



Universidad de Valladolid

FACULTAD DE CIENCIAS

TRABAJO FIN DE GRADO

Grado en Matemáticas

EL MODELO DE REGRESIÓN DE COX

Autor: Adrián Peña de la Fuente
Tutor: Eustasio del Barrio Tellado
2025

*Agradecer a
Eustasio del Barrio
el tiempo y la dedicación empleados
para la elaboración de este trabajo.
De la misma manera, agradecer a mi familia
y amigos el apoyo moral
durante toda mi trayectoria universitaria.*

Resumen

El trabajo aborda el estudio del modelo de regresión de Cox (bajo la hipótesis de riesgos proporcionales), que es de gran utilidad en el ámbito del análisis de supervivencia debido a su naturaleza semiparamétrica. Para ello, se exponen los fundamentos matemáticos que sustentan el modelo, adentrándonos en la teoría de procesos estocásticos, con un énfasis particular en las martingalas, que son herramientas cruciales para el modelado de eventos que evolucionan en el tiempo y para el tratamiento estadístico de la censura.

Por otro lado, se explora en profundidad el campo de los procesos de conteo, analizando distintos aspectos como el concepto de intensidad, la descomposición de Doob-Meyer o la inferencia tanto paramétrica como no paramétrica. Además, todos estos elementos han sido objeto de estudio en ciertos casos particulares, lo que ha permitido profundizar en su comprensión y aplicabilidad en contextos específicos.

Entender correctamente toda esta base teórica es esencial para comprender la utilidad y poder desarrollar las propiedades estadísticas del modelo de Cox, introducido en el último capítulo.

Palabras clave

Análisis de supervivencia, procesos estocásticos, martingalas, censura, procesos de conteo, descomposición de Doob-Meyer, inferencia, modelo de regresión de Cox, riesgos proporcionales.

Abstract

The work focuses on the study of the Cox regression model (under the proportional hazards assumption), which is highly useful in the field of survival analysis due to its semiparametric nature. To this end, the mathematical foundations underlying the model are presented, delving into the theory of stochastic processes, with particular emphasis on martingales, crucial tools for modeling time-evolving events and for the statistical treatment of censoring.

In addition, the field of counting processes is explored in depth, analyzing various aspects such as the concept of intensity, the Doob–Meyer decomposition, and both parametric and nonparametric inference. All these elements have also been studied in specific cases, which has helped to deepen the understanding of their theoretical basis and applicability in particular contexts.

A proper understanding of this theoretical background is essential to grasp the usefulness and to develop the statistical properties of the Cox model, introduced in the last chapter.

Key words

Survival analysis, stochastic processes, martingales, censoring, counting processes, Doob–Meyer decomposition, inference, Cox regression model, proportional hazards.

ÍNDICE GENERAL

Introducción	11
1. Preliminares matemáticos	13
1.1. Función de supervivencia y función de riesgo	13
1.2. Procesos estocásticos	14
1.3. Esperanza condicionada	15
1.4. Martingalas	16
1.4.1. Procesos de variación predecible y opcional	18
1.4.2. Teorema central del límite para las martingalas	20
2. Procesos de conteo	21
2.1. Introducción a los procesos de conteo	21
2.1.1. Intensidades	22
2.1.2. Compensadores y descomposición de Doob-Meyer	23
2.1.3. Variación y covariación en procesos de conteo	24
2.1.4. Casos particulares	26
2.1.4.1. El proceso de Poisson	27
2.1.4.2. Proceso de salto único	29
2.1.4.3. Modelo de supervivencia	30
2.1.4.4. Modelo de supervivencia con censuras	30
2.2. Inferencia en procesos de conteo	31
2.2.1. Inferencia no paramétrica	32
2.2.1.1. El estimador de Nelson-Aalen	32
2.2.1.2. El estimador de Kaplan-Meier	35
2.2.2. Inferencia paramétrica	38
2.2.2.1. Verosimilitud en procesos de conteo	38
3. El modelo de regresión de Cox	47
3.1. El modelo de Cox de riesgos proporcionales	47
3.1.1. La hipótesis de riesgos proporcionales	50
3.1.2. Función de verosimilitud parcial. Estimación de los coeficientes	51
Conclusiones	55
A. Demostraciones capítulo 1	57

Bibliografía

Introducción

Dentro del ámbito de la estadística, en áreas como la bioestadística y el análisis de fiabilidad, el análisis de supervivencia se ha convertido en una herramienta fundamental para examinar el tiempo que transcurre hasta que ocurre un determinado evento. Dicho evento puede referirse a diversos sucesos, como el fallecimiento de un paciente, la avería de una pieza mecánica o la lesión de un deportista. Para llevar a cabo este tipo de análisis, basta con contar con dos variables principales: el período de observación de cada individuo y una variable dicotómica que señale si el evento en cuestión ocurrió o no.

El análisis de datos de supervivencia ha sido un campo de investigación muy activo durante varias de las últimas décadas. Una contribución importante que impulsó todo el campo fue la formulación mediante procesos de conteo presentada por Aalen (1975) en su tesis doctoral en Berkeley. Desde entonces, se han escrito numerosos y excelentes libros de texto sobre análisis de supervivencia y procesos de conteo [11].

Todo este ámbito se caracteriza por la presencia de datos incompletos y variados. Los métodos desarrollados se han convertido en una herramienta sólida y flexible para investigar fenómenos que se desarrollan a lo largo del tiempo. Una de las principales particularidades de los datos de supervivencia es la frecuente aparición de censura. Esta situación se presenta cuando no se dispone de información completa sobre el tiempo hasta que ocurre el evento de interés para todos los participantes del estudio. Esto puede suceder, por ejemplo, si el estudio concluye antes de que algunos individuos experimenten el evento (censura por la derecha), o si algunos participantes se pierden durante el seguimiento. Si se ignoraran estos casos censurados, las estimaciones obtenidas serían inexactas y podrían llevar a conclusiones equivocadas. Por ello, es esencial contar con modelos capaces de incorporar y manejar adecuadamente esta información incompleta para garantizar resultados fiables y precisos.

En este contexto, el modelo de riesgos proporcionales de Cox [3], propuesto por Sir David Cox en 1972, se destaca como una de las aportaciones más relevantes y utilizadas dentro del análisis de supervivencia. Su éxito se debe principalmente a su flexibilidad y a la capacidad que ofrece para explorar la relación entre el tiempo hasta el evento y un conjunto de variables explicativas, sin exigir que se especifique previamente la forma exacta de la función de riesgo base. Esta naturaleza semiparamétrica lo diferencia de otros enfoques y le confiere una enorme utilidad en una amplia variedad de campos.

La base teórica del modelo de Cox está profundamente sustentada en la teoría de procesos estocásticos y martingalas, pilares fundamentales para el estudio matemático de fenómenos aleatorios que se desarrollan a lo largo del tiempo. Los procesos estocásticos permiten describir la evolución incierta de un sistema, mientras que las martingalas, en particular, proporcionan un marco sólido para el análisis probabilístico y el tratamiento formal de la censura y el riesgo en el contexto del análisis de supervivencia. Estos conceptos son esenciales para entender cómo se comporta el proceso de riesgo a medida que avanza el tiempo y para garantizar que las inferencias obtenidas a partir del modelo sean válidas.

Hay que destacar que el modelo de Cox ha sido implementado en múltiples paquetes estadísticos como R, SAS y Python, lo que ha contribuido a su amplia adopción en estudios empíricos. Sin embargo, para un uso adecuado de estas herramientas, es indispensable una comprensión rigurosa de los supuestos y fundamentos del modelo, por lo que se ha decidido centrar este trabajo en el desarrollo teórico del modelo de regresión de Cox, con el objetivo de entender en profundidad los fundamentos matemáticos que garantizan su solidez y validez. Más allá de su uso práctico,

nos interesa valorar el alcance del modelo, identificar sus posibles limitaciones y apreciar su rigor estadístico desde una perspectiva académica. Un conocimiento sólido de su base teórica es esencial para interpretar correctamente los resultados y entender el verdadero potencial del modelo en el análisis de supervivencia.

El trabajo comienza con un capítulo en el que se presentan los preliminares matemáticos necesarios para la elaboración de la memoria, donde destaca la introducción del concepto de martingala. El segundo capítulo se centra en el estudio de los procesos de conteo, tratándolos a partir de la descomposición de Doob-Meyer y estudiando tanto la inferencia paramétrica como no paramétrica en este tipo de procesos. Finalmente, en el último capítulo, se presenta el modelo de regresión de Cox bajo la hipótesis de riesgos proporcionales.

CAPÍTULO 1

PRELIMINARES MATEMÁTICOS

Comenzaremos este trabajo describiendo la terminología y conceptos fundamentales de los procesos estocásticos. En esta descripción intervendrán los elementos habituales de la Teoría de la Probabilidad [1].

De este modo, en este capítulo se van a recoger todas las herramientas matemáticas necesarias para llevar a cabo el trabajo. Dentro de los procesos estocásticos, uno de los aspectos en los que nos centramos son las martingalas, pues permiten tratar de forma razonable los datos censurados. Para introducir el concepto de martingala, será necesario tener claro el concepto de esperanza condicionada.

En definitiva, se trata de un apartado fundamental para poder comprender bien todo lo que se explicará en futuros capítulos. Debido a la gran cantidad de conceptos e información necesaria, no se entrará en demasiado detalle.

1.1. Función de supervivencia y función de riesgo

En esta sección exploraremos distintas pero equivalentes maneras de caracterizar la distribución de una variable aleatoria positiva. Este tipo de variables suele utilizarse para modelar el tiempo hasta que ocurre un determinado suceso, como, por ejemplo, la vida útil de un componente o el tiempo de espera hasta que se observa un evento. Estas representaciones alternativas serán especialmente útiles más adelante, cuando abordemos el concepto de intensidades en los procesos de conteo.

Supongamos que T es una variable aleatoria positiva, que interpretaremos como el tiempo transcurrido hasta que ocurre un evento, por ejemplo, el tiempo hasta que una persona fallece. Asumiremos que T tiene función de distribución F y función de densidad f , siendo esta última continua para simplificar la exposición.

Una herramienta fundamental en este contexto es la función de supervivencia, $S(t)$, que nos da la probabilidad de que el evento no haya ocurrido hasta al menos el instante de tiempo t , es decir:

$$S(t) = P(T > t) = 1 - F(t).$$

En virtud de las propiedades de la función de distribución se tiene que $S(t)$ es decreciente, es decir, $S(t_1) \geq S(t_2)$ si $t_1 < t_2$ y además $S(0) = 1$ y $\lim_{t \rightarrow \infty} S(t) = 0$.

Tras esto, ahora definimos la función de riesgo, $h(t)$, que representa la tasa instantánea de probabilidad de no sobrevivir en un instante inmediatamente posterior a t cuando se ha sobrevivido hasta el instante t . De manera formal,

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P(T \in (t, t + \Delta t] | T > t) = \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} \cdot \frac{1}{S(t)} = \frac{f(t)}{S(t)}.$$

En varias ocasiones se utiliza la función de riesgo acumulado, que se define como:

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

Podemos establecer una relación entre la función de supervivencia y la función de riesgo, siempre que se conozca la función de distribución, mediante la siguiente cadena de igualdades:

$$h(t) = H'(t) = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)} = (-\log S(t))'.$$

Y se cumple que $H(0) = -\log S(0) = 0$. Por tanto, obtenemos que $H(t) = -\log S(t)$, o, equivalentemente,

$$S(t) = \exp(-H(t)) = \exp\left(-\int_0^t h(s) ds\right).$$

1.2. Procesos estocásticos

1.1 Definición. Un proceso estocástico X es una familia de variables aleatorias $\{X(t), t \in \mathbb{T}\}$ ordenadas según el índice t , que suele identificarse con el tiempo. Puede interpretarse como una sucesión de variables aleatorias cuyas características pueden variar a lo largo del tiempo, pues para cada instante t se tendrá una variable aleatoria distinta representada. En otras palabras, un proceso estocástico es un modelo matemático para describir un fenómeno aleatorio que evoluciona con el tiempo.

Desde un punto de vista más matemático, considerando el espacio probabilístico (Ω, \mathcal{F}, P) , se puede entender X como una aplicación $X : \Omega \rightarrow \mathbb{R}^{\mathbb{T}}$ que manda $\omega \rightarrow X(\omega, t) = X_t(\omega)$ y donde $\mathbb{R}^{\mathbb{T}} = \{(X(t))_{t \in \mathbb{T}} : X(t) \in \mathbb{R}\}$ es el conjunto de funciones que van de \mathbb{T} en \mathbb{R} y $\mathbb{T} \subset [0, \infty)$ es el conjunto de índices de tiempo. A cada función del tiempo que se obtiene se le da el nombre de trayectoria. Notemos que cada variable aleatoria del proceso estocástico es una aplicación $X(t) : \Omega \rightarrow \mathbb{R}$ que es $\mathcal{F}|_{\beta}$ medible (con β la σ -álgebra de Borel).

Resumiendo todo lo dicho anteriormente, se puede interpretar un proceso estocástico como una función de dos variables $X : \Omega \times \mathbb{T} \rightarrow \mathbb{R}$ dada por $(\omega, t) \rightarrow X(\omega, t)$, de modo que para cada ω fijo tenemos una trayectoria.

Notemos que en $\mathbb{R}^{\mathbb{T}}$ consideraremos la σ -álgebra producto (la natural) que viene dada por $\beta^{\mathbb{T}} = \sigma(\{\pi_{t_1, \dots, t_k}^{-1}(B) : t_1, \dots, t_k \in \mathbb{T}, B \in \beta(\mathbb{R}^k), k \geq 1\})$ donde $\pi_{t_1, \dots, t_k} : \mathbb{R}^{\mathbb{T}} \rightarrow \mathbb{R}^k$ representa la función de proyección, de modo que $\pi_{t_1, \dots, t_k}((X(t))_{t \in \mathbb{T}})$ es el vector de componentes $(X(t_1), \dots, X(t_k))$. La σ -álgebra producto $\beta^{\mathbb{T}}$ es la menor σ -álgebra que hace medible las proyecciones.

1.2 Teorema. Un proceso estocástico $X(t)$ es $\mathcal{F}|_{\beta^{\mathbb{T}}}$ medible, si y sólo si $X(t)$ es $\mathcal{F}|_{\beta}$ medible para todo $t \in \mathbb{T}$.

Demostración. Ver apéndice A. □

Como consecuencia de este teorema, se tiene la siguiente definición.

1.3 Definición. La ley inducida por X es $P_X(B) = P(X \in B)$, con $B \in \beta^{\mathbb{T}}$. En concreto, P_X es una probabilidad en $(\mathbb{R}^{\mathbb{T}}, \beta^{\mathbb{T}})$.

También es un resultado básico de teoría de procesos estocásticos el siguiente teorema.

1.4 Teorema. Dados dos procesos estocásticos $X = (X(t))_{t \in \mathbb{T}}$ e $Y = (Y(t))_{t \in \mathbb{T}}$, entonces $X \stackrel{d}{=} Y$, si y sólo si $(X(t_1), \dots, X(t_k)) \stackrel{d}{=} (Y(t_1), \dots, Y(t_k))$ para todo $t_1, \dots, t_k \in \mathbb{T}$ con $k \geq 1$.

Demostración. Ver apéndice A. □

Como consecuencia de lo anterior, se tiene la siguiente definición.

1.5 Definición. Se dice que dos procesos estocásticos son equivalentes si tienen las mismas distribuciones finito-dimensionales.

Estos dos teoremas básicos que se acaban de ver constituyen la base de la teoría que se desarrolla en torno a los procesos estocásticos. Sigamos ahora con más definiciones.

1.6 Definición. Un proceso càdlàg es un proceso estocástico con trayectorias continuas por la derecha y con límite por la izquierda para casi todo $t \in \mathbb{T} = [0, T]$ (o $\mathbb{T} = [0, \infty)$).

El hecho de que un proceso sea càdlàg da mucha información y son con los que vamos a trabajar a lo largo del documento. Que un proceso sea càdlàg permite, por ejemplo, hablar de $\sup_{t \in \mathbb{T}} X(t)$ como una variable aleatoria aunque la familia sea no numerable.

1.7 Proposición. Sea $X(t)$ un proceso càdlàg, entonces $\sup_{t \in \mathbb{T}} X(t)$ es una variable aleatoria.

Demostración. Es claro que $\sup_{t \in \mathbb{T}} X(t) \geq \sup_{t \in \mathbb{T} \cap \mathbb{Q}} X(t)$ (pues disminuye el conjunto de tiempos t posibles). Veamos que $\sup_{t \in \mathbb{T}} X(t) = \sup_{t \in \mathbb{T} \cap \mathbb{Q}} X(t)$ para concluir (pues $\mathbb{T} \cap \mathbb{Q}$ es numerable).

Imaginemos que $\sup_{t \in \mathbb{T}} X(t) = a < \infty$ (si $a = \infty$ se razona parecido). Entonces, existe $t_n \in \mathbb{T}$ tal que $X(t_n) > a - \frac{1}{n}$. Por la densidad de \mathbb{Q} en \mathbb{R} , existen $t_{n,k} \in \mathbb{T} \cap \mathbb{Q}$ que decrecen a t_n cuando $k \rightarrow \infty$. Puesto que $X(t)$ es càdlàg, se verifica $X(t_{n,k}) \rightarrow X(t_n)$ si $k \rightarrow \infty$ por lo que $X(t_{n,k}) > a - \frac{2}{n}$ para un k suficientemente alto. Luego esto nos dice que $\sup_{t \in \mathbb{T} \cap \mathbb{Q}} X(t) \geq a - \frac{2}{n}$ para todo n y por tanto se concluye que $\sup_{t \in \mathbb{T}} X(t) = \sup_{t \in \mathbb{T} \cap \mathbb{Q}} X(t)$. □

1.3. Esperanza condicionada

Un tratamiento riguroso y completo del concepto de esperanza condicionada requiere el uso de algunos resultados no elementales de Teoría de la Medida. Incluimos esta sección con el objetivo de entender el sentido de condicionar por una σ -álgebra. Admitimos que en todo momento trabajamos con medidas signadas, es decir, con valores positivos o negativos. En primer lugar, introducimos diversos conceptos para posteriormente llegar al resultado clave conocido como Teorema de Radon-Nikodym, siguiendo [6].

1.8 Definición. Si λ y μ son dos medidas en el espacio medible (Ω, \mathcal{F}) , siendo μ una medida positiva, se dice que λ es absolutamente continua respecto de μ (y se denota $\lambda \ll \mu$) si $\mu(A) = 0$ implica que $\lambda(A) = 0$ para todo $A \subset \Omega$.

1.9 Definición. Se dice que una medida μ es σ -finita si existen conjuntos $A_n \subset \Omega$ medibles tales que $\mu(A_n) < \infty$ y $\bigcup_{n=1}^{\infty} A_n = \Omega$.

Notemos que, por ejemplo, la medida de Lebesgue en \mathbb{R} es σ -finita, considerando $A_n = [k, k+1]$ con $k \in \mathbb{Z}$. Previo al teorema de Radon-Nikodym, tenemos la siguiente proposición.

1.10 Proposición. Sea (Ω, \mathcal{F}, P) un espacio probabilístico. Sea $X : \Omega \rightarrow \mathbb{R}$ una variable aleatoria integrable. Consideremos $\mu(A) := \int_A X dP$ con $A \in \mathcal{F}$. Entonces, μ es una medida y $\mu \ll P$.

Como comentario, para demostrar esta proposición bastaría con ver que μ cumple con la propiedad de la aditividad numerable para uniones disjuntas, y esto viene garantizado por la aditividad de la integral de Lebesgue.

1.11 Teorema. (Radon-Nikodym). Si λ y μ son dos medidas tales que $\lambda \ll \mu$, siendo μ una medida positiva σ -finita, entonces existe una función f medible tal que

$$\lambda(A) = \int_A f d\mu \quad \forall A \in \mathcal{F}.$$

La función f es única, en el sentido de que si otra función, g , tuviese la misma propiedad, entonces $f = g$ en casi todo punto. A esta función, que se suele denotar $\frac{d\lambda}{d\mu}$, se la llama derivada de Radon-Nikodym de λ respecto de μ .

Para observar la demostración de este teorema, se puede recurrir a [6].

Considerando (Ω, \mathcal{F}, P) , sea $\mathcal{G} \subset \mathcal{F}$ una sub- σ -álgebra. Dada $\mu(A) := \int_A X dP$ con $A \in \mathcal{G}$ medida con P probabilidad en \mathcal{F} (para las condiciones de la Proposición 1.10), entonces P es probabilidad en \mathcal{G} . Cabe destacar que aunque X sea $\mathcal{F}|\beta$ medible, no tiene por qué ser $\mathcal{G}|\beta$ medible. Además, se tiene que μ es una medida en \mathcal{G} , verificando $\mu \ll P$.

En virtud de todo lo anterior, podemos definir la esperanza condicionada.

