesianos, ya que permiten obtener muestras aleatorias de las distribuciones *a posteriori*.

1953; Hastings, 1970). Finalmente, se pueden obtener muestras de la posteriori a través de un algoritmo Metropolis-within-Gibbs.

Algunas observaciones de este modelo:

- el modelo Gamma-poligonal permite modelizar cierta heterogeneidad en la población (otra manera de incorporar la heterogeneidad sería introduciendo “frailties”, que se presentarán más adelante en este capítulo),
- la función de riesgo baseline poligonal es apenas más compleja que el modelo constante a trozos (en sentido algorítmico) y posee la ventaja de ser continuo,
- considerando la componente paramétrica del modelo como una mezcla de distribuciones *Gamma* se logra evitar la restricción de la no-negatividad de $z'\beta \geq 0$,
- si los parámetros de regresión son independientes del tiempo, la componente paramétrica es constante en t , llevando a la conclusión de que nos encontraríamos con un modelo exponencial.

Por último, los autores aplicaron estos resultados a una base de datos de desempleo, en la cual se consideraron 559 cuestionarios de personas que se graduaron en la “Consellería de Cultura, Educación y Ciencia” o en la “Universidad de Valencia” (España) entre los años 1978 y 1993. La variable de interés es el tiempo (en meses) desde la graduación hasta que se obtiene el primer empleo.

Un trabajo con un enfoque que no presenta relación cercana con el trabajo previo fue desarrollado por Dunson & Herring (2005). En este caso, los autores proponen una modelización distinta. Motivados por los inconvenientes de seleccionar una priori truncada para ajustarse a la no-negatividad de la función de riesgo, redefinen el modelo acorde a ciertos indicadores aplicados a las covariables, los cuales son especificados de acuerdo a si se espera que dicha covariable esté asociada con una reducción o incremento (o si no tiene efecto) en el riesgo.

Este enfoque es el único en la literatura existente hasta el momento que considera como priori una distribución multinomial relacionada a lo que se espera de cada covariable. El indicador asociado a la k -ésima covariable se denota por M_k y puede tomar los valores:

$$M_k = \begin{cases} -1, & \text{si la } k\text{-ésima covariable está asociada con una disminución del riesgo,} \\ 0, & \text{si la } k\text{-ésima covariable no está asociada con un cambio en el riesgo,} \\ 1, & \text{si la } k\text{-ésima covariable está asociada con un incremento del riesgo.} \end{cases}$$

El vector de covariables z es “estandarizado” de manera que $|z_i| \leq 1, \forall i$. Esta estandarización puede realizarse, en el caso de variables acotadas, restando el mínimo posible valor (de manera que el mínimo sea 0) y dividiendo por el rango. En caso de variables no acotadas, los autores proponen considerar la variable restringida a un intervalo acotado, de manera de contener todos los valores de interés que se quieran incluir en el análisis, y realizar el mismo procedimiento, no descartando otras opciones que resulten coherentes para obtener la estandarización. Luego, implementando estas modificaciones, la función de riesgo adquiere la expresión:

$$\lambda(t) = \lambda_0^*(t) + \sum_{k=1}^p \{1_{(M_k=-1)}(1 - z_{ik}) + 1_{(M_k=1)}z_{ik}\} \beta_k^* = \lambda_0^*(t) + z_i^{*'} \beta^*,$$

donde $z_{ik}^* = 1_{(M_k=-1)}(1 - z_{ik}) + 1_{(M_k=1)}z_{ik}$, $\lambda_0^*(t) = \lambda_0(t) + \sum_{k=1}^p 1_{(M_k=-1)}\beta_k^*$ y β^* denota el vector con componentes $\beta_1^*, \dots, \beta_p^*$. La verosimilitud considerada por los autores está basada en la intensidad de un proceso de conteo $N(t)$, y está definida por:

$$\prod_{i=1}^n \left(\prod_{t \neq 0} [Y_i(t) \{\lambda_0^*(t) + z_i^{*'} \beta^*\}]^{dN_i(t)} \right) \exp \left(- \int_{t \neq 0} Y_i(t) \{\lambda_0^*(t) + z_i^{*'} \beta^*\} dt \right).$$

donde $dN_i(t)$ denota el incremento de $N_i(t)$ sobre el pequeño intervalo $[t, t + dt)$.

Dado que el incremento es infinitesimal, los $dN_i(t)$ contribuyen a la verosimilitud de la misma manera que variables Poisson independientes, aún cuando para cada t , $dN_i(t) \leq 1$.

Llamando t_1, \dots, t_{J-1} a los únicos tiempos de falla observados en la muestra, y considerando

el modelo *constante a trozos* con una grilla definida por estos tiempos (es decir, $s_j = t_j, j = 1, \dots, J - 1$), la verosimilitud puede ser re-escrita como:

$$\prod_{i=1}^n \prod_{j=1}^{J-1} \left(\prod_{t \in [s_{j-1}, s_j)} [Y_i(t) \{\lambda_0^*(t) + z_i^{*'} \beta^*\}]^{dN_i(t)} \right) \exp \left(- \int_{t \in [s_{j-1}, s_j)} Y_i(t) \{\lambda_0^*(t) + z_i^{*'} \beta^*\} dt \right).$$

Bajo el supuesto de que el riesgo acumulado en el intervalo $[s_{j-1}, s_j)$ es pequeño (es decir, no hay empates, y la muestra no es demasiado chica),

$$\forall i, j, \quad \int_{s_{j-1}}^{s_j} Y_i(t) \{\lambda_0^*(t) + z_i^{*'} \beta^*\} dt \approx 0.$$

Luego, la contribución a la verosimilitud de los individuos en riesgo a lo largo de este intervalo es aproximadamente

$$\{d\Lambda_{0j} + z_i^{*'} \beta^*(s_j - s_{j-1})\}^{dN_{i,j}} \exp(-\{d\Lambda_{0j} + z_i^{*'} \beta^*(s_j - s_{j-1})\}),$$

donde $d\Lambda_{0j} = \int_{s_{j-1}}^{s_j} \lambda_0^*(t) dt$ y $dN_{i,j} = 1$ (si el sujeto i representa una falla en el instante s_j), o resulta 0 en caso contrario.

Así, una nueva aproximación a la verosimilitud está dada por

$$\prod_{i=1}^n \prod_{j: Y_j=1} \{d\Lambda_{0j} + z_i^{*'} \beta^*(s_j - s_{j-1})\}^{dN_{i,j}} \exp(-\{d\Lambda_{0j} + z_i^{*'} \beta^*(s_j - s_{j-1})\}).$$

Las prioris seleccionadas para la baseline hazard $\lambda_0^*(t)$, los indicadores del modelo $M = (M_1, \dots, M_p)$, y el vector de coeficientes regresores β^* , son:

- distribuciones *Gamma* independientes para cada $A_j, j = 1, \dots, J$ en la baseline hazard, (las cuales fueron asumidas independientes en el modelo),
- distribuciones Multinomiales independientes para los M_k :

$$M_k = \begin{cases} -1, & \text{con probabilidad } \theta_{k,-1}, \\ 0, & \text{con probabilidad } \theta_{k,0}, \\ 1, & \text{con probabilidad } \theta_{k,1}, \end{cases}$$

donde $k = 1, \dots, p$, y $\theta_{k,-1} + \theta_{k,0} + \theta_{k,1} = 1$,

- distribuciones *Gamma* independientes para los parámetros regresores.

Los hiperparámetros de estas prioris son seleccionados en un contexto de elucidación de prioris (i.e. traducción matemática del conocimiento experto o de datos previos). Esta es una contribución importante de los autores en el contexto Bayesiano.

Los indicadores del modelo indican la dirección del efecto (positiva o negativa) de las diferentes covariables, pero no cuantifican la magnitud de dichos efectos. Dichas intensidades son medidas por los parámetros regresores.

Para simplificar y hacer más eficiente la generación del algoritmo, un enfoque de “data augmentation” es utilizado. Bajo los supuestos considerados, los dN_{ij} están distribuidos como variables aleatorias *Poisson* independientes:

$$dN_{ij} \stackrel{ind}{\sim} Poisson \left(d\Lambda_{0j} + \sum_{k=1}^p z_{ik}^* \beta_k^* \right), \quad \forall i, j : Y_{ij} = 1. \quad (2.2)$$

Esta expresión conduce a unas distribuciones a posteriori no estándar, dificultando no sólo un tratamiento analítico, sino que no es posible generar muestras aleatorias de ella. Por este motivo, variables independientes *Poisson* latentes son incluidas en el modelo, de la siguiente manera:

$$dN_{ij} = dN_{ij0} + \sum_{k=1}^p I_{(M_k \neq 0)} dN_{ijk}, \quad \forall i, j : Y_{ij} = 1,$$

donde

$$\begin{aligned} dN_{ij0} &\sim Poisson(d\Lambda_{0j}), \\ dN_{ijk} &\sim Poisson((s_j - s_{j-1})z_{ik}^* \beta_k^*). \end{aligned}$$

Usando la propiedad de la suma de variables aleatorias independientes *Poisson* es también *Poisson*, se sigue que integrando sobre las variables latentes obtenemos el equivalente a la expresión previa (2.2). Esta formulación tiene como ventaja el hecho de poder utilizar la conjugación *Poisson–Gamma* para obtener distribuciones posterioris condicionales simples.

En este punto, mediante un algoritmo de Gibbs se pueden generar muestras aleatorias de las condicionales a posteriori, y por ende de la full posteriori misma.

Los autores, además, generalizaron estas ideas para permitir un análisis Bayesiano en un modelo más general, el cual llamaron “Modelo de Riesgo Aditivo/Multiplicativo”, cuya función de riesgo resulta de la forma:

$$\lambda(t, z) = \lambda_0(t) \exp(z'\alpha) + z'\beta,$$

donde $\alpha = (\alpha_1, \dots, \alpha_p)'$ son los coeficientes de riesgo proporcionales, y el resto de los parámetros son los mismos que los definidos previamente. Este modelo es más general que los modelos tanto aditivo como multiplicativo, ya que contiene a ambos. Si $\alpha_i = 0$, para todo i , nos encontramos con el modelo aditivo y, si $\beta_i = 0$, para todo i , entonces nos encontramos con la expresión del modelo de riesgos proporcionales. Por conveniencia de notación, el modelo se define de manera que las mismas covariables sean incluidas tanto en las componentes aditiva como proporcional. Las prioris son adaptadas para incluir las componentes multiplicativas, y luego se generaliza el algoritmo de Gibbs propuesto previamente acorde a esto.

En la selección de la priori para $(\lambda_0, M, \alpha, \beta)$ (asumiendo independencia):

$$\tau(\lambda_0, M, \alpha, \beta) = \tau(\lambda_0)\tau(M, \beta)\tau(\alpha).$$

Las prioris $\tau(\lambda_0)$ y $\tau(M, \beta)$ elegidas son las mismas que en el modelo aditivo. Para inducir una priori para los componentes del modelo de riesgos proporcionales $\tau(\alpha)$, los autores consideran una priori definida previamente por ellos mismos para realizar inferencia Bayesiana en el modelo de Cox, a la que llamaron “one-inflated truncated gamma density”.

En una sección de Simulaciones, los autores verificaron que el nuevo modelo funciona correctamente, pudiendo discriminar entre las componentes aditivas y multiplicativas, mostrando que este nuevo enfoque goza de buenas propiedades (desde el enfoque frecuentista).

En una sección de aplicación a datos reales, los autores llevaron adelante un estudio sobre

una base de datos de pacientes con enfermedad coronaria (corazón) de 1571 individuos, quienes estuvieron libres de enfermedad hasta al menos los 45 años. El foco fue puesto en la edad en que se presentaba la enfermedad en pacientes mujeres con relación a ciertas covariables (hipertensión, niveles de colesterol, sobrepeso y obesidad). El estudio reveló que algunas covariables tienen un efecto distinto al aditivo o multiplicativo, concluyendo que el enfoque del modelo aditivo-multiplicativo parece ser el más razonable en este caso.

Una revisión de los dos artículos comentados hasta ahora fue incluida en el trabajo de tesis de Saad (2012), quien sugirió algunas modificaciones:

- En el modelo Gamma-poligonal de Beamonte & Bermúdez (2003), propone como alternativa una baseline hazard del tipo constante a trozos, y como priori para estos un proceso *Gamma*.
- En el modelo propuesto por Dunson & Herring (2005), sugiere un proceso *Gamma* como priori para los incrementos de la baseline. Además, adjunta códigos desarrollados en WinBUGS acordes al modelo. Luego, presenta una sección de análisis de datos reales en una base de pacientes de melanoma y cáncer de pulmón.

2.2.2. Inclusión de Frailties

Los primeros en incluir *frailties* en el modelo fueron Silva & Amaral Turkman (2005).

La inclusión de *frailties* (*fragilidades*) es motivada por el hecho de que en el análisis de Supervivencia se considera el supuesto de independencia (dadas las covariables) entre diferentes individuos, y que la heterogeneidad del modelo puede ser explicada en términos de las covariables observadas. Estas suposiciones subyacentes pueden no ser acordes debido a la influencia de cierta heterogeneidad no observada entre los individuos considerados para el estudio. Para tratar con estas potenciales limitaciones, un efecto aleatorio (*frailty*) es incluido en la función de riesgo. La incorporación de una *frailty* w ($w > 0$) en el modelo es

usualmente realizada de manera multiplicativa (es decir, la función de riesgo es multiplicada por w o por una función de w , como e^w). En este trabajo, las *frailties* son incorporadas de manera aditiva al AHM, donde los datos se consideran divididos en k grupos, incluyendo de esta manera el vector de *frailties* $w = (w_1, w_2, \dots, w_k)'$ al modelo, para luego proceder al análisis Bayesiano del nuevo “modelo de Supervivencia aditivo con *frailty*”.

De acuerdo con esta inclusión de *frailties*, la expresión de la función de intensidad resulta:

$$I_i(t|z_i, w_{l_i}) = Y_i(t) [\lambda_0(t) + \beta' z_i(t) + w_{l_i}], i = 1, \dots, p,$$

donde $N_i(t)$ es el número de ocurrencias de un evento de interés en el instante t , $l_i \in \{1, \dots, k\}$ y $Y_i(t)$ toma valores 1 o 0 de acuerdo a si el individuo i está en riesgo en el instante t , respectivamente. Prioris *Gamma* fueron consideradas para el vector de *frailties* w_1, w_2, \dots, w_k , las cuales se consideran i.i.d. Luego, la distribución conjunta de las *frailties* está dada por:

$$[\Gamma(a)]^{-k} b^{ka} \left(\prod_{l=1}^k w_l \right)^{a-1} \exp \left(-b \sum_{l=1}^k w_l \right),$$

donde a y b son los parámetros de forma y escala respectivamente. Estos hiperparámetros se pueden estimar de los datos. A los parámetros *frailty* se los suele considerar con media uno, para evitar problemas de identificabilidad con los modelos de *frailties* multiplicativas. La variabilidad de las *frailties* puede ser interpretada como el grado de heterogeneidad entre individuos. Los autores consideraron covariables dependientes del tiempo, aportando una flexibilidad extra al modelo. La misma idea que Dunson & Herring (2005) y Saad (2012) fue utilizada para la selección de prioris para la baseline y, de la misma manera, una función de regresión acumulada se define para cada covariable dependiente del tiempo, permitiendo considerar un proceso de incrementos a ellos.

Para modelizar la baseline hazard $\lambda_0(t)$, los autores adoptaron el modelo constante a trozos. Las “funciones de regresión acumuladas” fueron definidas acordes a esto y a una

grilla pre-seleccionada:

$$\Omega_q(t) = \int_0^t \beta_q(u) du, t \geq 0, q = 1, \dots, p.$$

Procesos a priori *Gamma* son considerados para los parámetros de la baseline y para las funciones de regresión acumuladas. O sea, para cada $\Lambda_{0j}, \Omega_{qj}$ (donde estos parámetros denotan los incrementos de Λ_0 y $\Omega_q, q = 1, \dots, p$ respectivamente en el intervalo $[s_{j-1}, s_j]$), es asignada una distribución *Gamma* con parámetros de forma y escala $c_0\Lambda_0^*$ y c_0 (en el caso de Λ_0), y con parámetros $c_q\Omega_q^*$ y c_q para cada Ω_q .

