PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

ESCUELA DE POSGRADO

Una aplicación de la regresión de Cox con puntos de cambio en las covariables

TESIS PARA OPTAR POR EL GRADO DE MAGISTER EN ESTADÍSTICA

Presentado por:

Lucía Inés Trujillo Angeles

Asesora: Dra. Elizabeth Doig Camino

Miembros del jurado:
Dr. Giancarlo Sal y Rosas Celi
Dr. Luis Valdivieso Serrano

Lima, Diciembre 2014

Dedicatoria

A la reina Valentina

Agradecimentos

Un agradecimiento sincero a la profesora Elizabeth Doig por su empuje, orientación y buena disposición durante todo el proceso de realización de la tesis. Sin su apoyo y sus consejos no hubiera podido culminar el desarrollo del estudio.

Resumen

El siguiente trabajo de tesis, estudiará el modelo de regresión de Cox con puntos de cambio en las covariables propuesto por Jensen y Lutkebohmert (2008), realizando el desarrollo y la aplicación para una base de líneas móviles postpago. El objetivo es obtener los parámetros de las covariables y el nuevo parámetro en el modelo que es el punto de cambio, para analizar la manera como estas covariables tienen influencia en la desactivación de una línea a solicitud del cliente.

Palabras-clave: Modelo de regresión de Cox, Modelo de riesgos proporcionales, Análisis de supervivencia, residuos de martingala, procesos de conteo, punto de cambio.

Abstract

The following thesis will study the Cox regression model with change-points in the covariates proposed by Jensen y Lutkebohmert (2008), completing the development and implementation for a postpaid mobile lines. The goal is to obtain the parameters of the covariates and the new parameter of the model which is the change-point, for analyzing the way how the covariates influence the deactivation of mobile line upon client request.

Keywords: Cox's regression model, Proportional Hazards Models, Survival Analisys, Martingale Residuals, Counting Process, Change-point.

Índice general

Lista de Abreviaturas

INEI	Instituto Nacional de Estadística e Informática.
OSIPTEL	Organismo Supervisor de Inversión Privada en Telecomunicaciones.
Mb	Megabytes.
Min	Minutos.

Lista de Símbolos

$\boldsymbol{\beta}$	Vector de parámetros de las covariables.
$\boldsymbol{X}$	Vector de covariables.
$\boldsymbol{I}$	Variable que indica la censura del individuo.
$\boldsymbol{\omega}$	Vector de puntos de cambio.
$\boldsymbol{\Omega}$	Rectángulo donde se encuentra el punto de cambio.
$\boldsymbol{\theta}$	Vector $(\widehat{\boldsymbol{\omega}}, \widehat{\boldsymbol{\beta}})$.
$\boldsymbol{\Theta}$	Conjunto de $\boldsymbol{\theta}$.
$\boldsymbol{N}$	Vector ocurrencias del evento.
μ	Media.
R	Vector de riesgo de los individuos en el tiempo.
$\lambda(t, \boldsymbol{\theta})$	Tasa de riesgo para un individuo en el tiempo t asociado al vector $\boldsymbol{\theta}$.
$\lambda_0(t)$	Función de riesgo base en el tiempo t.
$\Lambda_0(t)$	Función acumulada del riesgo base en el tiempo t.
$\boldsymbol{P^t}$	Medida de probabilidad.
$f(t)$	Función de distribución de probabilidad en el tiempo t.
$F(t)$	Función de distribución acumulada al tiempo t.
$S(t)$	Función de supervivencia en el tiempo t.
$M(t)$	Residuos de Martingala en el tiempo t.

Índice de figuras

Índice de cuadros

Capítulo 1

Introducción

El efecto de las covariables tienen un comportamiento diferente a lo largo del tiempo de observación del sujeto, esta característica no se refleja en el modelo de regresión de Cox (1972), por lo que se propone estudiar el modelo de Jensen y Lutkebohmert (2008), donde se observa la influencia de las covariables con puntos de cambios dentro de los parámetros de la regresión en el tiempo.

1.1. Antecedentes

Cuando se estudia el modelo de regresión de Cox clásico, asumimos que la influencia de la covariable es constante en el tiempo; sin embargo, cuando se analizan diferentes tipos de registros de eventos, esta asunción no encaja con algunas de las covariables de estudio, pues se observa que éstas tienen un comportamiento diferente en los cortes de tiempo para la observación del ensayo. Es a partir de esto donde el concepto de punto de cambio en la covariable interviene en el modelo de regresión de Cox; pues se especifica un umbral desconocido. Es decir, el parámetro de regresión puede cambiar a lo largo del intervalo de una covariable y la función de regresión subyacente es continua pero no diferenciable.

Estudios previos se desarrollaron intentando plasmar la idea del punto de cambio dentro del modelo de regresión de Cox, en Luo y Boyett (1997) se desarrolló el siguiente modelo:

$$\lambda(t) = \lambda_0(t) \exp\left\{\beta_0 I_{\{X \leq \theta_0\}} + \alpha_0 Z\right\}$$

Con covariables unidimensionales que no dependen del tiempo, X y Z, en donde interviene la constante β_0 según el comportamiento de X al alcanzar cierto umbral. En el modelo anterior las dos covariables no son dependientes del tiempo.

K. Liang y Liu (1990) propusieron:

$$\lambda(t) = \lambda_0(t) \exp\left\{(\beta + \theta I_{\{t \leq \tau\}})Z + \gamma X\right\}$$

Donde el punto de cambio se daba en un tiempo desconocido τ, en ambos modelos se observa que dependiendo del objeto de estudio se tiene que distinguir entre los puntos de cambio que se producen en las covariables y de aquellos que se producen en el tiempo.

Pons (2003) introduce el modelo

$$\lambda(t) = \lambda_0(t) \exp\left\{\boldsymbol{\alpha}' \boldsymbol{X}_1(t) + \boldsymbol{\beta}' \boldsymbol{X}_2(t) \boldsymbol{I}_{\{X_3 \leq \omega\}} + \boldsymbol{\gamma}' \boldsymbol{X}_2(t) \boldsymbol{I}_{\{X_3 > \omega\}}\right\}$$

Donde por la influencia de una covariable X_3 se genera un salto en ω. Kosorok y Song (2007) demostraron que el estimador del parámetro de punto de cambio es n-consistente, estableciendo consistencia y convergencia débil de los estimadores no paramétricos de máxima verosimilitud.

En Gandy et al. (2005) se propone el siguiente modelo:

$$\lambda(t, \boldsymbol{\theta}) = \lambda_0(t) R(t) \exp\left\{\boldsymbol{\beta}_1' \boldsymbol{X}_1(t) + \beta_2 X_2 + \beta_3 \left(X_2 - \boldsymbol{\omega}\right)\right\}$$

En el cual se admite una covariable que depende del tiempo $\boldsymbol{X}_1(t)$, otra que no depende X_2 además permite un punto de cambio en $\boldsymbol{\omega}$.

El modelo de estudio propuesto por Jensen y Lutkebohmert (2008) es una extensión del modelo previo en el cual se permite más de un punto de cambio en el modelo, y la posibilidad de que todas las covariables sean dependientes del tiempo y que el proceso de conteo pueda saltar más de una vez, es el que se muestra a continuación:

$$\lambda(t, \boldsymbol{\theta}) = \lambda_0(t) R(t) \exp\left\{\boldsymbol{\beta}_1' \boldsymbol{X}_1(t) + \boldsymbol{\beta}_2' \boldsymbol{X}_2(t) + \boldsymbol{\beta}_3' \left(\boldsymbol{X}_2(t) - \boldsymbol{\omega}\right)\right\}$$

Con $\boldsymbol{\omega} \in \boldsymbol{\Omega} \subset \mathbb{R}^m$ y $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \boldsymbol{\beta}_3'\right)' \in B \subset \mathbb{R}^{p+2m}$ Donde:

- p: Cantidad de covariables ordinarias
- m: Cantidad de puntos de cambio
- $\boldsymbol{\omega}$: Vector de puntos de cambio
- $\boldsymbol{\beta}$: Vector de parámetros de la regresión
- $\lambda_0(t)$: Riesgo base
- $\boldsymbol{\theta} = \left\{\boldsymbol{\omega}', \boldsymbol{\beta}'\right\}$
- $R(t)$: Vector de riesgo que toma solo valores de 0 y 1 para indicar que el sujeto está en riesgo o no. Los autores prefieren utilizar a la función $R(t)$ debido a que permite simplificar los cálculos. En el modelo que construye Cox (1972) no considera a la función $R(t)$ pero si incluye la diferencia entre el conjunto de individuos que están o no en riesgo para la realización de los cálculos.

La diferencia de este modelo con los mencionados en Pons (2003), Kosorok y Song (2007), es que la función es continua en el vector de punto de cambio, por lo tanto estos puntos de cambio son caracterizados como puntos de cambios suaves, en Chappell (1989) nos dice que los cambios suaves son mas apropiados que los saltos.

El trabajo de aplicación del modelo se basa en el tiempo de supervivencia de una base de líneas telefónicas móviles y se busca conocer la probabilidad de pérdida de un cliente en base al estudio de la influencia de las covariables. Para ello se considerará el modelo de Jensen y Lutkebohmert (2008) como una opción para el estudio de este problema de investigación en el cual definimos a la línea como el objeto de estudio y la pérdida del cliente como el evento de interés.

En base a mi experiencia por el trabajo que realizo con las líneas telefónicas móviles he observado que varias de las covariables que influyen en la desactivación de la línea son cambiantes en el tiempo y presentan puntos de quiebre, por lo cual se considera apropiado realizar el estudio con el modelo propuesto por Jensen y Lutkebohmert (2008)

1.2. Objetivos

El objetivo general de la tesis es estudiar las propiedades y aplicar el modelo de regresión de Cox con puntos de cambio en las covariables (Jensen y Lutkebohmert (2008)) a una base de usuarios de líneas telefónicas móviles y así identificar las covariables que impactan en la cancelación de la línea móvil. Los objetivos específicos son:

- Revisar la literatura acerca de los conceptos previos necesarios para el estudio del modelo.
- Estudiar la teoría del modelo de regresión de Cox con puntos de cambio en las covariables.
- Implementar el modelo haciendo uso tanto del software R como de Mathematica según sea necesario.
- Realizar la aplicación del modelo a una base de clientes de líneas telefónicas móviles.

1.3. Organización del Trabajo

En el capítulo 2, se presentan conceptos previos al estudio del modelo, en el capítulo 3 se desarrolla el modelo de regresión de Cox con puntos de cambio en las covariables, en el capítulo 4 se realiza la aplicación del modelo, finalmente se discuten algunas conclusiones obtenidas durante el estudio en el capítulo 5. En los anexos se presentan algunas pruebas de los resultados teóricos que se requieren para la aplicación del modelo así también los algoritmos y los códigos utilizados para la implementación.

Capítulo 2

Conceptos preliminares

En el siguiente capítulo, se presentan conceptos previos al estudio del modelo, la siguiente teoría es necesaria para entender los pasos usados en el modelo de regresión de Cox con puntos de cambio en las covariables.

2.1. Análisis de supervivencia

Según Cox y Oakes (1984) el análisis de supervivencia centra el interés en el estudio del tiempo de ocurrencia de un evento de uno o más individuos, para lo cual se define un evento de interés y un punto de inicio del estudio, el grupo de individuos será evaluado durante este tiempo hasta que ocurra el evento esperado.

El tiempo puede ser medido en días, meses, años, etc.; es decir, desde el inicio del estudio del individuo, hasta que ocurra el evento. El evento será, el hecho de interés que se espera observar.

Existen diferentes campos de aplicación para el análisis de supervivencia como son el estudio del tiempo de vida de una máquina, el tiempo en que una persona permanece haciendo uso de un servicio determinado, el tiempo en que un estudiante permanece en la universidad hasta que decide abandonarla, etc.

Puede ocurrir que durante la investigación, el tiempo hasta que ocurre el evento sea desconocido, estos casos serán conocidos como datos censurados.

Se definen 3 tipos de censura. Censuramiento a la derecha o tipo I es cuando el tiempo de supervivencia es superior o igual al tiempo definido para el estudio, mientras que en el censuramiento a la izquierda o tipo II, el tiempo de ocurrencia real del evento es menor o igual al tiempo observado. Finalmente el censuramiento aleatorio o tipo III cada individuo entra en el estudio de manera aleatoria y maneja un modo propio de censura.

Algunas de las circunstancias por las cuales podemos tener censuramiento son las siguientes:

1. No ocurre el evento de interés en el objeto de observación y acaba el tiempo de estudio establecido.

2. Que el individuo abandone el estudio antes que este acabe.

3. Que un evento ajeno al evento de estudio ocurra y se pierda el individuo.

De acuerdo a esta descripción previa, se procederán a definir las funciones principales para la medición cuantitativa del estudio.

2.1.1. Descripción de funciones para el análisis de supervivencia y funciones de riesgo

Para realizar un análisis cuantitativo de los datos que disponemos, debemos llevar esta información a variables que nos ayudarán a construir el modelo, por lo que definiremos las notaciones y conceptos.

T es una variable aleatoria positiva que representa el tiempo hasta que ocurre el evento de interés y está definida en el intervalo $[0, \infty)$. La función de distribución acumulada de T está definida como:

$$F(t) = P(T \leq t),\ t \geq 0$$

Donde $F(t)$ es la probabilidad de que un individuo muera antes del tiempo t. Dado que es acumulada, se puede obtener la función de distribución del tiempo, para el caso de que T sea continua, será la derivada:

$$f(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T \leq t + \Delta t)}{\Delta t} = \frac{dF(t)}{dt}$$

A partir de lo definido, se usará $S(t)$ como la función de supervivencia del individuo, que es la probabilidad que un individuo sobreviva al tiempo t o más y se formula como sigue:

$$S(t) = P(T > t) = 1 - P(T \leq t) = 1 - F(t) = \int_t^\infty f(\mu)d\mu,\ t \geq 0$$

La función $S(t)$ es monótona decreciente; si un $\mu > t$ entonces $S(\mu) \leq S(t)$. Además $S(t) \leq 1$.

La función de riesgo $\lambda(t)$ es la tasa instantánea de falla u ocurrencia del evento esperado en el intervalo $[t, t + \Delta t]$, dado que el individuo está vivo hasta el tiempo t. De manera específica, $\lambda(t)$ es definida como:

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

A la vez puede ser expresada con las funciones descritas en lineas previas:

$$\lambda(t) = \frac{f(t)}{S(t)} = -\frac{dS(t)/dt}{S(t)} = -\frac{dlog(S(t))}{dt}$$

También definimos la función de riesgo acumulada $\Lambda(t)$ como:

$$\Lambda(t) = \int_0^t \lambda(\mu)d\mu = -logS(t)$$

De la expresión anterior, puede obtener la función de supervivencia, a partir de la función de riesgo o de la función de riesgo acumulada, mediante la siguiente fórmula:

$$S(t) = \exp\left\{\int_0^t \lambda(\mu)d\mu\right\} = \exp\{-\Lambda(t)\}$$

2.2. Procesos de conteo y Residuos de Martingala

Las martingalas y los procesos de conteo son herramientas básicas en el análisis matemático de supervivencia y otros procesos estocásticos. En Fleming y Harrington (1991) y Andersen et al. (1993) encontramos la teoría necesaria para entender el modelo de Jensen y Lutkebohmert (2008) que está basado en estos dos conceptos.

2.2.1. Variable Aletoria

Se asume que Ω es el espacio de probabilidad de un evento aleatorio, con cada resultado denotado por ω, F y P son el σ-algebra del evento y la medida de probabilidad de Ω respectivamente.

Una función real Z definida sobre Ω es llamada variable aleatoria si:

$$Z \leq x \equiv \{\omega \in \Omega : Z(\omega) \leq x \in F, \forall x\}$$

Esto es, Z es el mapeo medible de (Ω, F, P) de la recta real equipada con la σ-algebra de Borel B

2.2.2. Integral de Stieltjes

La integral de Stieltjes es una generalización de la integral de Riemann. Es una integral de una función según otra función; en la construcción de sumas para distintas particiones se emplea, en lugar de la amplitud de cada subintervalo, $(t_i - t_{i-1})$, la diferencia de valor de la función según la cual se integra entre los extremos del subintervalo $(N(t_i) - N(t_{i-1}))$. Cuando esta función es la identidad, la integral de Stieltjes es una integral de Riemann.

Casos particulares:

- Si la función es escalonada, la integral de Stieltjes es una suma.
- Si la función es diferenciable, la integral de Stieltjes se transforma en una integral de Riemann.
- Si la función es diferenciable a tramos, la integral de Stieltjes se puede calcular como una suma después de transformar cada tramo a una integral de Riemann.

Se cita el siguiente ejemplo:

Se define una partición P de un intervalo como una selección de puntos $T = \{t_1, t_2, ..., t_m\}$. La integral de Stieltjes se define como:

$$\theta(f, P, T, N) = \sum_{i=1}^{m} f(t_i)(N(t_i) - N(t_{i-1}))$$

La función f es integrable según N en $[a, b]$ y la integral es $\int_a^b f(t)dN(t)$ si para todo $\epsilon > 0$ existe una partición suficientemente fina, tal que para cualquier otra partición P más fina que ésta y cualquier elección de T se cumpla que:

$$|\theta(f, P, T, N) - \int_a^b f(t)dN(t)| < \epsilon$$

2.2.3. Procesos Estocásticos

En Andersen et al. (1993) se define a un proceso estocástico, como una familia de variables aleatorias $\{X = X(t) : t \in \Gamma\}$, indexadas por un conjunto Γ, todas definidos en el mismo espacio de probabilidad (Ω, F, P).