1.12 Definición. La esperanza condicionada de la variable aleatoria X respecto de la σ -álgebra \mathcal{G} se define como:

$$E(X|\mathcal{G}) = \frac{d\mu}{dP}.$$

Podemos destacar las siguientes dos propiedades de la esperanza condicionada:

- $Y = E(X|\mathcal{G})$ es la única variable aleatoria $\mathcal{G}|\beta$ medible y $\int_A X dP = \int_A Y dP \quad \forall A \in \mathcal{G}$. Esta es la propiedad definitoria de la esperanza condicionada.
- $E(E(X|\mathcal{G})) = E(X)$.

1.4. Martingalas

En esta sección introduciremos los conceptos necesarios hasta llegar a la definición de martingala, y también comentaremos algunas de sus propiedades básicas.

Un proceso estocástico X induce una familia creciente de sub- σ -álgebras dada por $\mathcal{F}_t^X = \sigma(\{X(s) : 0 \leq s \leq t\})$ denominada historia interna de X . A la hora de condicionar respecto a una información, generalmente se condiciona a más información que la de un único proceso estocástico. Como consecuencia de esto, conviene destacar el concepto de historia o filtración $(\mathcal{F}_t; t \geq 0)$.

1.13 Definición. Sea (Ω, \mathcal{F}, P) un espacio probabilístico. Una historia o filtración de X , $\mathbb{F} = (\mathcal{F}_t; t \geq 0)$, es una familia de sub- σ -álgebras de \mathcal{F} tales que $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$ si $s \leq t$ (la familia de sub- σ -álgebras es creciente). Además, se asume que la filtración cumple también las llamadas “condiciones usuales”:

1. $\mathcal{F}_s = \bigcap_{t>s} \mathcal{F}_t \quad \forall s$.
2. $A \subset B \in \mathcal{F}, P(B) = 0 \Rightarrow A \in \mathcal{F}_0$.

La condición 1 se puede interpretar como que la familia es continua por la derecha y a la condición 2 nos referimos diciendo que la familia es completa.

El nombre de historia se debe a que la σ -álgebra \mathcal{F}_t puede interpretarse como la que engloba todos los eventos (información) que han ocurrido o no hasta el instante de tiempo t .

1.14 Definición. Una pre- t σ -álgebra \mathcal{F}_{t-} es la σ -álgebra más pequeña tal que $\mathcal{F}_s \subset \mathcal{F}_{t-}$ si $s < t$; esto es, recoge todos los eventos anteriores al instante de tiempo t .

1.15 Definición. Se dice que un proceso estocástico X está adaptado a una filtración \mathbb{F} si $\forall t \geq 0$, $X(t)$ es \mathcal{F}_t -medible y $\mathcal{F}_t^X \subset \mathcal{F}_t$.

Un ejemplo claro de filtración es aquella inducida por un proceso estocástico X . En este caso, consideramos la colección de σ -álgebras definida como $\mathcal{F}_t^X = \sigma(\{X(s) : s \leq t\})$, la cual hemos denominado como historia interna de X . Todo proceso es naturalmente adaptado a su historia interna. A menudo se trabaja con una filtración más general \mathbb{F} , donde cada sub- σ -álgebra de \mathcal{F} está dada por la expresión $\mathcal{F}_t = \mathcal{F}_0 \vee \mathcal{F}_t^X := \sigma(\mathcal{F}_0 \cup \mathcal{F}_t^X)$.

Un resultado importante (ver [7]) establece que dado un proceso estocástico, su filtración natural (historia interna) puede completarse y hacerse continua por la derecha (se consigue que se cumplan las “condiciones usuales”). A la nueva filtración se la conoce como filtración aumentada. Este resultado, junto con la construcción del párrafo anterior, garantiza que $\mathbb{F} = (\mathcal{F}_t; t \geq 0)$ forme una familia creciente de σ -álgebras que además es continua por la derecha.

Un proceso de salto continuo por la derecha es aquel en el que, para cualquier instante t y ω , la función $X(s, \omega)$ permanece constante para todo $s \in [t, t + \epsilon)$, con algún $\epsilon > 0$. Gracias al resultado citado, podemos concluir que si los eventos de interés ocurren a lo largo de un tiempo de forma continua, entonces la filtración correspondiente también evolucionará de forma continua hacia adelante (continuas por la derecha).

Con todo esto, ya podemos introducir el concepto de martingala, proceso de gran importancia en el que se va a sustentar el trabajo.

1.16 Definición. Una martingala con respecto a una filtración \mathcal{F}_t es un proceso estocástico M , continuo por la derecha y con límite por la izquierda (càdlàg) que cumple:

1. M es adaptado a \mathcal{F}_t .
2. $E|M(t)| < \infty$.
3. $E(M(t)|\mathcal{F}_s) = M(s) \forall s \leq t$.

Se dice que $M(t)$ es una submartingala si en vez de verificarse la condición 3 se tiene que $E(M(t)|\mathcal{F}_s) \geq M(s) \forall s \leq t$.

Por otro lado, se dice que una martingala es de cuadrado integrable si $\sup_{t \in \mathbb{T}} E(M(t)^2) < \infty$ donde $\mathbb{T} = [0, T]$.

Cabe destacar que, normalmente, en la definición de martingala no se pide que las trayectorias sean càdlàg. Sin embargo, esto no es un problema, pues existe un resultado que garantiza que dada una martingala M , siempre se puede encontrar una versión de M que tenga trayectorias càdlàg con probabilidad uno [8].

Intuitivamente, la condición 3 de la definición de martingala (Definición 1.16) se interpreta como que la media de $M(t)$ dada información hasta el instante de tiempo s es $M(s)$. Visto de otra forma:

$$E(dM(t)|\mathcal{F}_{t-}) = 0 \forall t > 0.$$

Esto es, una martingala tiene incrementos de media 0 dados por los eventos pasados, y sin necesidad de condicionamiento.

Es importante tener en cuenta que una martingala M puede ser considerada un proceso de error en el siguiente sentido [11]:

- Dado que $E(M(t)) = E(M(0))$, una martingala tiene media constante como función del tiempo y si $M(0) = 0$, entonces $E(M(t)) = 0 \forall t$ (martingala de media 0).
- Las martingalas tienen incrementos no correlacionados, es decir, para un martingala M se cumple:

$$\text{Cov}(M(t) - M(s), M(v) - M(u)) = 0 \quad \forall 0 \leq s \leq t \leq u \leq v.$$

Demostración. La demostración de esta segunda propiedad es sencilla. Teniendo en cuenta que $s \leq t \leq u \leq v$ entonces sabemos que $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}_u \subset \mathcal{F}_v$. Además, podemos suponer sin pérdida de generalidad que $E(M(t)) = 0$ (si no fuera así, bastaría con restar el valor $E(M(t))$ a cada término). Con esto, usando que como $s \leq t$, entonces $(M(t) - M(s))$ es \mathcal{F}_t -medible y las propiedades de la esperanza condicionada:

$$\begin{aligned} \text{Cov}(M(t) - M(s), M(v) - M(u)) &= E((M(t) - M(s))(M(v) - M(u))) = \\ &= E(E((M(t) - M(s))(M(v) - M(u)) | \mathcal{F}_t)) = E((M(t) - M(s))E((M(v) - M(u)) | \mathcal{F}_t)) = 0. \end{aligned}$$

Donde la última igualdad se justifica de acuerdo con la definición de martingala, pues se tiene que $E((M(v) - M(u)) | \mathcal{F}_t) = M(t) - M(t) = 0$ al ser $v \geq u \geq t$. \square

Una clase importante de procesos que complementan a las martingalas son los llamados procesos predecibles, los cuales vamos a definir a continuación [2].

Si tomamos $\mathbb{T} = [0, T]$ o $\mathbb{T} = [0, \infty)$, entonces la σ -álgebra correspondiente será $\beta([0, T])$ o $\beta([0, \infty))$, respectivamente. Definimos una sub- σ -álgebra de $\beta([0, T]) \otimes \mathcal{F}$ generada por los conjuntos producto de la forma $(s, t] \times U$, donde $U \in \mathcal{F}_s$, $s \leq t$ y $t \in \mathbb{T}$. Esta σ -álgebra se llama la σ -álgebra predecible y la denotamos por \mathcal{G} .

Entonces, dado un proceso estocástico $X : \mathbb{T} \times \Omega \rightarrow \mathbb{R}$ definido por $(t, \omega) \mapsto X(t, \omega)$, diremos que X es medible si es $\beta([0, T]) \otimes \mathcal{F}$ -medible.

1.17 Definición. Un proceso estocástico X se dice predecible si es $\mathcal{G} | \beta$ medible.

De forma más intuitiva, los procesos predecibles son aquellos que están definidos a partir de procesos adaptados, acotados y continuos por la izquierda. En otras palabras, un proceso es predecible si su valor es conocido en un tiempo justo anterior a t .

1.4.1. Procesos de variación predecible y opcional

En este epígrafe vamos a introducir dos conceptos asociados a las martingalas, los cuales son de especial importancia para el desarrollo de nuestro trabajo. La referencia para las demostraciones de este apartado es [8].

1.18 Proposición. Si $M(t)$ es una martingala, entonces $M(t)^2$ es una submartingala.

Demostración. Por la desigualdad de Jensen (la cual indica que $\phi(E(X)) \leq E(\phi(X))$ siendo ϕ una función convexa), es inmediato que:

$$E(M(t)^2 | \mathcal{F}_s) \geq E(M(t) | \mathcal{F}_s)^2 = M(s)^2,$$

dado que la función x^2 es convexa. \square

En base a la submartingala $M(t)^2$, suponiendo que $M(t)$ es una martingala de cuadrado integrable, podemos definir un nuevo proceso.

1.19 Definición. Se define el proceso de variación predecible de la martingala $M(t)$ como el siguiente límite en probabilidad:

$$\langle M \rangle(t) = Pr - \lim_{|\delta| \rightarrow 0, \delta = \{0=t_0 \leq t_1 \leq \dots \leq t_n=t\}} \sum_{i=1}^n E \left[(M(t_i) - M(t_{i-1}))^2 \mid \mathcal{F}_{t_{i-1}-} \right].$$

Se puede probar que $\langle M \rangle(t)$ es el único proceso predecible con $\langle M \rangle(0) = 0$ tal que $M(t)^2 - \langle M \rangle(t)$ es martingala. La demostración general tiene ciertas complicaciones y se puede encontrar en [8], y más adelante se probará un caso particular de este resultado.

De forma análoga, se habla de proceso de covariación predecible.

1.20 Definición. Se define el proceso de covariación predecible de las martingalas $M_1(t)$ y $M_2(t)$ como el siguiente límite en probabilidad:

$$\langle M_1, M_2 \rangle(t) = Pr - \lim_{|\delta| \rightarrow 0} \sum_{i=1}^n E \left[(M_1(t_i) - M_1(t_{i-1})) (M_2(t_i) - M_2(t_{i-1})) \mid \mathcal{F}_{t_{i-1}-} \right].$$

En particular, se cumple que $\langle M, M \rangle = \langle M \rangle$.

De nuevo, para este caso se tiene que $\langle M_1, M_2 \rangle(t)$ es el único proceso predecible con $\langle M_1, M_2 \rangle(0) = 0$ tal que $M_1(t)M_2(t) - \langle M_1, M_2 \rangle(t)$ es una martingala al ser diferencia de martingalas, pues se puede demostrar que:

$$M_1(t)M_2(t) - \langle M_1, M_2 \rangle(t) = \frac{1}{4} \{ (M_1 + M_2)^2 - \langle M_1 + M_2 \rangle(t) - [(M_1 - M_2)^2 - \langle M_1 - M_2 \rangle(t)] \}$$

haciendo uso de las identidades:

$$(M_1 + M_2)^2 - \langle M_1 + M_2 \rangle = M_1^2 + 2M_1M_2 + M_2^2 - \langle M_1 \rangle - 2\langle M_1, M_2 \rangle - \langle M_2 \rangle$$

$$(M_1 - M_2)^2 - \langle M_1 - M_2 \rangle = M_1^2 - 2M_1M_2 + M_2^2 - \langle M_1 \rangle + 2\langle M_1, M_2 \rangle - \langle M_2 \rangle,$$

y según lo comentado para la variación predecible.

Un concepto relacionado con todo lo anterior es el de variación y covariación opcional.

1.21 Definición. Se define el proceso de covariación opcional de las martingalas $M_1(t)$ y $M_2(t)$ como el siguiente límite en probabilidad:

$$[M_1, M_2](t) = \lim_{|\delta| \rightarrow 0} \sum_{i=1}^n (M_1(t_i) - M_1(t_{i-1})) (M_2(t_i) - M_2(t_{i-1})).$$

En particular, $[M](t) = [M, M](t)$ se llama el proceso de variación opcional. Se puede demostrar que el proceso $M^2(t) - [M](t)$ también es una martingala (más adelante se demostrará un caso particular), aunque $[M]$ no es, en general, un proceso predecible [8]. En la práctica, $[M]$ suele ser más fácil de calcular que $\langle M \rangle$.

1.4.2. Teorema central del límite para las martingalas

Los resultados sobre la normalidad asintótica son especialmente útiles para desarrollar métodos de inferencia aproximada. En el caso de variables aleatorias, esta teoría se basa en el Teorema Central del Límite junto con técnicas como el Δ -método [1]. Dado que en nuestro estudio trabajaremos con procesos estocásticos, es útil, aunque sea de manera general, entender la teoría asintótica en este contexto.

Dentro de esta teoría, las martingalas Gaussianas desempeñan un papel análogo al de la distribución normal.

1.22 Definición. Un proceso $\{U(t)\}_{t \geq 0}$ es una martingala Gaussiana si cumple que es una martingala, que sus trayectorias son continuas, que $U(0) = 0$, y que los vectores $(U(t_1), \dots, U(t_k))$ (con $k \geq 1$) tienen distribución normal multivariante.

Dado que una martingala tiene esperanza constante y $E(U(0)) = 0$, entonces $E(U(t)) = 0$ para todo t , lo que significa que una martingala Gaussiana es centrada. Además, los incrementos en una martingala son no correlacionados, lo que en el caso Gaussiano implica que son independientes. De este modo, si $U(t)$ es una martingala Gaussiana y $V(t) = \text{Var}(U(t))$, siendo V una función creciente, entonces se cumple que

$$\text{Cov}(U(s), U(t)) = V(\min(s, t)).$$

El equivalente al Teorema Central del Límite en el contexto de procesos es el Teorema Central del Límite para Martingalas.

1.23 Teorema. Si $\{M_n\}_{n \geq 0}$ es una sucesión de martingalas que cumple:

- (a) $\langle M_n \rangle(t) \rightarrow V(t)$ para todo $t \in [0, T]$ cuando $n \rightarrow \infty$.
- (b) el tamaño de los saltos (discontinuidades) de M_n tiende a cero cuando $n \rightarrow \infty$.

Entonces se tiene que:

$$M_n \rightarrow U$$

en distribución, donde U es una martingala Gaussiana con función de varianza V .

La consecuencia principal de este resultado es que para muchos funcionales de interés, denotados por H , se cumple que $H(M_n)$ converge en distribución a $H(U)$. Esto permite aproximar probabilidades del tipo $P(H(M_n) \leq x)$ por $P(H(U) \leq x)$, así como aproximar los percentiles de $H(M_n)$ por los de $H(U)$.

Un caso común es cuando $H(U)$ representa el supremo del proceso: $H(U) = \sup_{0 \leq t \leq T} |U(t)|$. Así, si queremos calcular $P(\sup_{0 \leq t \leq T} |M_n(t)| \leq x)$, esta se puede estimar mediante $P(\sup_{0 \leq t \leq T} |U(t)| \leq x)$, lo cual también se aplica para percentiles.

CAPÍTULO 2

PROCESOS DE CONTEO

En este capítulo se introducen los procesos de conteo como herramienta fundamental para modelar la aparición de eventos a lo largo del tiempo, especialmente en el contexto del análisis de supervivencia. Se estudian conceptos clave como la intensidad, la descomposición de Doob-Meyer y distintas formas de variación y covariación, que permiten analizar y descomponer estos procesos en componentes más manejables.

Además, se estudian casos particulares como el proceso de Poisson y los modelos de supervivencia, tanto con datos completos como con censura. Finalmente, se abordan técnicas de inferencia estadística, destacando los estimadores de Nelson-Aalen y Kaplan-Meier por un lado (esenciales para el análisis empírico de datos de supervivencia) y por otro la obtención de una verosimilitud general para procesos de conteo.

Este capítulo sienta así las bases prácticas y teóricas para la construcción del modelo de regresión de Cox en el capítulo siguiente, tomando como referencia [5].

2.1. Introducción a los procesos de conteo

2.1 Definición. Un proceso de conteo es un modelo matemático para la aparición de uno o varios tipos de sucesos de forma recurrente a lo largo del tiempo. Formalmente, dada una filtración $\mathbb{F} = (\mathcal{F}_t; t \geq 0)$ sobre un espacio probabilístico (Ω, \mathcal{F}, P) , un proceso de conteo multivariante $N(t) = (N_1(t), \dots, N_k(t))$ es un vector de procesos (càdlàg) tales que $N_j(0) = 0$, para $j = 1, \dots, k$, cuyas trayectorias son funciones constantes a trozos, no decrecientes, continuas por la derecha, y cuyas discontinuidades son saltos de magnitud $+1$, sin que dos componentes distintas tengan saltos simultáneamente.

De esta forma, como no se producen saltos simultáneos, la suma de sus componentes es también un proceso de conteo (proceso agregado)

$$N_{\bullet} = \sum_{j=1}^k N_j.$$

Es conveniente pensar en N_j como el proceso que registra la ocurrencia recurrente de sucesos de tipo j , y en N_{\bullet} como el proceso que refleja la aparición en el tiempo de cualquiera de los k tipos considerados.

Es fácil darse cuenta de que $N(t)$ es el número de sucesos que se producen en el intervalo $(0, t]$, y que dada una realización $\omega \in \Omega$, entonces $N_t(\omega)$ está representado como una función en t escalonada y continua por la derecha.

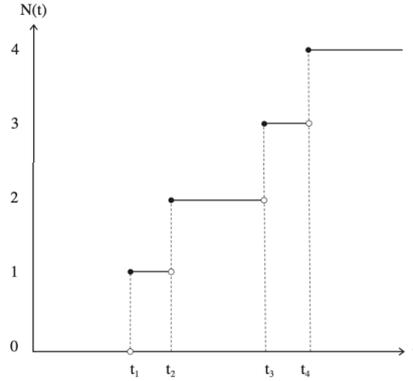


Figura 2.1: Ejemplo de un proceso de conteo $N(t)$.

2.1.1. Intensidades

En esta sección vamos a entender la construcción de un nuevo proceso, llamado intensidad asociada al proceso de conteo N . Para ello, se usarán resultados que se encuentran en [4].

Sea $N(t) = \sum_{n=1}^{\infty} I_{\{T_n \leq t\}}$ un proceso de conteo tal que $0 = T_0 < T_1 < \dots < T_n \rightarrow \infty$ son variables aleatorias positivas. Vamos a realizar la hipótesis simplificadora de que (T_1, \dots, T_n) es un vector aleatorio que tiene densidad conjunta, $f_{T_1, \dots, T_n}(t_1, \dots, t_n)$. Esta hipótesis excluye algunos procesos, pero estos no serán de nuestro interés. Vamos a definir la intensidad para procesos de conteo regulares.

Dado que existe $f_{T_1, \dots, T_n}(t_1, \dots, t_n)$, se puede hablar de la función de densidad condicionada a que $T_1 = t_1, \dots, T_{n-1} = t_{n-1}$, y lo denotamos como $f_{T_n | T_1=t_1, \dots, T_{n-1}=t_{n-1}}(t)$. Del mismo modo, se puede hablar de todas las funciones (supervivencia, riesgo...) condicionadas. En concreto, nos interesan la función de supervivencia condicionada,

$$\begin{aligned} S_{T_n | T_1=t_1, \dots, T_{n-1}=t_{n-1}}(t) &= P(T_n > t | T_1 = t_1, \dots, T_{n-1} = t_{n-1}) = \\ &= 1 - P(T_n \leq t | T_1 = t_1, \dots, T_{n-1} = t_{n-1}) = 1 - \int_{t_{n-1}}^t f_{T_n | T_1=t_1, \dots, T_{n-1}=t_{n-1}}(s) ds. \end{aligned}$$

y la función de riesgo condicionada, dada por:

$$h_n(t | T_1, \dots, T_{n-1}) = h_{T_n | T_1=t_1, \dots, T_{n-1}=t_{n-1}}(t) = \frac{f_{T_n | T_1=t_1, \dots, T_{n-1}=t_{n-1}}(t)}{S_{T_n | T_1=t_1, \dots, T_{n-1}=t_{n-1}}(t)}.$$

De manera habitual, se suele utilizar la función de riesgo acumulado que viene dada por

$$H_n(t | T_1, \dots, T_{n-1}) = H_{T_n | T_1=t_1, \dots, T_{n-1}=t_{n-1}}(t) = \int_0^t h_n(s | T_1, \dots, T_{n-1}) ds.$$

Con todo esto, dada una secuencia de tiempos $0 < t_1 < \dots < t_n < \dots$, se puede definir el siguiente proceso estocástico.