Con estas prioris, la posteriori resultante es proporcional a:

$$\prod_{j=1}^{J-1} \left[\prod_{i \in R_j} (I_{ij}^{N_{ij}} \exp(-I_{ij})) \Lambda_{0j}^{c_0\Lambda_{0j}^* - 1} \exp(-c_0\Lambda_{0j}) \prod_{q=1}^p \left(\Omega_{qj}^{c_q\Omega_{qj}^* - 1} \exp(-c_q\Omega_{qj}) \right) \right] \tau(w|\delta)\tau(\delta),$$

donde (llamando $ds_j = s_j - s_{j-1}$) $I_{ij} = I_i(s_j)ds_j = Y_{ij}(\Lambda_{0j} + \beta'_i\Omega_j + w_{li}ds_j), Y_{ij} = Y_i(s_j), \Omega_j = (\Omega_{1j}, \dots, \Omega_{pj})$ y $N_{ij} = dN_i(s_j)$. R_j es el conjunto de riesgo en el intervalo $(s_{j-1}, s_j]$, $i = 1, \dots, n, j = 1, \dots, J - 1, q = 1, \dots, p$, y $l_i \in \{1, \dots, k\}$. Los parámetros de interés son $(\Lambda_0, \Omega, a, b)$, donde las *frailties* w son consideradas como parámetros de estorbo, los cuales son eliminados mediante integración para obtener la posteriori marginal conjunta.

La expresión a posteriori resultante es complicada para trabajar. No obstante, las distribuciones marginales de Λ_0, Ω, a y b pueden ser estudiadas mediante métodos MCMC. Luego, mediante el algoritmo de Gibbs es posible obtener muestras aleatorias de la distribución posteriori conjunta.

Presentan además en el artículo una sección de ajuste de modelo. Al igual que en otras ramas de la estadística, tanto la comparación entre modelos como la bondad de ajuste son temas importantes en Análisis de Supervivencia. Una extensa revisión de métodos Bayesianos para el estudio de estos aspectos es detallada en el libro de Ibrahim, Chen & Sinha (2001) tales como el *criterio de información* (information criterion) o el *Conditional Predictive Ordinate* (CPO) para la comparación de modelos. Los autores presentan una manera de

estimar el CPO en este modelo, de manera de luego poder calcular cualquier medida para la comparación de modelos (por ejemplo, factores pseudo-Bayes).

En una sección de aplicación a datos reales, los autores analizaron dos bases de datos: la primera de ellas, compuesta por 90 pacientes varones con cáncer de laringe, y la segunda conformada por 42 pacientes de leucemia.

Dos años después, en *Additive Survival Models with Shared Frailty*, los autores presentaron una extensión de trabajo, considerando *frailties compartidas*. En este caso, el vector *frailty* w para cada individuo puede ser particionado en dos o más términos $w = w_1 + w_2 + \dots + w_k$ (donde w_j son los componentes de la *frailty* compartidos con otros individuos, $j = 1, \dots, k$) buscando poder considerar varios tipos de *frailties* para un mismo individuo. Con esta nueva expresión para las *frailties*, un vector indicador $a_i = (a_{i1}, a_{i2}, \dots, a_{ik})'$ es considerado para el i -ésimo individuo, en el cual cada $a_{ij}, j = 1, \dots, k$ indica si la componente j -ésima de la *frailty* w_j se encuentra presente o no.

La expresión resultante acorde a estas modificaciones es:

$$I_i(t|z_i, w) = Y_i(t) [\lambda_0(t) + z_i' \beta(t) + a_i' w].$$

El tratamiento Bayesiano propuesto para este modelo es análogo al del trabajo previo, conservando la selección de las distribuciones a priori para los distintos parámetros.

Como ilustración, los autores aplicaron este método a una base de datos de adopción compuesta por 125 familias con niños adoptados, buscando comparar la asociación entre la intensidad de muerte por infección (neumonía) y factores genéticos y de entorno, concluyendo que las condiciones de entorno resultan las más significativas.

Otra propuesta en el área de la estimación Bayesiana en el AHM con *frailties* fue exhibida por Chernoukhov (2013) en su tesis de maestría. En este trabajo, se consideraron *frailties* del tipo espacial, se asignaron distribuciones a priori *Gamma* para los distintos parámetros

de la baseline hazard (modelizada como constante a trozos), y uniforme impropia (es decir, constante e igual a uno en el espacio paramétrico) para los parámetros regresores. Tanto las covariables como las *frailties* fueron consideradas como dependientes del tiempo, y modeliza a ambos de la misma manera que a la baseline (a través del modelo constante a trozos). Las *frailties* fueron asumidas *Normales* con estructuras de covarianza de tipo *Geoestadística* o *conditional autoregressive* (CAR), dos estructuras dependientes del espacio ampliamente utilizadas.

En el modelo CAR, el autor desarrolló un método full Bayes (en los cuales se trabaja sobre la “full posteriori” obtenida de la función de verosimilitud y las respectivas distribuciones a priori sobre los parámetros, y se consideran tanto las distribuciones a priori como sus hiperparámetros conocidos antes de observar los datos) para estimar la totalidad de los parámetros del modelo. Las *frailties* fueron asumidas independientes, permitiendo la introducción de correlación parcial en cada intervalo de tiempo de manera independiente. La priori seleccionada para cada componente de la *frailty* es *Normal*, donde las varianzas son asumidas *Gamma Inversa*.

El procedimiento para generar muestras aleatorias de las condicionales a posteriori de la baseline, los parámetros de regresión y las *frailties* se realiza mediante un algoritmo del tipo Metropolis-within-Gibbs. Para esto, el autor realiza un estudio minucioso sobre el paso del algoritmo Metrópolis, de manera de obtener una “proposal density” (o densidad propuesta, que es la distribución mediante la cual se realiza el algoritmo) eficiente. Estas “proposal” son consideradas para cada distribución condicional por separado, y los hiperparámetros son seleccionados de manera que la proposal y la densidad condicional sean lo más parecidas posible. El autor propone dos métodos: uno utilizando la media de la condicional, y el otro utilizando el modo de dicha densidad. En el primer caso, las constantes normalizadoras son necesarias, por lo que los hiperparámetros son seleccionados de manera que las medias coincidan. Mediante el otro método (sugerido como más eficiente en el sentido numérico), los hiperparámetros son seleccionados usando el modo de la condicional, buscando obtener

un mejor ajuste. El método es testado a través de simulaciones, y luego es aplicado a una base de datos de cáncer de próstata.

De la misma manera, se pueden desarrollar técnicas de muestreo para casi todos las condicionales del modelo geostadístico. En este caso, la distribución considerada para las componentes *frailty* es una *Normal Multivariada* con vector de medias compuesto por ceros y matriz de varianzas y covarianzas acorde a las distancias entre locaciones. La priori asignada para las varianzas es una *Gamma Inversa*, y para los parámetros de correlación *Gamma*, con hiperparámetros elegidos acorde a cuán fuerte es la confianza en las prioris elegidas. La priori *Gamma Inversa* fue elegida de manera de utilizar un análisis conjugado. Con respecto a los parámetros de correlación, no logró obtener una “proposal density”, dejando un posible procedimiento para hallarlo como proyecto de investigación a futuro.

Los resultados obtenidos para el modelo CAR fueron publicados en Chernoukhov *et al* (2018).

2.2.3. Otras Propuestas

En un contexto diferente, un tratamiento Bayesiano empírico (el cual, a diferencia de los métodos full Bayes, procede a la elucidación de prioris y/o hiperparámetros utilizando la misma muestra) fue presentado por Sinha *et al* (2009). En este trabajo, se desarrollan estimadores Bayesianos empíricos para los parámetros de regresión, curvas de sobrevida y sus correspondientes desvíos estándar. Estos estimadores tienen la ventaja de ser fáciles de generar computacionalmente, y no requieren elucidación de sus hiperparámetros. El método garantiza un estimador monótono para la curva de sobrevida, y puede ser extendido fácilmente a covariables dependientes del tiempo. Las prioris seleccionadas son procesos *Gamma* para los incrementos de la cumulative baseline (la cual se modeliza mediante la formulación *constante a trozos*), y uniforme impropia para los parámetros de regresión. Los estimadores Bayesianos empíricos para el vector de hiperparámetros asociados al proceso a priori asig-

nado a $\Lambda_0(t)$ y a los parámetros de regresión son obtenidos mediante la maximización de la verosimilitud definida por el modelo (la cual es proporcional a un producto de distribuciones Poisson). Si bien la posteriori obtenida para $\Lambda_0(t)$ condicionada a los datos y al resto de los parámetros no presenta una expresión conocida, los autores hallan una expresión aproximada para su transformada de Laplace, por lo que los momentos a posteriori de $\Lambda_0(t)$ y $S_0(t)$ pueden ser evaluados.

Este procedimiento es aplicado a un estudio prospectivo de 205 pacientes de melanoma, cuya base de datos fue conformada entre los años 1962-1977. En este análisis, la única covariable considerada fue “género”, y los resultados son comparados con el estimador full Bayes de Dunson & Herring (2005), y con el estimador de Lin & Ying (1994). Una priori impropia no informativa fue utilizada para el método full Bayes, sugiriendo que el método full Bayes es apropiado cuando hay disponible información a priori precisa (de no ser así, el estimador Bayesiano empírico ajustó mejor al modelo). En un estudio mediante simulaciones, los autores concluyen que es preferible utilizar el método Bayesiano empírico cuando se desconoce la forma de λ_0 . El estimador full Bayes presenta un alto compromiso de sesgo comparado con los métodos Bayesiano empírico y con el estimador de momentos de Lin & Ying, especialmente cuando hay una alta presencia de datos censurados en la muestra, y bajo prioris no informativas. Adicionalmente, extienden el análisis a covariables dependientes del tiempo y estudian propiedades asintóticas.

Una nueva clase de modelos es desarrollada por Yin & Ibrahim (2005a-c), la cual contiene a los modelos proporcional y aditivo como casos particulares.

Considerando censura por la derecha, esta clase de modelos están basados en la transformación de Box-Cox (Box & Cox, 1964), la cual está dada por la siguiente expresión:

$$\phi(Y) = \begin{cases} \frac{(Y^\gamma - 1)}{\gamma}, & \text{if } \gamma \neq 0, \\ \log(Y), & \text{if } \gamma = 0, \end{cases}$$

donde

$$\lim_{\gamma \downarrow 0} \frac{(Y^\gamma - 1)}{\gamma} = \log(Y).$$

En el caso de los modelos semiparamétricos con baseline hazard λ_0 , la transformación propuesta es:

$$\phi(\lambda(t|z_i)) = \phi(\lambda_0(t)) + z_i(t)' \beta,$$

donde ϕ es la función de enlace definida a través de la transformación de Box-Cox.

Los modelos aditivo y multiplicativo están incluidos en esta clase de transformaciones (el modelo aditivo cuando $\gamma = 1$, y el modelo multiplicativo cuando $\gamma = 0$).

El primer objetivo es seleccionar el modelo eligiendo el γ , lo cual se realiza ajustando diferentes modelos con distintos valores de γ y evaluándolos a través de algún criterio de selección de modelo. El valor de γ que proporcione un mejor ajuste es seleccionado. Luego, se procede a la estimación del resto de los parámetros del modelo.

El modelo constante a trozos es considerado sobre $\lambda_0(t)$, denotando a λ_0 como el vector de dimensión J : $(A_1, \dots, A_J)'$. En la selección de prioris, los autores deben lidiar con el problema de la restricción de la función de riesgo (la cual debe ser no negativa). Para tratar este problema, se opta por la utilización de prioris truncadas. Se selecciona entonces una distribución *Normal Multivariada* para $(\beta|\lambda_0)$. Aplicando esto, la distribución conjunta de las prioris es:

$$\tau(\beta, \lambda_0) = \tau(\beta|\lambda_0)\tau(\lambda_0)1_{[0,\infty)}(A_j^\gamma + \gamma Z_i'), \quad \forall i, j.$$

Buscando obtener la distribución condicional de λ_0 , es necesario el cálculo de la constante normalizadora, lo que implica tratar con una complicada integral multivariada sujeta a restricciones en un complicado espacio paramétrico acotado no lineal. Para evitar este problema, los autores optan por modificar la priori seleccionada, buscando reducir el problema al caso univariado. Este procedimiento es una herramienta a tener en cuenta cuando uno considera una distribución multivariada truncada como priori para evitar el problema de la no-negatividad de la función de riesgo. Denotando por $Z_{i(-k)}$ y $\beta_{(-k)}$ a los vectores Z_i y β

con la componente k -ésima omitida, los autores definen:

$$h_\gamma(\lambda_0, \beta_{(-k)}, Z) = \min_{i,j} \left\{ \frac{A_j^\gamma + \gamma Z'_{i(-k)} \beta_{(-k)}}{\gamma Z_{ik}} \right\}.$$

Luego, se considera una priori conjunta para (β, λ_0) de la forma:

$$\tau(\beta, \lambda_0) = \tau(\beta_k | \beta_{(-k)}, \lambda_0) 1_{[-h_\gamma(\lambda_0, \beta_{(-k)}, Z), \infty)}(\beta_k).$$

Esta especificación de la priori sólo incluye una restricción para el parámetro univariado β_k , dejando todos los demás parámetros libres.

Es usual (aunque no necesario) considerar $\beta_{(-k)}$ y λ_0 independientes a priori, obteniendo $\tau(\beta_{(-k)}, \lambda_0) = \tau(\beta_{(-k)})\tau(\lambda_0)$. Los autores consideran a los componentes de la baseline hazard λ_0 independientes a priori, con distribución *Gamma*. Una priori *Normal* es considerada para cada componente de $\beta_{(-k)}$. Con esta construcción, es posible obtener muestras aleatorias de la posteriori empleando algoritmos Metropolis-within-Gibbs.

Además, los autores presentan una sección de ajuste de modelo, en la cual presentan tanto el DIC (Deviance Information Criterion) como el estadístico CPO correspondientes al modelo.

En otros artículos, los mismos autores presentan extensiones a esta propuesta, considerando *frailties*, y adaptando estas ideas a los modelos “cure fractions” (modelos en los cuales se considera que un porcentaje de la población se “cura”, no lográndose observar nunca el evento de interés). En todos ellos, presentan simulaciones y aplicaciones a datos reales, en los cuales el mejor ajuste de modelo se presenta para algún $0 < \gamma < 1$, justificando la importancia de este enfoque.

3. ANÁLISIS BAYESIANO DEL MODELO DE RIESGO ADITIVO

En este capítulo, presentamos una propuesta de estimación Bayesiana híbrida, distinta a las existentes en la literatura. Su nombre se debe a que los estimadores Bayesianos de los parámetros regresores no se obtienen mediante un método full Bayes, sino a través de una ecuación de estimación. Esto es posible gracias a una expresión de la verosimilitud construida a través de un algoritmo (el cual permite calcular los coeficientes de un polinomio de grado n obtenido como producto de polinomios de grado 1), logrando de esta manera una expresión para la verosimilitud proporcional a una mezcla de distribuciones *Gamma*. De esta manera, asignando prioris flexibles hallamos estimadores Bayesianos explícitos para la totalidad de los parámetros del modelo. Finalizamos el capítulo con simulaciones para testear el modelo, y una aplicación a datos reales.

3.1. *Expresión para la verosimilitud*

Teniendo en cuenta lo desarrollado hasta el momento, a continuación presentamos una propuesta en la cual los estimadores Bayesianos se hallarán de forma explícita (recordemos que, hasta ahora, las propuestas se basan en su totalidad en algoritmos Metrópolis y de tipo Gibbs para generar muestras de la posteriori, para luego obtener el estimador Bayesiano de manera aproximada). Comenzaremos presentando una expresión de la verosimilitud obtenida en base al cálculo de una expansión polinómica.