El conjunto Γ está asociado al tiempo tiempo y es usualmente discreto o continuo. En el presente estudio esto se restringirá a un conjunto discreto $\Gamma = \mathbb{N}$

Un proceso estocástico es:

- Integrable si, $\sup_{0\leq t<\infty} E|X(t)| < \infty$;
- Integrable al cuadrado si, $\sup_{0\leq t<\infty} E\left\{X(t)\right\}^2 < \infty$;
- Acotada si existe una constante finita γ de modo que $P\left\{\sup_{0\leq t<\infty} |X(t)| < \gamma\right\} = 1$

2.2.4. Procesos de Conteo

Los datos que cuentan el número de eventos de distintos tipos que se producen en el tiempo se pueden modelar con los procesos de conteo.

Un proceso puntual es una colección aleatoria enumerable de puntos en la recta real. Se considera a $\boldsymbol{N}(t)$ como el número de eventos de procesos puntuales que ocurren en el intervalo $[0, t]$ (Aalen (1978)).

Un proceso estocástico de conteo $\boldsymbol{N}(t) = (N_1(t), ..., N_n(t))$ definido en el intervalo $[0, t]$, será llamado proceso de conteo multivariado cuando se cumplan las siguientes condiciones:

- Las N_i son funciones continuas a la derecha con valor 0, en $t = 0$ y con un número finito de pasos, cada uno positivo y de tamaño 1.
- Dos componentes del proceso N_i y N_j $(i \neq j)$ no pueden saltar al mismo tiempo

En el proceso de conteo generalmente se observa el valor $N(t) - N(s)$, es decir que denota, el número de eventos de cierto tipo que se producen en el intervalo $]s, t]$, se puede mencionar como un ejemplo del proceso de conteo a la distribución de Poisson.

Si N es un proceso de conteo, f es una función (posiblemente aleatoria) de tiempo y $0 \leq s \leq t \leq \infty$, luego $\int_s^t f(\mu)dN(\mu)$ o de manera concisa $\int_s^t fdN$, es la representación con la integral de Stieltjes, de la suma de valores de f a los saltos de tiempo de N en el intervalo $]s, t]$

2.2.5. Residuos de Martingala

Sea $X_0, X_1, X_2, ...$ una cadena de Markov. La cadena será una martingala si para todo $n = 0, 1, 2, ...$, tenemos que $E(X_{n+1} - X_n|X_n) = 0$. Es decir, en promedio el valor de la cadena no varía, sin importar cual sea el valor de X_n en un momento dado.

Una de las ventajas que surgen del enfoque del análisis de supervivencia, es la posibilidad de efectuar el análisis de los residuos de martingala Fleming y Harrington (1991)

Los residuos son útiles para:

- Validar el supuesto de riesgo proporcional.

- Identificar los puntos o individuos de influencia.
- Identificar las covariables o sujetos que no están correctamente ajustados al modelo.
- Descubrir la forma funcional correcta de un predictor continuo.

Los residuos de martingala son muy asimétricos y con una cola muy larga hacia la derecha, particularmente para datos de supervivencia de un solo evento. Estos residuos son usados para estudiar la forma funcional de una variable.

El presenta trabajo hará uso de la martingala basado en un proceso de conteo para el *i-ésimo* individuo definido como:

$$M_i(t) = N_i(t) - \int_0^t R_i(s) \exp\left\{\beta' X_i(s)\right\} d\widehat{\Lambda}_0(\beta, s)$$

En donde $N_i(t)$ indica si ocurrió el evento del sujeto i en el tiempo t, $R_i(t)$ indica si el individuo i está en riesgo en el tiempo t, además $\widehat{\Lambda}_0(\beta, s)$ es el estimador de Breslow definido como:

$$\widehat{\Lambda}_0(\beta, t) = \int_0^t \frac{\sum_{i=1}^n dN_i(s)}{\sum_{i=1}^n R_i(s) \exp\left\{\beta' X_i(s)\right\}}$$

Fleming y Harrington (1991) muestran con mayor detalle la definición de esta martingala, en el apéndice B se presenta un resumen.

2.3. El modelo de regresión de Cox y sus características

El modelo de Cox para datos censurados Cox (1972) especifica que la función de riesgo como $\lambda(t) = lim_{\Delta t \to 0} P\left[T \leq t + \Delta t \,|T > t\right]$ con un tiempo potencial de falla T de un individuo con vector de covariables $\boldsymbol{X} = (X_1, ..., X_p)$ que dependen del tiempo tiene la forma:

$$\lambda(t|\boldsymbol{X}) = \lambda_0(t) \exp\left\{\boldsymbol{\beta}' \boldsymbol{X}(t)\right\} \;;\; t \geq 0 \tag{2.1}$$

Aquí $\boldsymbol{\beta}$ es un vector de dimensión p; $\lambda_0(t)$ es la función de riesgo base desconocida. El problema estadístico es el de estimar $\boldsymbol{\beta}$ y la función λ_0 respecto a una muestra de tamaño n, con posibles tiempos de supervivencia con censura a la derecha $t_1, ..., t_n$ y las covariables correspondientes $X_p = \{x_{pi}, ..., x_{pn}\}$ con p igual a la cantidad de covariables del sujeto.

El modelo de Cox tiene la propiedad de que si todas las covariables son iguales a 0, la fórmula se reduce a la función de riesgo base, por lo que esta función puede considerarse como punto de inicio, la cual no es especificada, por lo cual, el modelo de regresión de Cox, se considera semiparamétrico.

Cox (1975) propuso usar la verosimilitud parcial para estimar $\boldsymbol{\beta}$

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n \left\{ \frac{\exp\left\{\boldsymbol{\beta}' X_i(T_i)\right\}}{\sum_{j \in R_i} \exp\left\{\boldsymbol{\beta}' X_j(T_i)\right\}} \right\}^{\delta_i} \tag{2.2}$$

Donde $R_i = \{j : T_j \geq T_i\}$ y δ_i es el indicador de ocurrencia del evento. En el artículo Cox (1975) derivó la ecuación (2.2) como una función de verosimilitud parcial. Denotando a $\widehat{\boldsymbol{\beta}}$

como el valor que maximiza (2.2), entonces el estimador continuo obtenido por interpolación lineal para la función de riesgo es:

$$\widehat{\Lambda}(t) = \sum_{T_i \leq t} \frac{\delta_i}{\sum_{j \in R_i} \exp\left\{\widehat{\boldsymbol{\beta}}' X_j(T_i)\right\}} \tag{2.3}$$

para la subyacente función acumulada de riesgo $\Lambda_0 = \int_0^{'} \lambda_0(s)ds$ sugerida por Breslow (1974).

2.3.1. Formulación del modelo de Cox por proceso de conteo y sus propiedades

En Andersen y Gill (1982) se formula la regresión de Cox en el marco de procesos de conteo multivariado y se bosqueja como probar las propiedades asintóticas de $\widehat{\boldsymbol{\beta}}$ y $\widehat{\Lambda}$.

Debido al interés en las propiedades asintóticas, se considerará una secuencia de modelos indexados para $n = 1, 2....$ Se generalizará la posibilidad de observaciones censuradas del tiempo de vida de n individuos a la observación (en el enésimo modelo) de un proceso de conteo multivariado de n-componentes $\boldsymbol{N} = (N_1, ..., N_n)$, donde N_i cuenta eventos observados en la vida del *i-ésimo* individuo, $i = 1, ..., n$, sobre un intervalo de tiempo $[0, t]$. Así la trayectoria de las muestras de $N_1, ..., N_n$ son funciones escalonadas, cero en el tiempo cero con saltos de tamaño $+1$.

La asunción básica es que para cada n, N tiene un proceso aleatorio de intensidad $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_n)$ de tal manera que:

$$\lambda_i(t) = R_i(t)\lambda_0(t) \exp\left\{\boldsymbol{\beta}_0' \boldsymbol{X}_i(t)\right\} \tag{2.4}$$

Aquí $\boldsymbol{\beta}_0$ es un vector con columna fija de p componentes, λ_0 es una función de riesgo fijo subyacente y $R_i(t)$ es un proceso predictivo tomando valores en $\{0, 1\}$ en donde el valor de 1 indica que el *i-ésimo* individuo está bajo observación. De tal manera que, N_i solo salta cuando $R_i(t)$ es igual a 1. Finalmente $\boldsymbol{X}_i = (X_{i1}, ..., X_{ip})$ es el vector de covariables de tamaño p para el *i-ésimo* individuo.

2.3.2. Modelo de regresión de Cox con puntos de cambio

Los modelos con puntos de cambio, fueron desarrollados, basándose en observaciones en donde un valor está fuera de los intervalos regulares. De manera general se puede describir a los puntos de cambio como un proceso aleatorio indexado durante el tiempo que se observa y se quiere investigar si se produce un cambio en la distribución de los elementos aleatorios. Como se mencionó en la sección 1.1, los puntos de cambio fueron introducidos en el modelo de Cox primero con K. Liang y Liu (1990) en el modelo:

$$\lambda(t) = \lambda_0(t) \exp\left\{(\beta + \theta I_{\{t \leq \tau\}})X + \boldsymbol{\gamma}' Z\right\}$$

Donde X y Z son las covariables unidimensionales del modelo y según el valor de τ hará que la variable de punto de cambio θ intervenga en el modelo.

2.4. Consistencia del estimador

La consistencia es una propiedad de los estimadores, se dice que éste es consistente cuando converge al valor verdadero, mientras más se acerque el tamaño de la muestra al tamaño real de los datos, este estimador se acerca al valor verdadero (θ). El concepto está relacionado con la definición del sesgo de los estimadores. Un estimador puede presentar cierto sesgo, pero si es consistente dicho sesgo decrece conforme crece el tamaño muestral.

Se puede denotar lo anterior de la siguiente manera:

$$\lim_{n\to\infty}\left[\left|\widehat{\theta}_n - \theta\right| > \epsilon\right] = 0 \ ; \ \epsilon > 0$$

Se llamará un estimador $\widehat{\theta}_n$, $\sqrt{n}$-consistente, si la secuencia $\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right)$ se ajusta uniformemente. La interpretación es que $\widehat{\theta}_n$ determina el valor de θ_0 dentro de un rango $n^{-1/2}$.

2.5. Normalidad asintótica

Según Lehmann y Casella (1998) muchas veces es difícil evaluar cuan próximos están los estimadores obtenidos respecto de los valores verdaderos. Esta dificultad se puede superar en su mayoría por el empleo de las computadoras. Sin embargo, esto no es suficiente para lograr la robustez y eficiencia del estimador.

Una distribución asintótica es una distribución hipotética que sirve como distribución límite de una sucesión de distribuciones. Una de las principales ideas de una distribución asintótica es la de proveer aproximaciones a las funciones de distribución acumuladas de los estimadores estadísticos. Una sucesión de distribuciones corresponde a una sucesión de variables aleatorias X_i para $i = 1, 2, ...$ en el caso mas simple. Existe una distribución asintótica, si la distribución de probabilidad X_i, converge a una distribución de probabilidad cuando i crece. Un caso especial de una distribución asintótica es cuando una sucesión de variables aleatorias siempre converge a 0; esto es, el X_i va a 0, cuando i va a infinito. Aquí la distribución asintótica es una distribución degenerada correspondiente al valor cero.

Si una distribución asintótica existe, no es necesariamente cierto que un resultado de una secuencia de variables aleatorias es una sucesión convergente de números. Es la sucesión de distribuciones de probabilidad la que converge. Una de las distribuciones más comunes que surge como distribución asintótica es la distribución normal. En particular el teorema central del límite provee un ejemplo donde la distribución asintótica es la distribución normal.

Cuando existe una distribución asintótica, la sucesión de distribuciones de probabilidad de la variable aleatoria X_i, es la que converge.

2.6. *Op* y *op*

La notación de orden en probabilidad se ocupa de convergencia de conjuntos de variables aleatorias, donde la convergencia es en el sentido de la convergencia en probabilidad.

Para un conjunto de variables aleatorias X_n y un conjunto de constantes a_n la notación $X_n = op(a_n)$, significa que un grupo de valores X_n/a_n converge a cero en probabilidad, cuando n se acerque a un límite apropiado. De manera equivalente, $X_n = op(a_n)$ puede ser escrito como $X_n/a_n = op(1)$ donde $X_n = op(1)$ se define como:

$$\lim_{n\to\infty} P\left(|\boldsymbol{X}_n| \geq \epsilon\right) = 0 \quad \forall\, \epsilon \geq 0$$

En el caso de la notación $X_n = Op(a_n)$ significa que un grupo de valores X_n/a_n está limitada estocásticamente, esto es, para cualquier $\epsilon > 0$, existe un $M > 0$ de modo que:

$$P\left(|\boldsymbol{X}_n/a_n| > M\right) < \epsilon, \quad \forall n$$

Capítulo 3

Modelo de regresión de Cox con puntos de cambio en las covariables

Una variación del modelo de regresión de Cox, se presenta al utilizar puntos de cambio en las covariables para describir y ajustar mejor el modelo a un conjunto de datos, este cambio se da en un umbral desconocido del tiempo y se observa que la influencia de la covariable cambia suavemente, el modelo de estudio propuesto por Jensen y Lutkebohmert (2008) acepta que en una covariable haya más de un punto de cambio y que la regresión tenga más de una covariable con puntos de cambio.

Al final del estudio del modelo, se desarrollará una aplicación con datos reales de supervivencia de líneas telefónicas móviles.

3.1. Introducción

Se considerará un proceso de conteo multivariado $\boldsymbol{N}(t) = (N_1(t), ..., N_n(t))$, donde $N_i(t)$ es la cantidad de eventos observados en la vida del *i-ésimo* individuo, $i = 1, ..., n$, para un tiempo t sobre el intervalo $[0, \tau]$ las trayectorias de las muestras de $\boldsymbol{N}(t)$ son funciones escalonadas, cero en el tiempo cero con salto de tamaño uno. El proceso de conteo $\boldsymbol{N}(t)$ admite una función de riesgo $\lambda(t) = (\lambda_1(t), ..., \lambda_n(t))$ de tal manera que el proceso $M_i(t) = N_i(t) - \int_0^t \lambda_i(\mu)d\mu, i = 1, ..., n$ y $t \in [0, \tau]$ son martingalas (2.2). La tasa de riesgo del modelo de Cox con riesgo base $\lambda_0(t)$ y el vector de covariables $\boldsymbol{X}(t)$, está dado por $\lambda(t) = \lambda_0(t) \exp\left\{\boldsymbol{\beta}_0' \boldsymbol{X}(t)\right\}$. En este modelo se está asumiendo que la influencia de la covariable es constante en el tiempo sobre un rango de covariables. Gandy et al. (2005) construye un modelo que se basó en el análisis que se realizó a un conjunto de datos actuariales, es ahí donde identificó que no todas las covariables presentan un comportamiento constante, proponiendo así una variante del modelo de Cox, agregando un umbral desconocido $\boldsymbol{\omega}$ como parámetro del modelo.

Ahora en el modelo propuesto por Jensen y Lutkebohmert (2008) se permite más de un punto de cambio, que las covariables sean dependientes del tiempo y el proceso de conteo pueda saltar mas de una vez; esto es, el evento en observación pueda ocurrir más de una vez en el sujeto. El modelo involucra m puntos de cambio y p covariables ordinarias (sin puntos de cambio) esta dado como sigue:

$$\lambda_i(t, \boldsymbol{\theta}) = \lambda_0(t) R_i(t) exp\left\{\boldsymbol{\beta}_1' \boldsymbol{X}_{1i}(t) + \boldsymbol{\beta}_2' \boldsymbol{X}_{2i}(t) + \boldsymbol{\beta}_3' \left(\boldsymbol{X}_{2i}(t) - \boldsymbol{\omega}\right)\right\}$$

Donde $\boldsymbol{\theta} = (\boldsymbol{\omega}', \boldsymbol{\beta}')$ con $\boldsymbol{\omega} \in \Omega \subset \mathbb{R}^m$ y $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \boldsymbol{\beta}_3'\right)' \in B \subset \mathbb{R}^{p+2m}$

Aquí $\boldsymbol{\omega}$ y $\boldsymbol{\beta}$ son vectores que corresponden a los puntos de cambio y parámetros de la regresión respectivamente, $\lambda_0(t)$ es la intensidad base y $R_i(t)$ indica si el sujeto esta en riesgo. Para brevedad se considerará de la siguiente manera:

$$\lambda_i(t, \boldsymbol{\theta}) = \lambda_0(t) R_i(t) exp\left\{\boldsymbol{\beta}' \widetilde{\boldsymbol{X}}_i(t; \boldsymbol{\omega})\right\} \tag{3.1}$$

Donde:

$$\widetilde{\boldsymbol{X}}_i(t; \boldsymbol{\omega}) = (\boldsymbol{X}_{1i}'(t), \boldsymbol{X}_{2i}'(t), ((\boldsymbol{X}_{2i}(t) - \boldsymbol{\omega}))')'$$

El objetivo principal de este trabajo es estudiar el modelo propuesto por Jensen y Lutkebohmert (2008) y a través de una aplicación real estimar al vector de parámetros $\boldsymbol{\theta}_0$ el cual denotaremos con $\widehat{\boldsymbol{\theta}}_n$, que maximiza al logaritmo de la verosimilitud parcial. Jensen y Lutkebohmert (2008) han demostrado que las estimaciones de los puntos de cambio son $\sqrt{n}$-consistentes. A la vez , la función acumulada de riesgo $\Lambda_0(t) = \int_0^t \lambda_0(u)du$ se calculará aproximadamente a través de Breslow $\widehat{\Lambda}_0(t)$.