2.2 Definición. La función de intensidad condicionada de un proceso de conteo regular en \mathbb{R}_+ es la función representada por $\lambda(\cdot)$ y definida por

$$\lambda(t) = \begin{cases} h_1(t) & \text{si } 0 < t \leq T_1. \\ h_n(t \mid T_1, \dots, T_{n-1}) & \text{si } T_{n-1} < t \leq T_n \text{ y } n \geq 2. \end{cases}$$

Es claro que $\lambda(t)$ es un proceso bien definido, constante a trozos y con trayectorias continuas por la izquierda [4]. Se tiene que $\lambda(t)$ es \mathcal{F}_t -medible pues consideramos que $\mathcal{F}_t = \sigma(N(s) : s \leq t)$ y entonces nos encontramos entre el suceso $N(t)$ y el $N(t) + 1$, es decir,

$$\lambda(t) = h_n(t \mid T_1, \dots, T_{N(t)}) = h_{N(t)+1}(t \mid T_1, \dots, T_{N(t)}),$$

luego depende de $N(t)$ y sucesos anteriores, y podemos concluir que es adaptado.

Esta definición de intensidad de un proceso de conteo N hace referencia a añadir un punto adicional dentro de una extensión de la región de observación (lo que nos da la idea de que se trata de un proceso predecible). De este modo, el contenido intuitivo de la noción de una función de intensidad condicional está bien expresado a través de la sugerente relación:

$$\begin{aligned} \lambda(t) &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} E(N(t + \Delta t) - N(t) \mid \mathcal{F}_{t-}) & (2.1) \\ &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P(N(t + \Delta t) - N(t) \geq 1 \mid \mathcal{F}_{t-}) \\ &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P(N(t + \Delta t) - N(t) = 1 \mid \mathcal{F}_{t-}), \end{aligned}$$

donde \mathcal{F}_{t-} es la σ -álgebra de los eventos que ocurren en tiempos anteriores a t sin incluirlo de acuerdo con la Definición 1.14. Así, la intensidad condicional puede interpretarse como el riesgo condicional de la ocurrencia de un evento en t , dada la realización del proceso sobre el intervalo $[0, t)$. Esto se justificará de forma intuitiva más adelante cuando se introduzca la descomposición de Doob-Meyer.

Estrictamente, la notación debería reflejar el hecho de que $\lambda(\cdot)$ es una función $\lambda(\cdot \mid T_1, \dots, T_{N(t)})$ de la historia de los sucesos, o, aún más generalmente, que es en sí misma un proceso estocástico $\lambda(t, \omega)$ que depende de ω a través de la realización $\{T_1(\omega), \dots, T_{N(\omega)}\}$ de la historia hasta el tiempo t .

2.3 Definición. Se define la llamada intensidad acumulada como:

$$\Lambda_j(t) = \int_0^t \lambda_j(s) ds.$$

La intensidad del proceso de conteo agregado es, lógicamente, $\lambda_\bullet = \sum_{j=1}^k \lambda_j$ y la intensidad acumulada $\Lambda_\bullet = \sum_{j=1}^k \Lambda_j$.

2.1.2. Compensadores y descomposición de Doob-Meyer

2.4 Definición. Supongamos que $N(t)$ es un proceso de conteo adaptado a la filtración \mathcal{F} y que admite intensidad $\lambda(t)$ continua por la derecha. Definimos el compensador $\Lambda(t)$ como la integral

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

Dicha integral verifica que $M(t) = N(t) - \Lambda(t)$ es una martingala. Este es un resultado cuya prueba no es para nada sencilla, y se puede encontrar en [7].

En teoría de procesos estocásticos se conoce la ecuación $N(t) = M(t) + \Lambda(t)$ como descomposición de Doob-Meyer. A la diferencia entre el proceso que admite tal descomposición y la martingala se le conoce como compensador, tal y como hemos definido anteriormente. En un proceso de conteo, la intensidad acumulada es, precisamente, el compensador.

La clase de procesos que admite la descomposición de Doob-Meyer en suma de una martingala más un proceso predecible tiene gran importancia en la teoría de procesos estocásticos. Esa clase coincide, esencialmente, con las semimartingalas, que son (olvidando detalles técnicos) los procesos que se pueden escribir como diferencia de submartingalas.

De este modo, una vez sabemos que $M(t) = N(t) - \Lambda(t)$ es martingala, entonces, a partir de esto y que $N(t) = M(t) + \Lambda(t)$:

$$\frac{1}{\Delta t} E(N(t + \Delta t) - N(t) | \mathcal{F}_t) = \frac{1}{\Delta t} E(M(t + \Delta t) - M(t) | \mathcal{F}_t) + \frac{1}{\Delta t} E(\Lambda(t + \Delta t) - \Lambda(t) | \mathcal{F}_t).$$

Puesto que $M(t)$ es martingala, el primer término del segundo miembro será nulo, y dado que $\Lambda(t)$ es una función continua al proceder de una integral:

$$\frac{1}{\Delta t} E(N(t + \Delta t) - N(t) | \mathcal{F}_t) = \frac{1}{\Delta t} E\left(\int_t^{t+\Delta t} \lambda(s) ds | \mathcal{F}_t\right),$$

y asumiendo que $\lambda(s)$ es una función continua y de saltos (constante a trozos), se deduce:

$$\frac{1}{\Delta t} E(N(t + \Delta t) - N(t) | \mathcal{F}_t) \approx \frac{\Delta t}{\Delta t} \lambda(t) = \lambda(t).$$

Por tanto, vemos que se justifica las ecuaciones (2.1).

2.1.3. Variación y covariación en procesos de conteo

Cabe destacar que los procesos de variación y covariación de una martingala, tanto predecible como opcional, son herramientas fundamentales en el cálculo estocástico. Esta área se enfoca en el estudio de expresiones integrales y diferenciales en las que aparecen martingalas.

Un caso típico de interés es la integral $\int_0^t Y(s) dM(s)$, donde $Y(t)$ es un proceso predecible y $M(t)$ una martingala. En este tema, basta con considerar integrales respecto a martingalas provenientes de la descomposición de un proceso de conteo N con intensidad λ , es decir $M(t) = N(t) - \Lambda(t)$, con $\Lambda(t) = \int_0^t \lambda(s) ds$. En este caso se tiene:

$$\int_0^t Y(s) dM(s) = \int_0^t Y(s) dN(s) - \int_0^t Y(s) d\Lambda(s).$$

La integral respecto de $dN(s)$ es una integral de Stieltjes, y se interpreta como una suma sobre los puntos de salto del proceso N :

$$\int_0^t Y(s) dN(s) = \sum_{j: T_j \leq t} Y(T_j),$$

donde T_1, T_2, \dots son los tiempos de salto. Como $\frac{d\Lambda(s)}{ds} = \lambda(s)$ (asumimos que las trayectorias de $\Lambda(s)$ son derivables), se deduce que la otra integral de Stieltjes es:

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

En consecuencia, si M tiene la forma anterior, se obtiene:

$$\int_0^t Y(s) dM(s) = \sum_{j:T_j \leq t} Y(T_j) - \int_0^t Y(s)\lambda(s) ds.$$

Una propiedad importante viene dada en la siguiente proposición.

2.5 Proposición. Toda integral de un proceso predecible con respecto a una martingala produce también una martingala.

Demostración. Proporcionamos una demostración válida únicamente para procesos simples. Por un argumento de densidad, el resultado se extiende a todos los procesos predecibles. De este modo, consideremos un proceso simple dado por $Y(s) = \sum_{i=1}^k I_{(s_i, s_{i+1}]} Y_i$ donde suponemos que Y_i es \mathcal{F}_{s_i} -medible. Por tanto, se verifica que $Y(s)$ es predecible (pues conozco el valor del proceso al inicio del intervalo).

Queremos ver que $V(t) = \int_0^t Y(s) dM(s)$ es martingala. Para ello, teniendo en cuenta que

$$V(t) = \int_0^t Y(s) dM(s) = \sum_{i=1}^k \int_{s_i \wedge t}^{s_{i+1} \wedge t} Y_i dM(s) = \sum_{i=1}^k Y_i (M(s_{i+1} \wedge t) - M(s_i \wedge t)),$$

Buscamos ver que $E(V(t)|\mathcal{F}_s) = V(s)$:

$$E(V(t)|\mathcal{F}_s) = \sum_{i=1}^k E[Y_i (M(s_{i+1} \wedge t) - M(s_i \wedge t)) | \mathcal{F}_s],$$

Distintguimos tres casos (notemos que $s_{i+1} \wedge t = s_{i+1}$ y $s_i \wedge t = s_i$ siempre):

- Si $s \geq s_{i+1} \geq s_i$, entonces como Y_i es \mathcal{F}_{s_i} -medible también es \mathcal{F}_s -medible. Por otro lado, es claro que $M(s_i)$ y $M(s_{i+1})$ son \mathcal{F}_s -medibles, luego:

$$\begin{aligned} E[Y_i (M(s_{i+1} \wedge t) - M(s_i \wedge t)) | \mathcal{F}_s] &= Y_i E[(M(s_{i+1} \wedge t) - M(s_i \wedge t)) | \mathcal{F}_s] = \\ &= Y_i [(M(s_{i+1} \wedge t) - M(s_i \wedge t))] = Y_i [(M(s_{i+1}) - M(s_i))] = Y_i [(M(s_{i+1} \wedge s) - M(s_i \wedge s))]. \end{aligned}$$

- Si $s_{i+1} \geq s \geq s_i$, entonces como Y_i es \mathcal{F}_{s_i} -medible también es \mathcal{F}_s -medible. Luego, usando la propiedad de martingala:

$$\begin{aligned} E[Y_i (M(s_{i+1} \wedge t) - M(s_i \wedge t)) | \mathcal{F}_s] &= Y_i E[(M(s_{i+1} \wedge t) - M(s_i \wedge t)) | \mathcal{F}_s] = \\ &= Y_i (M(s_{i+1} \wedge s) - M(s_i)) = Y_i (M(s_{i+1} \wedge s) - M(s_i \wedge s)). \end{aligned}$$

- Si $s_{i+1} \geq s_i \geq s$. Usando que se cumple $E[f \cdot g | \mathcal{F}_s] = E[f \cdot E[g | \mathcal{G}] | \mathcal{F}_s]$ cuando f \mathcal{G} -medible y $\mathcal{F}_s \subset \mathcal{G}$:

$$E[Y_i (M(s_{i+1} \wedge t) - M(s_i \wedge t)) | \mathcal{F}_s] = E[Y_i E[(M(s_{i+1} \wedge t) - M(s_i \wedge t)) | \mathcal{F}_{s_i}] | \mathcal{F}_s].$$

Notemos que $E[(M(s_{i+1} \wedge t) - M(s_i \wedge t)) | \mathcal{F}_{s_i}]$ es nulo por la propiedad de los incrementos de media 0 dados por los eventos pasados, por tanto:

$$E[Y_i (M(s_{i+1} \wedge t) - M(s_i \wedge t)) | \mathcal{F}_s] = E[Y_i \cdot 0 | \mathcal{F}_s] = 0.$$

Por otro lado, $Y_i (M(s \wedge s_{i+1}) - M(s_i \wedge s)) = Y_i (M(s) - M(s_i)) = 0$. Luego se deduce:

$$E[Y_i (M(s_{i+1} \wedge t) - M(s_i \wedge t)) | \mathcal{F}_s] = Y_i (M(s \wedge s_{i+1}) - M(s_i \wedge s)).$$

Por tanto, en los 3 casos tratados se llega a que:

$$E(V(t)|\mathcal{F}_s) = \sum_{i=1}^k E[Y_i(M(s_{i+1} \wedge t) - M(s_i \wedge t))|\mathcal{F}_s] = \sum_{i=1}^k Y_i[(M(s_{i+1} \wedge s) - M(s_i \wedge s))] = V(s).$$

como queríamos ver. □

Por otro lado, las covariaciones son simétricas y bilineales (es decir, $\langle X, Y \rangle = \langle Y, X \rangle$ y $\langle a_1 X_1 + a_2 X_2, Y \rangle = a_1 \langle X_1, Y \rangle + a_2 \langle X_2, Y \rangle$, y lo mismo para la covariación opcional). Además, si un proceso X es derivable, de forma que $X(t) = \int_0^t x(s)ds$, y M es una martingala, entonces X y M son ortogonales ($\langle X, M \rangle = 0$). Intuitivamente, esto es consecuencia de que $\Delta X(t) \simeq x(t)\Delta t$ se comporta como un proceso predecible y, por tanto,

$$\begin{aligned} \langle X, M \rangle(t) &= \sum_{i=1}^n E((X(t_i) - X(t_{i-1}))(M(t_i) - M(t_{i-1}))|\mathcal{F}_{t_{i-1}}) \\ &\simeq \sum_{i=1}^n E((t_i - t_{i-1})x(t_{i-1})(M(t_i) - M(t_{i-1}))|\mathcal{F}_{t_{i-1}}) \\ &= \sum_{i=1}^n (t_i - t_{i-1})x(t_{i-1})E((M(t_i) - M(t_{i-1}))|\mathcal{F}_{t_{i-1}}) = 0. \end{aligned}$$

Además, se puede comprobar que si X e Y son dos procesos derivables (con derivada continua) entonces también $\langle X, Y \rangle = 0$. Todo esto se encuentra demostrado de forma rigurosa en [8].

Teniendo en cuenta esto, y centrándonos ahora en el caso de los procesos de conteo, concretamente en las martingalas $M_j(t) = N_j(t) - \Lambda_j(t)$, vamos a calcular la variación predecible de M_j . Usando las reglas de cálculo anteriores $\langle M_j \rangle = \langle M_j, M_j \rangle = \langle N_j - \Lambda_j, M_j \rangle = \langle N_j, M_j \rangle - \langle \Lambda_j, M_j \rangle = \langle N_j, M_j \rangle$, porque M_j es martingala y Λ_j derivable. Observamos también que $\langle N_j, \Lambda_j \rangle = \langle \Lambda_j, \Lambda_j \rangle + \langle M_j, \Lambda_j \rangle = 0$, de nuevo porque Λ_j es derivable. Como consecuencia $\langle N_j \rangle = \langle N_j, N_j \rangle = \langle N_j, M_j \rangle + \langle N_j, \Lambda_j \rangle = \langle N_j, M_j \rangle$ y, por lo tanto, $\langle M_j \rangle = \langle N_j \rangle$.

Para calcular $\langle N_j \rangle$ empleamos la notación diferencial:

$$d\langle N_j \rangle(t) = E((dN_j(t))^2 | \mathcal{F}_{t-}) = E(dN_j(t) | \mathcal{F}_{t-}) = d\Lambda_j(t).$$

Aquí hemos tenido en cuenta que N_j es un proceso de conteo, por lo que $dN_j(t)$ toma únicamente los valores 0 ó 1 y $(dN_j(t))^2 = dN_j(t)$. Por lo tanto:

$$\langle N_j \rangle(t) = \langle M_j \rangle(t) = \Lambda_j(t).$$

Con un razonamiento similar para la covariación entre martingalas M_i, M_j con $i \neq j$ se obtiene

$$\langle M_i, M_j \rangle(t) = 0 \quad \text{si } i \neq j,$$

También se pueden emplear cálculos parecidos para probar que:

$$[M_j](t) = N_j(t), \quad [M_i, M_j](t) = 0 \quad \text{si } i \neq j.$$

2.1.4. Casos particulares

En esta sección vamos a hablar de ciertos ejemplos y modelos de conteo.

2.1.4.1. El proceso de Poisson

Un proceso de Poisson con función de tasa $\xi(t) \geq 0$ se define como un proceso de conteo con incrementos independientes, tal que la diferencia $N(t) - N(s)$ sigue una distribución de Poisson con parámetro $\int_s^t \xi(r) dr$. La función acumulada de tasas se define como $\Xi(t) = \int_0^t \xi(r) dr$, y representa el número esperado de eventos ocurridos en el intervalo de tiempo $[0, t]$.

Este tipo de modelo es comúnmente utilizado en el análisis de fiabilidad, donde $N(t)$ puede interpretarse como el número de fallos (y correspondientes reparaciones) hasta el tiempo t . En este contexto, se puede expresar como $N(t) = \sum_{i=1}^{\infty} I_{\{T_i \leq t\}}$, donde T_1, T_2, \dots representan los tiempos de ocurrencia de los eventos (por ejemplo, fallos).

Una característica clave de este proceso es la independencia de sus incrementos, lo que simplifica notablemente el cálculo de su intensidad. En efecto, la intensidad del proceso se obtiene como:

$$\lambda(t) = \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} E[N(t + \Delta t) - N(t) | \mathcal{F}_{t-}] = \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} E[N(t + \Delta t) - N(t)] = \xi(t).$$

Esto muestra que en un proceso de Poisson, la intensidad no es un proceso estocástico propiamente dicho, sino una función determinista del tiempo. La posibilidad de elegir libremente la forma de $\xi(t)$ permite adaptar el modelo de Poisson a una amplia variedad de contextos y fenómenos.

En lo que resta de este apartado vamos a hablar de un proceso de Poisson particular, el homogéneo. Un proceso de conteo $\{N(t)\}_{t \geq 0}$ se denomina proceso de Poisson homogéneo con tasa $\lambda > 0$ si cumple con dos propiedades: tiene incrementos que son independientes y estacionarios, y además $N(t)$ sigue una distribución de Poisson con parámetro λt .

Calcular la intensidad en este caso es muy sencillo. Debido a la independencia de los incrementos, la variable $N(t + \Delta t) - N(t)$ es independiente de \mathcal{F}_{t-} . Por lo tanto,

$$E[N(t + \Delta t) - N(t) | \mathcal{F}_{t-}] = E[N(t + \Delta t) - N(t)] = \lambda(t + \Delta t - t) = \lambda \Delta t,$$

ya que, por la propiedad de incrementos estacionarios, $N(t + \Delta t) - N(t)$ tiene la misma distribución que $N(\Delta t)$, es decir, una distribución de Poisson con parámetro $\lambda \Delta t$.

A partir de esto, se concluye que la intensidad instantánea es:

$$\lambda(t) = \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} E[N(t + \Delta t) - N(t) | \mathcal{F}_{t-}] = \lambda.$$

Finalmente, la intensidad acumulada, es decir, el compensador, está dada simplemente por:

$$\Lambda(t) = \lambda t.$$

Con todo esto, se tiene que de acuerdo con la descomposición de Doob-Meyer, $N(t) = M(t) + \Lambda(t)$ y por tanto, $M(t) = N(t) - \lambda t$. Verifiquemos que $M(t)$ es una martingala. Gracias a la propiedad de que los incrementos del proceso de Poisson son independientes, si tomamos dos instantes $s \leq t$, la diferencia

$$M(t) - M(s) = N(t) - N(s) - \lambda(t - s),$$

es independiente de la filtración \mathcal{F}_s . Por lo tanto, el valor esperado condicional es:

$$E[M(t) - M(s) | \mathcal{F}_s] = E[M(t) - M(s)] = E[N(t) - N(s) - \lambda(t - s)] = 0.$$

Esto implica:

$$E[M(t) | \mathcal{F}_s] = M(s),$$

lo cual demuestra que $M(t)$ es una martingala.

Ahora, calculemos la variación predecible asociada al proceso, utilizando la propiedad de que los incrementos del proceso de Poisson son independientes. Para dos tiempos consecutivos t_{i-1} y t_i , se cumple que:

$$\begin{aligned} E[(M(t_i) - M(t_{i-1}))^2 | \mathcal{F}_{t_{i-1}-}] &= E[(N(t_i) - N(t_{i-1}) - \lambda(t_i - t_{i-1}))^2] = \\ &= \text{Var}[(N(t_i) - N(t_{i-1}))] = \lambda(t_i - t_{i-1}). \end{aligned}$$

Donde hemos utilizado que $E(N(t)) = \lambda t$, $\text{Var}(N(t)) = \lambda t$ y que $\text{Var}(X) = E[(X - E(X))^2]$ (y las propiedades del proceso de Poisson). Así, el límite que define la variación predecible se convierte en:

$$\langle M \rangle(t) = \lim_{|\delta| \rightarrow 0} \sum_{i=1}^n E[(M(t_i) - M(t_{i-1}))^2 | \mathcal{F}_{t_{i-1}-}] = \lambda t = \Lambda(t).$$

Además, se puede demostrar que $M^2(t) - \Lambda(t)$ es una martingala. Para ello, basta probar que $E[M^2(t) - \lambda t | \mathcal{F}_s] = M^2(s) - \lambda s$, para todo $s \leq t$, o lo que es lo mismo (usando la descomposición de Doob-Meyer), $E[(N(t) - \lambda t)^2 - \lambda t | \mathcal{F}_s] = (N(s) - \lambda s)^2 - \lambda s$, para todo $s \leq t$. Se verifica que:

$$\begin{aligned} E[(N(t) - \lambda t)^2 - \lambda t | \mathcal{F}_s] &= E[(N(s) - \lambda s + N(t) - N(s) - \lambda(t - s))^2 | \mathcal{F}_s] - \lambda t \\ &= E[(N(s) - \lambda s)^2 + 2(N(s) - \lambda s)(N(t) - N(s) - \lambda(t - s)) \\ &\quad + (N(t) - N(s) - \lambda(t - s))^2 | \mathcal{F}_s] - \lambda t \\ &= (N(s) - \lambda s)^2 + 2E[(N(s) - \lambda s)(N(t) - N(s) - \lambda(t - s)) | \mathcal{F}_s] \\ &\quad + E[(N(t) - N(s) - \lambda(t - s))^2] - \lambda t \\ &= (N(s) - \lambda s)^2 + 0 + \lambda(t - s) - \lambda t \\ &= (N(s) - \lambda s)^2 - \lambda s. \end{aligned}$$

donde usamos la linealidad de la esperanza matemática, que los incrementos son independientes y que $E((N(t) - N(s) - \lambda(t - s))) = 0$

NOTA: Usamos que como $N(s) - \lambda s$ es \mathcal{F}_s -medible y que $N(t) - N(s) - \lambda(t - s)$ es independiente de \mathcal{F}_s con esperanza nula, entonces $E[(N(s) - \lambda s)(N(t) - N(s) - \lambda(t - s)) | \mathcal{F}_s] = (N(s) - \lambda s)E((N(t) - N(s) - \lambda(t - s))) = 0$.