Considerando la expresión de la verosimilitud obtenida para el AHM en el capítulo anterior:

$$\mathcal{L}_n(\boldsymbol{\beta}, \lambda_0) = \exp \left\{ -\boldsymbol{\beta}' \sum_{i=1}^n t_i z_i \right\} \exp \left\{ -\sum_{i=1}^n \Lambda_0(t_i) \right\} \prod_{i=1}^n [\lambda_0(t_i) + \boldsymbol{\beta}' z_i]^{\delta_i},$$

nuestra propuesta inicial será desarrollar esta expresión, modelizando a la baseline hazard como constante a trozos, expandiendo el producto para cada valor de la función de riesgo asociado a cada uno de los J intervalos de la grilla. Llamaremos n_j , $j \in \{1, \dots, J\}$, al número de observaciones que pertenecen a cada intervalo, y (z_{k_j}, δ_{k_j}) a los respectivos valores de las *covariables* e *indicadores de no-censura*, respectivamente, para cada observación. Para un tratamiento Bayesiano, consideramos a los parámetros como realizaciones de un vector aleatorio \mathbf{B} y una *curva aleatoria* \mathbf{L}_0 , que denotaremos por $\boldsymbol{\beta}$ y $\lambda_0(\cdot)$ repectivamente. Esto lleva a

$$\begin{aligned} \mathcal{L}_n(\boldsymbol{\beta}, \lambda_0) &= \exp \left\{ -\boldsymbol{\beta}' \sum_{i=1}^n t_i z_i \right\} \exp \left\{ -\sum_{i=1}^n \Lambda_0(t_i) \right\} \prod_{j=1}^J \prod_{k_j=1}^{n_j} [a_j + \boldsymbol{\beta}' z_{k_j}]^{\delta_{k_j}} \\ &=: f_{\mathcal{X}|\mathbf{L}_0, \mathbf{B}}(\{(t_i, \delta_i, z_i)\}_{i=1, \dots, n}), \end{aligned}$$

donde la última expresión enfatiza la perspectiva Bayesiana, expresando la densidad conjunta de la muestra dados los parámetros. Luego, llamando $\tau(\mathbf{L}_0)$ a la priori de la baseline, la posteriori resultante es

$$f_{\mathbf{L}_0|\mathbf{B}, \mathcal{X}} \propto \exp \left\{ -\sum_{i=1}^n \Lambda_0(t_i) \right\} \prod_{j=1}^J \left(\prod_{k_j=1}^{n_j} [a_j + \boldsymbol{\beta}' z_{k_j}]^{\delta_{k_j}} \right) \times \tau(\mathbf{L}_0).$$

Notemos que la fórmula obtenida no corresponde a ninguna densidad estándar de la cual conozcamos sus momentos, ni algoritmos para simular muestras de ella. Nuestro objetivo actual será estimar sus dos primeros momentos. Expandiendo el producto

$$\prod_{k_j=1}^{n_j} [a_j + \boldsymbol{\beta}' z_{k_j}]^{\delta_{k_j}} = d_0 + d_1 a_j + \dots + d_{N_j} a_j^{N_j}, \quad (3.1)$$

obtenemos un polinomio de orden $N_j = \sum_{k_j=1}^{n_j} \delta_{k_j}$, donde los coeficientes dependen de $\{(z_i, \delta_i)\}_{i=1, \dots, n}$ y de $\boldsymbol{\beta}$. Para estimar estos coeficientes, Chernoukhov (2013) propuso un método que esencialmente consiste en dos pasos:

(i) evaluar el polinomio en $(N_j + 1)$ puntos diferentes; y

(ii) resolver el sistema de ecuaciones resultante.

La debilidad de este método es la inestabilidad numérica que presenta, más aún al aumentar el tamaño muestral. El método que proponemos resulta más eficiente y estable, y está basado en la relación recursiva que hay entre los coeficientes polinomiales acorde al incremento del tamaño muestral.

Recursión: Si la cantidad de observaciones no censuradas en la muestra es n , definimos (para un intervalo fijo en la partición, donde el valor de la baseline en dicho intervalo es ξ):

$$P_n(\xi) = \sum_{j=0}^n d_j^{(n)} \xi^j.$$

El incremento del tamaño muestral en 1 unidad no censurada, es equivalente a multiplicar P_n por $(\xi + \beta' z_{n+1})$:

$$\begin{aligned} P_{n+1}(\xi) &= \left(\sum_{j=0}^n d_j^{(n)} \xi^j \right) (\xi + \beta' z_{n+1}) \\ &= \sum_{j=0}^n d_j^{(n)} \xi^{j+1} + \sum_{j=0}^n d_j^{(n)} \beta' z_{n+1} \xi^j \end{aligned}$$

Luego, los coeficientes $d_j^{(n+1)}$, $j \in \{1, \dots, n+1\}$ de P_{n+1} se calculan en base a los coeficientes $d_j^{(n)}$, $j \in \{1, \dots, n\}$ de P_n de la siguiente manera:

- $d_0^{(n+1)} = d_0^{(n)} \beta' z_{n+1}$,
- $d_j^{(n+1)} = d_{j-1}^{(n)} + d_j^{(n)} \beta' z_{n+1}$, si $j \in \{1, \dots, n\}$,
- $d_{n+1}^{(n+1)} = d_n^{(n)} = 1$, por ser coeficientes principales en (3.1).

En adelante, no será necesario el supraíndice que indica la cantidad de observaciones no censuradas (sólo lo utilizamos para diferenciar entre los coeficientes de P_n y P_{n+1}). Bajo esta

relación recursiva, podemos calcular los d_i asociados a la muestra. Con esta modificación, y considerando β conocido, la verosimilitud $\mathcal{L}_n(\lambda_0)$ adopta la siguiente forma:

$$\mathcal{L}_n(\lambda_0) \propto \prod_{j=1}^J \left(\sum_{k=0}^{N_j} d_k a_j^k \right) \exp \left\{ - \sum_{i=0}^n \Lambda_0(t_i) \right\}.$$

Ahora, para cada a_j , la expresión de la cumulative baseline en un elemento de la muestra se puede reescribir como:

$$\Lambda_0(t_i) = \begin{cases} \Lambda_0(t_i) & t_i \leq s_{j-1}, \\ \Lambda_0(s_{j-1}) + a_j(t_i - s_{j-1}) & s_{j-1} < t_i \leq s_j, \\ \Lambda_0(s_{j-1}) + a_j(s_j - s_{j-1}) + \int_{s_j}^{t_i} \lambda_0(u) du & s_j < t_i. \end{cases}$$

Luego, llamando $m_j = \#\{t_i > s_j\}_{i=1, \dots, n}$, y recordando (1.3) obtenemos la siguiente expresión para

$$\exp \left\{ - \sum_{i=1}^n \Lambda_0(t_i) \right\} \propto \exp \left\{ -a_j \left(\sum_{k_j=1}^{n_j} (t_{k_j} - s_{j-1}) + m_j(s_j - s_{j-1}) \right) \right\}.$$

De acuerdo a esto, $\mathcal{L}_n(\lambda_0)$ es proporcional a una mezcla de distribuciones Gamma:

$$\mathcal{L}_n(\lambda_0) \propto \prod_{j=1}^J \sum_{k=0}^{N_j} d_k a_j^k e^{-a_j \left(\sum_{k_j=1}^{n_j} (t_{k_j} - s_{j-1}) + m_j(s_j - s_{j-1}) \right)}. \quad (3.2)$$

3.2. Selección de prioris

Especificamos las distribuciones *a priori* como sigue:

Priori para los parámetros Euclideos. Consideramos

$$\mathbf{B} \sim N_k(\boldsymbol{\mu}_\beta, \mathbf{C}_\beta) |_{\mathbb{R}^{k+}}, \quad (3.3)$$

es decir, su distribución es el truncamiento de una distribución normal multivariada con vector de medias $\boldsymbol{\mu}_\beta$ y matriz de covarianzas definida positiva \mathbf{C}_β , en el octante positivo \mathbb{R}^{k+} . Si bien esta pueda parecer una restricción fuerte, Esto conduce a que la función de densidad para \mathbf{B} esté dada por

$$f_{\mathbf{B}}(\boldsymbol{\beta}) = \frac{\left(\frac{1}{\sqrt{2\pi}}\right)^k |\det(\mathbf{C}_\beta)|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta)' \mathbf{C}_\beta^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}_\beta)\right\} \prod_{j=1}^k 1(\beta_j \geq 0)}{\int_0^\infty \cdots \int_0^\infty \left(\frac{1}{\sqrt{2\pi}}\right)^k |\det(\mathbf{C}_\beta)|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta)' \mathbf{C}_\beta^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}_\beta)\right\} d\boldsymbol{\beta}} \quad (3.4)$$

El denominador no es más que la probabilidad $P(\beta_1 \geq 0; \dots; \beta_k \geq 0)$, es decir, la constante normalizadora necesaria para que la función de densidad integre 1 sobre el espacio paramétrico. Denotaremos a dicha constante $\Omega(\boldsymbol{\mu}_\beta, \mathbf{C}_\beta)$, notando que depende sólo de $\boldsymbol{\mu}_\beta$ y \mathbf{C}_β .

Priori para la baseline hazard. Con la motivación de considerar dependencia entre los parámetros de la baseline, optamos por introducir un proceso Gamma, con una función pre-especificada creciente y continua a izquierda $\alpha(t)$ y un parámetro de escala c . Acorde a esto, llamando η_i al incremento de la función $\alpha(t)$ en el intervalo $[s_{i-1}, s_i)$, la priori conjunta para los incrementos en los distintos intervalos de la grilla de la cumulative baseline $\boldsymbol{\Lambda}_0^+$ resulta en

$$f_{\boldsymbol{\Lambda}_0^+}(a_1^+, \dots, a_J^+) = \prod_{i=1}^J \frac{c^{\eta_i}}{\Gamma(\eta_i)} (a_i^+)^{\eta_i-1} \exp(-c a_i^+). \quad (3.5)$$

Los detalles y especificaciones serán expuestos más adelante.

3.3. Distribución a Posteriori.

Multiplicando por la priori de \mathbf{L}_0 dada en la ecuación (3.5) y la priori de \mathbf{B} dada en la ecuación (3.4), obtenemos la densidad conjunta para la muestra y los parámetros:

$$f_{L_0, \mathcal{X}, \mathbf{B}} = \exp \left\{ -\boldsymbol{\beta}' \sum_{i=1}^n t_i z_i \right\} \frac{\exp \left\{ -\frac{1}{2} (\boldsymbol{\beta} - \boldsymbol{\mu}_\beta)' \mathbf{C}_\beta^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}_\beta) \right\} \prod_{j=1}^k 1(\beta_j \geq 0)}{(\sqrt{2\pi})^k |\det(\mathbf{C}_\beta)|^{\frac{1}{2}} \Omega(\boldsymbol{\mu}_\beta, \mathbf{C}_\beta)} \\ \exp \left\{ -\sum_{i=1}^n \Lambda_0(t_i) \right\} \prod_{j=1}^J \left(\prod_{k_j=1}^{n_j} [a_j^+ + \boldsymbol{\beta}' z_{k_j}]^{\delta_{k_j}} \right) \frac{c^{\eta_i}}{\Gamma(\eta_i)} (a_i^+)^{\eta_i-1} \exp(-c a_i^+).$$

Desde aquí, cálculos estándar nos llevan a las ditribuciones marginales a posteriori para los parámetros

$$f_{L_0 | \mathcal{X}, \mathbf{B}} \propto \exp \left\{ -\sum_{i=1}^n \Lambda_0(t_i) \right\} \prod_{j=1}^J \left(\prod_{k_j=1}^{n_j} [a_j^+ + \boldsymbol{\beta}' z_{k_j}]^{\delta_{k_j}} \right) \frac{c^{\eta_i}}{\Gamma(\eta_i)} (a_i^+)^{\eta_i-1} \exp(-c a_i^+),$$

y

$$f_{\mathbf{B} | \mathcal{X}, L_0} \propto e^{\left\{ -\boldsymbol{\beta}' \sum_{i=1}^n t_i z_i - \frac{1}{2} (\boldsymbol{\beta} - \boldsymbol{\mu}_\beta)' \mathbf{C}_\beta^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}_\beta) \right\}} \prod_{j=1}^k 1(\beta_j \geq 0) \prod_{j=1}^J \left(\prod_{k_j=1}^{n_j} [a_j^+ + \boldsymbol{\beta}' z_{k_j}]^{\delta_{k_j}} \right).$$

Mediante el algoritmo de Gibbs, es posible generar muestras de las posterioris marginales de $[\mathbf{L}_0 | \mathcal{X}, \mathbf{B}]$ alternando con las posterioris marginales de $[\mathbf{B} | \mathcal{X}, \mathbf{L}_0]$ de manera de aproximar a la full posteriori $[\mathbf{B}, \mathbf{L}_0 | \mathcal{X}]$. Así, para implementar el tradicional algoritmo de Gibbs, debemos poder tomar observaciones de $f_{\mathbf{L}_0 | \mathcal{X}, \mathbf{B}}(a_1, \dots, a_J)$ y de $f_{\mathbf{B} | \mathcal{X}, L_0}(\boldsymbol{\beta})$. No obstante, mediante la nueva reformulación de la verosimilitud, bajo ciertas condiciones, podremos hallar estimadores Bayesianos sin necesidad de recurrir a simulaciones.

3.4. El Método Bayesiano Híbrido

A continuación, presentaremos un método alternativo al mencionado previamente (en el que se utilizaría el algoritmo de Gibbs). Buscaremos desarrollar un método Bayesiano que nos permita

(i) separar la estimación de β de la de la baseline $\lambda_0(\cdot)$, y

(ii) generar estimadores en base a ello.

Ambos objetivos son importantes debido a que el Análisis de Sobrevida usualmente tiene dos metas:

(i) predecir la sobrevida de un individuo dadas ciertas covariables, y/o

(ii) inferir en cuánto la sobrevida o el riesgo de un evento puede depender de alguna covariable.

Mientras que para el primer objetivo (i.e., predicción) es necesaria la estimación de la totalidad de los parámetros (los parámetros de la baseline y los parámetros Euclideos), para el segundo objetivo sólo son necesarios los parámetros Euclideos. Como ya hemos mencionado, siguiendo el método de máxima verosimilitud en los tres modelos mencionados en esta tesis (PHM, AHM o AFTM), estimadores de la baseline y de los parámetros Euclideos se logran en forma conjunta (información sobre esto se puede ver, por ejemplo, en Andersen *et al*, 1993). Esto ha presentado conflictos computacionales que han motivado a los investigadores a proponer métodos de estimación alternativos. Tal es el caso del enfoque desarrollado por Cox para el PHM, conocido como *Partial Likelihood* (o verosimilitud parcial), el cual hace posible estimar los parámetros Euclideos β sin necesidad de conocer $\lambda_0(\cdot)$.

En el contexto del AHM, una manera clásica de estimar β consiste en hacerlo mediante una ecuación de estimación, en vez de mediante la función de Score proveniente del método de máxima verosimilitud o máxima verosimilitud parcial. Esto es porque ninguno de estos métodos permitía separar la estimación de β de la estimación de λ_0 . Los pioneros en este área fueron Lin & Ying (1994), quienes propusieron la ecuación de estimación para β

$$U(\beta) := \left[\sum_{i=1}^n \delta_i [z_i - \tilde{z}_n(t_i)] \right] - \left[\sum_{i=1}^n \left(\int_0^{t_i} [z_i - \tilde{z}_n(u)]^{\otimes 2} du \right) \right] \beta = 0, \quad (3.6)$$

donde dado un vector columna a , definimos la matriz $a^{\otimes 2} = a a'$, y

$$\tilde{z}_n(u) := \frac{\sum_{i=1}^n z_i 1(t_i \geq u)}{\sum_{i=1}^n 1(t_i \geq u)}$$

es la función vectorial que promedia todos los valores de z correspondientes a valores de tiempo mayores o iguales a u . Notemos que, cuando esta función es evaluada en el mínimo o el máximo valor del tiempo, $\tilde{z}_n(t_{(1)}) = \bar{z}$, mientras que $\tilde{z}_n(t_{(n)}) = z_{j^*}$ donde j^* es el índice de $t_{(n)}$.