En contraste a Pons (2003) y Kosorok y Song (2007) el modelo considera una función de regresión que es continua en el vector de puntos de cambio $\boldsymbol{\omega}$, por lo tanto estos puntos de cambio son caracterizados como puntos de cambio suaves, modelos de este tipo son algunas veces llamados modelos *bentline* con puntos de cambio en las covariables. Para un determinado número de aplicaciones los cambios suaves son mas apropiados que los saltos (Chappell (1989)).

3.2. Presentación y estimación de parámetros del modelo

Sea $[0, \tau], 0 < \tau < \infty$, un intervalo de tiempo fijo en donde todos los procesos estocásticos están definidos. Se asume que el proceso de conteo $\boldsymbol{N}(t) = (N_i(t), i = 1, ...n)$ es $n - dimensional$ y tiene elementos independientes. De manera precisa, asumimos que $(N, \boldsymbol{X}_1, \boldsymbol{X}_2)$, $(N_i, \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i})$ con $i = 1, ..., n$ son vectores independientes e idénticamente distribuidos de valores aleatorios, donde $\boldsymbol{X}$ es un proceso continuo. Además $\boldsymbol{M}(t) = \boldsymbol{N}(t) - \int_0^t \lambda(s)ds$, es un vector de martingala en el intervalo de tiempo $[0, \tau]$ donde los componentes de λ son definidos como en la ecuación (3.1).

El vector de puntos de cambio $\boldsymbol{\omega}$ se asume que se encuentra en un conjunto $\Omega = [\omega_{11}, \omega_{21}] \times [\omega_{12}, \omega_{22}] \times, ..., \times [\omega_{1m}, \omega_{2m}]$. La asunción que los valores $\omega_{11}, \omega_{21}, \omega_{12}, \omega_{22}, ..., \omega_{1m}, \omega_{2m}$, son conocidos no es una buena restricción para aplicaciones reales. El parámetro subyacente $\boldsymbol{\theta}_0$ para el tipo de modelo de Cox puede ser estimado por el valor de $\widehat{\boldsymbol{\theta}}_n$ que maximiza el algoritmo de la verosimilitud parcial.

Los pasos que se siguen para la estimación del parámetro $\widehat{\boldsymbol{\theta}}_n$, consisten primero definir al modelo que se va a usar, como se observa en la introducción en la ecuación (3.1) y los conceptos previos ecuación (2.4) y de la ecuación (2.2) se formulará el modelo de regresión de Cox con el proceso de conteo:

$$logL(\boldsymbol{\theta}) = \sum_{i=1}^{n} \int_0^{\tau} \boldsymbol{\beta}' \widetilde{\boldsymbol{X}}_i(t;\boldsymbol{\omega}) d\boldsymbol{N}_i(t) - \sum_{i=1}^{n} \int_0^{\tau} log\left(\sum_{j=1}^{n} exp\left(\boldsymbol{\beta}' \widetilde{\boldsymbol{X}}_j(t;\boldsymbol{\omega}) \right) \right) d\boldsymbol{N}_i(t) \quad (3.2)$$

La maximización se llevará a cabo en dos fases: Para un valor fijo de $\boldsymbol{\omega}$, se estima $\widehat{\boldsymbol{\beta}}_n(\boldsymbol{\omega}) = argmax_{\boldsymbol{\beta}\in B} \log L(\boldsymbol{\omega},\boldsymbol{\beta})$, luego se define $\log L(\boldsymbol{\omega},\boldsymbol{\beta}) = \log L(\boldsymbol{\omega}, \widehat{\boldsymbol{\beta}}(\boldsymbol{\omega}))$. Entonces $\boldsymbol{\omega}_0$ puede ser estimado por $\widehat{\boldsymbol{\omega}}_n$ tal que satisfaga:

$$\widehat{\boldsymbol{\omega}}_n = \arg_{\boldsymbol{\omega}\in\Omega} \max \log L(\boldsymbol{\omega}) \quad (3.3)$$

El estimador de máxima verosimilitud $\boldsymbol{\theta}_0$ es $\widehat{\boldsymbol{\theta}}_n = (\widehat{\boldsymbol{\omega}}_n, \widehat{\boldsymbol{\beta}}_n)$. En Jensen y Lutkebohmert (2008) se considera el límite de $\log L(\boldsymbol{\theta})$ cuando $n \to \infty$. Pero $\log L(\boldsymbol{\theta})$ no converge a un valor finito cuando $n \to \infty$, contempla la siguiente transformación equivalente a $\log L(\boldsymbol{\theta})$:

$$\begin{aligned} Z_n(\boldsymbol{\theta}) : &= \frac{1}{n}\left(\log L(\boldsymbol{\theta}) + (\log n) \sum_{i=1}^{n} N_i(\tau) \right) \\ &= \frac{1}{n} \sum_{i=1}^{n} \int_0^{\tau} \boldsymbol{\beta}' \widetilde{\boldsymbol{X}}_i(t;\boldsymbol{\omega}) dN_i(t) - \int_0^{\tau} log\left\{ \frac{1}{n} \sum_{i=1}^{n} R_i(t) \exp\left(\boldsymbol{\beta}' \widetilde{\boldsymbol{X}}_i(t;\boldsymbol{\omega}) \right) \right\} d\overline{N}(t) \end{aligned}$$

Donde $\overline{N}(t) = \frac{1}{n}\sum_{i=1}^{n} N_i(t)$

El estimador $\widehat{\boldsymbol{\theta}}_n$ no solo maximiza $\log L(\boldsymbol{\theta})$ sino también $Z_n(\boldsymbol{\theta})$. Entonces $Z_n(\boldsymbol{\theta}) \to z(\boldsymbol{\theta})$, cuando $n \to \infty$, es decir:

$$Z_n(\boldsymbol{\theta}) \to z(\boldsymbol{\theta}) = E\left[\int_{\tau}^{0} \boldsymbol{\beta}' \widetilde{\boldsymbol{X}}(t;\boldsymbol{\omega}) \lambda(t,\boldsymbol{\theta}_0) dt \right] - E\left[\int_0^{\tau} \log(s(t;\boldsymbol{\theta})) \lambda(t,\boldsymbol{\theta}_0) dt \right]$$

Donde $s(t;\boldsymbol{\theta}) = E\left[R(t) \exp(\boldsymbol{\beta}' \widetilde{\boldsymbol{X}}(t;\omega)) \right]$. Además la función de riesgo acumulada $\Lambda_0(t) = \int_0^t \lambda_0(\mu) d\mu$, es estimada por el estimador de Breslow 2.2:

Las condiciones necesarias para establecer las propiedades asintóticas de los estimadores se muestran en el apéndice B.2.

3.3. Consistencia del estimador

Jensen y Lutkebohmert (2008) extienden los argumentos usados en Gandy y Jensen (2005) para demostrar la consistencia del estimador $\hat{\boldsymbol{\theta}}_n$ para el caso bajo consideración. La prueba está basada en la convergencia uniforme de Z_n a z y en las propiedades de z en una vecindad de $\boldsymbol{\theta}_0$. Los lemas y teoremas que se requieren para demostrar la consistencia, se muestran en el apéndice B.3.

3.4. Tasa de convergencia

La diferencia entre un salto y un punto de cambio suave es la continuidad de $Z_n(\boldsymbol{\theta})$ en $\boldsymbol{\omega}$ en el modelo de punto de cambio suave.

La continuidad causa que el límite de $Z_n(\boldsymbol{\theta})$ sea diferenciable en ω. Jensen y Lutkebohmert (2008) demuestran que el ratio de convergencia del estimador del punto de cambio de su

modelo difiere del estimador encontrado por Kosorok y Song (2007).

Sea $V_{\in}(\boldsymbol{\theta}_0) = \{\boldsymbol{\theta} \in \boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| < \epsilon\}$ una vecindad de $\boldsymbol{\theta}_0$ y sea W_n el siguiente proceso

$$W_n(\boldsymbol{\theta}) = \sqrt{n}(Z_n(\boldsymbol{\theta}) - z(\boldsymbol{\theta}))$$

haciendo uso de los resultados teóricos utilizados por Jensen y Lutkebohmert que se muestran en el apéndice B.4, se tiene que:

$$P\left[\hat{\boldsymbol{\theta}}_n \in V_{\in}(\boldsymbol{\theta}_0)\right] > 1 - \eta$$

para un n suficientemente grande y para algún $\eta > 0$; o lo que sería lo mismo

$$\sqrt{n}\left\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right\| = Op(1)$$

3.5. Normalidad asintótica

Generalmente los métodos estándar en ete caso fallan desde que se asume la diferenciabilidad de la función de verosimilitud parcial con respecto a sus parámetros.

Jensen y Lutkebohmert (2008) demuestran la normalidad asintótica de los estimadores, a partir del hecho de que $\hat{\boldsymbol{\theta}}_n$ es un estimador consistente de $\boldsymbol{\theta}_0$; es decir:

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} N\left(0, \sigma_2\left(\boldsymbol{\theta}_0\right)\right)$$

Por otro lado, haciendo uso de la aproximación de Andersen y Gill (1982), demuestran la convergencia débil de $\sqrt{n}(\widehat{\Lambda}_n(t) - \Lambda_0(t))$ a un proceso Gaussiano.

Las demostraciones de los resultados teóricos requeridos se presentan en el apéndice B.5.

Capítulo 4

Aplicación del modelo de regresión de Cox con puntos de cambio en las covariables

En la actualidad el mercado peruano de telefonía móvil se encuentra cubierto en su totalidad, podemos encontrar incluso a una persona con más de un celular en su poder. Debido a las leyes del consumidor los usuarios pueden cancelar el contrato que los unía con la empresa o cambiar su línea, conservando su número, a una operadora de la competencia en cualquier momento. Es por eso, que las empresas de telecomunicaciones están preocupadas en mantener a sus clientes actuales, identificando sus necesidades y atendiendo a sus reclamos de manera más eficiente. El modelo de regresión de Cox clásico permite asociar el estado de cancelación de la línea con covariables que se cree influencian en la baja, pero el modelo propuesto por Jensen y Lutkebohmert va más allá del modelo clásico estudiando covariables que cambian drásticamente en el tiempo definiendo un punto de cambio; así, si antes los parámetros de la regresión eran constantes en el tiempo para el modelo clásico, en este caso no, lo que nos permitirá ampliar los tipos de covariables de estudio y ajustar mejor el modelo a los datos.

4.1. Introducción al problema y descripción de los datos de estudio

A partir de Enero de 2010 los usuarios de teléfonos móviles tienen la opción de poder cambiar de operadora conservando su número telefónico, esto ha generado grandes expectativas en las empresas como preocupación, pues de acuerdo a las estadísticas disponibles en la página oficial de OSIPTEL, el mercado está casi al 100 % de ser ocupado. En el cuadro 4.1 se muestra la cantidad de líneas por departamento de las diferentes operadoras, se observa que a diciembre de 2013 existen 29'953,933 de usuarios de teléfonos móviles que están conformadas por líneas prepago y postpago. INEI a Diciembre 2013 estimó una población de 30'644,128 por lo cual la inserción de telefonía móvil a nivel nacional representa el 97.75 % a esa fecha.

Si bien en un comienzo la estrategia de las operadoras era captar clientes, ahora esta ha cambiado, ahora se busca conservarlos o captar clientes de la competencia a través de la portación de líneas. Es por esto que se han creado ofertas y nuevos servicios de acuerdo al avance de la tecnología en equipos y de la exclusividad de estos en las empresas. Estos cambios en tecnología y el libre mercado administrados por un ente regulador como OSIPTEL, juegan un papel importante en la decisión del usuario móvil, que ahora tiene más opciones de operadoras móviles para decidir si continuar con la línea activa o darle de baja. Actualmente la oferta comercial de las diferentes empresas que operan en Perú consta de los siguientes

	Jul-13	Ago-13	Sep-13	Oct-13	Nov-13	Dic-13
Amazonas	201,452	198,349	198,633	197,388	195,156	208,308
Ancash	836,647	838,298	835,155	826,864	820,995	837,312
Apurímac	297,741	292,744	294,251	291,246	288,815	286,771
Arequipa	1,336,436	1,351,436	1,344,259	1,326,244	1,347,617	1,365,147
Ayacucho	470,320	472,416	471,556	467,149	466,386	456,900
Cajamarca	897,399	891,866	891,975	877,833	873,632	872,576
Callao	126,325	292,198	119,979	152,255	107,156	113,472
Cusco	948,810	947,473	952,537	940,818	936,663	973,631
Huancavelica	222,603	208,072	215,156	215,041	202,818	205,513
Huánuco	491,829	492,413	485,811	482,625	482,253	493,410
Ica	686,667	698,240	682,609	684,454	688,262	723,105
Junín	961,727	955,078	930,458	922,574	912,406	945,650
La Libertad	1,380,194	1,400,378	1,396,586	1,393,066	1,402,930	1,435,850
Lambayeque	881,959	890,422	881,865	874,095	878,238	921,677
Lima	9,871,994	9,763,319	10,037,937	9,873,237	9,961,596	10,026,096
Loreto	349,814	353,224	354,405	353,645	355,242	367,603
Madre de Dios	160,641	161,338	157,187	154,342	158,482	153,884
Moquegua	169,648	165,695	168,610	166,722	166,172	173,495
Pasco	187,102	182,372	182,128	182,287	175,421	174,647
Piura	1,138,455	1,138,278	1,145,701	1,153,342	1,136,448	1,179,317
Puno	983,080	977,572	985,834	991,834	970,106	994,682
San Martín	472,666	470,663	472,371	464,258	464,032	478,669
Tacna	310,787	312,443	311,196	304,807	311,700	316,136
Tumbes	167,018	164,943	163,666	164,822	163,088	177,042
Ucayali	293,011	290,439	293,472	291,455	294,177	300,376
SIN LAC	5,293,781	5,295,605	5,357,127	5,670,677	5,733,958	5,772,664
Total Perú	**29,138,106**	**29,205,274**	**29,330,464**	**29,423,080**	**29,493,749**	**29,953,933**

Cuadro 4.1: Líneas móviles a nivel nacional

productos:

- Planes Prepago: Aquellos planes sin un cargo fijo o recibo que paguen de manera mensual, su consumo se basa en recargas.
- Planes Postpago: Planes que tienen un cargo fijo mensual y una cantidad de minutos de voz, cantidad de megas y mensajes a utilizar que dependen de la cantidad que se paga de manera mensual. El mercado esta segmentado como lo muestra la figura 4.1, el 28 % de las líneas son postpago, a este segmento es el que se están orientando las empresas de telecomunicaciones y es ahí donde trabajaremos la aplicación del modelo en estudio.

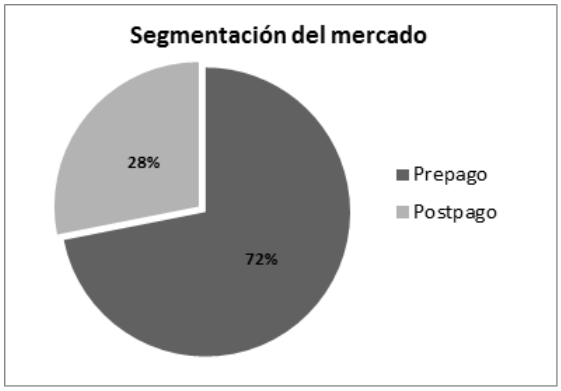


Figura 4.1: Segmentación del mercado Móvil

En el segmento postpago existen clientes Masivos (personas naturales con línea móvil) y clientes corporativos (Empresas). Dentro del segmento Postpago Masivo tenemos los siguientes segmentos:

- Postpago Voz: Líneas móviles sin servicio adicional de datos, solo comunicación por voz y sms.
- Smartphone: Líneas móviles que aparte de comunicación por voz tienen disponible el servicio de datos o internet móvil.
- BAM: Líneas que solo tienen el servicio de internet y solo se ofrecen con equipos modem.