Concluimos que, como $E[M^2(t) - \lambda t | \mathcal{F}_s] = M^2(s) - \lambda s$, se cumple la propiedad de martingala.

Pasamos ahora a calcular la variación opcional. Para ello, observamos que $(M(t_i) - M(t_{i-1}))^2 = (N(t_i) - N(t_{i-1}))^2 - 2\lambda(t_i - t_{i-1})(N(t_i) - N(t_{i-1})) + \lambda^2(t_i - t_{i-1})^2$. En un intervalo $(t_{i-1}, t_i]$ suficientemente pequeño (como para considerar $t_i - t_{i-1} \approx 0$), el incremento $N(t_i) - N(t_{i-1})$ será 0 salvo en los puntos de salto de N , donde valdrá 1. Entonces,

$$\lim_{|\delta| \rightarrow 0} \sum_{i=1}^n (M(t_i) - M(t_{i-1}))^2 = \lim_{|\delta| \rightarrow 0} \sum_{i=1}^n (N(t_i) - N(t_{i-1}))^2 = \sum_{s \leq t} (\Delta N(s))^2 = \sum_{s \leq t} \Delta N(s) = N(t).$$

De este modo, concluimos que: $[M](t) = N(t)$.

De nuevo, también se puede ver que $M^2(t) - [M](t) = M^2(t) - N(t)$ es una martingala. Queremos demostrar que $E[M^2(t) - N(t) | \mathcal{F}_s] = M^2(s) - N(s)$, para todo $s \leq t$. Teniendo en

cuenta que $E[M(t)^2 | \mathcal{F}_s] = M(s)^2 + E[(M(t) - M(s))^2 | \mathcal{F}_s]$ (en virtud de la Proposición 2.6 enunciada y demostrada posteriormente) y que $Var(M(t) - M(s)) = Var(N(t) - N(s) - \lambda(t-s)) = \lambda(t-s)$ se verifica:

$$\begin{aligned} E[M(t)^2 - N(t) | \mathcal{F}_s] &= E[M(t)^2 | \mathcal{F}_s] - E[N(t) | \mathcal{F}_s] \\ &= M(s)^2 + E[(M(t) - M(s))^2 | \mathcal{F}_s] - E[N(t) | \mathcal{F}_s] \\ &= M(s)^2 + Var[(M(t) - M(s)) | \mathcal{F}_s] - E[N(t) + N(s) - N(s) | \mathcal{F}_s] \\ &= M(s)^2 + \lambda(t-s) - (N(s) + E[N(t) - N(s)]) \\ &= M(s)^2 + \lambda(t-s) - (N(s) + \lambda(t-s)) \\ &= M(s)^2 - N(s). \end{aligned}$$

Se concluye, por tanto, que $M^2(t) - [M](t) = M^2(t) - N(t)$ es una martingala.

2.6 Proposición. Sean X , A y B variables aleatorias. Cuando $X = A + B$, A es \mathcal{F}_s -medible, y B es independiente de \mathcal{F}_s con $E(B | \mathcal{F}_s) = 0$, entonces: $E(X^2 | \mathcal{F}_s) = A^2 + E(B^2 | \mathcal{F}_s)$.

Demostración. Desarrollando $E(X^2 | \mathcal{F}_s)$:

$$E(X^2 | \mathcal{F}_s) = E((A + B)^2 | \mathcal{F}_s) = E(A^2 | \mathcal{F}_s) + E(B^2 | \mathcal{F}_s) + 2E(AB | \mathcal{F}_s).$$

Puesto que A es \mathcal{F}_s -medible, $E(A^2 | \mathcal{F}_s) = A^2$ y $E(AB | \mathcal{F}_s) = AE(B | \mathcal{F}_s)$. Finalmente, usando que $E[B | \mathcal{F}_s] = 0$:

$$E(X^2 | \mathcal{F}_s) = A^2 + E(B^2 | \mathcal{F}_s) + 2AE(B | \mathcal{F}_s) = A^2 + E(B^2 | \mathcal{F}_s).$$

□

2.1.4.2. Proceso de salto único

Sea T una variable aleatoria positiva con densidad f , función de supervivencia S y función de riesgo h . Definimos el proceso

$$N(t) = I_{\{T \leq t\}},$$

que toma el valor 0 hasta el instante aleatorio T , momento en el cual salta a 1 y permanece constante desde entonces.

En este contexto, la información contenida en \mathcal{F}_{t-} (es decir, la historia hasta justo antes de t) equivale a conocer si el evento $\{T \geq t\}$ ha ocurrido o no. Por lo tanto, se tiene que

$$\lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P(N(t + \Delta t) - N(t) = 1 | \mathcal{F}_{t-}) = h(t)Y(t),$$

donde $Y(t) = I_{\{T \geq t\}}$. Así, el proceso de intensidad del salto está dado por

$$\lambda(t) = Y(t)h(t),$$

y su correspondiente intensidad acumulada es

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

2.1.4.3. Modelo de supervivencia

Supongamos que X_1, \dots, X_n son variables aleatorias independientes e idénticamente distribuidas (i.i.d.), todas positivas, y con una función de riesgo asociada $h(t)$. Podemos interpretar a X_i como el tiempo de fallecimiento del individuo i .

Nos interesa analizar el vector de procesos de conteo $(N_1(t), \dots, N_n(t))$, donde cada proceso se define como

$$N_i(t) = I_{\{X_i \leq t\}}.$$

Es decir, $N_i(t)$ indica si el individuo i ha fallecido antes o en el instante t . Gracias a la independencia entre las variables aleatorias, los procesos N_i no presentan saltos simultáneos (no hay muertes en el mismo instante entre diferentes individuos).

Definimos ahora el proceso de conteo total como:

$$N(t) = \sum_{i=1}^n N_i(t),$$

el cual representa el número total de muertes ocurridas hasta el tiempo t y se conoce como el proceso de supervivencia.

A continuación, vamos a determinar tanto la intensidad como la intensidad acumulada de este proceso. Según lo visto anteriormente, la intensidad del proceso $N_i(t)$ está dada por

$$\lambda_i(t) = Y_i(t)h(t),$$

donde $Y_i(t) = I_{\{X_i \geq t\}}$ representa el hecho de que el individuo i aún está vivo justo antes del instante t . Por lo tanto, la intensidad total del proceso $N(t)$ se expresa como:

$$\lambda(t) = \sum_{i=1}^n \lambda_i(t) = Y(t)h(t),$$

donde

$$Y(t) = \sum_{i=1}^n Y_i(t) = \sum_{i=1}^n I_{\{X_i \geq t\}},$$

es decir, $Y(t)$ cuenta el número de individuos que siguen con vida justo antes del tiempo t , también llamados individuos “en riesgo”.

Finalmente, la intensidad acumulada del proceso se calcula con la integral:

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

2.1.4.4. Modelo de supervivencia con censuras

Supongamos ahora que X_1, \dots, X_n son variables aleatorias independientes e idénticamente distribuidas (v.a.i.i.d.), todas positivas, con una función de riesgo asociada $h(t)$. Como en el contexto del análisis de supervivencia, interpretamos X_i como el tiempo hasta el fallecimiento del individuo i .

Sin embargo, en un escenario más realista, puede que no siempre sea posible observar directamente dicho tiempo, debido a diversas razones como la finalización anticipada del estudio o la pérdida de seguimiento del paciente. Por ello, introducimos un instante aleatorio de censura U_i : observamos X_i solo si $X_i \leq U_i$, y si no, solo sabemos que $X_i \geq U_i$. En consecuencia, los datos observables son:

$$\tilde{X}_i = \min(X_i, U_i), \quad D_i = I_{\{X_i = \tilde{X}_i\}},$$

donde D_i es un indicador de censura que señala si el dato observado es completo ($D_i = 1$) o censurado ($D_i = 0$).

El proceso de salto único asociado al individuo i , denotado por $N_i(t)$, solo puede ser observado mientras $t \leq U_i$, es decir, si $C_i(t) = I_{\{U_i \geq t\}} = 1$. Definimos el proceso de censura como $C(t) = (C_1(t), \dots, C_n(t))$, el cual es predecible. El proceso efectivamente observable para cada individuo se define como:

$$N_i^C(t) = \int_0^t C_i(s) dN_i(s).$$

Este proceso coincide con $N_i(t)$ para $t \leq U_i$, y permanece constante en $N_i(U_i)$ para $t > U_i$. En consecuencia, $N_i^C(t)$ también es un proceso de conteo, cuya intensidad viene dada por:

$$\lambda_i^C(t) dt = E[dN_i^C(t) | \mathcal{F}_{t-}] = E[C_i(t)dN_i(t) | \mathcal{F}_{t-}] = C_i(t)E[dN_i(t) | \mathcal{F}_{t-}] = C_i(t)\lambda_i(t) dt.$$

Por lo tanto, la intensidad del proceso censurado es:

$$\lambda_i^C(t) = C_i(t)\lambda_i(t) = C_i(t)Y_i(t)h(t) = Y_i^C(t)h(t),$$

donde

$$Y_i^C(t) = I_{\{U_i \geq t, X_i \geq t\}} = I_{\{\tilde{X}_i \geq t\}},$$

es el indicador de que el individuo i está en riesgo en el instante t .

La descomposición de $N_i^C(t)$ en términos de una martingala más su compensador es:

$$N_i^C(t) = \int_0^t C_i(s) dN_i(s) = \int_0^t C_i(s) d\Lambda_i(s) + \int_0^t C_i(s) dM_i(s) = \Lambda_i^C(t) + M_i^C(t).$$

El proceso agregado $N_\bullet^C(t)$, que representa el proceso de supervivencia con censura, tiene una intensidad total dada por

$$\lambda^C(t) = \sum_{i=1}^n \lambda_i^C(t) = Y^C(t)h(t),$$

donde

$$Y^C(t) = \sum_{i=1}^n Y_i^C(t)$$

representa el número total de individuos en riesgo en el instante t , es decir, aquellos que aún no han fallecido ni han sido censurados antes de dicho instante.

2.2. Inferencia en procesos de conteo

La inferencia estadística es el conjunto de métodos y técnicas que permiten inducir, a partir de la información empírica proporcionada por una muestra, cuál es el comportamiento de una determinada población con un riesgo de error medible en términos de probabilidad. En este apartado nos vamos a centrar en los métodos de estimación de parámetros y no en los métodos de contraste de hipótesis.

2.2.1. Inferencia no paramétrica

Procedemos a hacer una descripción de las herramientas no paramétricas más importantes para el análisis de datos de procesos de conteo. Las técnicas no paramétricas se han consolidado como herramientas fundamentales en el análisis de supervivencia debido a su simplicidad y al hecho de que sus propiedades están bien estudiadas y comprendidas.

2.2.1.1. El estimador de Nelson-Aalen

El estimador de Nelson-Aalen es un método no paramétrico utilizado para estimar la función de riesgo acumulado dentro del marco del modelo de intensidades multiplicativas. En términos más específicos, se considera que $N = (N_1, \dots, N_k)$ es un proceso de conteo que presenta intensidades de tipo multiplicativo de la forma $\lambda_j(t) = h_j(t)Y_j(t)$, donde $h_j(t)$ es una función determinista e $Y_j(t)$ es un proceso predecible (es decir, observable). La función $h_j(t)$ puede interpretarse como la fuerza de transición, y comúnmente $Y_j(t)$ representa el número de individuos en riesgo en el instante t .

Estimar directamente la función $h_j(t)$ resulta más complejo que estimar su función de riesgo acumulado, definida como

$$H_j(t) = \int_0^t h_j(s) ds,$$

de forma similar a cómo estimar una función de densidad es más difícil que estimar su función de distribución. Por esta razón, se enfoca la atención en la estimación de $H_j(t)$.

La idea detrás del estimador de Nelson-Aalen se basa en la descomposición de los procesos de conteo como la suma de un proceso predecible y una martingala. A partir de la expresión diferencial:

$$dN_j(t) = h_j(t)Y_j(t) dt + dM_j(t),$$

se obtiene:

$$\frac{dN_j(t)}{Y_j(t)} = h_j(t) dt + \frac{dM_j(t)}{Y_j(t)}.$$

Dado que los incrementos de la martingala representan ruido aleatorio (esto es, $\frac{dM_j(t)}{Y_j(t)}$ es ruido), el término $\frac{dN_j(t)}{Y_j(t)}$ sugiere una forma natural de estimar $h_j(t) dt$, lo cual motiva la siguiente definición del estimador.

2.7 Definición. Se define el estimador de Nelson-Aalen como:

$$\hat{H}_j(t) := \int_0^t \frac{dN_j(s)}{Y_j(s)}.$$

Observemos que, en realidad, la función $\hat{H}_j(t)$ se calcula como una suma discreta:

$$\hat{H}_j(t) = \sum_{i: T_i^{(j)} \leq t} \frac{1}{Y_j(T_i^{(j)})},$$

donde $T_i^{(j)}$ son los tiempos en los que el proceso N_j experimenta saltos.

¿Es $\hat{H}_j(t)$ un buen estimador de $H_j(t)$? Para responder a esta pregunta, analizaremos su sesgo y varianza. Para estudiar el sesgo, definimos

$$J_j(t) := I_{\{Y_j(t) > 0\}}, \text{ de modo que } H_j^*(t) = \int_0^t h_j(s)J_j(s) ds = \int_0^t \frac{J_j(s)}{Y_j(s)} \lambda_j(s) ds = \int_0^t \frac{J_j(s)}{Y_j(s)} d\Lambda_j(s).$$

Es claro que se cumple que $\hat{H}_j(t) \leq H_j^*(t)$ (pues para $\hat{H}_j(t)$ solo se consideran los tiempos de salto). Además, intuitivamente se observa que $\hat{H}_j(t) \approx H_j^*(t)$ siempre que $P(Y_j(t) > 0) \approx 1$. Por otro lado, se puede expresar el estimador como:

$$\hat{H}_j(t) = \int_0^t \frac{J_j(s)}{Y_j(s)} dN_j(s),$$

por lo que la diferencia $\hat{H}_j(t) - H_j^*(t) = \int_0^t \frac{J_j(s)}{Y_j(s)} dM_j(s)$ es una martingala con valor inicial cero. Como consecuencia, su esperanza es constante e igual a cero:

$$E(\hat{H}_j(t)) = E(H_j^*(t)) = \int_0^t h_j(s) P(Y_j(s) > 0) ds.$$

Esto indica que $\hat{H}_j(t)$ es un estimador sesgado de $H_j(t)$, ya que tiende a subestimarlos. Sin embargo, el sesgo,

$$\text{sesgo}(\hat{H}_j(t)) = \int_0^t h_j(s) P(Y_j(s) = 0) ds,$$

es usualmente pequeño. Por ejemplo, en modelos de supervivencia (con o sin censura), al inicio del estudio es muy probable que haya individuos en riesgo, es decir, $P(Y_j(t) = 0) \approx 0$, por lo que el estimador será prácticamente insesgado.

Veamos ahora la varianza de $\hat{H}_j(t)$. Como $\hat{H}_j(t) - H_j^*(t)$ es una martingala, también lo es su cuadrado menos su variación predecible, como se mencionó en una sección anterior,

$$(\hat{H}_j(t) - H_j^*(t))^2 - \langle \hat{H}_j - H_j^* \rangle(t),$$

y tiene valor inicial cero. Esto implica que

$$\text{Var}(\hat{H}_j(t) - H_j^*(t)) = E[(\hat{H}_j(t) - H_j^*(t))^2] - E[\langle \hat{H}_j - H_j^* \rangle(t)] = E[(\hat{H}_j(t) - H_j^*(t))^2],$$

$$\text{Var}(\hat{H}_j(t) - H_j^*(t)) = E[\langle \hat{H}_j - H_j^* \rangle(t)].$$

donde hemos usado $E[(\hat{H}_j(t) - H_j^*(t))^2 - \langle \hat{H}_j - H_j^* \rangle(t)] = 0$ (al ser martingala con valor inicial cero) y que $E(\hat{H}_j(t)) = E(H_j^*(t))$.

En el caso (frecuente) en que $H_j^*(t) \approx H_j(t)$, se puede aproximar la varianza de $\hat{H}_j(t)$ mediante el estudio de la variación predecible de la martingala $\hat{H}_j(t) - H_j(t)$, $\text{Var}(\hat{H}_j(t)) = \text{Var}(\hat{H}_j(t) - H_j(t)) \approx \text{Var}(\hat{H}_j(t) - H_j^*(t))$. Las propiedades de martingalas estudiadas previamente nos indican que:

$$\langle \hat{H}_j - H_j \rangle(t) = \int_0^t \frac{J_j(s)}{Y_j(s)^2} d\langle M_j \rangle(s) = \int_0^t \frac{J_j(s)}{Y_j(s)^2} \lambda_j(s) ds = \int_0^t \frac{J_j(s)}{Y_j(s)} h_j(s) ds = \int_0^t \frac{J_j(s)}{Y_j(s)} dH_j(s).$$

En resumen:

$$\text{Var}(\hat{H}_j(t)) \approx \text{Var}(\hat{H}_j(t) - H_j^*(t)) = E[\langle \hat{H}_j - H_j^* \rangle(t)] = E\left(\int_0^t \frac{J_j(s)}{Y_j(s)} dH_j(s)\right).$$

Esta expresión podría sugerir el uso de $\langle \hat{H}_j - H_j^* \rangle(t)$ como un estimador para la varianza de $\hat{H}_j(t)$. Sin embargo, $\langle \hat{H}_j - H_j^* \rangle(t) = \int_0^t \frac{J_j(s)}{Y_j(s)} dH_j(s)$, lo cual depende de H_j , que no es conocido. En su lugar, $(\hat{H}_j(t) - H_j^*(t))^2 - [\hat{H}_j - H_j^*](t)$ también forma una martingala y, nuevamente,

$$\text{Var}(\hat{H}_j(t)) \approx \text{Var}(\hat{H}_j(t) - H_j^*(t)) = E[(\hat{H}_j(t) - H_j^*(t))^2] = E[(\hat{H}_j - H_j^*](t)),$$

pero ahora

$$[\hat{H}_j - H_j^*](t) = \int_0^t \frac{J_j(s)}{Y_j(s)^2} d[M_j](s) = \int_0^t \frac{J_j(s)}{Y_j(s)^2} dN_j(s).$$

Esto lleva a proponer el siguiente estimador:

$$\hat{\sigma}_j^2(t) := [\hat{H}_j - H_j^*](t) = \int_0^t \frac{J_j(s)}{Y_j(s)^2} dN_j(s),$$

como una estimación de $Var(\hat{H}_j(t))$. Nótese que $\hat{\sigma}_j^2(t)$ es un estimador insesgado de $Var(\hat{H}_j(t) - H_j^*(t))$ (considerando $N_j = M_j + \Lambda_j$ y la expresión de $Var(\hat{H}_j(t) - H_j^*(t))$),

$$E(\hat{\sigma}_j^2(t)) = E\left(\int_0^t \frac{J_j(s)}{Y_j(s)^2} dN_j(s)\right) = E\left(\int_0^t \frac{J_j(s)}{Y_j(s)^2} d\Lambda_j(s)\right) = E\left(\int_0^t \frac{J_j(s)}{Y_j(s)} dH_j(s)\right)$$

y además que, de acuerdo con la Definición 2.7:

$$\hat{\sigma}_j^2(t) = \int_0^t \frac{J_j(s)}{Y_j(s)} d\hat{H}_j(s).$$

Aunque la fórmula de la varianza obtenida a partir de la variación predecible no resulta útil de forma directa para estimar la varianza de $\hat{H}_j(t)$, sí resulta más adecuada para derivar una expresión de la varianza en el régimen asintótico.