3.4.1. Cálculo del estimador Bayesiano de β

Buscando una expresión de la función de estimación $U(\beta)$ de la ecuación (3.6) en un contexto Bayesiano, presentamos, basándonos en los trabajos de Lin & Ying, los estadísticos

$$\begin{aligned} V_1 &:= \sum_{i=1}^n \delta_i [z_i - \tilde{z}_n(t_i)], \\ V_2 &:= \sum_{i=1}^n \left(\int_0^{t_i} [z_i - \tilde{z}_n(u)]^{\otimes 2} du \right), \\ V_3 &:= \sum_{i=1}^n [z_i - \tilde{z}_n(t_i)]^{\otimes 2}, \end{aligned}$$

donde V_1 es un vector fila, mientras que V_2 y V_3 son matrices simétricas. Notemos que, bajo esta notación, $U(\beta) = V_1 - V_2 \beta$. Consideremos ahora la función

$$g(\beta) = \exp \left\{ -\frac{1}{2} [(\beta - V_2^{-1} V_1)' (V_2^{-1} V_3 V_2^{-1})^{-1} (\beta - V_2^{-1} V_1)] \right\}.$$

Tomando logaritmos y derivadas con respecto a β advertimos que

$$\begin{aligned} \log g(\beta) &= -\frac{1}{2} [\beta' (V_2^{-1} V_3 V_2^{-1})^{-1} \beta - 2 \beta' (V_2^{-1} V_3 V_2^{-1})^{-1} (V_2^{-1} V_1) \\ &\quad - (V_2^{-1} V_1)' (V_2^{-1} V_3 V_2^{-1})^{-1} (V_2^{-1} V_1)], \\ \frac{\partial}{\partial \beta} \log g(\beta) &= -(V_2^{-1} V_3 V_2^{-1})^{-1} \beta + (V_2^{-1} V_3 V_2^{-1})^{-1} (V_2^{-1} V_1). \end{aligned}$$

Notemos que $U(\boldsymbol{\beta}) = 0$ siempre que $(\partial/\partial\boldsymbol{\beta}) \log g(\boldsymbol{\beta}) = 0$:

$$\begin{aligned} -(V_2^{-1} V_3 V_2^{-1})^{-1} \boldsymbol{\beta} + (V_2^{-1} V_3 V_2^{-1})^{-1} (V_2^{-1} V_1) &= 0, \\ V_2 \boldsymbol{\beta} - V_1 &= 0, \\ U(\boldsymbol{\beta}) &= 0. \end{aligned}$$

De esta manera se aprecia que el estimador puntual de Lin & Ying (en ocasiones, los referiremos como LY) de $\boldsymbol{\beta}$ corresponde a la media (o el modo) de una distribución normal multivariada con media $\mathbf{m} = (V_2^{-1} V_1)$. Además, podemos estimar su varianza mediante $\mathbf{D} = (V_2^{-1} V_3 V_2^{-1})$. De esta manera, desde un contexto Bayesiano, podríamos considerar al estimador de LY como el modo o la media de una distribución a posteriori normal multivariada, donde la priori asignada a los parámetros Euclideos es impropia y de forma plana (priori uniforme de Laplace¹).

No obstante, en esta tesis hemos optado por una priori informativa normal multivariada, truncada al octante positivo dado por (3.3). Luego, proponemos una distribución pseudo-marginal a posteriori para $[\mathbf{B}|\mathcal{X}, \mathbf{L}_0]$

$$f_{\mathbf{B}|\mathcal{X}, \mathbf{L}_0}(\boldsymbol{\beta}) \propto \exp \left\{ -\frac{1}{2} [(\boldsymbol{\beta} - \mathbf{m})' \mathbf{D}^{-1} (\boldsymbol{\beta} - \mathbf{m})' + (\boldsymbol{\beta} - \boldsymbol{\mu}_\beta)' \mathbf{C}_\beta^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}_\beta)'] \right\} \prod_{j=1}^k 1(\beta_j \geq 0), \quad (3.7)$$

lo que luego de cálculos estándar se convierte en

$$[\mathbf{B}|\mathcal{X}, \mathbf{L}_0] \sim N_k \left[(\mathbf{D}^{-1} + \mathbf{C}_\beta^{-1})^{-1} (\mathbf{D}^{-1} \mathbf{m} + \mathbf{C}_\beta^{-1} \boldsymbol{\mu}_\beta), (\mathbf{D}^{-1} + \mathbf{C}_\beta^{-1})^{-1} \right] \Big|_{\mathbb{R}^{k+}}.$$

Es importante remarcar, en este punto, porqué llamamos a la distribución en la ecuación de arriba una “pseudo” posteriori, en lugar de distribución (plana) posteriori. Esto es debido a que los cálculos Bayesianos que dan origen a la densidad $f_{\mathbf{B}|\mathcal{X}, \mathbf{L}_0}$ en la ecuación (3.7) surgen de $g(\boldsymbol{\beta})$, la cual NO es la distribución condicional de la muestra dada los parámetros, sino

¹ Se verá con más detalle en el Capítulo 4.

por una transformación conveniente inspirada por una ecuación de estimación. Es por esto que nuestro método se llama “*híbrido*”. Dentro del desarrollo Bayesiano, nuestro método comparte la misma motivación que el de la verosimilitud *parcial* de Cox, y la ecuación de estimación de Lin y Ying desde un enfoque frecuentista de inferencia.

Resulta conveniente el hecho de que, bajo nuestra formulación, $[\mathbf{B}|\mathcal{X}, \mathbf{L}_0] = [\mathbf{B}|\mathcal{X}]$. Es decir, la distribución marginal a posteriori es independiente de \mathbf{L}_0 . Esto permite separar la estimación de $\boldsymbol{\beta}$ de la estimación de $\lambda_0(\cdot)$. Además, veremos que esta formulación nos permite estimar $\boldsymbol{\beta}$ en forma explícita, lo cual es un desarrollo paralelo al método de LY. Con respecto al estimador puntual Bayesiano, hemos optado en este trabajo por el *MAP*, es decir, modo de la distribución a posteriori. Esto sería,

$$\begin{aligned}\hat{\boldsymbol{\beta}}^{\text{B}} &= (\mathbf{D}^{-1} + \mathbf{C}_{\beta}^{-1})^{-1}(\mathbf{D}^{-1}\mathbf{m} + \mathbf{C}_{\beta}^{-1}\boldsymbol{\mu}_{\beta}) \\ &= (V_2 V_3^{-1} V_2 + \mathbf{C}_{\beta}^{-1})^{-1} (V_2 V_3^{-1} V_1 + \mathbf{C}_{\beta}^{-1}\boldsymbol{\mu}_{\beta})\end{aligned}\quad (3.8)$$

en el caso de que este sea un valor positivo, o cero en caso contrario. Es importante remarcar que, ante una priori no informativa, es decir, cuando $\|\mathbf{C}_{\beta}\| \rightarrow \infty$, el estimador puntual $\hat{\boldsymbol{\beta}}^{\text{B}} \rightarrow V_2^{-1} V_1$, el cual coincide cuando es positivo con el estimador clásico de LY. No obstante, un estimador Bayesiano alternativo basado en la esperanza de la distribución pseudo-posteriori podría presentar un serio problema: diverger cuando $\|\mathbf{C}_{\beta}\| \rightarrow \infty$, efecto provocado por el hecho de truncar la distribución normal multivariada al primer octante. Por este motivo, optamos no estimar la varianza de $\hat{\boldsymbol{\beta}}^{\text{B}}$ mediante la varianza de la distribución pseudo-posteriori truncada, por lo que proponemos un método diferente, el cual explicamos a continuación.

Estimación de la Varianza. Procederemos de la siguiente manera:

1. Estimar $\hat{\boldsymbol{\beta}}^{\text{B}}$ de acuerdo a la ecuación (3.8) en el caso positivo, o cero en caso contrario.
2. Para estimar la precisión de cada componente $\hat{\boldsymbol{\beta}}_k^{\text{B}}$ construiremos un intervalo para la HPD (highest posterior density) de nivel $(1 - \alpha)$ (esto sería, de todos los IC de nivel

$(1 - \alpha)$ considerando la densidad a posteriori, el más estrecho). Dicho intervalo será de la forma $[b_l, b_u]$. Es importante notar que en la formulación de Lin & Ying, dado que los intervalos de confianza están basados aproximadamente en una distribución Normal, toman la forma $\hat{\beta}^{\text{LY}} \pm z_{1-\alpha/2} \hat{\sigma}_{(\hat{\beta}^{\text{LY}})}$. En vista de esto, buscando que sean comparables de alguna manera, propondremos

$$\tilde{\sigma}_{(\tilde{\beta}^{\text{B}})} := \frac{b_u - b_l}{2 z_{1-\alpha/2}}. \quad (3.9)$$

Notemos que el hecho de que estemos trabajando con una distribución multivariada truncada, los estimadores puntuales no son necesariamente el punto medio de los intervalos de confianza.

Este método propuesto para la estimación de $\tilde{\sigma}_{(\tilde{\beta}^{\text{B}})}$ también tiene consecuencias importantes en el testeo de la significancia de las covariables. Basándonos en la relación entre intervalos de confianza y test de hipótesis, podremos indicar que el hecho de que el cero pertenezca al intervalo de confianza asociado a una covariable en particular es un indicador de que dicha covariable no resulta significativa. Considerando esto, es necesario destacar que la construcción de intervalos de la “highest pseudo-posterior density” es esencial para propósitos de testeo. Esto es porque, de lo contrario, forzar un intervalo bilateral bajo una distribución normal truncada siempre implicaría que $b_l(\alpha) > 0$ para todo $\alpha > 0$. Es decir, concluiríamos que ninguna variable resulta significativa.

3.4.2. Cálculo de los estimadores Bayesianos de los parámetros de la *baseline hazard*

Buscando implementar una priori informativa conjunta para los parámetros de la *baseline hazard*, recordaremos primero al usualmente implementado *proceso Gamma*.

Definición. Denotemos por $\mathcal{G}(\alpha, \beta)$ a la distribución Gamma con parámetro de forma $\alpha > 0$ y parámetro de escala $\beta > 0$. Sea $\alpha(t), t \geq 0$, una función continua a izquierda creciente tal que $\alpha(0) = 0$, y sea $Z(t), t \geq 0$, un proceso estocástico con las propiedades:

(i) $Z(0) = 0$;

(ii) $Z(t)$ tiene incrementos independientes en intervalos distintos; y

(iii) para $t > s$, $Z(t) - Z(s) \sim \mathcal{G}(c(\alpha(t) - \alpha(s)), c)$.

Entonces $\{Z(t) : t > 0\}$ es llamado un **proceso Gamma** y es denotado por $Z(t) \sim \mathcal{GP}(c\alpha(t), c)$.

3.5. Proceso Gamma en la cumulative baseline $\Lambda_0(\cdot)$

Para incorporar un proceso Gamma como priori en la baseline para el modelo de riesgos aditivos, necesitamos hallar una expresión alternativa para la verosimilitud en la cual ésta dependa de los incrementos de la cumulative baseline $\Lambda_0^+ = (\Lambda_{0_1}^+, \Lambda_{0_2}^+, \dots, \Lambda_{0_J}^+ := \infty)$ asociados a cada intervalo en la grilla. Considerando la expresión de la verosimilitud obtenida en (3.2), definida en función de $A_j, j = 1, \dots, J$, utilizaremos la relación entre A_j y $\Lambda_{0_j}^+$. Por definición, $\Lambda_{0_j}^+ := A_j \times (s_j - s_{j-1})$. Luego, reemplazando para cada j , $A_j = \frac{\Lambda_{0_j}^+}{(s_j - s_{j-1})}$, obtenemos:

$$\mathcal{L}_n(a_{0_j}^+) \propto \sum_{k=0}^{N_j} d_k \left(\frac{a_{0_j}^+}{s_j - s_{j-1}} \right)^k e^{-\left(\frac{a_{0_j}^+}{s_j - s_{j-1}} \right) \left(\left(\sum_{k_j=1}^{n_j} (t_{k_j} - s_{j-1}) \right) + m_j (s_j - s_{j-1}) \right)}. \quad (3.10)$$

Luego, seleccionando una función continua a izquierda y creciente α y un parámetro de escala c , la priori para cada $\Lambda_{0_j}^+$ es $\mathcal{G}(c(\alpha(s_j) - \alpha(s_{j-1})), c)$.

Es importante en este punto notar que el parámetro $\Lambda_{0_j}^+$ no es posible estimarlo (ya que no es finito), por lo que con el objetivo de poder realizar inferencias sobre este parámetro nos vemos forzados a truncar el eje real positivo en cualquier punto mayor que el último evento observado. Esto no traerá problemas en su implementación, ya que no existen estudios que no estén truncados por una fecha límite (en el caso de estudios clínicos en pacientes, no tendría

sentido que ninguna variable supere, por ejemplo, los 200 años). Por este motivo, con el objetivo de poder implementar un proceso Gamma en la cumulative baseline, consideraremos $s_J < \infty$.

Por construcción, los diferentes intervalos en la grilla son disjuntos, y siguiendo la estructura del proceso Gamma, son independientes. Acorde a esto, la priori para Λ_0^+ es:

$$\tau(\Lambda_0^+) = \prod_{j=1}^J f_{\mathcal{G}(c(\alpha_j), c)}(a_0^+),$$

donde $\alpha_j = \alpha(s_j) - \alpha(s_{j-1})$, y $f_{\mathcal{G}(a,b)}(t)$ denota la función de densidad de una distribución Gamma con parámetros a y b evaluada en t .

Multiplicando (3.10) por la priori, la posteriori resultante implementando el proceso Gamma es:

$$f_{\Lambda_0^+ | \mathcal{X}, \beta}(a_0^+) \propto \prod_{j=1}^J \sum_{k=0}^{N_j} d_k \left(\frac{a_{0j}^+}{(s_j - s_{j-1})} \right)^k e^{-\left(\frac{a_{0j}^+}{(s_j - s_{j-1})} \right) \left(\left(\sum_{k_j=1}^{n_j} (t_{k_j} - s_{j-1}) \right) + m_j (s_j - s_{j-1}) \right)}$$

$$\times \prod_{j=1}^J \frac{(a_{0j}^+)^{c\alpha_j - 1} e^{-a_{0j}^+ c} c^{c\alpha_j}}{\Gamma(c\alpha_j)}.$$

Por practicidad, llamaremos:

$$d_k^{(j)} = \frac{d_k}{(s_j - s_{j-1})^k \times \Gamma(c\alpha_j)},$$

$$c_j = \frac{\left(\sum_{k_j=1}^{n_j} (t_{k_j} - s_{j-1}) \right) + m_j (s_j - s_{j-1})}{(s_j - s_{j-1})} + c.$$

Con esta notación:

$$\begin{aligned}
f_{\Lambda_{0_j}^+ | \mathcal{X}, \boldsymbol{\beta}}(a_0^+) &= \frac{\sum_{k=0}^{N_j} d_k^{(j)} (a_0^+)^{(k+c\alpha_j-1)} e^{-a_0^+ c_j}}{\int_{-\infty}^{\infty} \sum_{k=0}^{N_j} d_k^{(j)} (a_0^+)^{(k+c\alpha_j-1)} e^{-a_0^+ c_j} da_0^+} \\
&= \frac{\sum_{k=0}^{N_j} d_k^{(j)} (a_0^+)^{(k+c\alpha_j-1)} e^{-a_0^+ c_j}}{\sum_{k=0}^{N_j} d_k^{(j)} \int_{-\infty}^{\infty} (a_0^+)^{(k+c\alpha_j-1)} e^{-a_0^+ c_j} da_0^+} \\
&= \frac{\sum_{k=0}^{N_j} \frac{d_k^{(j)}}{c_j^k} \Gamma(k+c\alpha_j) \cdot f_{\mathcal{G}(k+c\alpha_j, c_j)}(a_0^+)}{\sum_{k=0}^{N_j} \frac{d_k^{(j)}}{c_j^k} \Gamma(k+c\alpha_j)}.
\end{aligned}$$

Luego, la posteriori resultante es proporcional a una mezcla de distribuciones Gamma, permitiendo obtener el estimador Bayesiano explícito (utilizando función de pérdida cuadrática) mediante su esperanza.