Diariamente se realizan ventas de estos productos por diferentes canales o centros de ventas, para enfocar nuestro punto de estudio y que todos tengan las mismas condiciones, es que usaremos una parte de estas ventas. La base a trabajar para el estudio del modelo, consta de 7,862 registros de altas de Julio del 2011 vendidas en los puntos de venta de Cadenas ubicadas en Lima y solo el segmento Smartphone postpago. Se hizo el seguimiento en un periodo 25 meses para ver el comportamiento de pago, el tráfico de voz y de datos realizados de estas líneas. Dentro de las características que poseen estas líneas telefónicas, podemos encontrar los datos personales del usuario, como los datos propios de la línea adquirida, las características con las que se cuentan para el estudio son las siguientes:

- Línea: Línea móvil en observación.
- Tiempo: Mes de evaluación de la línea.
- Edad: Edad del cliente que adquirió la línea en el tiempo observado.
- Monto Facturado: Monto facturado en el mes.
- Tráfico de Voz: Minutos consumidos en el mes por el cliente.
- Tráfico de datos: Cantidad de tráfico de internet, medido en Mb, consumido por el cliente en el mes.

- Estado: El estado de la línea en el mes de evaluación, puede estar activa o desactiva.

De acuerdo a las características disponibles de la línea Smartphone postpago, utilizaremos el modelo propuesto por Jensen y Lutkebohmert para identificar cuáles de estás influyen en la decisión de dar de baja o desactivar la línea a **solicitud del cliente** (pedido del cliente), por **migración de postpago a prepago** (cambia su línea postpago a prepago) o por **portación a otra operadora**. Así podremos identificar los tipos de clientes que toman esa decisión y la empresa podrá desarrollar estrategias para conservar a los usuarios.

4.2. Análisis descriptivo de los datos

La base consta de líneas vendidas y activadas en Julio del 2011 para los puntos de ventas de Cadenas (puntos de venta que se encuentran en tiendas como Saga, Ripley y similares) de la región Lima, el segmento a estudiar es Smartphone postpago y el periodo de observación fue de 25 meses. Durante este tiempo para cada baja se observó el estado de las covariables a investigar y resumimos de manera descriptiva el comportamiento de estas variables.

Datos	Cantidades	Porcentaje
Estado de la línea		
Activa	3,613	46.0 %
Desactivas	4,249	54.0 %
Detalle de motivo de desactivación		
A solicitud Cliente	474	6.0 %
Desactivacion Port OUT	27	0.3 %
F02	1	0.0 %
F03	14	0.2 %
F07	194	2.5 %
Migración de Postpago a Prepago BAM	8	0.1 %
Migración Postpago a Prepago	1,154	14.7 %
Moroso	2,363	30.1 %
Transferencia	14	0.2 %
Sexo		
Hombre	4,214	53.6 %
Mujer	3,648	46.4 %

Cuadro 4.2: Descripción de la población

Al inicio del estudio, se trabajó con la base completa de la población seleccionada, sin embargo durante el proceso de implementación de los algoritmos para el modelo, se observó que requería mucho tiempo de procesamiento, trabajando con un procesador Core i5 de 4 Gb de memoria por lo cual se decidió tomar una muestra de la base que conservara las mismas características de la base original, se optó por usar el método de muestreo estratificado Lohr (1999), dado que los planes vendidos forman segmentos debido al cargo fijo contratado, se usó un nivel de confianza al 95 % y un error de estimación de 0,25 en el cargo fijo, según el procedimiento se obtuvo el siguiente cuadro comparativo de la muestra versus los datos de la población:

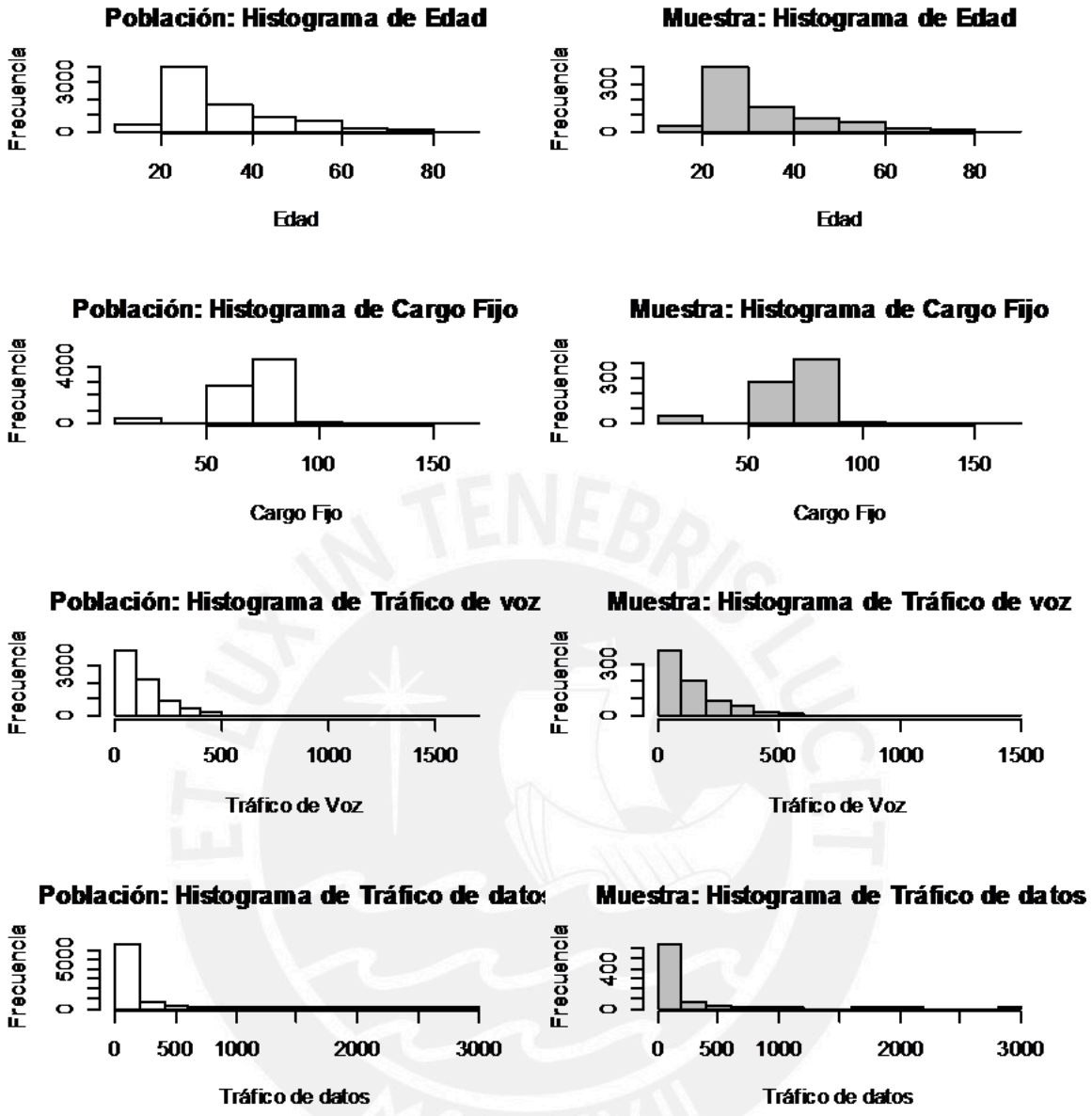


Figura 4.2: Histograma de las variables *edad*, *monto facturado*, *trafico de datos* y *minutos* para la muestra y población

La muestra seleccionada, según muestreo estratificado, de manera aleatoria consta de 755 observaciones, esto representa el 9,8 % de la base real y las características son las siguientes:

Según los 755 sujetos seleccionados se observan sus características iniciales; esto es, cuando se dieron de alta o se activaron, para lo cual podemos observar de acuerdo histograma y la tabla de datos descriptivos que la mayor cantidad de usuarios tiene una edad entre los 20 y 30 años, esto representa el 41 % del total de sujetos observados, en el caso del monto que se facturará y que pasará en observación durante 25 meses, se observa que el monto con más cantidad de líneas está en 83 soles, para el tráfico de voz que se mide en minutos, se tiene que las cantidades de minutos consumidos se encuentra entre 0 y 100, por último para el tráfico de datos se encuentra entre 0 y 80 Mb.

Datos	Cantidades	Porcentaje
Estado de la línea		
Activa	347	46.2 %
Desactivadas	408	53.8 %
Detalle de motivo de desactivación		
A solicitud Cliente	33	4.3 %
Desactivacion Port OUT	3	0.4 %
F03	1	0.1 %
F07	21	2.8 %
Migración de Postpago a Prepago BAM	1	0.1 %
Migración Postpago a Prepago	122	16.1 %
Moroso	227	29.9 %
Sexo		
Hombre	396	52.6 %
Mujer	359	47.4 %

Cuadro 4.3: Descripción de la muestra

De acuerdo a la muestra seleccionada se tiene mayor detalle de los estadísticos en el cuadro 4.4

	Edad	Monto Facturado	Tráfico de Voz	Tráfico de datos
n	759.00	759.00	759.00	759.00
Media	32.46	69.79	156.90	87.87
Error típico	0.72	6.08	7.92	7.92
Mediana	28.00	83.19	105	1.03
Moda	23.00	83.19	0.00	0.00
sd	11.64	19.73	167.55	218.09
Var	135.54	389.19	2,8074.72	47,565.98
Curtosis	1.23	2.05	13.00	60.29
Asimetria	1.34	-1.40	2.87	6.35
Rango	58.00	95.79	1442.00	2,965.00
Min	18.00	12.61	0.00	0.00
Max	76.00	108.40	1442.00	2,965.00

Cuadro 4.4: Estadísticas descriptivas de las covariables de la muestra al inicio del estudio

4.3. Resultados de la aplicación del modelo

De acuerdo a los algoritmos desarrollados para la implementación del modelo en Mathematica y la identificación de los residuales en R, se tiene que seguir una secuencia; en cada uno de estos pasos, se obtienen resultados que luego serán usados para el paso siguiente. Según esto, se empieza desde la identificación del punto de cambio hasta la obtención del modelo propuesto. A continuación se mostrarán los resultados obtenidos en cada uno de ellos.

4.3.1. Identificación del punto de cambio

Se dispone de la información histórica de las diferentes covariables descritas, el desarrollo e implementación del modelo de regresión con punto de cambio para la aplicación a estudiar trabaja con dos covariables una de ellas será la que contiene el punto de cambio (Algoritmo **??**). Se dispone de las variables *edad*, *monto_recibo*, *mb_datos* y *min_voz*, para las cuales se realizó un estudio previo para identificar cuál de ellas es la que tenía el punto de cambio, en la figura 4.3 se observan las corridas de las variables para visualizar donde es que se da el

cambio.

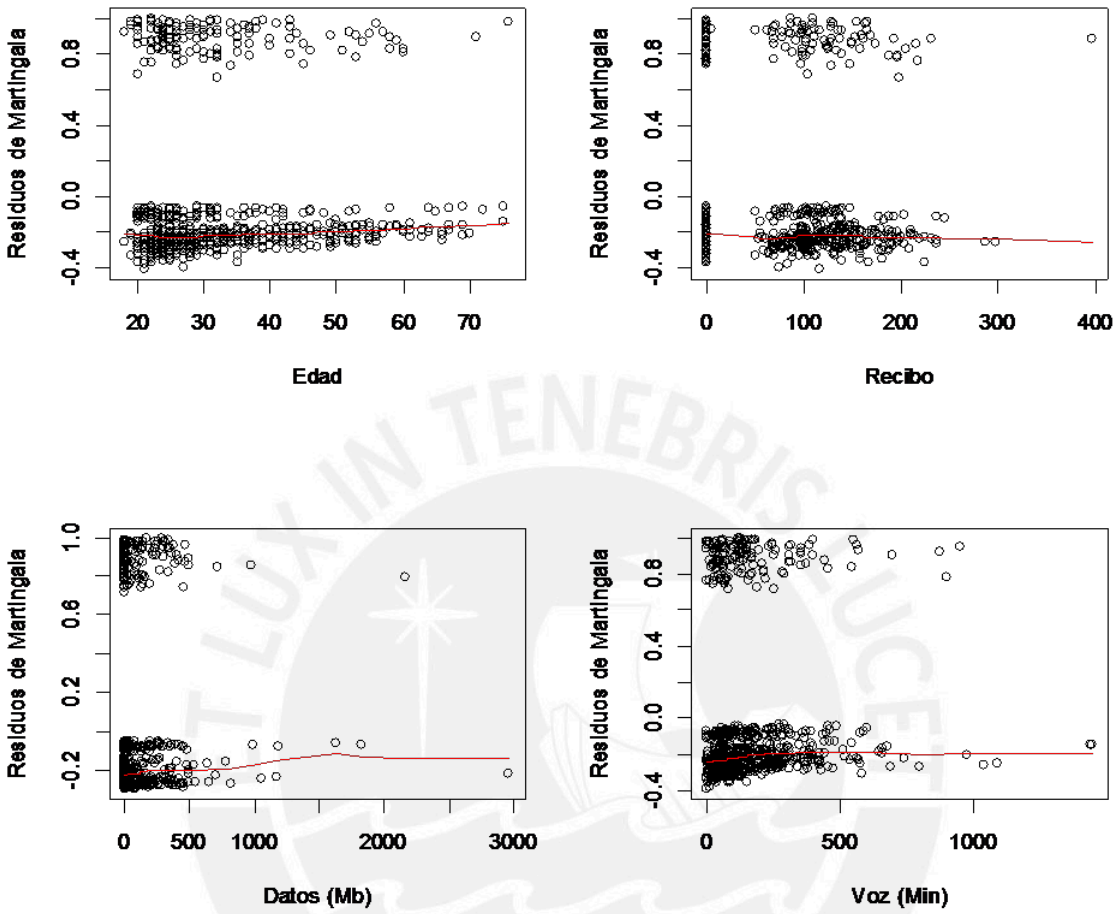


Figura 4.3: Residuos de Martingala para las variables *edad*, *monto facturado*, *trafico de datos* y *minutos*

Por experiencia, se observó de manera anticipada a la implementación, que la variable con punto de cambio es *monto_recibo* y la covariable ordinaria con la que trabajaremos es *edad*, a partir de esto se puede definir el modelo de regresión como:

$$\lambda(t,\boldsymbol{\theta}) = \lambda_0(t)\mathbf{R}(t)\exp\left\{\boldsymbol{\beta}_1'\boldsymbol{edad}(t) + \boldsymbol{\beta}_2'\boldsymbol{monto\_facturado}(t) + \boldsymbol{\beta}_3'\left(\boldsymbol{monto\_facturado}(t) - \boldsymbol{\omega}\right)\right\} \tag{4.1}$$

Se debe tener descrito a la vez el modelo de regresión de Cox (1972):

$$\lambda(t,\boldsymbol{\theta}) = \lambda_0(t)\mathbf{R}(t)\exp\left\{\boldsymbol{\beta}_1'\boldsymbol{edad}(t) + \boldsymbol{\beta}_2'\boldsymbol{monto\_facturado}(t)\right\} \tag{4.2}$$

Una vez que se tienen definidos los modelos, se podrá continuar con el siguiente algoritmo (A.2), se obtendrán las estadísticas del modelo de supervivencia de Cox por las instrucciones

ya construidas en la librería Survival (2013) del modelo de regresión de Cox, a partir de esto se obtienen valores iniciales para los estimadores de los parámetros que participan en el modelo con puntos de cambio.

En la figura 4.4 se muestra la función de supervivencia y función de riesgo acumulada del modelo regresión Cox (1972) para la ecuación (4.2)

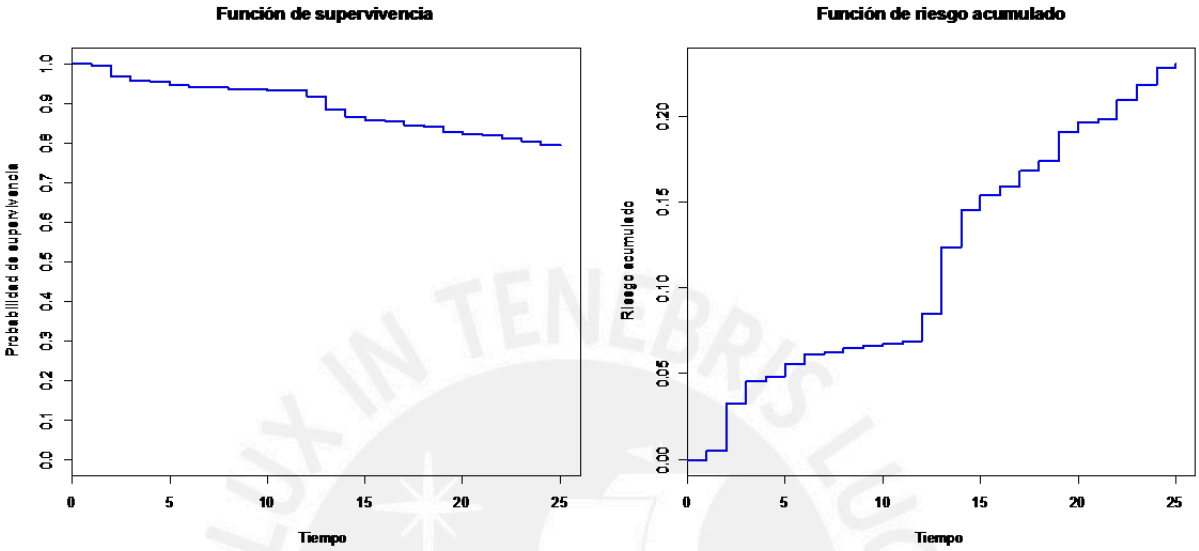


Figura 4.4: Función de supervivencia y riesgo acumulado del modelo de regresión de estudio

También se obtienen los datos del modelo de regresión (4.2) que se muestran en el cuadro 4.5, a partir del cuál se podría decir que el efecto de la covariable *monto_recibo* al incrementarse tiene un riesgo relativo de darse de baja o desactivarse de 0,992, mientras que para la variable *edad*, si es que ésta incrementase, el riesgo de que la línea se de baja es de 0,991, la diferencia entre el efecto de ambas covariables no es tan lejana una de la otra, por lo cual son dos covariables que impactan mucho en la decisión del cliente para dar de baja su línea. Finalmente se tienen los p-valores respectivos, que para las dos covariables resultan menores de 0.05, por lo que se rechaza la hipótesis de que los coeficientes sean iguales a cero.