Supongamos que $Y_j(t)/n \rightarrow y_j(t)$ cuando $n \rightarrow \infty$. Entonces, a partir de la expresión para $\langle \hat{H}_j - H_j^* \rangle(t)$, se deduce que:

$$Var\left(\sqrt{n}(\hat{H}_j(t) - H_j^*(t))\right) \rightarrow E\left[\int_0^t \frac{I(Y_j(s) > 0)}{y_j(s)} h_j(s) ds\right] = \int_0^t \frac{P(Y_j(s) > 0)}{y_j(s)} h_j(s) ds.$$

La condición $Y_j(t)/n \rightarrow y_j(t)$ es válida en muchos modelos. Por ejemplo, en modelos de supervivencia, se cumple que $\frac{Y(t)}{n} = \frac{1}{n} \sum_{i=1}^n I(X_i \geq t) \rightarrow P(X > t) = S(t)$, y algo similar ocurre cuando hay censura. Definamos entonces $V_j(t) = \int_0^t \frac{P(Y_j(s) > 0)}{y_j(s)} h_j(s) ds$. Bajo condiciones generales como la convergencia uniforme $\frac{Y_j(s)}{n} \xrightarrow{P} y_j(s)$ en $[0, t]$ con $\inf_{0 \leq s \leq t} y_j(s) \geq 0$, es posible aplicar el Teorema Central del Límite para martingalas (Teorema 1.23 del documento) y concluir que:

$$\sqrt{n}(\hat{H}_j - H_j) \xrightarrow{d} U_j,$$

donde U_j es una martingala gaussiana con varianza $V_j(t)$. Además, el estimador $\hat{V}_j(t) = n\hat{\sigma}_j^2(t)$ es consistente para $V_j(t)$. Estas condiciones se cumplen en la mayoría de los modelos de nuestro interés.

Una consecuencia importante de esta aproximación es que, al fijar el instante t , se tiene la siguiente convergencia marginal:

$$\frac{\hat{H}(t) - H^*(t)}{\hat{\sigma}(t)} \sim_{\text{aprox}} N(0, 1).$$

Como en general $H^*(t) \approx H(t)$, esto permite construir un intervalo de confianza aproximado de nivel β para $H(t)$:

$$\left[\hat{H}(t) \pm z_{\beta/2} \hat{\sigma}(t) \right].$$

Aquí, $z_{\beta/2}$ representa el percentil superior $\beta/2$ de la distribución normal estándar, es decir, $\Phi(z_{\beta/2}) = \beta/2$, donde Φ es la función de distribución de la normal estándar.

Estos son intervalos de confianza puntuales. Normalmente, interesa estudiar el comportamiento de $H(t)$ a lo largo de un intervalo completo $[0, T]$, más que en un único valor de t , pero esto no lo vamos a tratar.

2.2.1.2. El estimador de Kaplan-Meier

El modelo de supervivencia, tanto con como sin censura, es uno de los más empleados dentro de la clase de modelos con intensidades multiplicativas. En este contexto, aunque puede interesar estimar $H(t)$, la atención suele centrarse en la estimación de la función de supervivencia $S(t)$. Ambas funciones están relacionadas mediante la expresión $S(t) = \exp(-H(t))$, lo cual permite obtener un estimador de S directamente a partir de uno de H , y viceversa. Por ejemplo, si $\hat{H}(t)$ es el estimador de Nelson-Aalen para la función de riesgo acumulado, entonces

$$\hat{S}(t) = \exp(-\hat{H}(t)),$$

es un estimador para la función de supervivencia. Este estimador es conocido como el estimador de Breslow o también como el estimador de Fleming-Harrington. En esta sección, nos enfocamos en otro estimador de $S(t)$: el estimador de Kaplan-Meier, considerando sus propiedades y su vínculo con los anteriores. Este es el método no paramétrico más comúnmente usado para estimar la función de supervivencia a partir de datos censurados.

Para introducirlo, partimos de la relación entre $S(t)$ y $H(t)$, y aproximamos la integral en el exponente dividiéndola en pequeños intervalos (argumento heurístico). Así, obtenemos:

$$\begin{aligned} S(t) = \exp(-H(t)) &= \exp\left(-\int_0^t h(x) dx\right) = \prod_{s \leq t} \exp\left(-\int_s^{s+\Delta s} h(x) dx\right) \approx \\ &\approx \prod_{s \leq t} \exp(-\Delta H(s)) \approx \prod_{s \leq t} (1 - \Delta H(s)), \end{aligned}$$

usando que $e^{-x} \approx 1 - x$ cuando x es pequeño. Esta aproximación motiva la definición del estimador de Kaplan-Meier.

2.8 Definición. Se define el estimador de Kaplan-Meier como:

$$\hat{S}(t) := \prod_{s \leq t} (1 - \Delta \hat{H}(s)).$$

Cabe destacar que en esta expresión solo aparecen un número finito de factores distintos de 1, que corresponden a los tiempos en los que hay saltos en $\hat{H}(t)$, es decir, cuando ocurren eventos (muertes). De hecho, el estimador de Kaplan-Meier puede escribirse de forma aún más simple. Como se cumple que $d\hat{H}(t) = \frac{dN(t)}{Y(t)}$, entonces se obtiene:

$$\hat{S}(t) := \prod_{s \leq t} \left(1 - \frac{\Delta N(s)}{Y(s)}\right) = \prod_{j: T_j \leq t} \left(\frac{n_j - d_j}{n_j}\right),$$

donde los valores T_j representan los momentos en los que se registran fallecimientos (observaciones no censuradas), n_j indica cuántos individuos están en riesgo justo en T_j y d_j cuenta cuántas muertes ocurren en ese mismo instante. Generalmente $d_j = 1$, pero si se trata de datos agrupados (como cuando se mide en días o semanas), pueden ocurrir varios eventos simultáneamente. En tales casos, la misma fórmula del estimador sigue siendo válida.

El estimador $\hat{S}(t)$ es una función escalonada: presenta saltos en los momentos en que ocurren fallecimientos y se mantiene constante a partir del último de ellos.

El mismo enfoque utilizado anteriormente para estudiar sesgo y varianza del estimador de Nelson-Aalen puede aplicarse aquí. Definimos

$$S^*(t) = \exp(-H^*(t)), \quad \text{con} \quad H^*(t) = \int_0^t h(s)J(s) ds, \quad J(s) = I_{\{Y(s)>0\}}.$$

Entonces, puede demostrarse que:

$$\frac{\hat{S}(t)}{S^*(t)} - 1 = - \int_0^t \frac{\hat{S}(s^-) J(s)}{S^*(s) Y(s)} dM(s).$$

Para verlo, estudiamos el cociente $R(t) = \frac{\hat{S}(t)}{S^*(t)}$. Derivemos $R(t)$ usando esta expresión:

$$R(t) = \exp\left(H^*(t) - \hat{H}(t)\right).$$

Luego, haciendo algunos cálculos:

$$dR(t) = R(t^-) \cdot d\left(H^*(t) - \hat{H}(t)\right) = R(t^-) \cdot \left(dH^*(t) - d\hat{H}(t)\right).$$

Sabemos que:

$$dH^*(t) = h(t) J(t) dt.$$

$$d\hat{H}(t) = \frac{J(t)}{Y(t)} dN(t) = \frac{J(t)}{Y(t)} (h(t)Y(t) dt + dM(t)) = h(t) J(t) dt + \frac{J(t)}{Y(t)} dM(t).$$

Por tanto:

$$dR(t) = R(t^-) \left(h(t)J(t) dt - \left[h(t)J(t) dt + \frac{J(t)}{Y(t)} dM(t) \right] \right) = -R(t^-) \cdot \frac{J(t)}{Y(t)} dM(t).$$

Recordando que:

$$R(t^-) = \frac{\hat{S}(t^-)}{S^*(t)}.$$

Entonces:

$$d\left(\frac{\hat{S}(t)}{S^*(t)}\right) = -\frac{\hat{S}(t^-)}{S^*(t)} \cdot \frac{J(t)}{Y(t)} dM(t).$$

Ahora, integrando desde 0 hasta t :

$$\frac{\hat{S}(t)}{S^*(t)} - \frac{\hat{S}(0)}{S^*(0)} = - \int_0^t \frac{\hat{S}(s^-)}{S^*(s)} \cdot \frac{J(s)}{Y(s)} dM(s).$$

Dado que $\hat{S}(0) = S^*(0) = 1$, se tiene:

$$\frac{\hat{S}(t)}{S^*(t)} - 1 = - \int_0^t \frac{\hat{S}(s^-) J(s)}{S^*(s) Y(s)} dM(s).$$

Esto implica que el proceso $\hat{S}(t)/S^*(t) - 1$ es una martingala. Como consecuencia, tenemos que

$$E \left[\frac{\hat{S}(t)}{S^*(t)} \right] = 1, \quad \text{para todo } t \geq 0.$$

Además, dado que $H^*(t) \leq H(t)$, se deduce que $S^*(t) \geq S(t)$, lo cual implica:

$$E \left[\frac{\hat{S}(t)}{S(t)} \right] \geq E \left[\frac{\hat{S}(t)}{S^*(t)} \right] = 1,$$

es decir,

$$E \left[\hat{S}(t) \right] \geq S(t).$$

Por tanto, el estimador de Kaplan-Meier tiende a sobreestimar $S(t)$, aunque el sesgo suele ser pequeño, especialmente en las primeras etapas, al igual que ocurre con el estimador de Nelson-Aalen.

Nos enfocamos ahora en la varianza de $\hat{S}(t)$. Usando que $[\frac{\hat{S}(t)}{S^*(t)} - 1]^2 - \langle \frac{\hat{S}(t)}{S^*(t)} - 1 \rangle$ es martingala, se deduce:

$$E \left[\left(\frac{\hat{S}(t)}{S^*(t)} - 1 \right)^2 \right] = E \left[\left\langle \frac{\hat{S}(t)}{S^*(t)} - 1 \right\rangle (t) \right].$$

Como $d\langle M \rangle(s) = d\Lambda(s) = Y(s) dH(s)$, se obtiene:

$$\left\langle \frac{\hat{S}}{S^*} - 1 \right\rangle (t) = \int_0^t \left(\frac{\hat{S}(s^-)}{S^*(s)} \right)^2 \frac{J(s)}{Y^2(s)} d\langle M \rangle(s) = \int_0^t \left(\frac{\hat{S}(s^-)}{S^*(s)} \right)^2 \frac{J(s)}{Y(s)} dH(s).$$

Un estimador natural de esta cantidad, a partir de la variación opcional, es:

$$\hat{\sigma}^2(t) = \int_0^t \frac{J(s)}{Y(s)} d\hat{H}(s) = \int_0^t \frac{1}{Y(s)^2} dN(s).$$

Usualmente se cumple que $S^*(t) \approx S(t)$, y por lo tanto, también se tiene que

$$\text{Var}(\hat{S}(t)/S(t) - 1) \approx \text{Var}(\hat{S}(t)/S^*(t) - 1),$$

lo cual sugiere usar el siguiente estimador de varianza:

$$\widehat{\text{Var}}(\hat{S}(t)) := \left(\hat{S}(t) \right)^2 \hat{\sigma}^2(t).$$

Como alternativa, se puede estimar directamente la expresión anterior sustituyendo $S^*(s)$ por $\hat{S}(s)$ en:

$$\left\langle \frac{\hat{S}}{S^*} - 1 \right\rangle (t) = \int_0^t \left(\frac{\hat{S}(s^-)}{S^*(s)} \right)^2 \frac{J(s)}{Y(s)} (1 - \Delta H(s)) dH(s),$$

teniendo en cuenta que: $\frac{\hat{S}(s)}{\hat{S}(s^-)} = \frac{n_j - d_j}{n_j} = 1 - \frac{\Delta N(s)}{Y(s)}$. Con esto, se obtiene un estimador alternativo de la varianza:

$$\widetilde{\text{Var}}(\hat{S}(t)) := \left(\hat{S}(t) \right)^2 \tilde{\sigma}^2(t), \quad \text{con} \quad \tilde{\sigma}^2(t) = \int_0^t \frac{1}{Y(s)(Y(s) - \Delta N(s))} dN(s).$$

Este último es conocido como el estimador de Greenwood. Diversos estudios muestran que $\widetilde{\text{Var}}(\hat{S}(t))$ suele ofrecer mejores resultados que $\widehat{\text{Var}}(\hat{S}(t))$ en la práctica.

Como comentario final, cabe destacar que para los cálculos de esta sección correspondiente al estimador de Kaplan-Meier se ha empleado el siguiente teorema.

2.9 Teorema. Sea V un proceso de variación finita con trayectorias continuas por la derecha, y sea f tal que f' existe y es continua. Entonces el proceso $(f(V(t)))_{t \geq 0}$ también es de variación finita, y se cumple que:

$$f(V(t)) - f(V(0)) = \int_{0+}^t f'(V(s-)) dV(s) + \sum_{0 < s \leq t} [f(V(s)) - f(V(s-)) - f'(V(s-))\Delta V(s)],$$

La notación \int_{0+}^t denota la integral en el intervalo semiabierto $(0, t]$ y por proceso de variación finita entendemos que para casi toda trayectoria de V :

$$\sup_{\Pi} \sum_{i=1}^n |V(t_i) - V(t_{i-1})| < \infty,$$

donde $\Pi = \{0 = t_0 < t_1 < \dots < t_n = T\}$ es una partición arbitraria de $[0, T]$.

Concretamente, este resultado es el Teorema 31 del capítulo 2 en [12].

2.2.2. Inferencia paramétrica

En algunos contextos es útil describir el comportamiento de un proceso mediante una forma funcional específica. Por ejemplo, podríamos tener razones para suponer que cierta característica del proceso sigue un patrón determinado. Supongamos, de manera más general, que observamos un proceso de conteo $N(t)$ en el intervalo temporal $[0, T]$. Podemos asumir inicialmente que $N(t)$ es un proceso univariado, con una intensidad dependiente del tiempo y de un parámetro $\theta \in \Theta \subset \mathbb{R}^k$, el cual es desconocido. ¿Cómo se realiza inferencia estadística en este tipo de modelos paramétricos? El enfoque más común es el método de máxima verosimilitud. Para poder aplicarlo, es necesario contar con una expresión clara de la función de verosimilitud.

2.2.2.1. Verosimilitud en procesos de conteo

De forma general, la verosimilitud se refiere a la probabilidad o densidad de los datos observados, entendida como una función del parámetro. Para un proceso de conteo $N(t)$ en el intervalo $[0, T]$, la información observada consiste en que han ocurrido n eventos en los tiempos $t_1 < \dots < t_n \leq T$. Al igual que en un modelo de variables i.i.d. con función de densidad, este tipo de observación tiene probabilidad cero, por lo que se utiliza la densidad en lugar de la probabilidad para definir la verosimilitud. Usaremos las notaciones $L(\theta)$ y $\ell(\theta)$ para representar la verosimilitud y la log-verosimilitud, respectivamente. Cabe destacar que llegaremos a las expresiones de $L(\theta)$ y $\ell(\theta)$ de forma heurística (a pesar de que existe un razonamiento formal que las demuestra) y que esto es útil, pues permite tratar de manera unificada muchas situaciones.

La idea principal para obtener una expresión de la verosimilitud es considerar que, en un intervalo pequeño $(t, t + \Delta t)$, el incremento $\Delta N(t)$ se comporta aproximadamente como una variable de Bernoulli con parámetro $P(\Delta N(t) = 1 \mid \mathcal{F}_{t-}) \approx \lambda(t, \theta)\Delta t$. Por lo tanto, la probabilidad de observar $\Delta N(t)$ en este intervalo, dado el historial hasta el tiempo $t-$, es aproximadamente

$$(\lambda(t, \theta)\Delta t)^{\Delta N(t)} (1 - \lambda(t, \theta)\Delta t)^{1 - \Delta N(t)}.$$

Dividiendo el intervalo $[0, T]$ en segmentos de longitud Δt , se puede expresar que la probabilidad del conjunto de observaciones es el producto de las probabilidades de cada segmento, condicionadas sucesivamente, lo cual conduce al límite siguiente:

$$\begin{aligned} & \lim_{\Delta t \downarrow 0} \prod_{t \leq T} (\lambda(t, \theta) \Delta t)^{\Delta N(t)} (1 - \lambda(t, \theta) \Delta t)^{1 - \Delta N(t)} \\ &= \lim_{\Delta t \downarrow 0} (\Delta t)^{N(T)} \prod_{t \leq T} \lambda(t, \theta)^{\Delta N(t)} (1 - \lambda(t, \theta) \Delta t)^{1 - \Delta N(t)}. \end{aligned}$$

De los cálculos que haremos a continuación se deduce que el límite anterior es cero. Sin embargo, si eliminamos el término $(\Delta t)^{N(T)}$ (que no depende del parámetro θ , sino únicamente del número de sucesos observados en el intervalo), obtendremos un límite no nulo. Es más fácil verlo tomando logaritmos. Entonces, usando que $\log(1 - x) \approx -x$ para x pequeño,

$$\begin{aligned} & \lim_{\Delta t \downarrow 0} \log \left(\prod_{t \leq T} \lambda(t, \theta)^{\Delta N(t)} (1 - \lambda(t, \theta) \Delta t)^{1 - \Delta N(t)} \right) = \\ &= \lim_{\Delta t \downarrow 0} \sum_{t \leq T} \log \lambda(t, \theta) \Delta N(t) + \sum_{t \leq T} (1 - \Delta N(t)) \log(1 - \lambda(t, \theta) \Delta t) \\ &= \lim_{\Delta t \downarrow 0} \sum_{t \leq T} \log \lambda(t, \theta) \Delta N(t) - \sum_{t \leq T} (1 - \Delta N(t)) \lambda(t, \theta) \Delta t \\ &= \lim_{\Delta t \downarrow 0} \sum_{t \leq T} \log \lambda(t, \theta) \Delta N(t) - \int_0^T \lambda(t, \theta) dt \\ &= \int_0^T \log \lambda(t, \theta) dN(t) - \int_0^T \lambda(t, \theta) dt. \end{aligned}$$

En este sentido, podemos decir que la densidad de probabilidad de observar $N(t)$ en $[0, T]$ es

$$\exp \left(\int_0^T \log \lambda(t, \theta) dN(t) - \int_0^T \lambda(t, \theta) dt \right).$$

Esto nos lleva a considerar la verosimilitud:

$$L(\theta) := \exp \left(\int_0^T \log \lambda(t, \theta) dN(t) - \int_0^T \lambda(t, \theta) dt \right). \quad (2.2)$$

y la log-verosimilitud:

$$\ell(\theta) := \int_0^T \log \lambda(t, \theta) dN(t) - \int_0^T \lambda(t, \theta) dt. \quad (2.3)$$

Lógicamente, si observamos un proceso $(N_1(t), \dots, N_n(t))$ en $[0, T]$ y los procesos N_i son independientes, con intensidades $\lambda_i(t, \theta)$, la log-verosimilitud será la suma de las log-verosimilitudes:

$$\ell(\theta) = \sum_{i=1}^n \int_0^T \log \lambda_i(t, \theta) dN_i(t) - \int_0^T \lambda_i(t, \theta) dt. \quad (2.4)$$

Ahora supongamos que el proceso $N(t)$ es observado hasta un tiempo aleatorio $T(\omega)$, definido sobre el mismo espacio de probabilidad. La clave es que la observación está restringida al intervalo aleatorio $[0, T(\omega)]$. Toda la información observable sobre el proceso está contenida en esta trayectoria truncada.

La heurística usada se basa en aproximar localmente el proceso por incrementos tipo Bernoulli, y construir una densidad como producto de términos individuales. Esa idea se puede extender directamente al caso con T aleatorio si condicionamos dado el valor observado de T .

La log-verosimilitud condicional al valor observado $T = t$ es:

$$\ell(\theta | T = t) = \int_0^t \log \lambda(s, \theta) dN(s) - \int_0^t \lambda(s, \theta) ds. \quad (2.5)$$

Y la verosimilitud condicional correspondiente es:

$$L(\theta | T = t) = \exp \left(\int_0^t \log \lambda(s, \theta) dN(s) - \int_0^t \lambda(s, \theta) ds \right). \quad (2.6)$$

Esto es idéntico a la fórmula original, pero ahora T es aleatorio y simplemente se reemplaza por su valor observado t en cada trayectoria.

Si queremos modelar la verosimilitud conjunta de N y de T (por ejemplo, si T tiene una densidad $f_T(t)$), entonces la verosimilitud completa sería:

$$L_{completa}(\theta) = f_T(t) \cdot \exp \left(\int_0^t \log \lambda(s, \theta) dN(s) - \int_0^t \lambda(s, \theta) ds \right). \quad (2.7)$$

La parte que depende del modelo de conteo es la segunda. La función $f_T(t)$ se incluye si queremos también estimar parámetros del modelo de censura, o si T no es independiente de θ .

A continuación, procedemos a estudiar algunos casos particulares.

■ Ejemplo 1: Modelo de supervivencia sin censuras

Supongamos que estamos analizando, durante el intervalo $[0, T]$, un conjunto de n individuos cuya duración de vida está descrita por una función de riesgo $h(t, \theta)$. Sea X_i el tiempo de vida del individuo i , y supongamos que los individuos son independientes. Entonces, la log-verosimilitud total es $\ell(\theta) = \sum_{i=1}^n \ell_i(\theta)$, donde $\ell_i(\theta)$ representa la contribución del individuo i a la log-verosimilitud.