Siguiendo esto, y llamando $e_k^{(j)} := \frac{d_k^{(j)}}{c_j^k} \Gamma(k+c\alpha_j)$:

$$\begin{aligned}
\widehat{a_{0_j}^+}^B &= E[\Lambda_{0_j}^+ | \mathcal{X}, \boldsymbol{\beta}] = \int_{-\infty}^{\infty} a_0^+ \left[\frac{\sum_{k=0}^{N_j} e_k^{(j)} f_{\mathcal{G}(k+c\alpha_j, c_j)}(a_0^+)}{\sum_{k=0}^{N_j} e_k^{(j)}} \right] da_0^+, \\
&= \frac{\sum_{k=0}^{N_j} e_k^{(j)} \int_{-\infty}^{\infty} a_0^+ [f_{\mathcal{G}(k+c\alpha_j, c_j)}(a_0^+)] da_0^+}{\sum_{k=0}^{N_j} e_k^{(j)}} \\
&= \frac{\sum_{k=0}^{N_j} e_k^{(j)} (k+c\alpha_j)}{\sum_{k=0}^{N_j} e_k^{(j)} c_j}.
\end{aligned}$$

Para estimar su varianza, lo haremos directamente a través de la distribución a posteriori:

$$\begin{aligned}
V[\Lambda_{0_j}^+ | \mathcal{X}, \boldsymbol{\beta}] &= \left(\int_{-\infty}^{\infty} (a_0^+)^2 \left[\frac{\sum_{k=0}^{N_j} e_k^{(j)} f_{\mathcal{G}(k+c\alpha_j, c_j)}(a_0^+)}{\sum_{k=0}^{N_j} e_k^{(j)}} \right] da_0^+ \right) - E[\Lambda_{0_j}^+ | \mathcal{X}, \boldsymbol{\beta}]^2, \\
&= \left(\frac{\sum_{k=0}^{N_j} e_k^{(j)} \int_{-\infty}^{\infty} (a_0^+)^2 [f_{\mathcal{G}(k+c\alpha_j, c_j)}(a_0^+)] da_0^+}{\sum_{k=0}^{N_j} e_k^{(j)}} \right) - \left(\widehat{a_{0_j}^+}^B \right)^2 \\
&= \frac{\sum_{k=0}^{N_j} e_k^{(j)} [(k+c\alpha_j) c_j^2 + ((k+c\alpha_j) c_j)^2]}{\sum_{k=0}^{N_j} e_k^{(j)}} - \left(\widehat{a_{0_j}^+}^B \right)^2.
\end{aligned}$$

Observación 1: Cuando $c \rightarrow \infty$, cada $\widehat{a_{0_j}^+}^B$ converge de manera casi segura a α_j (el valor esperado de la priori).

Demostración:

$$\widehat{a_{0_j}^+}^B = \frac{\sum_{k=0}^{N_j} e_k^{(j)} (k + c \alpha_j)}{\sum_{k=0}^{N_j} e_k^{(j)} c_j} = \frac{\sum_{k=0}^{N_j} e_k^{(j)} k}{\sum_{k=0}^{N_j} e_k^{(j)} c_j} + \frac{\sum_{k=0}^{N_j} e_k^{(j)} c \alpha_j}{\sum_{k=0}^{N_j} e_k^{(j)} c_j} = \frac{\sum_{k=1}^{N_j} e_k^{(j)} k}{e_0^{(j)} c_j + \sum_{k=1}^{N_j} e_k^{(j)} c_j} + \frac{c \alpha_j}{c_j}.$$

Dado que $e_k^{(j)} > 0, \forall k \in \{0, 1, \dots, N_j\}$,

$$0 \leq \frac{\sum_{k=1}^{N_j} e_k^{(j)} k}{e_0^{(j)} c_j + \sum_{k=1}^{N_j} e_k^{(j)} c_j} \leq \frac{\sum_{k=1}^{N_j} e_k^{(j)} N_j}{\sum_{k=1}^{N_j} e_k^{(j)} c_j} = \frac{N_j}{c_j},$$

el primer término tiende a 0 cuando $c \rightarrow \infty$. Luego:

$$\lim_{c \rightarrow \infty} \widehat{a_{0_j}^+}^B = \lim_{c \rightarrow \infty} \frac{c \alpha_j}{c_j} = \alpha_j.$$

Observación 2: Cuando $c \rightarrow 0$, $\widehat{a_{0_j}^+}^B$ no depende de α_j .

Demostración: Es claro dado que cada vez que α_j aparece en la expresión, aparece multiplicado por c .

Es este otro sentido por el que nuestro método es *híbrido*. En el caso de la baseline hazard, al contrario de lo realizado con los parámetros Euclideos, no hemos necesitado lidiar con los efectos del truncamiento. Esto hace posible estimar $\lambda_0(\cdot)$ (modelizada como constante a trozos) utilizando la función de pérdida cuadrática. Naturalmente, podríamos haber optado también en este caso por el modo de la distribución a posteriori, pero el hecho de presentarse como una mezcla de distribuciones Gamma (lo que supone una forma multimodal) requeriría un proceso de maximización numérico, impidiendo la obtención de estimadores Bayesianos de manera explícita. Por esta segunda razón, hemos optado por la formulación híbrida.

Una ventaja importante de nuestro método en comparación con los existentes actualmente en la literatura sobre estimación Bayesiana en el modelo de riesgos aditivos, es su flexibilidad acorde a los objetivos del análisis estadístico. Es decir, si sólo necesitamos una interpretación de los resultados, o comparación de riesgo entre grupos (por ejemplo, entre un tratamiento

y placebo), entonces sólo la estimación de β es necesaria. Esto lo podemos lograr de manera eficiente con nuestro método sin la necesidad de estimar $\lambda_0(\cdot)$. Solo en el caso de necesitar una predicción, estimaremos $\lambda_0(\cdot)$ usando el algoritmo desarrollado previamente.

3.6. Simulaciones

A continuación, presentaremos una sección de simulaciones implementando los métodos desarrollados para la estimación Bayesiana de parámetros en el modelo aditivo, donde C denota a la variable de censura. Generaremos 1000 muestras de tamaños 100 y 500 respectivamente. Proponemos los siguientes parámetros:

- $\lambda(t) = 1 + 0.5z$ (es decir, $\lambda_0(t) = 1$, $\beta = 0.5$),
- $k = 1$, $C \sim \mathcal{E}(1/2)$, $Z \sim \chi_1^2$.

3.7. Simulaciones de los parámetros de regresión β

Para estimar el desvío estándar, construimos el HPD intervalo de $(1 - \alpha) 100\%$, es decir, $[b_l(\alpha), b_u(\alpha)]$. Luego, calculamos el estimador obtenido en (3.9):

$$\hat{\sigma}^B(\hat{\beta}_k^B) := \frac{b_u(\alpha) - b_l(\alpha)}{2 z_{1-\alpha/2}}.$$

Observaciones:

- $\hat{\beta}_k^B \geq 0$.
- Si $\beta_k > 0$ y $\|\mathbf{C}_\beta\| \rightarrow \infty$ o $n \rightarrow \infty$,

$$\hat{\beta}_k^B \approx \hat{\beta}_k^{\text{LY}} \quad \text{y} \quad \hat{\sigma}^B(\hat{\beta}_k^B) \approx \hat{\sigma}^{\text{LY}}(\hat{\beta}_k^{\text{LY}}).$$

En las tablas 3.1 y 3.2 hemos expuesto los resultados obtenidos al variar los parámetros de posición μ_β y el parámetro de escala (que llamaremos w) de la priori de β . Observamos que

las estimaciones se parecen a las del estimador clásico de Lin & Ying (expuesto en el margen superior de cada tabla) cuando:

- el parámetro de escala aumenta (es decir, la priori se hace menos informativa),
- la posición del parámetro a priori de posición se acerca al verdadero valor,
- el tamaño de la muestra aumenta.

Tab. 3.1: $n = 100$, LY = 0.5569 (0.2457)

μ_β	w						
	0.1	0.5	1	2	5	10	1000
0	0.3241 (0.1625)	0.4813 (0.2122)	0.5155 (0.2211)	0.5352 (0.2259)	0.5479 (0.2291)	0.5524 (0.2301)	0.5569 (0.2312)
0.25	0.4157 (0.1784)	0.5093 (0.2154)	0.5306 (0.2227)	0.5430 (0.2268)	0.5512 (0.2294)	0.5540 (0.2303)	0.5569 (0.2312)
0.5	0.5073 (0.1851)	0.5372 (0.2182)	0.5457 (0.2242)	0.5509 (0.2275)	0.5544 (0.2297)	0.5556 (0.2304)	0.5569 (0.2312)
0.75	0.5989 (0.1872)	0.5652 (0.2203)	0.5608 (0.2256)	0.5587 (0.2283)	0.5576 (0.2300)	0.5573 (0.2306)	0.5569 (0.2312)
1	0.6905 (0.1879)	0.5932 (0.2221)	0.5759 (0.2268)	0.5666 (0.2290)	0.5608 (0.2303)	0.5589 (0.2307)	0.5569 (0.2312)
2	1.0570 (0.1884)	0.7051 (0.2262)	0.6362 (0.2304)	0.5981 (0.2314)	0.5737 (0.2315)	0.5654 (0.2313)	0.5570 (0.2312)
10	3.9886 (0.1884)	1.6002 (0.2290)	1.1189 (0.2365)	0.8496 (0.2394)	0.6770 (0.2378)	0.6175 (0.2354)	0.5575 (0.2312)

Tab. 3.2: $n = 500$, $LY = 0.5106$ (0.1036)

μ_β	w						
	0.1	0.5	1	2	5	10	1000
0	0.4585 (0.0982)	0.4992 (0.1024)	0.5048 (0.1030)	0.5077 (0.1033)	0.5094 (0.1035)	0.5100 (0.1035)	0.5106 (0.1036)
0.25	0.4829 (0.0982)	0.5045 (0.1024)	0.5075 (0.1030)	0.5090 (0.1033)	0.5100 (0.1035)	0.5103 (0.1035)	0.5106 (0.1036)
0.5	0.5074 (0.0982)	0.5099 (0.1024)	0.5102 (0.1030)	0.5104 (0.1033)	0.5105 (0.1035)	0.5106 (0.1035)	0.5106 (0.1036)
0.75	0.5319 (0.0982)	0.5152 (0.1024)	0.5129 (0.1030)	0.5118 (0.1033)	0.5111 (0.1035)	0.5108 (0.1035)	0.5106 (0.1036)
1	0.5564 (0.0982)	0.5205 (0.1024)	0.5156 (0.1030)	0.5131 (0.1033)	0.5116 (0.1035)	0.5111 (0.1035)	0.5106 (0.1036)
2	0.6543 (0.0982)	0.5419 (0.1024)	0.5264 (0.1030)	0.5186 (0.1033)	0.5138 (0.1035)	0.5122 (0.1035)	0.5106 (0.1036)
10	1.4375 (0.0982)	0.7128 (0.1024)	0.6128 (0.1030)	0.5620 (0.1033)	0.5312 (0.1035)	0.5209 (0.1035)	0.5107 (0.1036)

3.8. Simulaciones para los parámetros de la baseline

A continuación, presentamos un nuevo estudio de simulaciones para mostrar el rendimiento de los estimadores hallados, sobre el mismo modelo $\lambda(t) = 1 + 0.5z$. Para el desvío estándar de los estimadores Bayesianos de los incrementos en la cumulative baseline, proponemos el estimador $\hat{\sigma}^B(\Lambda_{0_j}^+) = \sqrt{V[\Lambda_{0_j}^+ | \mathcal{X}, \beta]}$.

- La grilla sobre \mathbb{R}_0^+ es seleccionada acorde a los cuantiles aproximados 0.2, 0.4, 0.6 y 0.8 (los cuales resultan 0.125, 0.3, 0.6 y 1.15 respectivamente) de la variable en base a la cual se generó la muestra. La cantidad de observaciones “en riesgo” disminuye a

medida que el tiempo avanza, por lo que se espera una mayor precisión en la estimación de los parámetros correspondientes a los intervalos más cercanos al 0.

- De acuerdo a esta grilla, $\Lambda_0^+ = (0.125, 0.175, 0.3, 0.55)$.
- La priori elegida es una función continua a izquierda $\alpha(t)$, que toma valores:
 - $\alpha(0) = 0$,
 - $\alpha(0.125) = 5$,
 - $\alpha(0.3) = 6$,
 - $\alpha(0.6) = 6.3$,
 - $\alpha(1.15) = 6.31$
 - cuando $t \rightarrow \infty$, $\alpha(t) \rightarrow \infty$.
- La priori resultante es $(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (5, 1, 0.3, 0.01)$. Notemos que la priori marca una tendencia hacia la derecha en los primeros dos intervalos de la grilla, y hacia la izquierda en el cuarto intervalo.
- El “peso” de la priori es regulado por el parámetro c . Presentamos una simulación para los valores $c = 10$, $c = 1$ y $c = 0.1$.
- En esta simulación, estimamos $\hat{\beta}$ (en este caso resulta univariado, por lo que C_{β} es de dimensión 1, y lo llamamos σ_{β}) con el estimador Bayesiano híbrido, con parámetros a priori $\mu_{\beta} = 0.5$ y $\sigma_{\beta} = 10000$. El valor del parámetro de dispersión se eligió de manera de minimizar el impacto de la priori. Los estimadores de los parámetros de la cumulative baseline se obtienen en base a esta estimación.

Tab. 3.3: $n = 100$, $\hat{\beta} = 0.5569$

c	$\Lambda_{0_1}^+ = 0.125$	$\Lambda_{0_2}^+ = 0.175$	$\Lambda_{0_3}^+ = 0.3$	$\Lambda_{0_4}^+ = 0.55$
10	0.6527 (0.0824)	0.2967 (0.0657)	0.2991 (0.0825)	0.3354 (0.1127)
1	0.1880 (0.0495)	0.1845 (0.0587)	0.2967 (0.0928)	0.5167 (0.1762)
0.1	0.1265 (0.0431)	0.1688 (0.0577)	0.2951 (0.0943)	0.5384 (0.1850)

Tab. 3.4: $n = 500$, $\hat{\beta} = 0.5106$

c	$\Lambda_{0_1}^+ = 0.125$	$\Lambda_{0_2}^+ = 0.175$	$\Lambda_{0_3}^+ = 0.3$	$\Lambda_{0_4}^+ = 0.55$
10	0.2510 (0.0247)	0.2041 (0.0269)	0.2991 (0.0406)	0.4919 (0.0732)
1	0.1384 (0.0198)	0.1771 (0.0260)	0.2990 (0.0416)	0.5437 (0.0799)
0.1	0.1257 (0.0192)	0.1742 (0.0259)	0.2990 (0.0417)	0.5494 (0.0807)

Observaciones:

- Como era esperado, el valor de c tiene un impacto en el estimador (otorgando mayor o menor peso a la priori),
- el desvío estándar aumenta conforme el intervalo donde se realiza la estimación se aleja del 0,
- cuando el tamaño de la muestra aumenta, la priori pierde relevancia.

3.9. Datos Reales

A continuación, aplicamos el estimador Bayesiano híbrido de β a la base de datos de “Welsh Nickels refinery”, la cual fue introducida por Doll *et al* (1970), y luego analizada por Breslow & Day (1987), Lin & Yin (1994) y Álvarez & Ferrario (2016), entre otros. Esta base contiene datos completos de 679 trabajadores de una refinería de níquel en el sur de Gales, buscando determinar el riesgo de contraer carcinoma de bronquios o senos nasales asociados al refinamiento de níquel. Los datos fueron recolectados de los recibos de pago semanales de la empresa entre los años 1934 y 1981. Las variables involucradas son AFE:= Age at first employment (*edad del primer empleo*), YFE:= Year at first employment (*año del primer empleo*) y EXP:= Exposure level (*nivel de exposición*), y las transformaciones de las variables se realizan de manera que los resultados sean comparables con los análisis previos. Cuando la priori tiene más peso (el caso de $\sigma_{\beta} = 0.01$), se puede observar un corrimiento hacia el valor central $\mu_{\beta} = 0$. Notemos que, al disminuir el peso ($\sigma_{\beta} = 1000$), el estimador Bayesiano híbrido coincide (con este nivel de precisión) con el estimador clásico.

Tab. 3.5: Los resultados se multiplicaron por 10^3 .

	$\hat{\beta}_{LY}$	$\mu_{\beta} = 0 \ \sigma_{\beta} = 0.01$	$\mu_{\beta} = 0 \ \sigma_{\beta} = 1000$
$\log(\text{AFE}-10)$	2.5126	2.4183	2.5126
$(\text{YFE}-1915)/10$	0.2179	0.1868	0.2179
$-(\text{YFE}-1915)^2/100$	2.6005	2.3829	2.6005
$\log(\text{EXP}+1)$	1.7172	1.7081	1.7172

4. PRIORIS NO INFORMATIVAS

Una priori no informativa busca tener el menor impacto posible sobre la distribución a posteriori. Se las suele llamar también vagas, difusas, planas o de referencia. El área de las prioris no informativas es amplia y polémica, ya que se suelen utilizar dichas prioris para representar ignorancia. Podemos clasificarlas en dos grupos:

Definición. Diremos que una priori π es propia si

$$\int_{\Theta} \pi(\theta) d\theta = 1.$$

Definición. Diremos que una priori es impropia si

$$\int_{\Theta} \pi(\theta) d\theta = \infty.$$

Cabe destacar que una distribución a priori impropia puede conducir a una posteriori impropia, en cuyo caso no se podrán hacer inferencias.