Covariables	beta	exp(beta)	error estándar	z	p
monto_recibo	-0.00762	0.992	0.00150	-5.08	3.9E-07
Edad	-0.00875	0.991	0.00754	-1.16	2.5E-01

Cuadro 4.5: Resultados del modelo de regresión de Cox Clásico

Según los resultados anteriores y el modelo (4.2) con los mismo datos que se usarán para el modelo (4.1), se puede continuar con el algoritmo A.3 e identificar en qué parte es que ocurre el cambio en la curva del gráfico de residuos de martingala. Como se identificó que en la covariable *monto_recibo* es donde ocurre el punto de cambio, de la gráfica 4.4, se toma solo la que pertenece a los montos facturados. Se hará un acercamiento al rango donde se observa el punto de cambio, por lo cual, en el eje X se limitarán los valores de 50 a 150; ya que en la figura 4.5 se observa un ligero cambio entre los valores de 95 y 110.

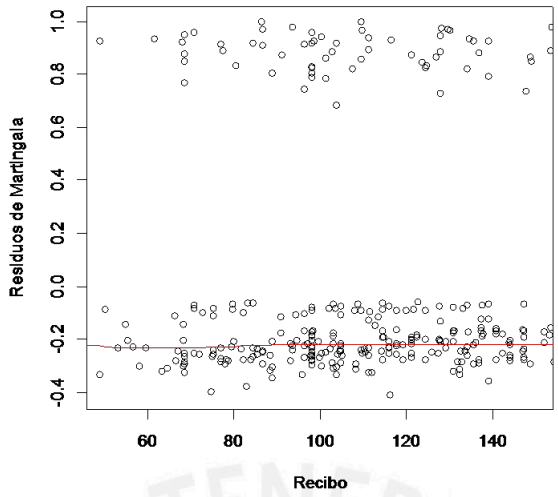


Figura 4.5: Residuos de Martingala en función de la covariable *monto facturado*

4.3.2. Cálculo de la función de Log-Verosimilitud, vector de parámetros $\boldsymbol{\beta}$ y punto de cambio $\boldsymbol{\omega}$

Los resultados obtenidos en la sección anterior, nos servirán como parámetro de entrada para el algoritmo A.3, que fue desarrollado en el software Mathematica (2012). El código desarrollado se limita solo a la estimación de los parámetros considerando un punto de cambio, pero hace la salvedad de que el modelo de Jensen y Lutkebohmert (2008), nos permite más de un punto de cambio.

La información que se utilizó para la construcción del modelo fue: los datos de la muestra, la covariable ordinaria y la covariable con punto de cambio; así como la observación inicial donde ocurre el punto de cambio y el tiempo de observación del estudio.

De acuerdo a los resultados obtenidos en R aplicando el modelo de Cox (1972) que se definen los intervalos en los cuales se maximizarán los valores del vector $\boldsymbol{\beta}$ y de acuerdo a la figura 4.5 se define el intervalo del valor del punto de cambio $\boldsymbol{\omega}$.

Definidos los puntos donde se buscarán los valores que maximicen la función, se ejecutará el proceso iterativo de cálculo de los parámetros $\boldsymbol{\beta}$ y $\boldsymbol{\omega}$ en dos fases:

Se evalúa la función definida en (3.2) ingresando como aproximación inicial para el parámetro del punto de cambio al valor que se observó en la figura 4.5, esto retornará valores aproximados para el vector $\hat{\boldsymbol{\beta}}$

Con los $\hat{\boldsymbol{\beta}}$ calculados se continua con la siguiente fase del cálculo, ahora se estimará $\hat{\boldsymbol{\omega}}$

4.3.3. Iteraciones para el cálculo de los parámetros $\boldsymbol{\beta}$ y $\boldsymbol{\omega}$

Se implementará el algoritmo A.4, para la aplicación del modelo de punto de cambio. Se realizarán varias iteraciones para ver la convergencia de los valores, a fin de obtener a las estimaciones de $\hat{\boldsymbol{\beta}}$ y $\hat{\boldsymbol{\omega}}$. Este proceso el proceso consistirá en fijar primero el valor del punto de cambio $\boldsymbol{\omega}$ en los intervalos definidos, este proceso se calculó que demora alrededor 8 minutos, luego de terminado el proceso con el vector $\boldsymbol{\beta}$ calculado, este pasará como parámetros a la

siguiente ecuación donde se fijan los valores de $\boldsymbol{\beta} = \{\beta_1, \beta_2, \beta_3\}$ para obtener el valor del punto de cambio, este proceso demora cerca de 11 minutos, por lo cual se decidió realizar diferentes ensayos para lograr valores iniciales apropiados a fin de generar una sucesión convergente para los parámetros: $\beta_1, \beta_2, \beta_3$; cabe resaltar que la duración del proceso de las dos fases es de aproximadamente 20 minutos, en solo una iteración. Finalmente, se logró obtener una sucesión convergente que se muestra en el cuadro 4.6.

Como primer paso se fija $\hat{\boldsymbol{\omega}}$ y las componentes del vector $\hat{\boldsymbol{\beta}}$ se evaluarán en cierto rango para obtener al máximo.

$$\widehat{\boldsymbol{\beta}} = \arg\max \log L(\boldsymbol{\beta}, \boldsymbol{\omega})$$

Como segundo paso se fija al vector de $\hat{\boldsymbol{\beta}}$ de acuerdo a lo obtenido en el paso 1 y el valor de $\hat{\boldsymbol{\omega}}$ se evaluará en cierto rango para obtener el máximo valor.

$$\widehat{\boldsymbol{\omega}} = \arg\max \log L(\widehat{\boldsymbol{\beta}}, \boldsymbol{\omega}) \ \textit{donde} \ 95 \leq \boldsymbol{\omega} \leq 110$$

Iteración	ω	β_1	β_2	β_3
1	100.00	-0.0117798	-0.0092819937	-0.0082828
2	105.58	-0.0117798	-0.0092819937	-0.0082828
3	105.58	-0.0117798	-0.0092819937	-0.0082828

Cuadro 4.6: Resultados de Iteraciones con el modelo de Jensen y Lutkebohmert (2008)

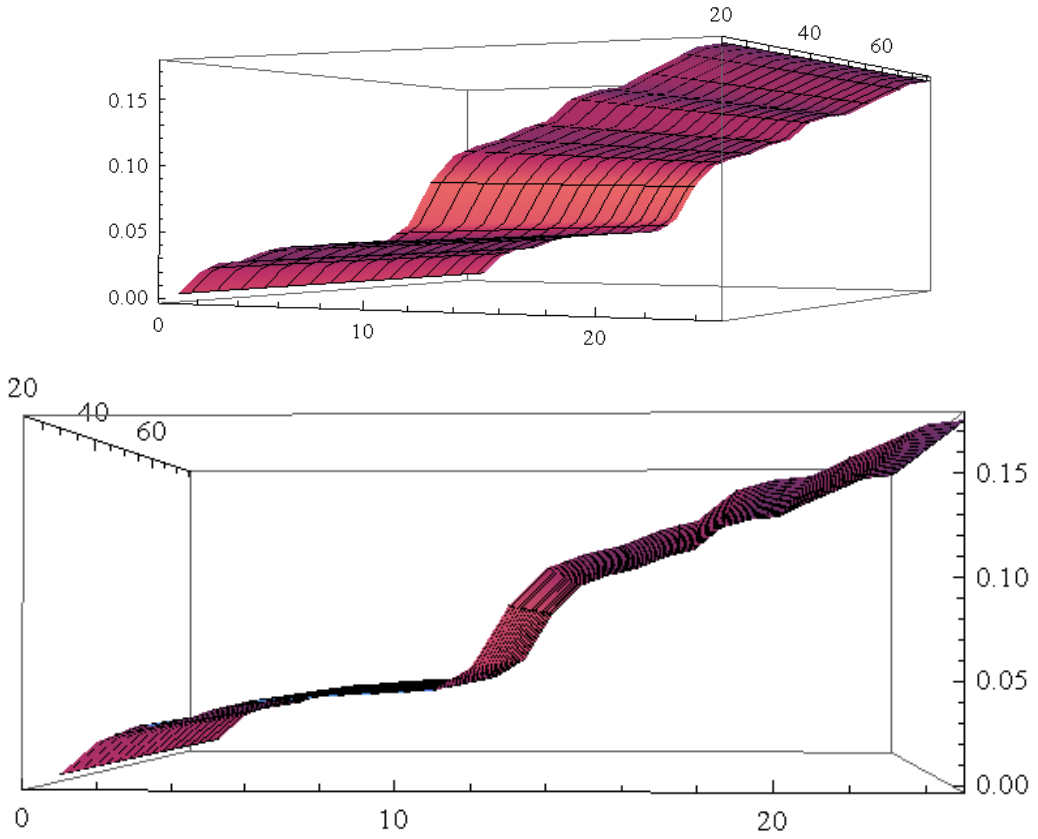
Ahora se tienen los parámetros del modelo de regresión de Cox con puntos de cambio y los parámetros del modelo de regresión de Cox Clásico los cuales se resumen en el cuadro 4.7

Cox (1972)		
Covariables	$\boldsymbol{\beta}$	$\exp\{\boldsymbol{\beta}\}$
edad	-0.00875	0.991
monto_facturado	-0.00762	0.992
Jensen y Lutkebohmert (2008)		
Covariables	$\boldsymbol{\beta}$	$\exp\{\boldsymbol{\beta}\}$
edad	-0.0117798000	0.98828931
monto_facturado	-0.0092819937	0.99076095
$\{monto\_facturado - \boldsymbol{\omega}\}$	-0.0082828000	0.99175141

Cuadro 4.7: Resultados de aplicación de los modelos Cox (1972) y Jensen y Lutkebohmert (2008)

En el cuadro 4.8, se muestra la gráfica en R_3 del riesgo respecto de las variables *edad*, y *tiempo*, se observa que a medida que transcurre el tiempo aumenta el riesgo y que también los valores de riesgo para personas jóvenes son ligeramente mayores que para personas de mayor edad. Entre los meses 12 y 14 es que se observa un mayor incremento del riesgo que se mantiene en incremento para todas las edades, lo mismo sucede entre los meses 18 y 20 pero el incremento no se compara con la primera observación. Además se ve que el riesgo al final de los 25 meses tiene un valor de 0.18.

En el cuadro 4.9, muestra dos ángulos diferentes de la función de riesgo cuando este se evalúa respecto a las variables *monto facturado* y *tiempo*, al igual que en el gráfico de la *edad*, para el periodo de 12 a 14 meses el riesgo incrementa, sin embargo esta función de riesgo que



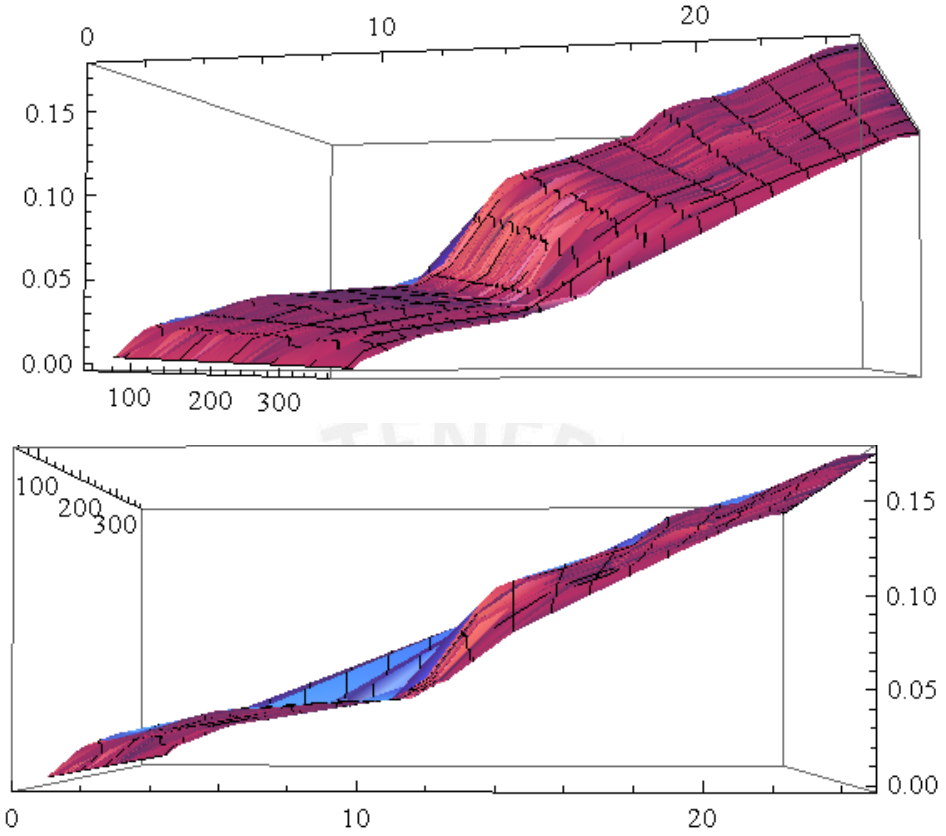
Cuadro 4.8: Modelo con la función de riesgo, *edad* y *tiempo*

se forma no es la misma a partir de los $montos_facturados$ de 200 soles en adelante, donde se observa un crecimiento del riesgo constante, además que para valores de montos menores a 200 el riesgo es ligeramente mayor a valores mayores.

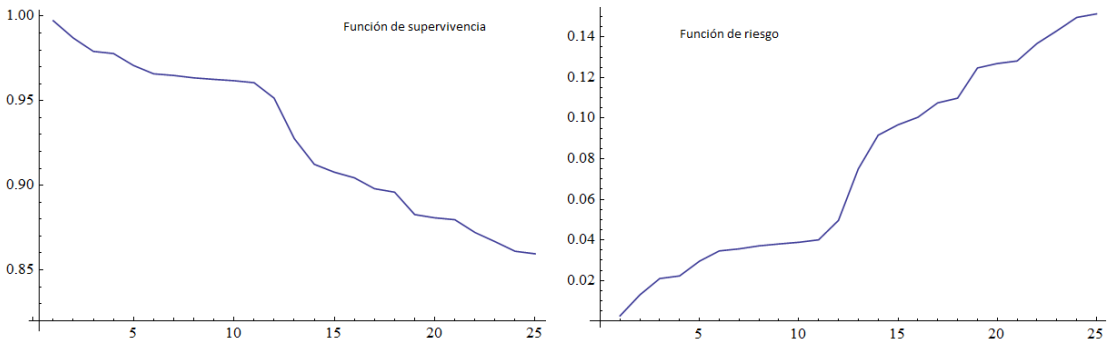
En el cuadro 4.10 se muestra la función de supervivencia y riesgo para el modelo de regresión Jensen y Lutkebohmert (2008) con puntos de cambio en las covariables, en la cual se observa lo dicho en líneas previas, para ambos casos el riesgo se incrementa o la supervivencia cae entre los meses 12 y 14, tambíen notamos que el riesgo es bajo para los 15 meses evaluados.

Para un tiempo $T = t$ fijo, se grafican las funciones de riesgo para valores continuos de la *edad*, se observa que para el caso que se fije la *edad* en 20 el riesgo es el más alto para montos facturados que se encuentren entre 100 y 200. Cuando la edad se fija en 70 el riesgo es mucho menor comparado al riesgo de clientes de 20 años, pero al igual que el anterior, para montos facturados entre 100 y 200 son los más altos comparados con montos altos.