En este modelo, definimos la función de intensidad como $\lambda_i(t, \theta) = Y_i(t)h(t, \theta)$, donde $Y_i(t) = I_{\{X_i \geq t\}}$ es un indicador que toma valor 1 si el individuo sigue en riesgo en el instante t . Recordando que $h(t, \theta) = f(t, \theta)/S(t, \theta)$, con $f(t, \theta)$ la densidad y $S(t, \theta)$ la función de supervivencia, la expresión para $\ell_i(\theta)$ queda dada por:

$$\ell_i(\theta) = \int_0^T \log \lambda_i(t, \theta) dN_i(t) - \int_0^T \lambda_i(t, \theta) dt,$$

lo cual se puede escribir como:

$$\ell_i(\theta) = \begin{cases} \log h(X_i, \theta) - H(X_i, \theta), & \text{si } X_i \leq T. \\ -H(T, \theta), & \text{si } X_i > T. \end{cases}$$

De forma equivalente, la verosimilitud para el individuo i es:

$$L_i(\theta) = \exp(\ell_i(\theta)) = \begin{cases} h(X_i, \theta) \exp(-H(X_i, \theta)) = f(X_i, \theta), & X_i \leq T. \\ \exp(-H(T, \theta)) = S(T, \theta), & X_i > T. \end{cases}$$

Por lo tanto, la verosimilitud total se expresa como:

$$L(\theta) = \prod_{i=1}^n f(X_i, \theta)^{I_{\{X_i \leq T\}}} S(T, \theta)^{I_{\{X_i > T\}}}.$$

En este modelo, las observaciones completas (cuando se conoce $X_i \leq T$) contribuyen con la densidad, mientras que las observaciones incompletas (censuradas) lo hacen mediante la función de supervivencia.

Si se dispone de todos los datos completos ($X_i \leq T$ para todo i), la verosimilitud se simplifica a:

$$L(\theta) = \prod_{i=1}^n f(X_i, \theta),$$

es decir, la forma habitual de la verosimilitud bajo observaciones independientes e igualmente distribuidas.

■ **Ejemplo 2: Modelo de supervivencia con censuras**

En este caso, los cálculos deben realizarse en función del proceso $N_i^C(t)$, siendo $\lambda_i(t, \theta) = h(t, \theta)Y_i^C(t)$, donde $Y_i^C(t) = I_{\{\tilde{X}_i \geq t\}}$ representa el indicador de riesgo considerando la censura. Aquí, \tilde{X}_i es el tiempo de observación del individuo i , que puede corresponder a un tiempo censurado o no.

La contribución del individuo i a la log-verosimilitud se expresa como:

$$\ell_i(\theta) = \int_0^T \log \lambda_i^C(t, \theta) dN_i^C(t) - \int_0^T \lambda_i^C(t, \theta) dt,$$

lo que se puede simplificar como:

$$\ell_i(\theta) = \begin{cases} \log h(\tilde{X}_i, \theta) - H(\tilde{X}_i, \theta), & \text{si } \tilde{X}_i = X_i \text{ (evento observado).} \\ -H(\tilde{X}_i, \theta), & \text{si } \tilde{X}_i \neq X_i \text{ (censura).} \end{cases}$$

A partir de esto, obtenemos la expresión de la verosimilitud total:

$$L(\theta) = \prod_{i=1}^n f(\tilde{X}_i, \theta)^{D_i} S(\tilde{X}_i, \theta)^{1-D_i},$$

donde D_i es el indicador de evento (1 si el evento fue observado, 0 si está censurado).

■ **Ejemplo 3: Procesos de Poisson**

Interpretamos $N(t)$ como el número de fallos (y reparaciones) que se producen en un sistema durante el intervalo $[0, t]$. Suponemos que los tiempos de fallo están dados por $T_1 < \dots < T_n$.

El modelo de Poisson (posiblemente no homogéneo) se considera un modelo de “reparación imperfecta” (el estado del sistema tras la reparación es el mismo estado que antes de la reparación). En este modelo, la intensidad se expresa como $\lambda(t) = \xi(t, \theta)$, donde ξ es una función determinista. Que la intensidad no cambie antes y después del fallo implica que el sistema vuelve al mismo estado tras la reparación.

La log-verosimilitud se puede obtener definiendo la función de intensidad acumulada $\Lambda(t, \theta) = \Xi(t, \theta) = \int_0^t \xi(s, \theta) ds$, con lo cual se tiene:

$$\ell(\theta) = \int_0^T \log \xi(t, \theta) dN(t) - \Xi(T, \theta).$$

Un caso particular es el modelo de Poisson homogéneo, donde $\theta = \lambda > 0$ y $\lambda(t, \lambda) = \lambda$, resultando:

$$\ell(\lambda) = \int_0^T \log \xi(t, \lambda) dN(t) - \int_0^T \xi(t, \lambda) dt = N(T) \log \lambda - \lambda T.$$

Para este caso sencillo, se puede hallar el estimador máximo verosímil, EMV. De este modo, derivando la log-verosimilitud:

$$\frac{\partial \ell}{\partial \lambda} = \frac{1}{\lambda} N(T) - T.$$

Imponiendo que esta ecuación sea igual a 0, se llega a que el EMV es:

$$\hat{\lambda} = \frac{N(T)}{T}.$$

Ahora, veamos que se trata de un buen estimador; es decir, veamos que es insesgado y consistente.

- Para ver que el estimador es insesgado, hay que demostrar que $E[\hat{\lambda}] = \lambda$. Sabemos que si $N(t)$ es un proceso de Poisson homogéneo con intensidad λ , entonces $N(T) \sim \text{Poisson}(\lambda T)$, lo que implica que $E[N(T)] = \lambda T$. Evaluamos la esperanza del estimador:

$$E[\hat{\lambda}] = E\left[\frac{N(T)}{T}\right] = \frac{1}{T} E[N(T)] = \frac{1}{T} \cdot \lambda T = \lambda.$$

Por lo tanto, el estimador $\hat{\lambda}$ es insesgado.

- Para demostrar que $\hat{\lambda}$ es consistente, analizamos el comportamiento cuando $T \rightarrow \infty$. Busco ver que $\hat{\lambda} = \frac{N(T)}{T} \xrightarrow{p} \lambda$. Sabemos que $E[N(T)] = \lambda T$ y que $\text{Var}(N(T)) = \lambda T$. En base a esta información:

$$\text{Var}(\hat{\lambda}) = \text{Var}\left(\frac{N(T)}{T}\right) = \frac{\lambda T}{T^2} = \frac{\lambda}{T} \rightarrow 0 \quad \text{cuando } T \rightarrow \infty.$$

Esto implica que $\hat{\lambda}$ converge en probabilidad a λ cuando $T \rightarrow \infty$, pues utilizando la desigualdad de Chebyshev:

$$P(|\hat{\lambda} - \lambda| \geq \varepsilon) \leq \frac{\text{Var}(\hat{\lambda})}{\varepsilon^2}.$$

Como $\text{Var}(\hat{\lambda}) \rightarrow 0$, entonces esa probabilidad también tiende a cero. Por tanto, $\hat{\lambda}$ es consistente.

Otro caso particular dentro de los procesos de Poisson es el llamado proceso “power law”, en el que $\theta = (\phi, \mu)$ con $\phi, \mu > 0$ y $\lambda(t) = \xi(t, \theta) = \frac{\phi}{\mu} \left(\frac{t}{\mu}\right)^{\phi-1}$ para $t > 0$, resultando:

$$\begin{aligned}
\ell(\phi, \mu) &= \int_0^T \log \xi(t, \phi, \mu) dN(t) - \int_0^T \xi(t, \phi, \mu) dt \\
&= \int_0^T \log \left(\frac{\phi}{\mu} \left(\frac{t}{\mu} \right)^{\phi-1} \right) dN(t) - \int_0^T \frac{\phi}{\mu} \left(\frac{t}{\mu} \right)^{\phi-1} dt \\
&= \int_0^T \left(\log \frac{\phi}{\mu} + \log \frac{t^{\phi-1}}{\mu^{\phi-1}} \right) dN(t) - \left(\frac{T}{\mu} \right)^\phi \\
&= (\phi - 1) \int_0^T \log \frac{t}{\mu} dN(t) + \log \frac{\phi}{\mu} N(T) - \left(\frac{T}{\mu} \right)^\phi.
\end{aligned}$$

y resolviendo la integral de Stieltjes $\int_0^T \log(s) dN(s) = \sum_{i:T_i \leq T} \log(T_i)$:

$$\ell(\phi, \mu) = (\phi - 1) \left[\sum_{i:T_i \leq T} \log \frac{T_i}{\mu} \right] + N(T) \log \frac{\phi}{\mu} - \left(\frac{T}{\mu} \right)^\phi.$$

Derivamos para tratar de obtener el EMV. Fijamos ϕ y tratamos de maximizar en μ . Derivando con respecto a μ obtenemos:

$$\begin{aligned}
\frac{\partial \ell(\phi, \mu)}{\partial \mu} &= (\phi - 1) \left[\sum_{i:T_i \leq T} \frac{-1}{\mu} \right] - \frac{N(T)}{\mu} + \phi \frac{T^\phi}{\mu^{\phi+1}}. \\
\frac{\partial \ell(\phi, \mu)}{\partial \mu} &= \frac{-N(T)\phi}{\mu} + \phi \frac{T^\phi}{\mu^{\phi+1}}.
\end{aligned}$$

Igualando a cero obtenemos que $\ell(\phi, \mu)$ es máximo si $\mu = T/N(T)^{1/\phi}$. Sustituyendo esta expresión en ℓ obtenemos la log-verosimilitud perfil, usando las propiedades de los logaritmos:

$$\tilde{\ell}(\phi) = (\phi - 1) \sum_{i:T_i \leq T} \log T_i + N(T) \log \phi - N(T)\phi \log T + N(T) \log N(T) - N(T).$$

Derivando,

$$\frac{\partial \tilde{\ell}(\phi)}{\partial \phi} = \sum_{i:T_i \leq T} \log T_i + \frac{N(T)}{\phi} - N(T) \log(T).$$

Igualando a cero, obtenemos que $\tilde{\ell}(\phi)$ alcanza su valor máximo en $\phi = \frac{N(T)}{\sum_{i:T_i \leq T} \log \frac{T}{T_i}}$.

Con todo esto, el EMV es:

$$(\hat{\phi}, \hat{\mu}) = \left(\frac{N(T)}{\sum_{i:T_i \leq T} \log \frac{T}{T_i}}, \frac{T}{N(T)^{1/\phi}} \right)$$

Notemos que el proceso “power law” incluye el proceso de Poisson homogéneo de tasa $\frac{1}{\mu} = \lambda$ sin más que tomar $\phi = 1$.

■ **Ejemplo 4: Procesos de renovación**

Los procesos de renovación son una alternativa a los procesos de Poisson dentro del ámbito de la fiabilidad. Supongamos, como en el ejemplo anterior, que $N(t)$ representa el número de reparaciones en el intervalo $[0, t]$, de modo que $N(t) = \sum_{i=1}^{\infty} I_{\{T_i \leq t\}}$, donde T_1, \dots, T_n indican los tiempos de fallo (y reparación), considerando también $T_0 = 0$. Se dice que $N(t)$ es un proceso de renovación si los intervalos entre fallos, $S_i = T_i - T_{i-1}$ con $i \geq 1$, son variables aleatorias independientes e idénticamente distribuidas. Dado que estas variables son siempre positivas, es útil caracterizar su distribución mediante la función de riesgo, $h(t)$.

El comportamiento del sistema puede entenderse como si, en el instante $t = 0$, comenzara a operar una máquina con una duración aleatoria descrita por una función de riesgo $h(t)$. Cada vez que la máquina falla, se reemplaza inmediatamente por otra con la misma función de riesgo, y así sucesivamente. En el instante t , la única información relevante para predecir el tiempo restante de funcionamiento de la máquina actual es el tiempo que ha estado en funcionamiento, es decir, $t - T_{N(t)}$. Por tanto, se tiene que:

$$\begin{aligned} \lambda(t) &= \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} P(N(t + \Delta t) - N(t) \geq 1 | \mathcal{F}_{t-}) = \\ &= \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} P(S_{N(t)+1} \in (t - T_{N(t)}, t - T_{N(t)} + \Delta t] | \mathcal{F}_{t-}) = h(t - T_{N(t)}). \end{aligned}$$

El modelo paramétrico de renovación que se presenta es un caso de reparación perfecta: tras un fallo (y su respectiva reparación), el sistema queda en las mismas condiciones que al inicio. Este modelo incluye como caso particular al proceso de Poisson homogéneo, donde $h(t) = \lambda > 0$, ya comentado anteriormente. Este es el único proceso que también es de Poisson y de renovación. Existen muchas otras opciones posibles para la función de riesgo.

Por ejemplo, si se considera el caso $h(t) = \frac{\phi}{\mu} \left(\frac{t}{\mu}\right)^{\phi-1}$, se obtiene el proceso de renovación de tipo Weibull. Es importante notar que el proceso “power law” y el proceso de renovación Weibull son distintos, incluso cuando $h(t) = \frac{\phi}{\mu} \left(\frac{t}{\mu}\right)^{\phi-1}$. En la Figura 2.2, se ilustra el comportamiento de la intensidad en ambos casos, utilizando $\phi = 3$ y $\mu = 1$. El gráfico de la izquierda corresponde al caso de Poisson, en el cual $\lambda(t) = h(t) = 3t^2$, es decir, una función determinista. En contraste, en el modelo de renovación, la intensidad es un proceso estocástico genuino, que presenta saltos en los momentos en que ocurren fallas. En cada uno de estos instantes, el proceso reinicia desde el principio, repitiendo la forma de la función $3t^2$ sobre intervalos aleatorios.

Centrándonos ahora en la inferencia, supondremos que $T_1 < \dots < T_n$ son los tiempos de fallo. Ahora es conveniente denotar $T_0 = 0$ y $T_{n+1} = T$. Entonces, a partir de la expresión general de la log-verosimilitud,

$$\ell(\theta) = \int_0^T \log \lambda(t, \theta) dN(t) - \int_0^T \lambda(t, \theta) dt,$$

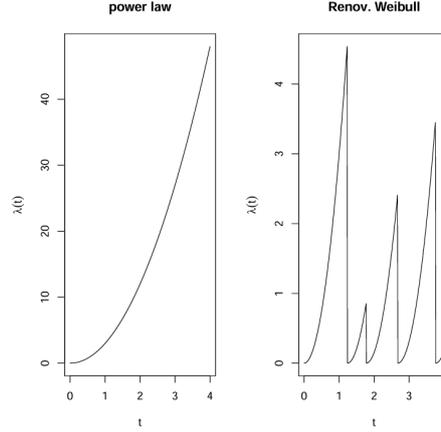


Figura 2.2: Modelo “power law” y modelo de renovación.

nos queda:

$$\begin{aligned}
 \ell(\theta) &= \int_0^T \log h(t - T_{N(t)}, \theta) dN(t) - \int_0^T h(t - T_{N(t)}, \theta) dt \\
 &= \sum_{i=1}^n \log h(T_i - T_{N(T_{i-})}, \theta) - \sum_{i: T_i \leq T} \int_{T_{i-1}}^{T_i} h(t - T_{i-1}, \theta) dt \\
 &= \sum_{i=1}^n \log h(T_i - T_{i-1}, \theta) - \sum_{i=1}^{n+1} \int_0^{S_i} h(s, \theta) ds \\
 &= \sum_{i=1}^n \log h(S_i, \theta) - \sum_{i=1}^{n+1} \int_0^{S_i} h(s, \theta) ds.
 \end{aligned}$$

Por tanto, tras resolver la primera integral de Stieltjes teniendo en cuenta que T_i es el momento del salto (considero que el proceso es càdlàg y que no se ha dado el salto en T_i-) y resolver la segunda integral descomponiéndola en suma de integrales:

$$\ell(\theta) = \sum_{i=1}^n \log h(S_i, \theta) - \sum_{i=1}^{n+1} H(S_i, \theta),$$

donde $H(t, \theta) = \int_0^t h(s, \theta) ds$ representa la función de riesgo acumulado asociada a los tiempos entre fallos.

Por otro lado, sabemos que $\log h(t, \theta) = \log f(t, \theta) + H(t, \theta)$ (donde f es la densidad del tiempo hasta el fallo). Para verlo, basta con tener en cuenta que $S(t, \theta) = \exp(-H(t, \theta))$ y que $h(t, \theta) = \frac{f(t, \theta)}{S(t, \theta)} = f(t, \theta) \exp(H(t, \theta))$. Tomando logaritmos:

$$\log h(t, \theta) = \log (f(t, \theta) \cdot \exp(H(t, \theta))) = \log f(t, \theta) + H(t, \theta).$$

Usando esta igualdad se llega a la expresión final de la log-verosimilitud:

$$\ell(\theta) = \sum_{j=1}^n \log f(S_j, \theta) - H(T - T_n, \theta).$$

Esto nos permite ver que la función de verosimilitud del modelo de renovación es equivalente a la de un modelo de supervivencia para los tiempos entre fallos, donde el tiempo transcurrido desde la última falla se interpreta como un dato censurado. Este resultado es natural, dado que los intervalos entre fallos son variables aleatorias i.i.d. en un proceso de renovación, y la única consideración adicional es incluir la información de que S_{n+1} debe ser al menos $T - T_n$.

CAPÍTULO 3

EL MODELO DE REGRESIÓN DE COX

En este capítulo se presenta el modelo de regresión de Cox, una de las herramientas más utilizadas en el análisis de supervivencia. Este modelo permite estudiar cómo diferentes covariables afectan al riesgo de ocurrencia de un evento a lo largo del tiempo, sin necesidad de especificar la forma exacta de la función de riesgo base.

Se introduce la formulación del modelo de riesgos proporcionales y se desarrolla la función de verosimilitud parcial, clave para la estimación de los coeficientes del modelo.

3.1. El modelo de Cox de riesgos proporcionales

El modelo de Cox describe la función de riesgo $h(t)$ en función del tiempo t y de un conjunto de covariables o variables explicativas/predictoras/factores de riesgo o de confusión, $\mathbf{X} = (X_1, \dots, X_p)$, que caracterizan al individuo bajo estudio de la siguiente manera:

$$h(t, \boldsymbol{\beta}, \mathbf{X}) = h(t, \boldsymbol{\beta}, X_1, \dots, X_p) = h_0(t) \exp \left(\sum_{j=1}^p \beta_j X_j \right). \quad (3.1)$$

Los parámetros del modelo son los coeficientes β_j . Suponemos como hipótesis inicial que los tiempos de supervivencia siguen distribuciones continuas, capturados de manera precisa, sin posibilidad de empates. Para cada sujeto i donde $i = 1, \dots, n$, se conoce el tiempo de evento (falla o muerte) t_i , el estado del evento o censura D_i (codificado como 1 si el evento fue observado y 0 si está censurado), y sus covariables fijas $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})$. Si incorporamos el subíndice i para denotar a cada individuo, el modelo (3.1) puede reescribirse como:

$$h(t, \boldsymbol{\beta}, \mathbf{X}_i) = h(t, \boldsymbol{\beta}, X_{i1}, \dots, X_{ip}) = h_0(t) \exp \left(\sum_{j=1}^p \beta_j X_{ij} \right).$$

La función $h_0(t)$ recibe el nombre de “función de riesgo basal” y representa el riesgo asociado a un sujeto cuyo valor de todos los predictores es cero [10]; es decir, el “individuo de referencia” para la posterior interpretación del análisis:

$$h(t, \boldsymbol{\beta}, X_1 = 0, \dots, X_p = 0) = h_0(t) \exp \left(\sum_{j=1}^p \beta_j 0 \right) = h_0(t) \exp(0) = h_0(t) 1 = h_0(t).$$

De manera equivalente, se puede entender que la función de riesgo basal es aquella función “básica” del modelo en ausencia de predictores.

La función de riesgo basal, $h_0(t)$, es la única parte de la expresión del modelo de Cox que depende del tiempo t . En cambio, la otra parte, $\exp\left(\sum_{j=1}^p \beta_j X_j\right)$, únicamente depende del conjunto de covariables $\mathbf{X} = (X_1, \dots, X_p)$ correspondientes a los sujetos, que nosotros asumiremos como “independientes del tiempo”.

Por otro lado, es posible trabajar con predictores que sí cambian con el tiempo. A estos los denotamos como $\mathbf{X}(t) = (X_1(t), \dots, X_p(t))$. En este contexto, se puede aplicar el modelo de Cox, pero ya no se cumpliría estrictamente la condición de “proporcionalidad de riesgos” que describiremos más adelante. Cuando los predictores dependen del tiempo, la regresión se conoce como “modelo de Cox extendido”.

El modelo de Cox (3.1) se considera “semiparamétrico” porque combina una parte paramétrica con otra no paramétrica:

- a) La **parte paramétrica** corresponde a $\exp\left(\sum_{j=1}^p \beta_j X_j\right)$, es decir, a la exponencial del predictor lineal $\sum_{j=1}^p \beta_j X_j$. En esta sección del modelo se estiman los parámetros de regresión $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \dots, \hat{\beta}_p)$ mediante la maximización de la llamada “función de verosimilitud parcial”, la cual analizaremos en detalle más adelante.
- b) La **parte no paramétrica** corresponde a la función de riesgo basal $h_0(t)$. Esta función es arbitraria y no especificada previamente, y su estimación se realiza en una segunda etapa, una vez obtenida la estimación de los parámetros $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \dots, \hat{\beta}_p)$. Es precisamente debido a esta componente no paramétrica que el modelo de Cox se clasifica como un modelo “semiparamétrico”.