Algunos motivos por los que puede resultar conveniente utilizar prioris no informativas son:

- * minimizar el impacto de la información a priori disponible,

- * dificultad para una correcta elucidación de las prioris a utilizar,

- * el objetivo del estudio radica en evaluar la bondad del ajuste de un modelo. En estos casos, la priori no informativa podría ser utilizada como “referencia”,

* en problemas de alta dimensión, es usual utilizar prioris informativas para los parámetros de interés, y prioris no informativas para los parámetros de estorbo del modelo, buscando neutralizar su impacto,

* el uso de prioris no informativas en un análisis Bayesiano puede resultar de utilidad para construir procedimientos clásicos buenos.

En muchas situaciones, una priori impropia puede interpretarse como un límite de distribuciones propias. A continuación, detallaremos algunos casos de prioris no informativas clásicas, implementándolas en el AHM.

4.1. Priori Uniforme de Laplace

La primer priori impropia en la historia fue propuesta por Laplace, y consiste en utilizar como priori para el parámetro θ la función $\pi(\theta) = 1$ sobre el dominio paramétrico. De esta manera, la distribución a posteriori resulta la función de verosimilitud.

En el caso del modelo aditivo, y siempre modelizando a la baseline hazard como constante a trozos, recordamos el desarrollo obtenido en (3.1) donde desarrollamos la productoria

$$\prod_{k_j=1}^{n_j} [a_j + \boldsymbol{\beta}' z_{k_j}]^{\delta_{k_j}} = d_0 + d_1 a_j + \dots + d_{N_j} a_j^{N_j} \quad (4.1)$$

en búsqueda de obtener una expresión para la verosimilitud. Considerando una priori impropia de Laplace, y asumiendo independencia entre los parámetros de la baseline y los parámetros regresores, la distribución a posteriori de la baseline condicionada a los parámetros regresores resulta igual a la función de verosimilitud (3.2):

$$f_{A_1, \dots, A_J | \mathcal{X}, \mathbf{B}} \propto \prod_{j=1}^J \sum_{k=0}^{N_j} d_k a_j^k e^{-a_j \left(\sum_{k_j=1}^{n_j} (t_{k_j} - s_{j-1}) + m_j (s_j - s_{j-1}) \right)}. \quad (4.2)$$

Análogamente a lo desarrollado para los parámetros de la baseline, desarrollamos la expansión polinómica previa acorde a un parámetro β_q fijo (considerando a los demás parámetros de regresión conocidos, donde z_{-q} y $\boldsymbol{\beta}_{-q}$ representan a los vectores de covariables y parámetros regresores respectivamente omitiendo la q -ésima componente). La expresión resultante es:

$$\prod_{j=1}^J \prod_{k_j=1}^{n_j} [a_j + \boldsymbol{\beta}' z_{k_j}]^{\delta_{k_j}} = \prod_{l=1}^n [\lambda_0(t_l) + \boldsymbol{\beta}_{-q} z_{l-q} + \beta_q z_{l_q}]^{\delta_l} = q_0 + q_1 \beta_q + \dots + q_N \beta_q^N, \quad (4.3)$$

donde, al igual que en Capítulo 3, $N = \sum_{l=1}^n \delta_l$. Advertimos que los coeficientes $\{q_k\}_{k=0, \dots, N}$ se pueden generar mediante el mismo algoritmo utilizado para construir la función de verosimilitud (3.2). Luego, la distribución a posteriori de β_q condicionada al resto de los parámetros es:

$$f_{\beta_q | \mathcal{X}, L_0, \mathbf{B}_{-q}} \propto \sum_{k=0}^N q_k \beta_q^k e^{-\beta_q (\sum_{l=1}^n z_{l_q} t_l)}. \quad (4.4)$$

En este punto, la totalidad de las distribuciones condicionales son conocidas y con la misma distribución (mezcla de distribuciones Gamma), por lo que es posible generar muestras aleatorias de ellas. De esta manera, implementando el algoritmo de Gibbs, podremos hallar estimadores Bayesianos de la totalidad de los parámetros del modelo. Es importante notar que en este caso, los estimadores Bayesianos no los podremos hallar de forma explícita, como lo hicimos mediante el método híbrido.

4.2. Priori de Jeffreys

La priori no informativa de Laplace podrá aportar ignorancia cuando uno desee obtener información del parámetro θ al que se le asignó la distribución, pero no es así cuando la inferencia se realiza sobre una transformación del parámetro (por ejemplo, θ^2). Esto llevó a Jeffreys a proponer la hoy famosa “priori de Jeffreys”:

$$\pi(\theta) = \sqrt{\det(I_F(\theta))},$$

donde $I_F(\theta)$ denota la información de Fisher del parámetro θ . Este método posee la ventaja de ser *invariante* ya que tiene la propiedad de conservarse mediante transformaciones, y ha probado ser un éxito notable en problemas unidimensionales. El mismo Jeffreys, sin embargo, advirtió la vulnerabilidad del método en el caso multidimensional.

Procederemos a realizar los cálculos buscando obtener la información de Fisher de los parámetros del modelo aditivo:

$$f_{\mathbf{L}_0, \mathbf{B}, \mathcal{X}} = \exp \left\{ - \sum_{l=1}^n (\Lambda_0(t_l) + \boldsymbol{\beta}' t_l z_l) \right\} \prod_{j=1}^J \left(\prod_{k_j=1}^{n_j} [a_j + \boldsymbol{\beta}' z_{k_j}]^{\delta_{k_j}} \right),$$

Luego, calculamos su logaritmo:

$$\log(f_{\mathbf{L}_0, \mathbf{B}, \mathcal{X}}) = - \sum_{l=1}^n \Lambda_0(t_l) - \boldsymbol{\beta}' \sum_{l=1}^n t_l z_l + \sum_{j=1}^J \left(\sum_{k_j=1}^{n_j} \delta_{k_j} \log [a_j + \boldsymbol{\beta}' z_{k_j}] \right).$$

Para cada a_i , consideramos los conjuntos $T^- = \{t_l : t_i < s_{i-1}\}$, $T = \{t_l : s_{i-1} \leq t_l < s_i\}$ y $T^+ = \{t_l : s_i \leq t_l\}$. Bajo esta notación:

$$\sum_{l=1}^n \Lambda_0(t_l) = \sum_{t_l \in T^-} \Lambda_0(t_l) + \sum_{t_l \in T} \Lambda_0(t_l) + \sum_{t_l \in T^+} \Lambda_0(t_l).$$

Conforme a esto:

$$\frac{\partial \log f_{\mathbf{L}_0, \mathbf{B}, \mathcal{X}}}{\partial a_i} = - \sum_{t_l \in T} (t_l - s_{i-1}) - m_i (s_i - s_{i-1}) + \left(\sum_{k_i=1}^{n_i} \frac{\delta_{k_i}}{a_i + \boldsymbol{\beta}' z_{k_i}} \right),$$

$$\frac{\partial \log f_{\mathbf{L}_0, \mathbf{B}, \mathcal{X}}}{\partial \beta_q} = - \sum_{i=1}^n t_i z_{i_q} + \sum_{j=1}^J \left(\sum_{k_j=1}^{n_j} \frac{\delta_{k_j}}{a_j + \boldsymbol{\beta}' z_{k_j}} z_{k_{j_q}} \right).$$

Finalmente:

$$\frac{\partial^2 \log f_{\mathbf{L}_0, \mathbf{B}, \mathcal{X}}}{\partial a_i^2} = - \sum_{k_i=1}^{n_i} \frac{\delta_{k_i}}{(a_i + \boldsymbol{\beta}' z_{k_i})^2},$$

$$\frac{\partial^2 \log f_{\mathbf{L}_0, \mathbf{B}, \mathcal{X}}}{\partial \beta_q^2} = - \sum_{j=1}^J \left(\sum_{k_j=1}^{n_j} \frac{\delta_{k_j}}{(a_j + \boldsymbol{\beta}' z_{k_j})^2} z_{k_{j_q}}^2 \right).$$

En este punto, notamos que obtener el determinante de la matriz de información de Fisher conlleva un esfuerzo formidable. En el caso de los parámetros de la baseline, se la modelizó como constante a trozos (de no ser así, deberíamos hallar la información de Fisher sobre el funcional que representa la curva, aspecto que no se ha encontrado en la literatura). En el caso del análisis de Supervivencia, la estimación se suele centrar en el parámetro regresor en sí, y no en una función de él. El cálculo de la priori de Jeffreys queda de esta manera abierto a una posible investigación a futuro.

4.3. *Priori de Máxima Entropía*

La entropía es un concepto usualmente asociado a la física y la ingeniería, el cual es una medida de la variación de la incerteza de una variable aleatoria. Entre mediados y fines del siglo XIX surgieron los primeros trabajos alusivos a la entropía, con el físico alemán Carnot (quien brindó una formulación de la teoría en 1824), y el físico alemán Clausius (quien detalló una forma matemática rigurosa en 1850). Luego, el físico austríaco Boltzmann logró proporcionar una profunda interpretación de esta ley, concluyendo que la entropía de un sistema aislado está vinculado a la probabilidad de su estado actual.

Formalmente, Shannon (ingeniero eléctrico y matemático, recordado como “el padre de la teoría de la información”) define la entropía como una función que permite medir la cantidad de información asociada a un proceso aleatorio y que, a mayor probabilidad de que se produzca, menor será la información que aporte.

Consideremos a β una variable aleatoria que toma valores en $\mathbb{R}^+ := (0, +\infty)$. Sea f_β la función de densidad de β . Entonces, f_β cumple las siguientes condiciones:

- $f_\beta(t) \geq 0, \forall t \in \mathbb{R}^+,$
- $\int_{\mathbb{R}^+} f_\beta(t) dt = 1.$

Supongamos que disponemos de información a priori del primer momento de β (o sea, conocemos $E(\beta) = b$). Buscaremos, entonces, mediante herramientas de cálculo variacional (una recopilación de ellas puede encontrarse en Sagan (1992)), hallar la función de densidad f_β que maximice la entropía de Shannon $S(f_\beta)$:

$$S(f_\beta) := - \int_{\mathbb{R}^+} \log[f_\beta(t)] f_\beta(t) dt.$$

La interpretación de este resultado sería buscar, dentro de todas las funciones de densidades posibles con dominio en \mathbb{R}^+ , sujetas a la condición del primer momento, aquella que sea “menos informativa”.

Para resolver este problema, consideremos:

- $f(x, y, t) = -x \log(x)$
- $g_1(x, y, t) = x$
- $g_2(x, y, t) = tx$

Acorde a cómo están definidas estas funciones, existen λ_1 y λ_2 reales tales que f_β satisface la ecuación de Euler-Lagrange para la ecuación h definida por:

$$h(x, y, t) = f(x, y, t) - \lambda_1 g_1(x, y, t) - \lambda_2 g_2(x, y, t),$$

es decir:

$$\frac{\partial h}{\partial x}(f_\beta(t), f'_\beta(t), t) - \frac{\partial h}{\partial y}(f_\beta(t), f'_\beta(t), t) = 0, \forall t \in \mathbb{R}^+.$$

En virtud de que

$$\frac{\partial h}{\partial x} = -\log(x) - 1 - \lambda_1 - \lambda_2 t \quad \text{y} \quad \frac{\partial h}{\partial y} = 0$$

tenemos que:

$$-\log(x) - 1 - \lambda_1 - \lambda_2 t = 0.$$

Por consiguiente, si f_β es función extremal, obtenemos en consecuencia que:

$$f_\beta = C e^{-\lambda_2 t}, \quad C = e^{1+\lambda_1}.$$

Dado que f_β es una función de densidad,

$$\int_0^\infty C e^{-\lambda_2 t} dt = \frac{C}{\lambda_2} = 1,$$

por lo tanto, $C = \lambda_2$.

Con esto, llegamos a que la priori de máxima entropía de una variable con dominio en \mathbb{R}^+ , con primer momento conocido e igual a b , es la densidad de una v.a. exponencial con esperanza b .

Luego, considerando la expresión de la función de verosimilitud desarrollada en (4.4) acorde a cada parámetro de regresión

$$\mathcal{L}_n(\beta_j | \mathcal{X}, L_0, \boldsymbol{\beta}_{-j}) \propto \sum_{k=0}^N q_k \beta_j^k e^{-\beta_j (\sum_{i=1}^n z_{i,j} t_i)},$$

y multiplicándola por la priori de máxima entropía (suponiendo independencia sería el producto de las densidades), la expresión de la posteriori con respecto a cada parámetro regresor resulta en

$$f_{\beta_j | \mathcal{X}, L_0, \boldsymbol{\beta}_{-j}} \propto \sum_{k=0}^N q_k \beta_j^k e^{-\beta_j (\sum_{i=1}^n z_{i,j} t_i)} \times \frac{1}{b_j} e^{-\frac{\beta_j}{b_j}} \propto \sum_{k=0}^N q_k \beta_j^k e^{-\beta_j \left(\sum_{i=1}^n z_{i,j} t_i + \frac{1}{b_j} \right)},$$

es decir, resulta en una mezcla de densidades Gamma.

Con respecto a los parámetros de la *baseline hazard*, la priori de máxima entropía para cada valor en los intervalos acotados es *Uniforme*. Por lo que la posteriori resultante considerando la totalidad de los parámetros del modelo y tomando como priori a la de máxima entropía (de acuerdo a la restricción del primer momento en los parámetros regresores) resulta proporcional a una mezcla de densidades Gamma. De esta manera, un análisis mediante simulaciones utilizando el algoritmo de Gibbs es posible, permitiendo obtener estimadores Bayesianos de forma numérica.

5. OTRAS OPCIONES DE PROCESOS A PRIORI PARA LA BASELINE

5.1. Proceso Gamma sobre la baseline hazard $\lambda_0(\cdot)$

A diferencia del proceso Gamma que incorporamos en la cumulative baseline utilizado en el Capítulo 3, otra opción es aplicar un proceso Gamma directamente sobre la baseline. Esto claramente será posible sólo en los casos donde podamos asumir que la baseline es monótona creciente. Dado que el proceso Gamma se aplica sobre los incrementos de un proceso estocástico, en este caso lo aplicaremos sobre los incrementos $\Delta_j = A_j - A_{j-1}$, donde los valores $\{A_j\}_{j=1,\dots,J}$ representan las variables que denotan los distintos valores de la baseline hazard bajo el modelo constante a trozos. Consideramos la reformulación trivial $A_j = \sum_{i=1}^j (A_i - A_{i-1})$, donde $A_0 = 0$, y reemplazando en (3.2), llegamos a una expresión similar como mezcla de distribuciones Gamma, donde sólo cambian los coeficientes del polinomio (en este caso, dependen de los otros valores de la baseline, lo que resulta natural dado la priori elegida).

El reemplazo a realizar resulta en:

$$A_j = \sum_{i=1}^j \Delta_i = \Delta_j + \sum_{i=1}^{j-1} \Delta_i.$$

Luego, el procedimiento es el mismo que el desarrollado anteriormente, incluyendo, para cada $j \in \{1, \dots, J\}$, el término $\sum_{i=1}^{j-1} \Delta_i$ en la construcción de los coeficientes del polinomio d_k .

5.2. Proceso Beta sobre la cumulative baseline $\Lambda_0(\cdot)$

Recordemos que, en el caso continuo, la cumulative baseline es

$$\Lambda(t) := -\log(S(t)).$$

Una expresión más general para la función de riesgo acumulado es definida acorde a la función de intensidad (definida previamente en (2.1)), acorde al proceso de conteo $N(t) = N(0, t)$, el cual si tomamos $N(0) = 0$ cuenta la cantidad de eventos ocurridos en el intervalo $[0, t)$. La misma es válida aún cuando la función de Supervivencia no es continua, y su expresión es

$$\Lambda(t) = \int_{[0, t]} \frac{dF(u)}{S(u)}.$$

Ambas expresiones son equivalentes cuando F es absolutamente continua, el cual será el caso en el que trabajaremos.

Utilizaremos la expresión de $F(t)$ bajo producto integral,

$$F(t) = 1 - S(t) = 1 - \prod_{[0, t]} \{1 - d\Lambda(t)\}.$$

Un análisis detallado sobre esta reformulación de los modelos de supervivencia puede encontrarse, por ejemplo, en Cook & Lawless (2007) o Aalen *et al* (2008).