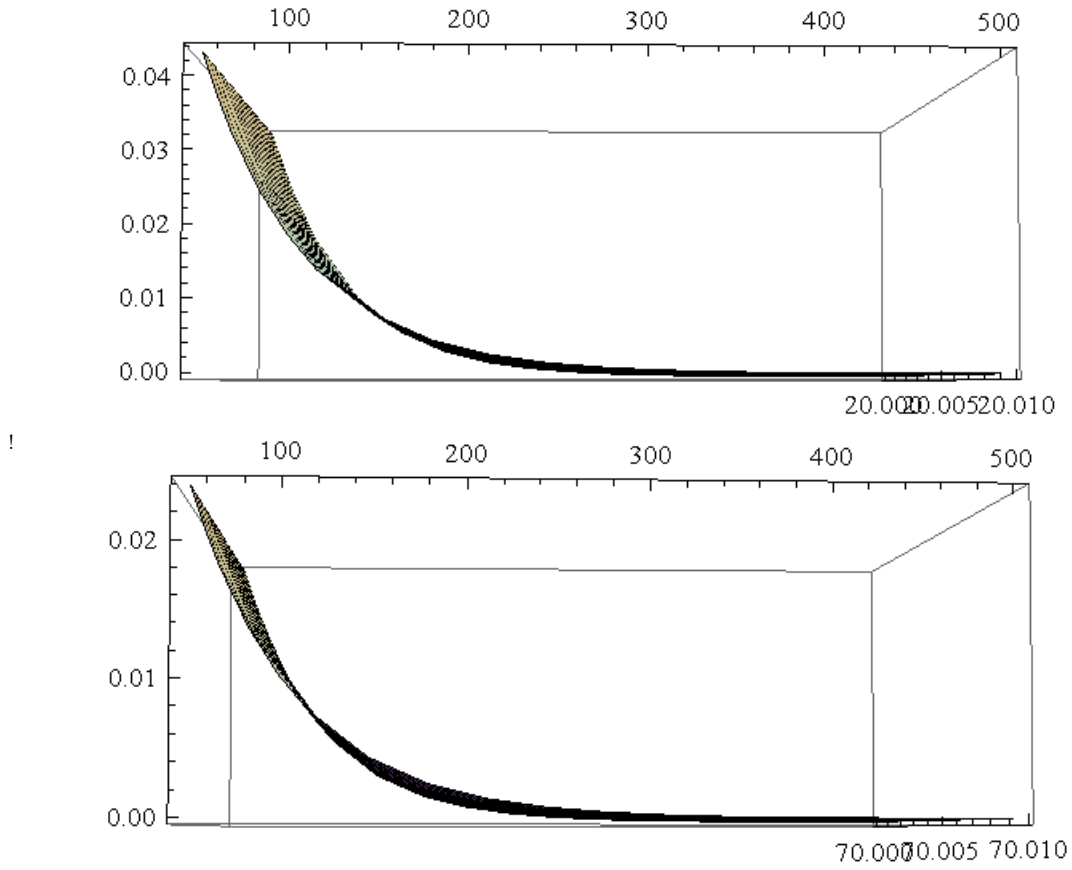
Se realiza el mismo ejercicio realizado para las gráficas del cuadro 4.11, pero ahora fijando los montos facturados continuos entre el valor central de 100 y 200 soles. Se observa el crecimiento del riesgo constante, haciéndolo lineal para ambos montos, pero es mayor para el monto de 100 soles y con edades entre los 20 años.



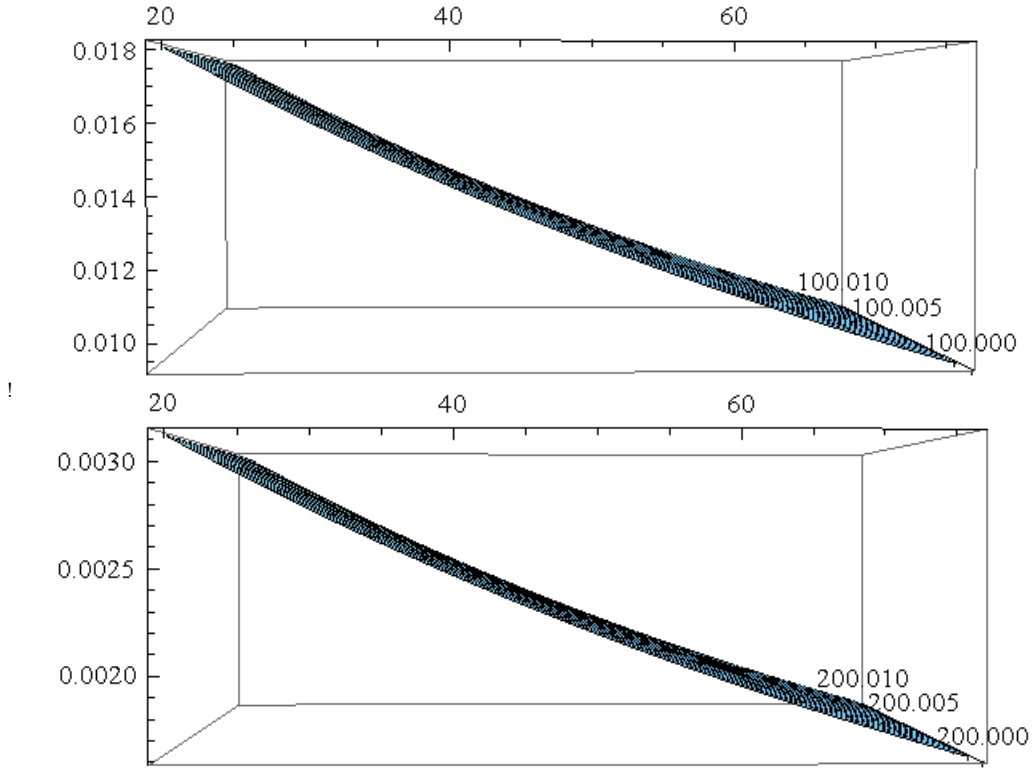
Cuadro 4.9: Modelo con la función de riesgo, *monto facturado* y *tiempo*



Cuadro 4.10: Función de supervivencia y riesgo



Cuadro 4.11: Función de riesgo en un tiempo t para la variable *edad* con valores de 20 y 70 años



Cuadro 4.12: Función de riesgo en un tiempo t para la variable $monto_facturado$ de 100 y 200 años

Capítulo 5

Conclusiones y Sugerencias

5.1. Conclusiones relacionados al modelo

Conclusiones en base al modelo de estudio

- El modelo en estudio adicionalmente a las ventajas que se puedan encontrar dentro del modelo de regresión de Cox clásico, permite ampliar el estudio a covariables que son dependientes del tiempo y que en un determinado momento cambian los parámetros del modelo, sin perder la consistencia de los estimadores. Al agregar esta característica de las covariables, se mejora el modelo haciéndolo mas completo para cualquier tipo de situación en estudio.
- En el modelo propuesto por Jensen y Lutkebohmert (2008) se incluyen varias covariables que consideran más de un punto de cambio a diferencia de otros propuestos previamente.
- Se puede realizar el estudio a procesos que tengan mas de un evento en observación.
- Durante el desarrollo del trabajo, se llego a entender la importancia del estudio de comportamiento funcional de las covariables, muchas veces este proceso se deja de lado y se procede a aplicar un modelo que encaje, en nuestro caso, previo a la aplicación del modelo, se estudió el comportamiento de las covariables en el tiempo, lo que permitió comprobar que estas no eran constantes y a la vez validar que si bien el modelo de Cox (1972) se aplica para el caso de estudio, existen muchas otras modelos derivados de este, que pueden dar mejores resultados en el cálculo de los parámetros.

5.2. Conclusiones respecto a la aplicación del modelo al conjunto de datos

Respecto a los resultados obtenidos del conjunto de datos de líneas móviles

- El modelo de regresión de Cox clásico, permite analizar la relación de las covariables con la situación final o estado del evento en estudio, de forma simple, sin necesidad de definir una distribución de los datos inicialmente, dada su naturaleza semi-paramétrica. Si bien esto es conocido en teoría, para nuestra aplicación resulta beneficioso puesto que no existen estudios precedentes al nuestro sobre el tema, por lo cual al no tener ninguna referencia estadística, esta característica del modelo resulta práctica para la aplicación.

- Se observa que la mayor cantidad de líneas en riesgo de darse de baja a solicitud del cliente, migraciones a prepago o portabilidad se da en las personas con edades entre el rango de 20 a 30 años para cargos fijos entre 90 y 110 soles. Si bien existen estudios dentro de la empresa que evalúan la baja de la línea por endeudamiento, este estudio nos revela que para los casos de solicitudes de baja del cliente el comportamiento no es el mismo. El factor de riesgo es menor para facturas emitidas de mayor cantidad de soles y para usuarios de mayor edad al rango dado, por lo cual se deduce que los clientes de mayor edad se comprometen con el pago de sus recibos y están conformes con el servicio. En el caso de las personas que están dentro del rango, la cantidad de bajas se eleva al llegar a los meses doce y dieciocho, lo que indica que muchas de esas personas se comprometen a pagar sus recibos por cierto periodo hasta completar el tiempo de contrato. Terminado el contrato, estos deciden dar de baja a la línea o migrar a un plan prepago, que nos los obliga al pago de una factura mensual.

  En la realidad se observa este escenario, pero no se cuantifica, dado que el riesgo de baja a solicitud del cliente es bajo comparado con la morosidad. Pero con el escenario actual de la competencia en telecomunicaciones, ayudará a preservar a los clientes. Se pueden crear estrategias de retención del cliente a un plan postpago al llegar al término del contrato, para esto el rango de edades y el cargo contratado servirá para crear planes móviles que se orienten a ese segmento.

- Inicialmente asumí por la experiencia laboral que existía un punto de cambio, esta suposición ha quedado corroborada a través de la aplicación del modelo, en que se aprecia un ajuste adecuado de los datos.

5.3. Sugerencias para investigaciones futuras

Según la investigación realizada se encontraron las siguientes sugerencia para estudios posteriores en base al modelo desarrollado

- Comparar los modelos previos como el de Luo y Boyett (1997) , Gandy et al. (2005) o Pons (2003) con el de estudio propuesto por Jensen y Lutkebohmert (2008) y analizar las estimaciones obtenidas.

- Estudiar otros modelos de regresión como por ejemplo el logístico que admita los puntos de cambio y compararlo con el modelo en estudio.

- Hacer el desarrollo de un algoritmo que identifique automáticamente los puntos de cambio con la información dada, ya que por el momento se realiza de manera gráfica.

- En base a la sugerencia anterior, crear un algoritmo genérico de manera que no existan parámetros de entrada en el algoritmo, solo bastaría ingresar los datos y que el desarrollo realizado de manera automática, proporcione los resultados deseados. En el código desarrollado en Mathematica, se necesitaría especificar las covariables con punto de cambio, el parámetro inicial del punto de cambio, ademas de que está trabajando solo con dos covariables y con un punto de cambio en el tiempo.

El modelo de Jensen y Lutkebohmert (2008) se presta para una aplicación con mayor complejidad en la cantidad de covariables, cantidad de eventos que suceden en el tiempo y mayor cantidad de puntos de cambio, el desarrollo es un código Ad-hoc.

- Por fines prácticos, dado que los principales objetivos de la tesis estaban en el estudio del modelo y la implementación de este en código, es que se tomaron dos covariables de todas las disponibles para la línea. Si se deseara realizar un estudio profundo del comportamiento de la línea a través del tiempo y las características de ésta para darse de baja, podrían considerarse más covariables e inclusive segmentar el modelo y estudiarlos en paralelo.

- Para el estudio se consideró como ocurrencia del evento a las líneas que se daban de baja a solicitud del cliente o porque migraron a otra empresa de telecomunicaciones, por lo cual si estudiáramos los otros motivos de baja (por morosidad, fraude, etc.) encontraríamos otro escenario donde quizás el comportamiento de las covariables varíe.

- Si bien los autores del modelo con punto de cambio indican la forma de realizar la prueba de bondad de ajuste a través de la referencia del artículo Gandy y Jensen (2006), ésta no se llegó a realizar para la muestra de 774 usuarios, puesto que el proceso implica el desarrollo del código para la teoría que exponen en su trabajo presentado en el 2006, en un artículo diferente.

  Con el deseo de ilustrar la prueba de hipótesis que presentan Gandy y Jensen (2006), se extrajo una muestra e 30 clientes de la población, haciendo uso del muestreo estratificado.

  Se demostrará con valores que el modelo de regresión propuesto por Jensen y Lutkebohmert (2008) será mas apropiado que el de el modelo de Cox (1972), para lo cual se define lo siguiente:

$$H_0 : Modelo\ de\ regresión\ Cox\ (1972)$$

$$H_1 : Modelo\ de\ regresión\ Jenseny\ Lutkebohmert\ (2008)$$

  El algoritmo fue implementado en Mathematica demoró un tiempo total de 22 minutos para realizar operaciones con las 30 observaciones.

  La teoría que se sugiere para la aplicación de la prueba de hipótesis se muestra en el apéndice B.6. Después de realizados los cálculos se obtuvo el p-valor:

Hipótesis Nula	**Modelo alternativo**	***p*-valor**
Modelo de regresión de Cox Clásico	Modelo con puntos de cambio	0.00003

Cuadro 5.1: La tabla contiene el p-valor que fue calculado basado en el artículo de Gandy y Jensen (2006)

Nos indica que se rechaza la hipótesis nula por lo cual rechazamos el modelo de regresión de Cox (1972)

Apéndice A

Diagramas de flujo para implementación de códigos

Para lograr el desarrollo de los algoritmos que nos ayuden a implementar el modelo de estudio, recurrimos a la construcción de diagramas de flujo que nos permitirá identificar una secuencia lógica para la codificación del modelo.

De acuerdo al modelo estudiado, se identificaron las siguientes tareas a desarrollar para llegar a la estimación final de los parámetros del modelo para una regresión de Cox con puntos de cambio en la covariables.

A.1. Algoritmo para identificación de puntos de cambio

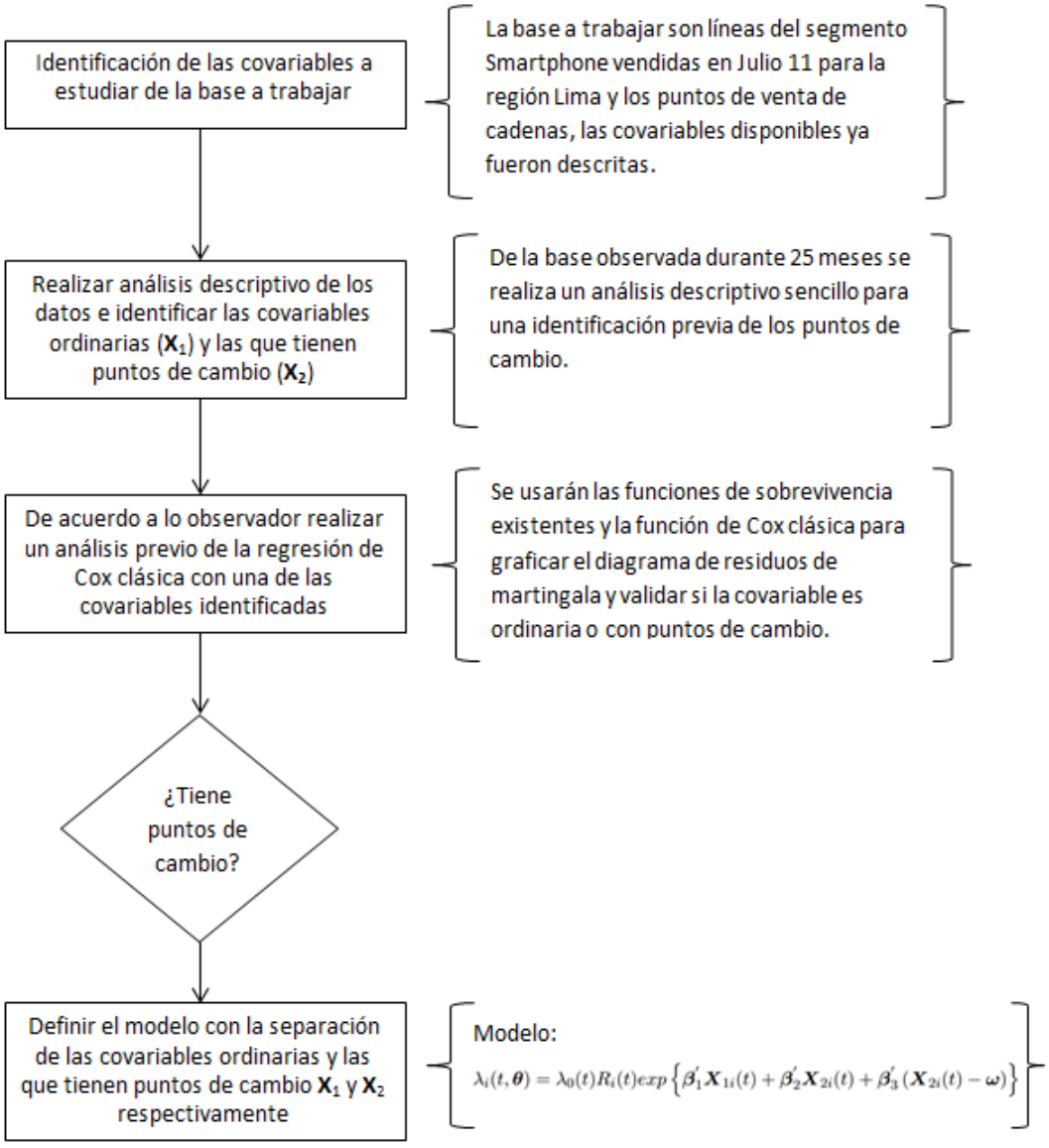


Figura A.1: Algoritmo para la identificación de la covariable con punto de cambio