Una vez estimada la parte paramétrica, $\exp\left(\sum_{j=1}^p \hat{\beta}_j X_j\right)$, y posteriormente la función de riesgo basal estimada, $\hat{h}_0(t)$, obtenemos el modelo semiparamétrico completo:

$$\hat{h}(t, \hat{\boldsymbol{\beta}}, \mathbf{X}) = \hat{h}_0(t) \exp\left(\sum_{j=1}^p \hat{\beta}_j X_j\right).$$

Tal y como ya hemos mencionado, en el modelo de Cox $h_0(t)$ no está especificada, y por tanto, la distribución del error tampoco lo está. En contraste, en un modelo paramétrico, la forma funcional se encuentra completamente determinada, excepto por los parámetros de la distribución asumida, los cuales deben ser estimados.

El carácter “semiparamétrico” del modelo de Cox lo convierte en una herramienta muy popular en el análisis de supervivencia. Al no requerir la especificación previa de la función de riesgo basal, permite estimar los coeficientes de regresión, calcular las razones de riesgo y modelar curvas de supervivencia en una amplia gama de escenarios. En este sentido, se puede considerar que el modelo de Cox es “robusto”, ya que los resultados obtenidos tienden a acercarse a los que se lograrían con el modelo paramétrico verdaderamente adecuado.

Si conociéramos con certeza cuál es el modelo paramétrico adecuado, optaríamos por dicho modelo en lugar del modelo de Cox, aunque deberíamos estar completamente convencidos de que el modelo paramétrico es el correcto. A pesar de que es posible emplear tests de bondad de ajuste, garantizar que la distribución paramétrica sea la correcta no siempre es sencillo. Por esta razón, el modelo de regresión de Cox representa una buena alternativa frente a los modelos paramétricos.

En otras palabras, el uso del modelo de Cox nos permite evitar la selección de un modelo paramétrico incorrecto para nuestro análisis.

Podríamos describir el modelo de regresión de Cox (3.1) de una manera más general en lo que refiere a su componente paramétrica:

$$h(t, \boldsymbol{\beta}, \mathbf{X}) = h_0(t) \psi(\boldsymbol{\beta}, \mathbf{X}). \quad (3.2)$$

La función $\psi(\boldsymbol{\beta}, \mathbf{X})$ se interpreta como el “riesgo relativo” en el instante t para un individuo con características $\mathbf{X} = (X_1, \dots, X_p)$, en comparación con un individuo con $\mathbf{X} = (0, \dots, 0)$.

El modelo (3.1) se correspondería con $\psi(\boldsymbol{\beta}, \mathbf{X}) = \exp\left(\sum_{j=1}^p \beta_j X_j\right)$. Esta forma parametrizada del modelo de Cox se conoce como forma “log lineal” y es la más utilizada. También se pueden considerar otras formas como la “lineal”: $\psi(\boldsymbol{\beta}, \mathbf{X}) = 1 + \sum_{j=1}^p \beta_j X_j$, y la “logística”: $\psi(\boldsymbol{\beta}, \mathbf{X}) = \log\left(1 + \exp\left(\sum_{j=1}^p \beta_j X_j\right)\right)$, las cuales corresponden a otras familias funcionales que, dependiendo de ciertos parámetros adicionales, abarcan los casos particulares lineales (log lineal y lineal).

Si examinamos la forma paramétrica del modelo de Cox presentada en (3.1), podemos notar que, al establecer una relación exponencial entre el predictor lineal y las funciones de riesgo (o de supervivencia), garantizamos que las estimaciones del modelo serán siempre positivas. Esta característica es altamente deseable, especialmente considerando que otros enfoques, como el modelo lineal simple, no siempre aseguran valores positivos.

Entre las propiedades más atractivas del modelo de Cox destacan [10]:

1. **No negatividad en las estimaciones:** la relación exponencial entre el predictor y el riesgo garantiza que las tasas de riesgo estimadas nunca serán negativas, lo cual es fundamental en la interpretación de los resultados.
2. **Estimación de los parámetros sin especificar la función de riesgo basal:** aunque la función de riesgo de referencia no esté determinada explícitamente, es posible estimar los coeficientes del predictor lineal de manera consistente.
3. **Análisis de los efectos de las variables explicativas:** una vez que se han estimado los parámetros, es factible estudiar la influencia de las variables independientes y calcular las razones de riesgo, sin necesidad de haber estimado previamente la función de riesgo.
4. **Estimación flexible de la función de riesgo:** en una segunda etapa, el modelo de Cox permite estimar la función de riesgo $h(t, \boldsymbol{\beta}, \mathbf{X})$ y las curvas de supervivencia $S(t, \boldsymbol{\beta}, \mathbf{X})$ con un mínimo de supuestos, dado que la función de riesgo basal $h_0(t)$ (o la supervivencia basal $S_0(t)$) no se especifica de antemano.

Un aspecto adicional que explica la “popularidad” del modelo de Cox es su preferencia frente al modelo de regresión logística cuando se analizan tiempos de supervivencia, ya sea con datos censurados o no. Esto se debe a que el modelo de Cox aprovecha mayor cantidad de información contenida en los datos de tiempo, mientras que el modelo logístico únicamente maneja variables de respuesta binarias (0 o 1), sin considerar la información relacionada con las censuras.

Cabe señalar que, en el modelo de riesgos proporcionales de Cox, el tratamiento y la codificación de los predictores se realizan de manera similar a como se hace en los modelos lineales o lineales generalizados. Sin embargo, una diferencia importante es que el término de “intercept” o medio no se modela explícitamente, ya que su función es absorbida por la función de riesgo basal.

3.1.1. La hipótesis de riesgos proporcionales

En el modelo de Cox, el primer objetivo es establecer la relación entre los riesgos de muerte de dos individuos que están sometidos a diferentes perfiles de covariables. Para ello, se asume una hipótesis central conocida como la proporcionalidad de riesgos.

Para entender mejor esta idea, introducimos el concepto de razón de riesgos (Hazard Ratio, HR), que compara el riesgo entre dos individuos caracterizados por vectores de covariables diferentes, $\mathbf{X} = (X_1, \dots, X_p)$ y $\mathbf{X}^* = (X_1^*, \dots, X_p^*)$, definiéndolo como [10]:

$$HR = \frac{h(t, \boldsymbol{\beta}, \mathbf{X}^*)}{h(t, \boldsymbol{\beta}, \mathbf{X})}. \quad (3.3)$$

En el numerador suele colocarse el grupo de mayor riesgo (\mathbf{X}^*) y en el denominador el de menor riesgo (\mathbf{X}). Bajo esta construcción, se espera que $HR > 1$, indicando que $h(t, \boldsymbol{\beta}, \mathbf{X}^*) > h(t, \boldsymbol{\beta}, \mathbf{X})$, es decir, que el riesgo de morir es mayor para el individuo con perfil \mathbf{X}^* que para el de perfil \mathbf{X} . La interpretación del HR es sencilla: valores superiores a uno indican un aumento del riesgo respecto al grupo de referencia; valores cercanos a uno sugieren riesgos similares.

Si sustituimos la expresión del modelo (3.1) en la fórmula (3.3), obtenemos [9]:

$$HR = \frac{h_0(t) \exp\left(\sum_{j=1}^p \beta_j X_j^*\right)}{h_0(t) \exp\left(\sum_{j=1}^p \beta_j X_j\right)} = \exp\left(\sum_{j=1}^p \beta_j (X_j^* - X_j)\right),$$

con lo que:

$$HR = \exp\left(\sum_{j=1}^p \beta_j (X_j^* - X_j)\right). \quad (3.4)$$

Notamos que el valor de la razón de riesgos en (3.4) no depende de la función de riesgo basal, sino únicamente de los valores de los predictores y de los coeficientes β estimados; es decir, no varía con el tiempo. Por esta razón, en el modelo de Cox se asume la hipótesis de riesgos proporcionales, lo cual implica que las covariables no dependen del tiempo.

La hipótesis de riesgos proporcionales establece explícitamente que la razón de riesgos (3.3) permanece constante a lo largo del tiempo, es decir:

$$h(t, \boldsymbol{\beta}, \mathbf{X}^*) = \text{constante} \cdot h(t, \boldsymbol{\beta}, \mathbf{X}).$$

Si aplicamos esta propiedad a la expresión del modelo de Cox (3.4) y llamamos θ a dicha constante, una vez estimados los coeficientes de regresión por máxima verosimilitud parcial, obtenemos que la razón de proporcionalidad es constante en el tiempo y se calcula como:

$$\hat{\theta} = \exp\left(\sum_{j=1}^p \hat{\beta}_j (X_j^* - X_j)\right). \quad (3.5)$$

Consideremos dos individuos, i y j , que únicamente difieren en la k -ésima covariable, de modo que $X_k = 0$ para el individuo i y $X_k = 1$ para el individuo j . Entonces, para cualquier instante de tiempo t , la razón de riesgos entre ambos sujetos viene dada por:

$$HR = \frac{h(t, \boldsymbol{\beta}, X_1, \dots, X_{k-1}, 1, X_{k+1}, \dots, X_p)}{h(t, \boldsymbol{\beta}, X_1, \dots, X_{k-1}, 0, X_{k+1}, \dots, X_p)}, \quad (3.6)$$

y se cumple que $HR = \exp(\beta_k)$, ya que:

$$\begin{aligned}
HR &= \frac{h_0(t) \exp\left(\sum_{j=1}^{k-1} \beta_j X_j + \beta_k \cdot 1 + \sum_{j=k+1}^p \beta_j X_j\right)}{h_0(t) \exp\left(\sum_{j=1}^{k-1} \beta_j X_j + \beta_k \cdot 0 + \sum_{j=k+1}^p \beta_j X_j\right)} \\
&= \frac{h_0(t) \exp\left(\beta_k + \sum_{j \neq k} \beta_j X_j\right)}{h_0(t) \exp\left(\sum_{j \neq k} \beta_j X_j\right)} \\
&= \exp(\beta_k).
\end{aligned}$$

Cuando las covariables son continuas, la interpretación de $HR = \exp(\beta_j)$ corresponde al cambio en la razón de riesgos al incrementar en una unidad la covariable continua X_j . Si deseamos calcular el efecto de un aumento de c unidades en X_j , la razón de riesgos sería $\exp(c\beta_j)$.

Finalmente, hay que tener en cuenta que al aplicar el modelo de regresión de Cox sobre un conjunto de datos, es importante comprobar que se satisface la hipótesis de proporcionalidad de riesgos. Normalmente, esto se realiza verificando que el efecto de cada variable se mantiene constante a lo largo del tiempo.

3.1.2. Función de verosimilitud parcial. Estimación de los coeficientes

En el modelo de regresión de Cox, los parámetros $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$ se obtienen maximizando el logaritmo de la llamada función de verosimilitud parcial. Dicha maximización se lleva a cabo mediante procedimientos numéricos, y nos proporciona la estimación $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \dots, \hat{\beta}_p)$. Con esta estimación, queda completamente especificada la parte paramétrica del modelo:

$$h(t, \hat{\boldsymbol{\beta}}, \mathbf{X}) = h_0(t) \exp\left(\sum_{j=1}^p \hat{\beta}_j X_j\right),$$

lo que nos permite realizar inferencias sobre el vector de parámetros y calcular los HR relevantes para nuestro análisis.

La función de verosimilitud parcial, que veremos a continuación, recibe el calificativo de “parcial” porque únicamente considera en su construcción las probabilidades asociadas a los tiempos de muerte o fallo, omitiendo las probabilidades correspondientes a los datos censurados. No obstante, en la evaluación de las probabilidades de los tiempos de muerte, se tienen en cuenta todos los individuos que están en riesgo en el momento inicial de cada tiempo de fallo, independientemente de si posteriormente fueron censurados o no.

Para llegar a la función de verosimilitud parcial vamos a usar algunos conceptos vistos en el capítulo 2, manteniendo la misma notación y seguiremos como referencia [11] y [9]. En primer lugar, hay que destacar que en procesos de supervivencia estudiados bajo el modelo de Cox, se suele asumir que la intensidad es de la forma:

$$\lambda(t, \boldsymbol{\beta}) = Y(t)h(t, \boldsymbol{\beta}, \mathbf{X}),$$

donde la función de riesgo viene dada por

$$h(t, \boldsymbol{\beta}, \mathbf{X}) = \lambda_0(t) \exp(\boldsymbol{\beta}' \cdot \mathbf{X}),$$

de modo que:

$$\lambda(t, \boldsymbol{\beta}) = \lambda_0(t)Y(t) \exp(\boldsymbol{\beta}' \cdot \mathbf{X}). \quad (3.7)$$

Supongamos ahora que se observan n copias independientes $(N_i(t), Y_i(t), \mathbf{X}_i)$, para $i = 1, \dots, n$, durante un cierto intervalo de tiempo $[0, T]$, con $T < \infty$, y que cada proceso $N_i(t)$ tiene una intensidad que sigue la forma dada en la ecuación (3.7):

$$\lambda_i(t, \boldsymbol{\beta}) = \lambda_0(t)Y_i(t) \exp(\boldsymbol{\beta}' \cdot \mathbf{X}_i). \quad (3.8)$$

Entonces, se puede hablar de la intensidad del proceso de conteo agregado N_{\bullet} , que será:

$$\lambda(t, \boldsymbol{\beta}) = \sum_{i=1}^n \lambda_i(t, \boldsymbol{\beta}) = \lambda_0(t) \sum_{i=1}^n Y_i(t) \exp(\boldsymbol{\beta}' \cdot \mathbf{X}_i) = \lambda_0(t) s_0(t, \boldsymbol{\beta}), \quad (3.9)$$

donde $s_0(t, \boldsymbol{\beta}) = \sum_{i=1}^n Y_i(t) \exp(\boldsymbol{\beta}' \cdot \mathbf{X}_i)$ es el proceso de individuos a riesgo.

El parámetro de regresión $\boldsymbol{\beta}$ se estima como el valor que maximiza la función de verosimilitud parcial de Cox (denotada por $L(\boldsymbol{\beta}) = L(\beta_1, \dots, \beta_p)$), dada por:

$$L(\boldsymbol{\beta}) = \prod_t \prod_i \left(\frac{\exp(\boldsymbol{\beta}' \cdot \mathbf{X}_i)}{s_0(t, \boldsymbol{\beta})} \right)^{\Delta N_i(t)}. \quad (3.10)$$

Atendiendo a [11], si el valor de $\boldsymbol{\beta}$ se fija, entonces un estimador natural de $\Lambda_0(t)$ es el estimador de tipo Nelson–Aalen:

$$\hat{\Lambda}_0(t, \boldsymbol{\beta}) = \int_0^t \frac{1}{s_0(s, \boldsymbol{\beta})} dN_{\bullet}(s), \quad (3.11)$$

donde $N_{\bullet}(t) = \sum_i N_i(t)$. El estimador (3.11) puede interpretarse como el maximizador, para un valor fijo de $\boldsymbol{\beta}$, de la función de verosimilitud dada por:

$$\prod_t \prod_i (d\Lambda_0(t) \exp(\boldsymbol{\beta}' \cdot \mathbf{X}_i))^{\Delta N_i(t)} \exp\left(-\int_0^T s_0(t, \boldsymbol{\beta}) d\Lambda_0(t)\right), \quad (3.12)$$

con respecto a los saltos $\Delta\Lambda_0(t)$. Esta expresión de la verosimilitud se puede deducir a partir de la expresión de la ecuación (2.4) del capítulo anterior. De este modo, partiendo de la expresión de la log-verosimilitud para un modelo de conteo con intensidades $\lambda_i(t)$:

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^n \int_0^T \log \lambda_i(t, \boldsymbol{\beta}) dN_i(t) - \int_0^T \lambda_i(t, \boldsymbol{\beta}) dt,$$

y sustituyendo la expresión de $\lambda_i(t, \boldsymbol{\beta})$ vista en (3.8):

$$\begin{aligned} \ell(\boldsymbol{\beta}) &= \sum_{i=1}^n \int_0^T \log[\lambda_0(t) Y_i(t) \exp(\boldsymbol{\beta}' \mathbf{X}_i)] dN_i(t) - \int_0^T \lambda_0(t) Y_i(t) \exp(\boldsymbol{\beta}' \mathbf{X}_i) dt \\ &= \sum_{i=1}^n \int_0^T [\log \lambda_0(t) + \log Y_i(t) + \boldsymbol{\beta}' \mathbf{X}_i] dN_i(t) - \int_0^T \lambda_0(t) Y_i(t) \exp(\boldsymbol{\beta}' \mathbf{X}_i) dt. \end{aligned}$$

Como $Y_i(t)$ es una función indicadora (1 si el individuo está en riesgo, 0 si no), y $dN_i(t) = 0$ si $Y_i(t) = 0$, el término $\log Y_i(t)$ no contribuye (es constante como función de $\boldsymbol{\beta}$) y puede ignorarse. Entonces:

$$\begin{aligned} \ell(\boldsymbol{\beta}) &= \sum_{i=1}^n \int_0^T \log \lambda_0(t) dN_i(t) + \sum_{i=1}^n \boldsymbol{\beta}' \mathbf{X}_i \int_0^T dN_i(t) - \int_0^T \lambda_0(t) \sum_{i=1}^n Y_i(t) \exp(\boldsymbol{\beta}' \mathbf{X}_i) dt + C \\ &= \int_0^T \log \lambda_0(t) \sum_{i=1}^n dN_i(t) + \sum_{i=1}^n \Delta N_i(T) \boldsymbol{\beta}' \mathbf{X}_i - \int_0^T \lambda_0(t) s_0(t, \boldsymbol{\beta}) dt + C. \end{aligned}$$

Aplicando la exponencial para obtener la verosimilitud:

$$\begin{aligned}
L(\boldsymbol{\beta}) &= \exp(\ell(\boldsymbol{\beta})) \\
&= \exp \left[\int_0^T \log \lambda_0(t) \sum_{i=1}^n dN_i(t) + \sum_{i=1}^n \Delta N_i(T) \boldsymbol{\beta}' \mathbf{X}_i - \int_0^T \lambda_0(t) s_0(t, \boldsymbol{\beta}) dt + C \right] \\
&= \exp \left[\sum_{i=1}^n \sum_{t \leq T} \log \lambda_0(t) \Delta N_i(t) + \sum_{i=1}^n \sum_{t \leq T} \boldsymbol{\beta}' \mathbf{X}_i \Delta N_i(t) - \int_0^T \lambda_0(t) s_0(t, \boldsymbol{\beta}) dt + C \right] \\
&= \exp \left[\sum_{i=1}^n \sum_{t \leq T} \log(\lambda_0(t)^{\Delta N_i(t)}) + \sum_{i=1}^n \sum_{t \leq T} \boldsymbol{\beta}' \mathbf{X}_i \Delta N_i(t) - \int_0^T \lambda_0(t) s_0(t, \boldsymbol{\beta}) dt + C \right].
\end{aligned}$$

Separando los factores y suprimiendo el término $\exp(C)$:

$$\begin{aligned}
L(\boldsymbol{\beta}) &= \exp \left[\sum_{i=1}^n \sum_{t \leq T} \log(\lambda_0(t)^{\Delta N_i(t)}) \right] \exp \left[\sum_{i=1}^n \sum_{t \leq T} \boldsymbol{\beta}' \mathbf{X}_i \Delta N_i(t) \right] \exp \left(- \left[\int_0^T \lambda_0(t) s_0(t, \boldsymbol{\beta}) dt \right] \right) \\
&= \prod_i \prod_{t \leq T} \exp \left[\log(\lambda_0(t)^{\Delta N_i(t)}) \right] \exp \left[\boldsymbol{\beta}' \mathbf{X}_i \Delta N_i(t) \right] \exp \left(- \left[\int_0^T \lambda_0(t) s_0(t, \boldsymbol{\beta}) dt \right] \right) \\
&= \prod_i \prod_{t \leq T} (\lambda_0(t)^{\Delta N_i(t)}) \exp \left[\boldsymbol{\beta}' \mathbf{X}_i \right]^{\Delta N_i(t)} \exp \left(- \left[\int_0^T \lambda_0(t) s_0(t, \boldsymbol{\beta}) dt \right] \right) \\
&= \prod_i \prod_{t \leq T} \left[\lambda_0(t) \exp(\boldsymbol{\beta}' \mathbf{X}_i) \right]^{\Delta N_i(t)} \exp \left(- \left[\int_0^T \lambda_0(t) s_0(t, \boldsymbol{\beta}) dt \right] \right).
\end{aligned}$$

Usando la función de riesgo acumulada $\Lambda_0(t)$, donde $\lambda_0(t) = \frac{d\Lambda_0(t)}{dt}$, se obtiene la expresión (3.12):

$$L(\boldsymbol{\beta}) = \prod_t \prod_i (d\Lambda_0(t) \exp(\boldsymbol{\beta}' \mathbf{X}_i))^{\Delta N_i(t)} \cdot \exp \left(- \int_0^T s_0(t, \boldsymbol{\beta}) d\Lambda_0(t) \right).$$

donde es importante tener en cuenta que en lugar de escribir $\lambda_0(t)$ en el primer factor, se usa que: $\lambda_0(t) = \frac{d\Lambda_0(t)}{dt} \Rightarrow \lambda_0(t) =$ derivada del salto en Λ_0 . Y como estamos evaluando justo en los tiempos de salto, lo que ocurre es que:

$$\lambda_0(t)^{\Delta N_i(t)} = \left(\frac{d\Lambda_0(t)}{dt} \right)^{\Delta N_i(t)} \quad \text{se puede pensar como} \quad d\Lambda_0(t)^{\Delta N_i(t)}.$$

Importante: $\prod_t (d\Lambda_0(t) \exp(\boldsymbol{\beta}' \mathbf{X}_i))^{\Delta N_i(t)}$ es un producto de a lo sumo un factor, es decir, si en $[0, T]$ el proceso N_i da un salto en el instante T_i , entonces $\prod_t (d\Lambda_0(t) \exp(\boldsymbol{\beta}' \mathbf{X}_i))^{\Delta N_i(t)} = d\Lambda_0(T_i) \exp(\boldsymbol{\beta}' \mathbf{X}_i)$. Si no hay ningún salto en el intervalo $[0, T]$, el producto vale simplemente 1.