En el caso del AHM, dado un vector de covariables z , la expresión de la función de Supervivencia de acuerdo a esta notación es

$$S(s_j) = \prod_{i=1}^j (1 - h_i) \exp \{ -\beta' z ds_i \}.$$

Bajo la formulación del producto integral, la expresión que adquiere la verosimilitud en el AHM es la siguiente:

$$\mathcal{L}_n(L_0, \beta) \propto \prod_{j=1}^J \left((1 - h_j) \exp \left\{ - \sum_{i \in \mathcal{R}_j - \mathcal{D}_j} \beta' z_i dt_i \right\} \right) \prod_{l \in \mathcal{D}_j} (1 - (1 - h_j) \exp \{ -\beta' z_l dt_l \}),$$

donde $h_j = \Lambda_0(s_j) - \Lambda_0(s_{j-1})$, \mathcal{R}_j es el conjunto de observaciones que aún se encuentran en riesgo en el instante s_{j-1} y \mathcal{D}_j es el conjunto de observaciones que presentan el evento en $(s_{j-1}, s_j]$. El proceso Beta con incrementos independientes como priori para $\Lambda_0(\cdot)$ fue presentado por Hjort (1990).

Un proceso Beta genera una función de distribución acumulada $F(t)$, con incrementos independientes

$$d\Lambda(s) \sim \mathcal{B}(c(s)d\Lambda^*(s), c(s)[1 - d\Lambda^*(s)]),$$

donde el asterisco como supraíndice enfatiza el carácter de hiperparámetro (de la priori) y $\mathcal{B}(a, b)$ denota una distribución Beta con parámetros (a, b) . Debido a la complicada propiedad de convolución entre distribuciones Beta independientes, la distribución exacta del incremento $\Lambda(s)$ es sólo aproximadamente Beta sobre un intervalo finito, independientemente de cuán pequeña sea la longitud del intervalo. Multiplicando por la priori

$$\tau(h_j) \propto \prod_{j=1}^J h_j^{c_{0j}\alpha_{0j}-1} (1 - h_j)^{c_{0j}(1-\alpha_{0j})-1},$$

o, equivalentemente, tomando $h_j \sim \mathcal{B}(c_{0j}\alpha_{0j}, c_{0j}(1 - \alpha_{0j}))$, hallamos la siguiente expresión para la posteriori evaluada en cada elemento h_j

$$\begin{aligned} f(h_j|\boldsymbol{\beta}, D) &\propto h_j^{c_{0j}\alpha_{0j}-1} (1 - h_j)^{c_{0j}(1-\alpha_{0j})} c_{\beta}(j) \prod_{l \in \mathcal{D}_j} (1 - (1 - h_j) \exp\{-\boldsymbol{\beta}' z_l dt_l\}) \\ &= h_j^{c_{0j}\alpha_{0j}-1} (1 - h_j)^{c_{0j}(1-\alpha_{0j})} c_{\beta}(j) \prod_{l \in \mathcal{D}_j} (1 - \exp\{-\boldsymbol{\beta}' z_l dt_l\} + h_j \exp\{-\boldsymbol{\beta}' z_l dt_l\}) \\ &= h_j^{c_{0j}\alpha_{0j}-1} (1 - h_j)^{c_{0j}(1-\alpha_{0j})} c_{\beta}(j) \prod_{l \in \mathcal{D}_j} (b_{\beta}(l) + a_{\beta}(l)h_j) \\ &= h_j^{c_{0j}\alpha_{0j}-1} (1 - h_j)^{c_{0j}(1-\alpha_{0j})} c_{\beta}(j) \sum_{k=0}^{\#\mathcal{D}_j} c_k h_j^k \\ &= \sum_{k=0}^{\#\mathcal{D}_j} c_k h_j^{c_{0j}\alpha_{0j}+k-1} (1 - h_j)^{c_{0j}(1-\alpha_{0j})}, \end{aligned}$$

donde $c_\beta(j) = \exp\left\{-\sum_{i \in \mathcal{R}_j - \mathcal{D}_j} \beta' z_i dt_i\right\}$, $\#\mathcal{D}_j$ es la cantidad de elementos de \mathcal{D}_j , $b_\beta(l) = 1 - \exp\{-\beta' z_l dt_l\}$, $a_\beta(l) = \exp\{-\beta' z_l dt_l\}$ y el conjunto $\{c_k\}_{k \in \{0, \dots, \#\mathcal{D}_j\}}$ se calcula mediante el algoritmo recursivo presentado en el Capítulo 3. Luego, a través de la expansión polinómica desarrollada previamente, la posteriori resultante es una mezcla de distribuciones Beta, lo que nos permite hallar el estimador Bayesiano (bajo función de pérdida cuadrática) de manera analítica, utilizando la linealidad de la esperanza.

5.3. Proceso Dirichlet bajo conocimiento específico de $S(t)$ para t_F fijo

El vector aleatorio $y = (y_1, \dots, y_{k-1})$ se dice que tiene distribución Dirichlet $k - 1$ dimensional, que denotaremos por $y \sim D_{k-1}(\alpha_1, \alpha_2, \dots, \alpha_k)$, si $(y_1, y_2, \dots, y_{k-1})$ tiene densidad conjunta

$$f(y_1, y_2, \dots, y_{k-1}) = \left(\frac{\Gamma(\alpha)}{\prod_{j=1}^k \Gamma(\alpha_j)} \right) \left(\prod_{j=1}^{k-1} y_j^{\alpha_j - 1} \right) \left(1 - \sum_{j=1}^{k-1} y_j \right)^{\alpha_k - 1},$$

donde

$$y_j \geq 0, \quad j = 1, 2, \dots, k-1, \quad \sum_{j=1}^{k-1} y_j \leq 1, \quad \alpha = \sum_{j=1}^k \alpha_j.$$

Un proceso estocástico P indexado por los elementos de una partición dada $B = (B_1, B_2, \dots, B_k)$ de un espacio muestral Ω se dice que es un proceso Dirichlet en (Ω, B) con vector de parámetros α , si para cada partición distinta B de Ω el vector aleatorio $(P(B_1), P(B_2), \dots, P(B_k))$ tiene distribución Dirichlet con parámetros $(\alpha(B_1), \alpha(B_2), \dots, \alpha(B_k))$.

En los estudios oncohematológicos de Linfoma de Hodgkin, por ejemplo, se asume en base a la experiencia que la curva de supervivencia global¹ entra en una suerte de plató del 40% luego de los tres años de seguimiento. Esta información puede ser utilizada, definiendo como

¹ La supervivencia global en este tipo de estudios se mide desde el diagnóstico, y el evento de interés es la muerte por enfermedad.

Λ_0^{++} al valor de la cumulative baseline en ese punto t_F (los 36 meses), y recuperándola del dato

$$S(36) = 0.4 \implies \Lambda_0^{++} := \Lambda_0(36) = -\log(0.4) = 0.9162907.$$

De esta manera, dada la grilla (s_0, \dots, s_J) ,

$$\Lambda_0^{++} := \sum_{j=1}^J \Lambda_{0j}^+,$$

y asumiendo Λ_0^{++} constante, definimos nuevas variables con el fin de que su suma sea 1, y así poder implementar un proceso Dirichlet. De esta manera, definimos, para cada j :

$$\Lambda_{0j}^{(D)} := \frac{\Lambda_{0j}^+}{\Lambda_0^{++}}.$$

Por la manera en que están definidas, es claro que $\sum_{i=1}^J \Lambda_{0i}^{(D)} = 1$. Luego, podemos reescribir la verosimilitud (3.2) en función de las variables $\Lambda_{0j}^{(D)}$, las cuales no son más que un producto escalar de las variables Λ_{0j}^+ (las cuales a su vez son un producto escalar de las variables a_j). Conforme a esta modificación, adaptamos los parámetros d_k y c_j (los cuales llamaremos $d_k^{(D)}$ y $c_j^{(D)}$), de modo de conservar la igualdad. Debido a esto, estamos en condiciones de proponer una priori Dirichlet para los parámetros de la cumulative baseline. La verosimilitud conjunta para los parámetros de la *cumulative baseline* estandarizada resulta en:

$$f_{\Lambda_0^{(D)}|\mathcal{X},\boldsymbol{\beta},\Lambda_0^{++}}(a_{0j}^{(D)}) \propto \prod_{j=1}^J \sum_{k=0}^{N_j} d_k^{(D)} \left(a_{0j}^{(D)}\right)^k e^{-\left(a_{0j}^{(D)} c_j^{(D)}\right)},$$

es decir, una mezcla de distribuciones Gamma.

Dado que buscamos implementar un *proceso Dirichlet*, procedemos a reescribir la priori conjunta, reemplazando la última componente sobre la partición de la *baseline hazard* $a_{0J}^{(D)}$ por $1 - \sum_{j=1}^{J-1} a_{0j}^{(D)}$. Realizando el producto y llamando $K_j := 1 - \sum_{s \neq j} a_{0s}^{(D)}$, para $j \in \{1, \dots, J-1\}$ llegamos a una expresión de la verosimilitud condicional para cada parámetro

$\Lambda_{0j}^{(D)}$ de la forma:

$$f_{\Lambda_{0j}^{(D)}|\mathcal{X},\boldsymbol{\beta},\Lambda_0^{++}}(a_{0j}^{(D)}) \propto \sum_{k=0}^{N_j} d_k^{(D)} \left(a_{0j}^{(D)}\right)^k e^{-\left(a_{0j}^{(D)}c_j^{(D)}\right)} \left[\sum_{k=0}^{N_j} d_k^{(D)} (K_j - a_{0j}^{(D)})^k e^{-\left((K_j - a_{0j}^{(D)})c_j^{(D)}\right)} \right].$$

Aunque la expresión no resulte amigable debido a la cantidad de variables involucradas, advertimos que (utilizando nuevamente el algoritmo para generar los coeficientes polinomiales de $(K_j - a_{0j}^{(D)})^k$) y realizando los productos correspondientes, la distribución resultante no es más que una nueva mezcla de distribuciones *Gamma*, la que expresamos:

$$f_{\Lambda_{0j}^{(D)}|\mathcal{X},\boldsymbol{\beta},\Lambda_0^{++}}(a_{0j}^{(D)}) \propto \sum_{k=0}^{N_j+N_j} d_k^{(D^*)} \left(a_{0j}^{(D)}\right)^k e^{-\left(a_{0j}^{(D)}c_j^{(D^*)}\right)},$$

donde $d_k^{(D^*)}$ y $c_j^{(D^*)}$ no son más que las nuevas expresiones de los coeficientes adaptados a la nueva expresión de la verosimilitud.

La priori conjunta resulta en:

$$\pi(a_{01}^{(D)}, a_{02}^{(D)}, \dots, a_{0_{J-1}}^{(D)}) = \left(\frac{\Gamma(\alpha)}{\prod_{j=1}^J \Gamma(\alpha_j)} \right) \left(\prod_{j=1}^{J-1} (a_{0j}^{(D)})^{\alpha_j-1} \right) \left(1 - \sum_{j=1}^{J-1} a_{0j}^{(D)} \right)^{\alpha_J-1}.$$

Advertimos que, multiplicando dicha priori por la verosimilitud, obtenemos una expresión que resulta en una mezcla de distribuciones de la forma

$$f(x) \propto x^A (D - x)^B e^{Cx}. \quad (5.1)$$

A continuación, presentamos la distribución *Beta Generalizada*, mediante la cual modelizaremos la expresión obtenida.

Recordamos la distribución Beta, cuya función de densidad está dada por

$$f(x|u, v) = \frac{\Gamma(u+v)}{\Gamma(u)\Gamma(v)} x^{u-1} (1-x)^{v-1}, x \in (0, 1),$$

donde la única restricción sobre los parámetros u y v es que sean positivos.

Diremos que una variable aleatoria tiene distribución *Beta Generalizada* con parámetros (u, v, a, b) , y la denotaremos por $BG(u, v, a, b)$, si su función de densidad es

$$f_{BG}(x|u, v, a, b) = \frac{\Gamma(u+v)}{h^{u+v-1}\Gamma(u)\Gamma(v)}(x-a)^{u-1}(h-x)^{v-1}, x \in (a, a+h),$$

donde u, v, a y b son positivos, $b > a$ y $h = b - a$.

Recordando la definición de función generadora de momentos de una variable aleatoria X , en la cual utilizamos la expresión en serie de Taylor de la función exponencial

$$e^x = \sum_{k=1}^{\infty} \frac{x^k}{k!}, \forall t \in \mathbb{R},$$

obtenemos la expresión de la función generadora de momentos para una variable aleatoria con distribución *Beta Generalizada*:

$$\begin{aligned} M_X(t) &:= E(e^{tx}) = E(e^{t(x-a+a)}) \\ &= E(e^{t(x-a)+ta}) \\ &= e^{ta} E(e^{t(x-a)}) \\ &= e^{ta} \int_{-\infty}^{\infty} e^{t(x-a)} f_X(x) dx \\ &= e^{ta} \int_{-\infty}^{\infty} \sum_{k=1}^{\infty} \frac{(t(x-a))^k}{k!} f_X(x) dx \\ &= e^{ta} \sum_{k=1}^{\infty} \frac{t^k}{k!} \int_a^{a+h} (x-a)^k \frac{\Gamma(u+v)}{h^{u+v-1}\Gamma(u)\Gamma(v)} (x-a)^{u-1} (h-x)^{v-1} dx \\ &= e^{ta} \sum_{k=1}^{\infty} \frac{t^k}{k!} \int_a^{a+h} \frac{\Gamma(u+v)}{h^{u+v-1}\Gamma(u)\Gamma(v)} (x-a)^{k+u-1} (h-x)^{v-1} dx \\ &= e^{ta} \frac{\Gamma(u+v)}{\Gamma(u)} \sum_{k=1}^{\infty} h^k \frac{\Gamma(u+k)}{\Gamma(k+u+v)} \frac{t^k}{k!} \int_a^{a+h} f_{BG}(x|u, v, a, h) dx \\ &= e^{ta} \frac{\Gamma(u+v)}{\Gamma(u)} \sum_{k=1}^{\infty} h^k \frac{\Gamma(u+k)}{\Gamma(k+u+v)} \frac{t^k}{k!}, \end{aligned}$$

donde el intercambio entre los símbolos de sumatoria e integral están justificados por la convergencia absoluta del integrando. Recordando la expresión de la función Beta matemática

$$\mathbb{B}(x, y) := \frac{\Gamma(x+y)}{\Gamma(x)\Gamma(y)},$$

y multiplicando por $\frac{\Gamma(v)}{\Gamma(v)}$, la podemos reescribir

$$M_X(t) = e^{ta} \mathbb{B}(u, v) \sum_{k=1}^{\infty} h^k \frac{1}{\mathbb{B}(u+k, v)} \frac{t^k}{k!}.$$

Teniendo en cuenta que el estimador Bayesiano considerando función de pérdida cuadrática es la esperanza de la distribución a posteriori, calcular la esperanza no sería más que multiplicar por la variable (lo que en este caso, sólo aumentaría en 1 el exponente de la variable). Luego, calcular esta integral no es más que evaluar esta función generadora de momentos y evaluarla en C (como está expresado en (5.1)).

Si bien la función generadora de momentos está dada en forma de serie, dado que la serie es convergente $\forall t \in \mathbb{R}$, podremos aproximar este valor con la precisión que deseemos.

Luego, por linealidad de la esperanza, podremos calcular el estimador Bayesiano de manera numérica, pero sin necesidad de recurrir a algoritmos MCMC.

6. ELUCIDACIÓN DE PRIORIS

La elucidación de prioris juega un rol crucial en el análisis Bayesiano. Consiste en el proceso mediante el cual se selecciona la distribución a priori que mejor se ajusta al estudio de acuerdo a la información previa disponible. A continuación, presentamos una propuesta para transformar el conocimiento previo en información sobre la media de la distribución a priori, con el fin de obtener información acerca de los hiperparámetros.

6.1. *Elucidación para los parámetros regresores*

En la práctica, es usual que la información a priori venga dada como un intervalo de confianza para la función de sobrevivencia en un tiempo fijo condicionada a cierto parámetro regresor. Comencemos con el caso más simple: sea Z_k una variable dicotómica que toma valores 0 y 1, de la cual conocemos los intervalos de confianza de nivel $(1 - \alpha)$ para $S(t_1|Z_k = 0)$ y $S(t_1|Z_k = 1)$, los cuales llamamos $[l_{0,1-\alpha}, u_{0,1-\alpha}]$ y $[l_{1,1-\alpha}, u_{1,1-\alpha}]$ respectivamente, para algún tiempo t_1 . Dicha información puede provenir de alguna opinión experta, o de una base de datos previa. Por ejemplo, es usual que la opinión experta en cierto ensayo clínico indique los intervalos de confianza en algún instante del tiempo para las distintas cohortes, de acuerdo a si se encuentra o no presente cierta característica inicial. Nuestro objetivo será utilizar dicha información de manera de traducirla como conocimiento sobre los hiperparámetros de las distribuciones a priori.