A.2. Punto de inicio donde ocurre el punto de cambio

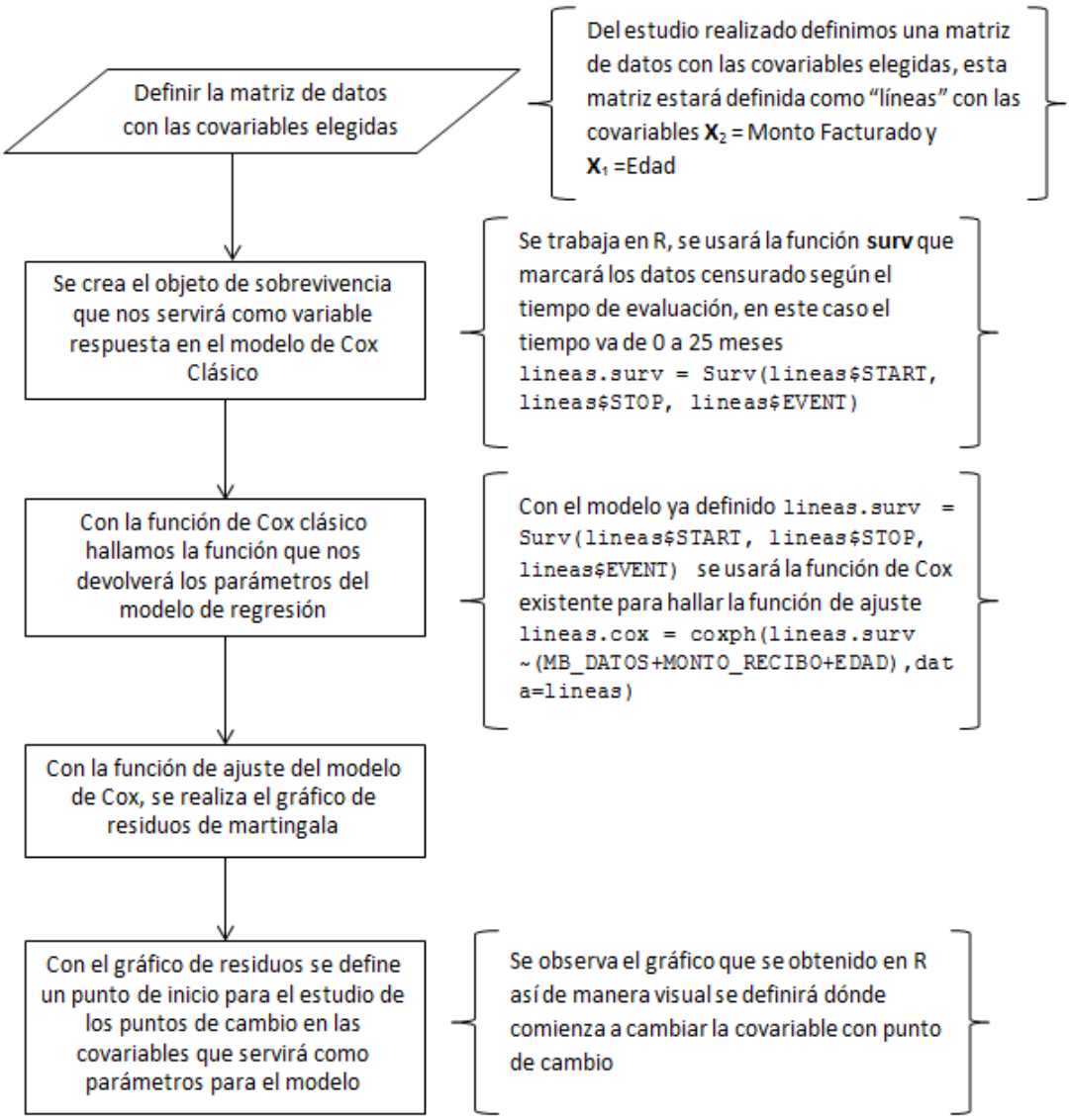


Figura A.2: Algoritmo para identificar el punto de cambio

A.3. Calculando la Log-Verosimilitud

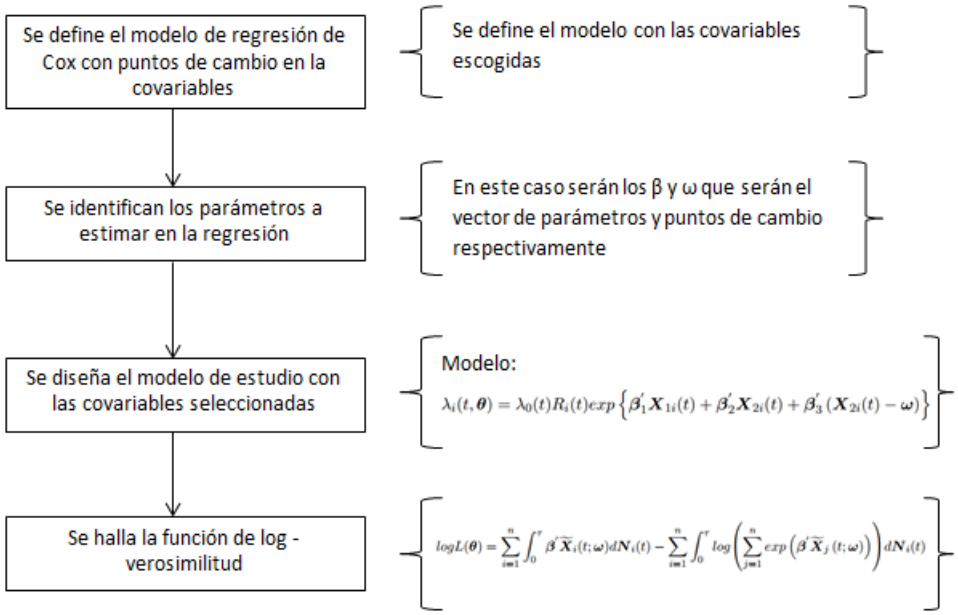


Figura A.3: Algoritmo para el cálculo de la log-verosimilitud

A.4. Calculando parámtros de la regresión

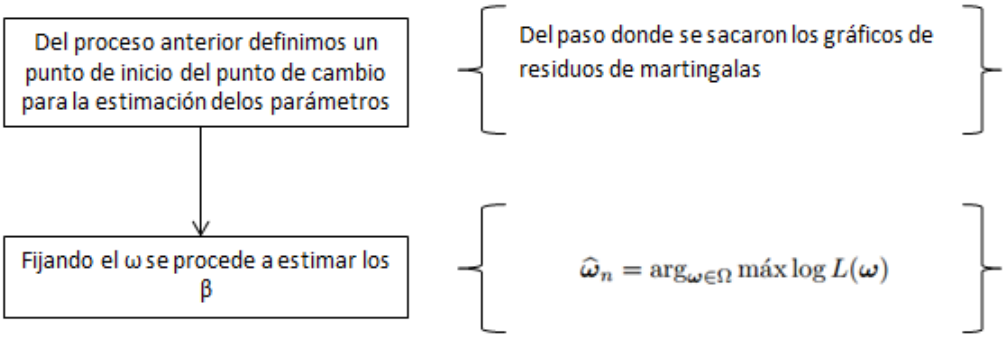


Figura A.4: Algoritmo para el cálculo del punto de cambio y los parámetros de las covariables

Apéndice B

Lemas y demostraciones del modelo propuesto

Las siguientes condiciones, lemas y teoremas, son parte del trabajo de demostración realizado por Jensen y Lutkebohmert (2008) para demostrar de manera teórica la ventaja del modelo frente al modelo Clásico, esto para datos que muestren que la forma funcional de la covariable no es constante en el tiempo.

B.1. Martingala basado en el proceso de conteo

Sean T_i tiempo potencial de falla del sujeto i y U_i un tiempo de censura. $Y_i = \text{mín}(T_i, U_i)$, $\delta_i = I_{\{T_i \leq U_i\}}$ es el indicador de ocurrencia del evento. Los procesos N_i y R_i serán:

$$N_i(t) = I_{\{Y_i \leq t, \delta_i = 1\}}$$

$N_i(t)$ indica si ocurrió el evento al sujeto i en el tiempo t

$$R_i(t) = I_{\{Y_i \geq t\}}$$

$R_i(t)$ indica si el sujeto i se encuentra en riesgo en el tiempo t.
Si $\overline{N} = \sum_i N_i$ y $\overline{R} = \sum_i R_i$ entonces:

$$\widehat{\Lambda}(t) = \int_0^t \frac{I_{\{\overline{R}(\mu)>0\}}}{\overline{R}(\mu)} d\overline{N}(\mu),$$

Donde se define $\widehat{\Lambda}(t) = 0$, si $\overline{R}(\mu) = 0$. Cuando un modelo no estadístico es asumido, la información acerca de Λ está disponible solo para $\{\mu : \overline{R}(\mu) > 0\}$, y de hecho $\widehat{\Lambda}(t)$ realmente estima la cantidad aleatoria

$$\Lambda^*(t) = \int_0^t I_{\{\overline{R}(\mu)>0\}} \lambda(\mu) d\mu$$

A partir de lo cual se obtiene:

$$\begin{aligned} \widehat{\Lambda}(t) \quad - \quad \Lambda^*(t) &= \int_0^t \frac{I_{\{\overline{R}(\mu)>0\}}}{\overline{R}(\mu)} d\overline{N}(\mu) - \int_0^t I_{\{\overline{R}(\mu)>0\}} \lambda(\mu) d\mu \\ &= \int_0^t \frac{I_{\{\overline{R}(\mu)>0\}}}{\overline{R}(\mu)} \left\{ d\overline{N}(\mu) - \overline{R}\lambda(\mu) d\mu \right\} \end{aligned}$$

Lo que también se puede escribir de la siguiente manera:

$$\sum_{i=1}^{n} \int_0^t \frac{I_{\{\overline{R}(\mu)>0\}}}{\overline{R}(\mu)} dM_i(\mu),$$

Con

$$M_i(\mu) = N_i(\mu) - \int_0^{\mu} R_i(s) d\Lambda(s)$$

$$\widehat{M}_i(\mu) = N_i(\mu) - \widehat{E}_i(\mu) = N_i(\mu) - \int_0^{\mu} R_i(s) \exp\left\{\beta' X_i(s)\right\} d\widehat{\Lambda}_0(\beta, s)$$

Donde $\widehat{\Lambda}_0(\beta, s)$ es el estimador riesgo de Breslow definido como:

$$\widehat{\Lambda}_0(\beta, s) = \int_0^t \frac{\sum_{i=1}^n dN_i(s)}{\sum_{i=1}^n R_i(s) \exp\left\{\beta' X_i(s)\right\}}$$

y están basados en la martingala de un proceso de conteo para el i-ésimo individuo , $M_i(\mu) = N_i(\mu) - E_i(\mu)$, definida mediante:

$$M_i(\mu) = N_i(\mu) - \int_0^{\mu} R_i(s) \exp\left\{\beta' X_i(s)\right\} d\widehat{\Lambda}_0(\beta, s)$$

B.2. Condiciones necesarias parar establecer las propiedades asintóticas de los estimadores

Las siguientes condiciones son necesarias para establecer las propiedades asintóticas de los estimadores. Para una fácil notación los autores definen la medida de probabilidad $\boldsymbol{P^t}$ con

$$d\boldsymbol{P^t} = q_t^{-1} dQ^t \ , \quad Q_t(A) = \int_A \exp(\boldsymbol{\beta}_0' \widetilde{\boldsymbol{X}}(t; \boldsymbol{\omega}_0)) dP \ y \ q_t = \int dQ^t$$

Dado que $\boldsymbol{q_t} < \infty$. Se asume de que existe un conjunto convexo y compacto $\boldsymbol{\Theta} \subset \mathbb{R}^{p+3m}$ con $\boldsymbol{\theta}_0$ en su interior de modo que se mantiene lo siguiente:

C1. (a) El vector aleatorio $\boldsymbol{X}_2(t)$ tiene una distribución continua absoluta con densidad $f_{\boldsymbol{X}_2(t)}$ que es estrictamente positivo, acotado y continua para cualquier $\boldsymbol{\omega}$ en una vecindad de $\boldsymbol{\omega}_0$ y para todo $t \in [0, \tau]$.

(b) $sup_{t\in[0,\tau]} \lambda_0(t) < \infty$.

C2. Para k = 0,1,2

$$E\left[\sup_{t\in[0,\tau]} \sup_{\boldsymbol{\theta}\in\Theta} \left\{ \left(\|\boldsymbol{X}_1(t)\|^k + \|\boldsymbol{X}_2(t)\|^k \right) \exp\left(\boldsymbol{\beta}' \widetilde{\boldsymbol{X}}(t; \boldsymbol{\omega}) \right) \right\}^2 \right] \ < \ \infty$$

C3. La función $s(t; \boldsymbol{\theta}) = E[\exp(\boldsymbol{\beta}' \widetilde{\boldsymbol{X}}(t; \boldsymbol{\omega}))]$, esta acotada lejos de cero en $[0, \tau] \times \Theta$ y las dos primeras derivadas parciales de $s(t; \boldsymbol{\theta})$ con respecto a $\boldsymbol{\beta}$ existen , son acotadas en $[0, \tau] \times \boldsymbol{\Theta}$ y continuas en $\boldsymbol{\Theta}$, uniformemente en $t \ \in [0, \tau]$.

C4. (a) Para todo $t \in [0,\tau]$ existe una vecindad $\boldsymbol{\Theta}_0$ de $\boldsymbol{\theta}_0$ de modo que la matriz de covarianza $Cov_{P^t}(\boldsymbol{Y}(t))$, es definido positivo, donde:

$$\boldsymbol{Y}(t) = \left(-\boldsymbol{\beta}_{30} I\left\{\boldsymbol{X}_2 > \boldsymbol{\omega}_0\right\},\ \boldsymbol{X}_1^{'}(t),\ \boldsymbol{X}_2^{'}(t), ((\ \boldsymbol{X}_2(t) - \boldsymbol{\omega}_0))^{'}\right)^{'}$$

(b) Para $k = 0, 1, 2$

$$\sup_{x\in\boldsymbol{\Omega}} E\left[\sup_{t\in[0,\tau]}\ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\{\left(\|\boldsymbol{X}_1(t)\|^k + \|\boldsymbol{X}_2(t)\|^k\right)\exp\left(\boldsymbol{\beta}^{'}\widetilde{\boldsymbol{X}}(t;\boldsymbol{\omega})\right)\right\}^j |\boldsymbol{X}_2(t) = x\right] \ < \ \infty,$$
$$j = 1, 2$$

$$\sup_{x,x^{'}}\ \sup_{t\in[0,\tau]}\ \sup_{\boldsymbol{\theta}\in\Theta}\left|E\left\{\exp\left(\boldsymbol{\beta}^{'}\widetilde{\boldsymbol{X}}(t;\omega)\right)|\boldsymbol{X}_2(t) = x\right\} - E\left\{\exp\left(\boldsymbol{\beta}^{'}\widetilde{\boldsymbol{X}}(t;\omega)\right)\left|\boldsymbol{X}_2(t) = x^{'}\right.\right\}\right|$$
$$\xrightarrow{\|x-x'\|\to 0} 0$$

Donde x y x' varían en $\boldsymbol{\Omega}$

B.3. Resultados teóricos necesarios para la consistencia del estimador

Lema 1. *Bajo las condiciones* $C1 - C3$, $\sup_{\boldsymbol{\theta}\in\Theta}|Z_n(\boldsymbol{\theta}) - z(\boldsymbol{\theta})|$ *converge en probabilidad a cero cuando* $n \to \infty$.