Al reemplazar $d\Lambda_0(t)$ en (3.12) por $d\hat{\Lambda}_0(t, \boldsymbol{\beta})$, se obtiene una derivación heurística de la función de verosimilitud parcial de Cox dada en la ecuación (3.10).

De este modo, partiendo de la expresión (3.12), que representa la verosimilitud completa aproximada para un valor fijo de $\boldsymbol{\beta}$, y sustituyendo lo comentado anteriormente, se obtiene:

$$\prod_t \prod_i \left(d\hat{\Lambda}_0(t) \exp(\boldsymbol{\beta}' \cdot \mathbf{X}_i) \right)^{\Delta N_i(t)} \exp \left(- \int_0^T s_0(t, \boldsymbol{\beta}) d\hat{\Lambda}_0(t) \right).$$

Sustituyendo $d\hat{\Lambda}_0(t)$ en la verosimilitud atendiendo a la expresión del estimador de tipo Nelson-Aalen, obtenemos:

$$\begin{aligned} & \prod_t \prod_i \left(\frac{dN_{\bullet}(t)}{s_0(t, \boldsymbol{\beta})} \exp(\boldsymbol{\beta}' \cdot \mathbf{X}_i) \right)^{\Delta N_i(t)} \exp \left(- \int_0^T s_0(t, \boldsymbol{\beta}) \cdot \frac{dN_{\bullet}(t)}{s_0(t, \boldsymbol{\beta})} \right) = \\ & = \prod_t \prod_i \left(\frac{\exp(\boldsymbol{\beta}' \cdot \mathbf{X}_i)}{s_0(t, \boldsymbol{\beta})} \right)^{\Delta N_i(t)} \cdot dN_{\bullet}(t)^{\Delta N_i(t)} \cdot \exp \left(- \int_0^T dN_{\bullet}(t) \right). \end{aligned}$$

Observamos que los dos últimos términos no dependen de $\boldsymbol{\beta}$, por lo que se pueden absorber en una constante durante la maximización. Así, nos quedamos con:

$$\prod_t \prod_i \left(\frac{\exp(\boldsymbol{\beta}' \cdot \mathbf{X}_i)}{s_0(t, \boldsymbol{\beta})} \right)^{\Delta N_i(t)}$$

Esta expresión coincide exactamente con la verosimilitud parcial propuesta por Cox:

$$L(\boldsymbol{\beta}) = \prod_t \prod_i \left(\frac{\exp(\boldsymbol{\beta}' \cdot \mathbf{X}_i)}{s_0(t, \boldsymbol{\beta})} \right)^{\Delta N_i(t)}.$$

Tomando logaritmos llegamos a la expresión:

$$\log L(\boldsymbol{\beta}) = \sum_{i=1}^n \int_0^T (\boldsymbol{\beta}' \cdot \mathbf{X}_i - \log s_0(t, \boldsymbol{\beta})) dN_i(t). \quad (3.13)$$

Una vez obtenida la expresión (3.13), se deriva en ella respecto a los parámetros:

$$\frac{\partial \log L}{\partial \beta_j}, \quad (3.14)$$

$$\frac{\partial^2 \log L}{\partial \beta_i \partial \beta_j}. \quad (3.15)$$

De la ecuación (3.14), igualando a cero $\frac{\partial \log L}{\partial \beta_j} = 0$ para $j = 1, \dots, p$, se obtienen los sistemas de ecuaciones necesarios para calcular las estimaciones $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \dots, \hat{\beta}_p)$, usualmente resueltas mediante métodos numéricos. La expresión (3.15) permite verificar que se alcanza un máximo.

Conclusiones

En este trabajo se ha intentado explorar uno de los modelos más importantes en el ámbito del análisis de supervivencia, que es el modelo de riesgos proporcionales de Cox. En este contexto, es importante manejar datos con censura, de modo que para evitar tener que hacer un tratamiento especial para cada situación, es conveniente e interesante usar una formulación en términos de martingalas y procesos de conteo.

Las martingalas ayudan a formalizar la incertidumbre en el tiempo, permitiendo describir el comportamiento de algunas funciones de forma matemática y coherente con la información disponible hasta cada instante.

En particular, uno de los principales aportes del trabajo ha sido mostrar cómo definir y analizar adecuadamente las intensidades y compensadores en procesos de conteo, lo cual es fundamental para la formulación de la verosimilitud parcial del modelo de Cox. Por otro lado, el estudio de los procesos de conteo y la descomposición de Doob-Meyer ha permitido entender el comportamiento subyacente a los mecanismos que rigen la aparición de eventos en contextos de tiempo continuo. En cuanto a la inferencia, cabe destacar que tras tratar los estimadores de Kaplan-Meier y de Nelson-Aalen, se ha obtenido una expresión general para la verosimilitud de procesos de conteo. Esta expresión ha sido aplicada a algunos casos particulares y ha sido de utilidad para justificar la verosimilitud correspondiente al modelo de Cox.

Una vez tratado todo este marco, se ha introducido el modelo de Cox bajo la hipótesis de riesgos proporcionales. Se han destacado las ventajas de que se trate de un modelo semiparamétrico y se ha obtenido una expresión para la verosimilitud a partir de la cual se pueden calcular estimaciones mediante procedimientos de optimización numérica.

En definitiva, el modelo de regresión de Cox se presenta como una herramienta semiparamétrica potente, flexible y ampliamente utilizada en diversas disciplinas. Este trabajo ha contribuido a cimentar esa comprensión y a mostrar la coherencia y el rigor del modelo desde un enfoque teórico, destacando su utilidad, sus condiciones de validez y su potencial analítico en el estudio de fenómenos que se desarrollan a lo largo del tiempo.

APÉNDICE A

DEMOSTRACIONES CAPÍTULO 1

A.1 Teorema. Un proceso estocástico $X(t)$ es $\mathcal{F}|\beta^{\mathbb{T}}$ medible, si y sólo si, $X(t)$ es $\mathcal{F}|\beta$ medible para todo $t \in \mathbb{T}$.

Demostración. Veamos en primer lugar que si $X(t)$ es $\mathcal{F}|\beta^{\mathbb{T}}$ medible, entonces $X(t)$ es $\mathcal{F}|\beta$ medible $\forall t \in \mathbb{T}$. Notemos que sabemos que $\beta^{\mathbb{T}} = \sigma(\xi_1)$, con

$$\xi_1 := \left\{ \Pi_{t_1, \dots, t_k}^{-1}(B) : t_1, \dots, t_k \in \mathbb{T}, B \in \beta(\mathbb{R}^k), k \geq 1 \right\}.$$

Veamos que también se verifica que $\beta^{\mathbb{T}} = \sigma(\xi_2)$, con

$$\xi_2 := \left\{ \Pi_{t_1}^{-1}(B_1) \cap \dots \cap \Pi_{t_k}^{-1}(B_k) : t_1, \dots, t_k \in \mathbb{T}, B_i \in \beta(\mathbb{R}), k \geq 1 \right\}.$$

Es claro que $\xi_2 \subseteq \xi_1$, pues

$$\Pi_{t_1}^{-1}(B_1) \cap \dots \cap \Pi_{t_k}^{-1}(B_k) = \Pi_{t_1, \dots, t_k}^{-1}(B_1 \times \dots \times B_k),$$

y $\beta(\mathbb{R}^k)$ es la menor σ -álgebra que contiene a $B_1 \times \dots \times B_k$ (conjunto medible si B_i es medible $\forall i$). Por tanto, $\sigma(\xi_2) \subseteq \sigma(\xi_1)$.

Para concluir que $\sigma(\xi_2) = \sigma(\xi_1)$, consideramos el conjunto

$$H = \left\{ B \in \beta(\mathbb{R}^k) : \Pi_{t_1, \dots, t_k}^{-1}(B) \in \sigma(\xi_2) \right\}.$$

Notemos que, por construcción,

$$\{B_1 \times \dots \times B_k : B_i \in \beta(\mathbb{R})\} \subseteq H.$$

Veamos ahora que H es una σ -álgebra:

1. $\mathbb{R}^k = \mathbb{R} \times \dots \times \mathbb{R} \in H$ por lo que acabamos de decir.
2. ¿Si $B \in H \Rightarrow B^c \in H$? Si $B \in H \Rightarrow \Pi_{t_1, \dots, t_k}^{-1}(B) \in \sigma(\xi_2)$. Por otro lado, $\Pi_{t_1, \dots, t_k}^{-1}(B^c) = (\Pi_{t_1, \dots, t_k}^{-1}(B))^c \in \sigma(\xi_2)$ (pues $\Pi_{t_1, \dots, t_k}^{-1}(B) \in \sigma(\xi_2)$ y las propiedades de σ -álgebra). Por tanto, $B^c \in H$.
3. Veamos que H es cerrado para uniones numerables. Consideramos $\bigcup_{i=1}^{\infty} B_i, B_i \in H$. Queremos ver que $\bigcup_{i=1}^{\infty} B_i \in H$, es decir, $\Pi_{t_1, \dots, t_k}^{-1}(\bigcup_{i=1}^{\infty} B_i) \in \sigma(\xi_2)$. Sabemos que como $B_i \in H$, $\Pi_{t_1, \dots, t_k}^{-1}(B_i) \in \sigma(\xi_2)$ y que $\Pi_{t_1, \dots, t_k}^{-1}(\bigcup_{i=1}^{\infty} B_i) = \bigcup_{i=1}^{\infty} \Pi_{t_1, \dots, t_k}^{-1}(B_i)$. Luego $\Pi_{t_1, \dots, t_k}^{-1}(B_i) \in \sigma(\xi_2)$ por ser unión numerable de elementos de $\sigma(\xi_2)$.

Concluimos que H es σ -álgebra. Además, $\mathbb{R}^k \in H$. Luego $\beta(\mathbb{R}^k) \subset H$. Pero $H \subset \beta(\mathbb{R}^k)$, luego $H = \beta(\mathbb{R}^k)$. De todo esto, se sigue que $\xi_1 \subset \sigma(\xi_2) \Rightarrow \sigma(\xi_1) \subset \sigma(\xi_2)$ y por tanto $\sigma(\xi_1) = \sigma(\xi_2)$.

Por la definición de ξ_2 , es claro que ξ_2 es una π -clase. Veamos, ahora sí, que $X(t)$ es $\mathcal{F}|\beta$ medible. $X(t)$ es $\mathcal{F}|\beta$ medible si $\forall B \in \beta(\mathbb{R})$, $(X(t) \in B) \in \mathcal{F}$. Pero $(X(t) \in B) = (\Pi_t(X) \in B) = (X \in \Pi_t^{-1}(B))$ y $(X \in \Pi_t^{-1}(B)) \in \xi_2 \subset \beta^{\mathbb{T}}$. Como $X(t)$ es $\mathcal{F}|\beta^{\mathbb{T}}$ medible $\Rightarrow (X \in \Pi_t^{-1}(B)) = (X(t) \in B) \in \mathcal{F}$. Luego $X(t)$ es $\mathcal{F}|\beta$ medible.

Veamos la otra implicación: supongamos que $X(t)$ es $\mathcal{F}|\beta$ medible. Sea $\mathcal{D} = \{B \in \beta^{\mathbb{T}} : X^{-1}(B) \in \mathcal{F}\}$. Si $t_1, \dots, t_k \in \mathbb{T}$ y $B_1, \dots, B_k \in \beta \Rightarrow (X(t_1) \in B_1, \dots, X(t_k) \in B_k) \in \mathcal{F}$ (al ser X $\mathcal{F}|\beta$ medible y ser intersección de conjuntos de \mathcal{F}). Pero ese conjunto es:

$$(X(t_1) \in B_1, \dots, X(t_k) \in B_k) = (\Pi_{t_1, \dots, t_k}(X) \in B_1 \times \dots \times B_k) = (X \in \Pi_{t_1, \dots, t_k}^{-1}(B_1 \times \dots \times B_k)) \in \mathcal{F}.$$

Por tanto, $\Pi_{t_1}^{-1}(B_1) \cap \dots \cap \Pi_{t_k}^{-1}(B_k) \in \mathcal{D}$ (pues su contraimagen está en \mathcal{F}). Esto significa que $\xi_2 \subset \mathcal{D}$. Como además ξ_2 es π -clase y \mathcal{D} es λ -clase, podemos aplicar el Teorema de Dynkin para concluir que $\sigma(\xi_2) = \beta^{\mathbb{T}} \subset \mathcal{D}$. Luego $\mathcal{D} = \beta^{\mathbb{T}}$ y por tanto $X^{-1}(B) \in \mathcal{F} \forall B \in \beta^{\mathbb{T}}$.

Anexo: Veamos que \mathcal{D} es λ -clase:

1. $\mathbb{R}^{\mathbb{T}} \in \mathcal{D}$ pues $\Omega = X^{-1}(\mathbb{R}^{\mathbb{T}}) \in \mathcal{F}$ (al ser X $\mathcal{F}|\beta$ medible).
2. Sea $B \in \mathcal{D} \Rightarrow X^{-1}(B) \in \mathcal{F}$. ¿ $X^{-1}(B^c) \in \mathcal{F}$? Sí, pues $X^{-1}(B^c) = (X^{-1}(B))^c \in \mathcal{F}$.
3. Sean $B_1, B_2, \dots \in \mathcal{D}$ disjuntos $\Rightarrow X^{-1}(B_i) \in \mathcal{F} \forall i$. ¿ $\bigcup_{i=1}^{\infty} B_i \in \mathcal{D}$? Sí, pues $X^{-1}(\bigcup_{i=1}^{\infty} B_i) = \bigcup_{i=1}^{\infty} X^{-1}(B_i) \in \mathcal{F}$.

□

A.2 Teorema. Dados dos procesos estocásticos $X = (X(t))_{t \in \mathbb{T}}$ e $Y = (Y(t))_{t \in \mathbb{T}}$, entonces $X \stackrel{d}{=} Y$ si y sólo si $(X(t_1), \dots, X(t_k)) \stackrel{d}{=} (Y(t_1), \dots, Y(t_k))$ para todo $t_1, \dots, t_k \in \mathbb{T}$ con $k \geq 1$.

Demostración. Veamos en primer lugar que la igualdad en distribución implica igualdad de distribuciones finito-dimensionales. Si $X \stackrel{d}{=} Y$, por definición las leyes de probabilidad inducidas por X y Y en el espacio de trayectorias coinciden:

$$P_X = P_Y.$$

Sea (t_1, \dots, t_k) una colección finita de tiempos en \mathbb{T} . Consideremos la proyección:

$$\pi_{t_1, \dots, t_k} : \mathbb{R}^{\mathbb{T}} \rightarrow \mathbb{R}^k, \quad \pi_{t_1, \dots, t_k}((X(t))_{t \in \mathbb{T}}) = (X(t_1), \dots, X(t_k)).$$

Entonces:

$$\mathcal{L}(X(t_1), \dots, X(t_k)) = P_X \circ \pi_{t_1, \dots, t_k}^{-1} = P_Y \circ \pi_{t_1, \dots, t_k}^{-1} = \mathcal{L}(Y(t_1), \dots, Y(t_k)).$$

Por lo tanto, las distribuciones finito-dimensionales de X y Y coinciden.

Veamos ahora que la igualdad de distribuciones finito-dimensionales implica igualdad en distribución. Supongamos que:

$$\forall k \in \mathbb{N}, \forall t_1, \dots, t_k \in \mathbb{T}, \quad \mathcal{L}(X(t_1), \dots, X(t_k)) = \mathcal{L}(Y(t_1), \dots, Y(t_k)).$$

Recordemos que en $\mathbb{R}^{\mathbb{T}}$ consideraremos la σ -álgebra natural que viene dada por $\beta^{\mathbb{T}} = \sigma(\xi)$ con $\xi = \{\pi_{t_1, \dots, t_k}^{-1}(B) : t_1, \dots, t_k \in \mathbb{T}, B \in \beta(\mathbb{R}^k), k \geq 1\}$. Por otro lado, para cualquier conjunto de la forma $C = \pi_{t_1, \dots, t_k}^{-1}(B)$ se verifica por hipótesis:

$$P_X(C) = P((X(t_1), \dots, X(t_k)) \in B) = P((Y(t_1), \dots, Y(t_k)) \in B) = P_Y(C).$$

Por lo tanto, $\forall C \in \xi \Rightarrow P_X(C) = P_Y(C)$. De acuerdo con la demostración del teorema anterior, sabemos que ξ es una π -clase que genera $\beta^{\mathbb{T}}$ y tenemos igualdad de medidas (probabilidades) en ξ . Sea $\mathcal{D} = \{A \in \beta^{\mathbb{T}} : P_X(A) = P_Y(A)\}$. Es claro que $\xi \subset \mathcal{D}$ y que \mathcal{D} es una λ -clase. Luego, por el teorema de Dynkin se concluye que $\sigma(\xi) = \mathbb{R}^{\mathbb{T}} \subset \mathcal{D}$.

Luego se tiene que $\beta^{\mathbb{T}} = \mathcal{D}$ y por tanto $P_X(A) = P_Y(A) \forall A \in \beta^{\mathbb{T}}$. Es decir,

$$P_X = P_Y.$$

Por lo tanto:

$$X \stackrel{d}{=} Y.$$

Anexo: Veamos que \mathcal{D} es λ -clase:

1. $\mathbb{R}^{\mathbb{T}} \in \mathcal{D}$ pues $P_X(\mathbb{R}^{\mathbb{T}}) = P_Y(\mathbb{R}^{\mathbb{T}}) = 1$.
2. Sea $A \in \mathcal{D} \Rightarrow P_X(A) = P_Y(A) \Rightarrow 1 - P_X(A) = 1 - P_Y(A) \Rightarrow P_X(A^c) = P_Y(A^c)$.
3. Sean $A_1, A_2, \dots \in \mathcal{D}$ disjuntos $\Rightarrow P_X(\bigcup_{i=1}^{\infty} A_i) = \bigcup_{i=1}^{\infty} P_X(A_i) = \bigcup_{i=1}^{\infty} P_Y(A_i) = P_Y(\bigcup_{i=1}^{\infty} A_i)$.

□

BIBLIOGRAFÍA

- [1] Billingsley, P. *Probability and Measure*. 3.^a ed. New York: Wiley, 1995.
- [2] Cohen, S. N. et al. *Stochastic Calculus and Applications*. 2.^a ed. Probability and its Applications. Cham: Springer, 2015. ISBN: 978-1-4939-2866-8, 978-1-4939-2867-5. DOI: [10.1007/978-1-4939-2867-5](https://doi.org/10.1007/978-1-4939-2867-5).
- [3] Cox, D. R. “Regression Models and Life Tables”. En: *Journal of the Royal Statistical Society: Series B (Methodological)* 34[2] 1972, págs. 187-220.
- [4] Daley, D. J. et al. *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*. 2.^a ed. Probability and Its Applications. New York: Springer, 2003, págs. 229-232.
- [5] Del Barrio, E. *Apuntes sobre procesos de conteo*. Apuntes personales. Material de clase, no publicado.
- [6] Folland, G. B. *Real Analysis: Modern Techniques and Their Applications*. 2.^a ed. New York: John Wiley & Sons, 1999. ISBN: 978-0471317166.
- [7] Kallenberg, O. *Foundations of Modern Probability*. 2.^a ed. Probability and Its Applications. New York: Springer, 2002.
- [8] Karatzas, I. et al. *Brownian Motion and Stochastic Calculus*. Vol. 113. Graduate Texts in Mathematics. Springer, 1988.
- [9] Klein, J. P. et al. *Survival Analysis: Techniques for Censored and Truncated Data*. 2.^a ed. New York: Springer, 2003. ISBN: 978-1-4757-2730-2.
- [10] Kleinbaum, D. G. et al. *Survival Analysis: A Self-Learning Text*. 3.^a ed. Statistics for Biology and Health. New York: Springer, 2012. ISBN: 978-1-4419-6646-9.
- [11] Martinussen, T. et al. *Dynamic Regression Models for Survival Data*. Statistics for Biology and Health. Springer, 2006. ISBN: 978-0-387-20274-7.
- [12] Protter, P. E. *Stochastic Integration and Differential Equations*. 2.^a ed. Vol. 21. Stochastic Modelling and Applied Probability. Berlin: Springer, 2005. ISBN: 978-3-642-05560-7.