Para esto, advertimos que, en el modelo aditivo:

$$\frac{S(t_1|Z_k = 1)}{S(t_1|Z_k = 0)} = e^{-\beta_k t_1}$$

Es decir, dados $\widehat{S}(t_1|Z_k = 1)$ y $\widehat{S}(t_1|Z_k = 0)$ (los cuales pueden estar dados a priori, o bien ser obtenidos de los intervalos de confianza) podemos obtener un estimador puntual para β_k :

$$\widehat{\beta}_k = -\frac{\log(\widehat{S}(t_1|Z_k = 1)) - \log(\widehat{S}(t_1|Z_k = 0))}{t_1}.$$

Luego, podemos utilizar esta información obtenida de manera de poder seleccionar los hiperparámetros de la distribución a priori que se ajusten a dicho parámetro de posición, y a la amplitud de dichos intervalos de confianza.

Este razonamiento puede ser extendido mediante manipulación algebraica al caso multinomial, o a variables continuas discretizadas. Tal es el caso de estudios clínicos en los que se considera alguna variable predictora continua, en la cual la información a priori suele estar asociada a ciertos intervalos de la variable, permitiendo una discretización de la misma.

6.2. Elucidación para los parámetros de la baseline

6.2.1. Susarla & Van Ryzin

En inferencia Bayesiana no paramétrica, una manera típica de introducir como priori un proceso Dirichlet es a través de una función de distribución $F(t) = 1 - S(t)$ (en el caso continuo), en base a la cual se calcula el vector de hiperparámetros. Susarla & Van Ryzin (1976) desarrollan un estimador Bayesiano para la función de Supervivencia implementando un proceso Dirichlet considerando censura por la derecha. Dada una distribución base $F_B(t)$, construyen un estimador $\widehat{S}^{S\&VR}(t)$ (el cual coincide con el estimador de Kaplan-Meier bajo una priori no informativa, o cuando el tamaño muestral tiende a ∞). En base a este estimador $\widehat{S}^{S\&VR}(t)$, proponemos una ligera modificación, de manera de traducir esta información a priori que tenemos sobre la sobrevivida $S(t)$, en información a priori sobre $S_0(t)$.

Dada una grilla de dimensión J , para cada $j \in \{1, \dots, J\}$ evaluamos el estimador $\widehat{S}^{S\&VR}(t)$

en cada punto de ella, obteniendo un estimador puntual $\hat{S}^{S\&VR}(s_j)$. Nuestro objetivo es obtener un estimador Bayesiano para $\hat{S}_0(t)$. Para ello, comenzamos definiendo las *proporciones de decrecimiento* de la sobrevida

$$p_j := S(s_{j-1}) - S(s_j). \quad (6.1)$$

Dada la naturaleza de estos nuevos parámetros, notemos que $\sum_{j=1}^J p_j = 1$. En base al estimador $\hat{S}^{S\&VR}(t)$, definimos

$$\hat{p}_j := \hat{S}^{S\&VR}(s_{j-1}) - \hat{S}^{S\&VR}(s_j),$$

de acuerdo a nuestra grilla. Recordando que bajo el modelo aditivo

$$S(t) = S_0(t) \exp(-z'\boldsymbol{\beta}t),$$

reescribimos la ecuación (6.1):

$$p_j = S_0(s_{j-1}) \exp(-z'\boldsymbol{\beta}s_{j-1}) - S_0(s_j) \exp(-z'\boldsymbol{\beta}s_j). \quad (6.2)$$

De acuerdo a como definimos la grilla, $s_0 = 0$ y $s_J = \infty$, lo que implica que $\hat{S}(0) = S(0) = 1$ y $\hat{S}(s_{j-1}) \geq S(s_j)$, ya que la función de Supervivencia es decreciente. En vista de esto, podemos hallar \hat{p}_j para cada j en $\{1, \dots, J-1\}$, calculando primero:

$$\begin{aligned} \hat{p}_1 &= \hat{S}^{S\&VR}(0) - \hat{S}^{S\&VR}(s_1) \\ &= 1 - \hat{S}_0(s_1) \exp(-z'\boldsymbol{\beta}s_1), \end{aligned}$$

logrando despejar de esta manera

$$\hat{S}_0(s_1) = \frac{1 - \hat{p}_1}{\exp(-z'\boldsymbol{\beta}s_1)}.$$

Repitiendo este procedimiento de manera recursiva, dados $\boldsymbol{\beta}$ y z , podemos recuperar $\hat{S}_0(s_j)$, para cada j en $\{1, \dots, J-1\}$ a través de la relación

$$S_0(s_j) = \frac{S_0(s_{j-1}) \exp(-z'\boldsymbol{\beta}s_{j-1}) - p_j}{\exp(-z'\boldsymbol{\beta}s_j)}$$

obtenida de la ecuación (6.2). De esta manera se puede asociar el conocimiento a priori de $S(t)$ a $S_0(t)$, suponiendo conocidos (o ya estimados) los parámetros regresores.

Para elucidar los hiperparámetros, una alternativa en este caso es, considerando un proceso a priori sobre los incrementos de la baseline (por ejemplo, un proceso Gamma), y disponiendo de alguna información de $S_0(t)$ (como una mediana t_{med}) despejamos los parámetros del proceso en función de $S_0(t_{med})$ y $S_0(t_F)$, donde t_F lo consideramos un tiempo conocido donde truncaremos el estudio. Esto lo podemos hacer ya que:

$$\begin{aligned}\Lambda_0(t_F) &= -\log(S_0(t_F)), \\ \Lambda_0(t_{med}) &= -\log(S_0(t_{med})) = -\log(1/2) = 0.6931472,\end{aligned}$$

Luego, en base a estos datos, obtenemos información sobre los parámetros de la curva $\alpha(t)$ que utilizaremos para generar un proceso a priori sobre $\lambda(t)$.

7. EXTENSIONES Y CONCLUSIONES

7.1. Extensiones

Si bien hemos presentado una propuesta de análisis Bayesiano detallado del AHM, las expresiones condicionales de la verosimilitud como una mezcla de distribuciones *Gamma* permite la incorporación de una gran variedad de prioris, permitiendo obtener estimadores Bayesianos explícitos mediante un análisis conjugado y, en los casos en los que no sea así, la expresión conocida de la verosimilitud en función de distribuciones flexibles (*Gamma*) facilitará el desarrollo de algoritmos de manera de realizar un análisis numérico.

Si bien optamos por un estimador híbrido para los parámetros regresores, en vista de la expresiones halladas para las distribuciones condicionales bajo la priori *Uniforme*, notamos que podemos obtener estimadores Bayesianos bajo función de pérdida cuadrática para estos parámetros, mediante el algoritmo de Gibbs.

Una extensión importante (la cual, bajos ciertas condiciones, es una consecuencia directa de algunos de los resultados obtenidos) es la opción de considerar a los parámetros regresores *dependientes del tiempo*. Suponiendo una grilla dada en el eje real, consideramos el modelo *constante a trozos* tanto para la *baseline hazard* como para los parámetros regresores, de la misma manera que en el trabajo de Silva & Amaral-Turkman (2005).

De esta manera, considerando a cada β_k como dependiente del tiempo (es decir, la consideramos como una función $\beta_k(t)$), la aproximamos a través del modelo *constante a trozos*:

$$\beta_k(t) = \begin{cases} \beta_{k_1}, & \text{si } 0 \leq t < s_1, \\ \beta_{k_2}, & \text{si } s_1 \leq t < s_2, \\ \vdots & \\ \beta_{k_J}, & \text{si } s_{J-1} \leq t, \end{cases}$$

permitiendo tratar a cada uno de ellos de la misma manera en la que venimos tratando a $\lambda_0(t)$. De esta manera, con el mismo procedimiento desarrollado para obtener las expresiones condicionales obtenidas en las ecuaciones (4.2), (4.3) y (4.4), podemos construir las distribuciones condicionales de acuerdo al modelo de Aalen (sin la restricción de que cada $\beta_k(t)$ sea constante), las cuales resultan nuevamente en una mezcla de distribuciones *Gamma*. Bajo esta nueva formulación, toda la sección de procesos a priori sobre la baseline es aplicable a cada parámetro regresor.

Desde el enfoque frecuentista, la ecuación (4.4) habilita a nuevas alternativas de estimadores clásicos.

Otra extensión posible es la inclusión de *frailties aditivas*, emulando nuevamente el trabajo de Silva & Amaral-Turkman (2005), con la importante ventaja que, bajo la nueva expresión de la verosimilitud, y acorde a las prioris seleccionadas, los estimadores Bayesianos se hallan de manera explícita.

Se pueden adaptar, además, las secciones de ajuste de modelo presentadas en trabajos previos, de manera de calcular el *CPO*, *BIC*, *DIC*, *medida L* y *riesgo de Bayes* (entre otros) acordes a este modelo.

Esta expresión se puede adaptar a otras áreas del análisis de supervivencia tales como eventos recurrentes, riesgos competitivos, y considerar otros tipos de censura.

7.2. Conclusiones

Hemos presentado un nuevo enfoque en el análisis del modelo del riesgo aditivo, a través de la nueva expresión de la verosimilitud mediante el cálculo de los coeficientes polinomiales en el desarrollo de

$$\prod_{i=1}^n (\lambda_0(t) + \beta x)^{\delta_i},$$

obteniendo una expresión en base a distribuciones conocidas, y facilitando su análisis e incorporación de *prioris*. Es remarcable el hecho de que esta distribución sea una mezcla de distribuciones *Gamma*, ya que esto ofrece una alternativa viable y simple de trabajar con el AHM, cuando se modeliza como *constante a trozos* a la componente no paramétrica de su función de riesgo.

Hasta el momento (y a nuestro entender) no hay trabajos desarrollados que permitan la obtención de estimadores Bayesianos de manera explícita, y esta expresión de la verosimilitud permite nuevas alternativas incluso desde una perspectiva clásica. El algoritmo desarrollado para generar los coeficientes polinomiales es numéricamente estable aún con muestras grandes, y en los casos en los que no sea posible calcularlo siempre es posible subdividir la grilla de manera de disminuir la cantidad de observaciones por intervalo.

Hemos trabajado en particular con el modelo AHM, y los métodos desarrollados son flexibles.

La sección de *prioris no informativas* ofrece una alternativa para disminuir el impacto de las *prioris* seleccionadas, y presentamos métodos de elucidación de manera de lograr traducir el conocimiento experto (el cual consideramos que será dado en términos de la función de sobrevivencia).

En las secciones de simulaciones, advertimos la relevancia de considerar una *priori* adecuada, justificando las ventajas y desventajas del enfoque.

Bibliografía

- Aalen, O. (1980). A model for nonparametric regression analysis of counting processes. In *Mathematical statistics and probability theory* (pp. 1-25). Springer, New York, NY.
- Aalen, O., Borgan, O., & Gjessing, H. (2008). *Survival and event history analysis: a process point of view*. Springer Science & Business Media.
- Álvarez, E. E. & Ferrario, J. (2012). Revisión de la Estimación Robusta en Modelos Semiparamétricos de Supervivencia. *IASI (Journal of the Interamerican Statistical Institute)* 64 (182-183), 85-106.
- Álvarez, E. E. & Ferrario, J. (2016). Robust estimation in the additive hazards model. *Communications in Statistics-Theory and Methods* 45 (4), 906-921.
- Andersen, P. K., Borgan, O., Gill, R. D., & Keiding, N. (2012). *Statistical models based on counting processes*. Springer Science & Business Media.
- Beamonte, E., & Bermúdez, J. D. (2003). A bayesian semiparametric analysis for additive hazard models with censored observations. *Test*, 12(2), 347-363.
- Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2), 211-243.
- Breslow, N. E., & Day, N. E. (1987). The design and analysis of cohort studies. Statistical methods in cancer research. *Lyon, France: International Agency for Research on Cancer*, 2.
- Chernoukhov, A. (2013). Bayesian Spatial Additive Hazard Model. *Electronic Theses and Dissertations. Windsor University*.

-
- Chernoukhov, A., Hussein, A., Nkurunziza, S., & Bandyopadhyay, D. (2018). Bayesian inference in time-varying additive hazards models with applications to disease mapping. *Environmetrics*, 29(5-6), e2478.
- Cook, R. J., & Lawless, J. (2007). The statistical analysis of recurrent events. Springer Science & Business Media.
- Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2), 187-202.
- Cox, R. T. (1946). Probability, frequency and reasonable expectation. *American journal of physics*, 14(1), 1-13.
- Doll, R., Morgan, L. G. and Speizer, F. E. (1970). Cancers of the lung and nasal sinuses in nickel workers. *British journal of cancer* 24 (4), 623-632.
- Dunson, D. B., & Herring, A. H. (2005). Bayesian model selection and averaging in additive and proportional hazards models. *Lifetime data analysis*, 11(2), 213-232.
- Fleming, Thomas R and Harrington, David P. Counting Processes and Survival Analysis. John Wiley & Sons. 2011.
- Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, (6), 721-741.
- Hastings, W. K. (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, 57(1), 97-109.
- Hjort, N. L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *the Annals of Statistics*, 18(3), 1259-1294.

-
- Ibrahim, J. G., Chen, M. H., & Sinha, D. (2014). Bayesian Survival Analysis. *Wiley StatsRef: Statistics Reference Online*.
- Ibrahim, J. G., Chen, M. and Sinha, D. (2001). *Bayesian Survival Analysis*. Springer Science & Business Media.
- Kalbfleisch, J. D., & Prentice, R. L. (1980). The Statistical Analysis of Failure Time Data. *John Wiley & Sons*.
- Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282), 457-481.
- Klein, J. P., & Moeschberger, M. L. (2006). *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media.
- Lawless, J. F. (2011). Statistical models and methods for lifetime data (Vol. 362). John Wiley & Sons.
- Lin, D. Y., & Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika*, 81(1), 61-71.
- Martinussen, T., & Scheike, T. H. (2007). *Dynamic regression models for survival data*. Springer Science & Business Media.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6), 1087-1092.
- Robert, C. (2007). *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Science & Business Media.
- Saad, A.A.M. (2012) Bayesian Non-parametric Estimation in Proportional Hazard Models: Survival Data Analysis with WinBUGS. LAP Lambert Academic Publishing.

-
- Sagan, H. (1992). *Introduction to the Calculus of Variations*. Courier Corporation.
- Silva, G. L., & Amaral-Turkman, M. A. (2005). Bayesian analysis of an additive survival model with frailty. *Communications in Statistics-Theory and Methods*, 33(10), 2517-2533.
- Silva, G. L. & Amaral-Turkman, M. A. (2007) Additive Survival Models with Shared Frailty. Proceedings of the 22nd International Workshop on Statistical Modelling. 544-548.
- Sinha, Debajyoti and Dey, Dipak K. Semiparametric Bayesian Analysis of Survival Data. *Journal of the American Statistical Association*. 1997;92(439):1195–1212.
- Sinha, Debajyoti, McHenry, M Brent, Lipsitz, Stuart R and Ghosh, Malay (2009) Empirical Bayes estimation for additive hazards regression models. *Biometrika*. 96(3):545–558.
- Susarla, V., & Van Ryzin, J. (1976). Nonparametric Bayesian estimation of survival curves from incomplete observations. *Journal of the American Statistical Association*, 71(356), 897-902.
- Yin, G., & Ibrahim, J. G. (2006). Bayesian transformation hazard models. In *Optimality* (pp. 170-182). Institute of Mathematical Statistics.
- Yin, Guosheng and Ibrahim, Joseph G. (2005a) Bayesian Frailty models based on Box-Cox transformed hazards. *Statistica Sinica*. 781–794.
- Yin, Guosheng and Ibrahim, Joseph G. (2005b) A Class of Bayesian Shared Gamma Frailty Models with Multivariate Failure Time Data. *Biometrics*. 61(1):208–216.
- Yin, Guosheng and Ibrahim, Joseph G. (2005c) A general class of Bayesian Survival models with zero and nonzero cure fractions. *Biometrics*. 61(2):403–412.

Zhang, J. and Lawson, A. B. (2011). Bayesian parametric accelerated failure time spatial model and its application to prostate cancer. *Journal of Applied Statistics* 38 (3), 591-603.