Demostración. $Z_n(\boldsymbol{\theta})$ está definido de la siguiente manera:

$$\begin{aligned}
Z_n(\boldsymbol{\theta}) &= (\boldsymbol{\beta}_1^{'}, \boldsymbol{\beta}_2^{'})\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\begin{pmatrix}\boldsymbol{X}_{1i}(t)\\ \boldsymbol{X}_{2i}(t)\end{pmatrix} dM_i(t)\\
&+ (\boldsymbol{\beta}_1^{'}, \boldsymbol{\beta}_2^{'})\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\begin{pmatrix}\boldsymbol{X}_{1i}(t)\\ \boldsymbol{X}_{2i}(t)\end{pmatrix} R_i(t)\exp\left(\boldsymbol{\beta}_0^{'}\widetilde{\boldsymbol{X}}\left(s;\boldsymbol{\omega}_0\right)\right) d\Lambda_0\left(t\right)\\
&+ \boldsymbol{\beta}_3^{'}\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\left(\boldsymbol{X}_{2i}\left(t\right) - \boldsymbol{\omega}\right)^{+} dN_i(t)\\
&- \int_0^{\tau}\log\left(n^{-1}\sum_{i=1}^{n} R_i(t)\exp\left(\boldsymbol{\beta}^{'}\widetilde{\boldsymbol{X}}_i\left(t;\boldsymbol{\omega}\right)\right)\right) d\overline{N}(t) \qquad \text{(B.1)}
\end{aligned}$$

De acuerdo a C1 y C2 desde que $\ominus$ es compacto el primer término tiene media cero por lo tanto converge en probabilidad cero por la ley de grandes números. De manera similar, el segundo término converge a:

$$(\boldsymbol{\beta}_1^{'}, \boldsymbol{\beta}_2^{'})E\left[R\left(t\right)\begin{pmatrix}\boldsymbol{X}_{1i}(t)\\ \boldsymbol{X}_{2i}(t)\end{pmatrix}\exp\left(\boldsymbol{\beta}_0^{'}\tilde{\boldsymbol{X}}\left(t;\boldsymbol{\omega}_0\right)\right) d\Lambda_0\left(t\right)\right]$$

El tercer término en la ecuación anterior puede ser tomado como sigue. Por la condición $C2$, para todo $\boldsymbol{\omega} \in \boldsymbol{\Omega}$,

$$E\left[\int_0^\tau \left(\boldsymbol{X}_2\left(t\right)-\boldsymbol{\omega}\right)\lambda\left(t,\theta_0\right)dt\right] \leq E\left[\int_0^\tau \left(\boldsymbol{X}_2\left(t\right)-\widetilde{\boldsymbol{\omega}}\right)\lambda\left(t,\boldsymbol{\theta}_0\right)dt\right] < \infty$$

Donde $\widetilde{\boldsymbol{\omega}} = \left(\boldsymbol{\omega}_{11},...,\boldsymbol{\omega}_{1m}\right)'$ es índice inferior izquierdo de $\boldsymbol{\omega}$. Por lo tanto $\int_0^t \left(\boldsymbol{X}_2\left(t\right)-\boldsymbol{\omega}\right)dM\left(s\right)$ es una martingala y

$$E\left[\int_0^\tau \left(\boldsymbol{X}_2\left(t\right)-\boldsymbol{\omega}\right)dN\left(t\right)\right] = E\left[\int_0^\tau \left(\boldsymbol{X}_2\left(t\right)-\widetilde{\boldsymbol{\omega}}\right)\lambda\left(t,\theta_0\right)dt\right] < \infty$$

Se aplicará el teorema de Glivenko-Cantelli dado con referencia en el teorema 19.4 y ejemplo 19.8 en der Vaart (1998). Claramente $\int_0^\tau \left(\boldsymbol{X}_2\left(t\right)-\widetilde{\boldsymbol{\omega}}\right)dN\left(t\right)$ es una función envolvente para $\int_0^\tau \left(\boldsymbol{X}_2\left(t\right)-\boldsymbol{\omega}\right)dN(t)$. Si $\int_0^\tau \left(\boldsymbol{X}_2\left(t\right)-\boldsymbol{\omega}\right)dN(t)$ es continuo en $\boldsymbol{\omega}$ se tiene:

$$\sup_{\boldsymbol{\omega}\in\boldsymbol{\Omega}}\left|\frac{1}{n}\sum_{i=1}^{n}\int\left(\boldsymbol{X}_{2i}-\boldsymbol{\omega}\right)dN_i(t) - E\left[\int_0^\tau \left(\boldsymbol{X}_2\left(t\right)-\widetilde{\boldsymbol{\omega}}\right)\lambda\left(t,\boldsymbol{\theta}_0\right)dt\right]\right| \xrightarrow{P} 0$$

Se multiplica por el parámetro de vector acotado $\boldsymbol{\beta}_3'$ da la convergencia del tercer término en B.1

Para demostrar una convergencia asintótica uniforme del cuarto término en B.1 se aplica la ley de los grandes números dado por Andersen y Gill (1982). Entonces:

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{t\in[0,\tau]}\left|\frac{1}{n}\sum_{i=1}^{n}R_i(t)\exp(\boldsymbol{\beta}'\widetilde{\boldsymbol{X}}_i(t,\boldsymbol{\omega})) - s(t,\boldsymbol{\theta})\right| \xrightarrow{P} 0$$

Donde se usa la la condición de integrabilidad C2. Dado que $s(t,\boldsymbol{\theta})$ es delimitada lejos de 0 por la condición C3, esto sigue:

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{t\in[0,\tau]}\left|\log\left(\frac{1}{n}\sum_{i=1}^{n}R_i(t)\exp(\boldsymbol{\beta}'\widetilde{\boldsymbol{X}}_i(t,\boldsymbol{\omega}))\right) - \log\left(s(t,\boldsymbol{\theta})\right)\right| \xrightarrow{P} 0$$

Desde

$$\frac{1}{n}\sum_{i=1}^{n}N_i(\tau) \xrightarrow{P} EN(\tau) = E\left[\int_0^\tau \lambda(t,\boldsymbol{\theta}_0)dt\right] < \infty$$

La diferencia entre

$$\int_0^\tau \log(s(t,\boldsymbol{\theta}))d\overline{N}(t)$$

y el cuarto término en (B.1) converge uniformemente a 0 en probabilidad. Usando el teorema de Glivenko-Cantelli se tiene:

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\int_0^\tau \log(s(t,\boldsymbol{\theta}))d\overline{N}(t) - E\left[\int_0^\tau \log(s(t,\boldsymbol{\theta}))\lambda(t,\boldsymbol{\theta}_0)dt\right]\right| \xrightarrow{P} 0$$

Donde la función envolvente

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{t\in[0,\tau]}\log(s(t,\boldsymbol{\theta}))N(\tau)$$

Está limitada por C2 y C3

□

Para mostrar la concavidad de $z(\boldsymbol{\theta})$ en una vecindad de $\boldsymbol{\theta}_0$, se usa la matriz Hessiana de $z(\boldsymbol{\theta})$ en $\boldsymbol{\theta}_0$.

Lema 2. *Bajo las condiciones $C1 - C4$ la matriz Hessiana $H(\boldsymbol{\theta}_0)$ de $z(\boldsymbol{\theta}_0)$ es dado por*

$$H(\boldsymbol{\theta}_0) = -\int_0^\tau q_s Cov_{P^s}(Y(s))\lambda_0(s)ds$$

,

Donde $\boldsymbol{Y}(t) = (-\boldsymbol{\beta}_{30} I\left\{\boldsymbol{X}_2(t) > \boldsymbol{\omega}_0\right\}, \boldsymbol{X}_1'(t), \boldsymbol{X}_2'(t), ((\boldsymbol{X}_2(t) - \boldsymbol{\omega}_0))')'$.

Demostración. Las derivadas de $z(\boldsymbol{\theta})$ con respecto a $\boldsymbol{\beta}$ puede ser calculado y derivada con respecto a $\boldsymbol{\omega}$ es el siguiente:

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{\omega}} z(\boldsymbol{\theta}) &= \int_0^\tau \left[E\left[R(t)(-\boldsymbol{\beta}_3) I\left\{\boldsymbol{X}_2(t) > \boldsymbol{\omega}\right\} \exp(\boldsymbol{\beta}_0' \widetilde{\boldsymbol{X}}(t, \boldsymbol{\omega}_0))\right]\right. \\ &- \left.\frac{s(t;\boldsymbol{\theta}_0)}{s(t;\boldsymbol{\theta})}\left(\frac{\partial}{\partial \boldsymbol{\omega}} s(t;\boldsymbol{\theta})\right)\lambda_0(t)dt\right]\end{aligned}$$

Donde

$$\frac{\partial}{\partial \boldsymbol{\omega}} s(t;\boldsymbol{\theta}) = E\left[R(t)(-\boldsymbol{\beta}_3 I\left\{\boldsymbol{X}_2(t) > \omega\right\} \exp(\boldsymbol{\beta}' \widetilde{\boldsymbol{X}}(t, \boldsymbol{\omega})))\right]$$

La diferenciación y la integración pueden intercambiarse debido a la condición $C4$. La segunda derivada de $z(\boldsymbol{\theta})$ con respecto a $\boldsymbol{\omega}$ que existen porque la condición $C4$ y el teorema de diferenciación de Lebesgue es dado por

$$\begin{aligned}\frac{\partial^2}{(\partial \boldsymbol{\omega})^2} z(\boldsymbol{\theta}) &= \int_0^\tau E\left[R(t)(-\boldsymbol{\beta}_3) \exp(\boldsymbol{\beta}_0' \widetilde{\boldsymbol{X}}(t, \boldsymbol{\omega}_0))\middle| \boldsymbol{X}_2(t) = \boldsymbol{\omega}\right] f_{\boldsymbol{X}_2(t)}(\boldsymbol{\omega})\lambda_0(t)dt \\ &+ \int_0^\tau \frac{s(t;\boldsymbol{\theta}_0)}{s(t;\boldsymbol{\theta})^2}\left(\frac{\partial}{\partial \boldsymbol{\omega}} s(t;\boldsymbol{\theta})\right)^2 \lambda_0(t)dt \\ &- \int_0^\tau \frac{s(t;\boldsymbol{\theta}_0)}{s(t;\boldsymbol{\theta})}\left(\frac{\partial}{(\partial \boldsymbol{\omega})^2} s(t;\boldsymbol{\theta})\right)\lambda_0(t)dt\end{aligned}$$

Donde

$$\begin{aligned}\frac{\partial^2}{(\partial \boldsymbol{\omega})^2} s(t;\boldsymbol{\theta}) &= E\left[R(t)(-\boldsymbol{\beta}_3 I\left\{\boldsymbol{X}_2(t) > \omega\right\} \exp(\boldsymbol{\beta}' \widetilde{\boldsymbol{X}}(t, \boldsymbol{\omega})))\right] \\ &- E\left[R(t)(-\boldsymbol{\beta}_3) \exp(\boldsymbol{\beta}' \widetilde{\boldsymbol{X}}(t, \boldsymbol{\omega}))\middle| \boldsymbol{X}_2(t) = \boldsymbol{\omega}\right] f_{\boldsymbol{X}_2(t)}(\boldsymbol{\omega})\end{aligned}$$

Para $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ la segunda derivada parcial con respecto a $\boldsymbol{\omega}$ puede ser reescrito como

$$\begin{aligned}\frac{\partial^2}{(\partial\boldsymbol{\omega})^2}z(\theta_0) &= \int_0^\tau \left[\frac{1}{q_s}\left(\int(-\boldsymbol{\beta}_{30})I\left\{\boldsymbol{X}_2(t)>\boldsymbol{\omega}_0\right\}dQ^s\right)^2\right.\\ &- \left.\int(-\boldsymbol{\beta}_{30})^2 I\left\{\boldsymbol{X}_2(t)>\boldsymbol{\omega}_0\right\}\right]\lambda_0(s)ds\\ &= -\int_0^\tau q_s Cov_{p^s}(-\boldsymbol{\beta}_{30}I\left\{\boldsymbol{X}_2(t)>\boldsymbol{\omega}_0\right\})\lambda_0(s)ds\end{aligned}$$

Con P^s y Q^s definido en la condición $C4$. Otros cálculos largos también incluyendo los otros elementos de H muestran que $H(\boldsymbol{\theta}_0)$ is dado como lo anterior. □

Teorema 1. *Bajo las condiciones $C1-C4$ hay una vecindad de $\boldsymbol{\Theta}_0$ de $\boldsymbol{\theta}_0$ de modo que si $\widehat{\boldsymbol{\theta}}_n$ radica en $\boldsymbol{\Theta}_0$, esto sigue que $\widehat{\boldsymbol{\theta}}_n$ converge en probabilidad a $\boldsymbol{\theta}_0$ cuando $n\to\infty$.*

Demostración. Por Lema 1 se sabe que $\boldsymbol{Z}_n$ converge uniformemente a z. Por lo tanto basta demostrar que z es estrictamente cóncava en una vecindad $\Theta_0\subset\Theta$ y alcanza un máximo en $\boldsymbol{\theta}_0$. Esto puede ser verificado que $\frac{\partial}{\partial\boldsymbol{\omega}}z(\boldsymbol{\theta}_0)=0$ y $\frac{\partial}{\partial\boldsymbol{\beta}}z(\boldsymbol{\theta}_0)=0$.

Además la condición $C4$ asigna que la matriz Hessiana $H(\boldsymbol{\theta}_0)$ es definida negativa. Desde que H es continua en $\boldsymbol{\theta}$ y definido negativo en $\boldsymbol{\theta}_0$ existe una vecindad Θ_0 de $\boldsymbol{\theta}_0$ en la cual H es definida negativa y por lo tanto z es estrictamente cóncavo en Θ_0 □

Definiciones que se necesitan para la validación de la convergencia 3.4

B.4. Resultados teóricos necesarios para el análisis de la tasa de convergencia

Lema 3. *Bajo las condiciones $C1$ y $C4$, para ϵ suficientemente pequeño, existe una constante $\alpha>0$ de modo que para todo $\boldsymbol{\theta}$ en $V_\epsilon(\boldsymbol{\theta}_0)$, $z(\boldsymbol{\theta})-z(\boldsymbol{\theta}_0)\leq-\alpha\left\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\right\|^2$.*

Demostración. Para $z(\boldsymbol{\theta})=E\left[\int_0^\tau(\boldsymbol{\beta}'\widetilde{\boldsymbol{X}}(t,\boldsymbol{\omega})-\log(s(t,\boldsymbol{\theta})))\lambda(t,\boldsymbol{\theta}_0)\right]dt$ se sabe que $\frac{\partial}{\partial\boldsymbol{\omega}}z(\boldsymbol{\theta}_0)=0$ y $\frac{\partial}{\partial\boldsymbol{\beta}}z(\boldsymbol{\theta}_0)=0$. Por lo tanto por la expansión de Taylor de $z(\boldsymbol{\theta})$ para $\in$ suficientemente pequeña y para $\boldsymbol{\theta}$ en $V_\in(\boldsymbol{\theta}_0)$

$$\begin{aligned}z(\boldsymbol{\theta})-z(\boldsymbol{\theta}_0) &= \frac{\partial}{\partial\boldsymbol{\omega}}z(\boldsymbol{\theta}_0)(\boldsymbol{\omega}-\boldsymbol{\omega}_0)+\frac{\partial}{\partial\boldsymbol{\beta}}z(\boldsymbol{\theta}_0(\boldsymbol{\beta}-\boldsymbol{\beta}_0))\\ &+ \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)'\boldsymbol{H}(\boldsymbol{\theta}_0)(\boldsymbol{\theta}-\boldsymbol{\theta}_0)+o(\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|^2)\\ &\leq -\alpha\left\|\boldsymbol{\theta}-\boldsymbol{\theta}\right\|^2\end{aligned}$$

Ya que $\boldsymbol{H}(\boldsymbol{\theta}_0)$ es definida negativa □

Teorema 2. *Bajo las condiciones $C1$ y $C4$, $\sqrt{n}\left\|\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\right\|=Op(1)$.*

Demostración. Dejemos $\epsilon>0$ ser suficientemente pequeño para asegurarnos que el Lema3 contiene $V_\epsilon(\boldsymbol{\theta}_0)$. Por el teorema 1 conocemos que $\widehat{\boldsymbol{\theta}}_n$ converge a $\boldsymbol{\theta}_0$ en una vecindad de $\boldsymbol{\theta}_0$, es decir, $P(\widehat{\boldsymbol{\theta}}_n\in V_\epsilon(\boldsymbol{\theta}_0))>1-\eta$ para un n suficientemente largo y algún $\eta>0$.

Ahora para cada n, el parámetro conjunto $V_\epsilon(\boldsymbol{\theta}_0)$ $\boldsymbol{\theta}_0$ puede ser particionado en subconjunto $H_{n,j} = \left\{\boldsymbol{\theta} \in V_\epsilon(\boldsymbol{\theta}_0) : 2^J \sqrt{n}\,\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq 2_{J+1}\right\}$, j $\in Z$. basado en las ideas de Ibragimov et al. (1981) y para un n suficientemente largo tenemos

$$
\begin{aligned}
& P \quad \left(\sqrt{n}\left\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right\| > M\right) \\
\leq & \quad P\left(\sup_{\substack{\boldsymbol{\theta}\in V_\epsilon(\boldsymbol{\theta}_0) \\ M\leq\sqrt{n}\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|}} Z_n(\boldsymbol{\theta}) \geq Z_n(\boldsymbol{\theta}_0)\right) + \eta \\
\leq & \quad \sum_{\{j:2^j>M\}} P\left(\sup_{H_{n,j}} Z_n(\boldsymbol{\theta}) - Z_n(\boldsymbol{\theta}_0) \geq 0\right) + \eta \\
= & \quad \sum_{\{j:2^j>M\}} P\left(\sup_{H_{n,j}} \left(W_n(\boldsymbol{\theta}) - W_n(\boldsymbol{\theta}_0)\right) \geq -\sqrt{n}\left(z(\boldsymbol{\theta}) - z(\boldsymbol{\theta}_0)\right)\right) + \eta \\
\leq & \quad \sum_{\{j:2^j>M\}} P\left(\sup_{H_{n,j}} \left(W_n(\boldsymbol{\theta}) - W_n(\boldsymbol{\theta}_0)\right) \geq \sqrt{n}\alpha\,\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2\right) + \eta \\
= & \quad \sum_{\{j:2^j>M\}} P\left(\sup_{H_{n,j}} \sqrt{n}\left(W_n(\boldsymbol{\theta}) - W_n(\boldsymbol{\theta}_0)\right) \geq n\alpha\,\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2\right) + \eta \\
\leq & \quad \sum_{\{j:2^j>M\}} P\left(\sup_{H_{n,j}} \sqrt{n}\left(W_n(\boldsymbol{\theta}) - W_n(\boldsymbol{\theta}_0)\right) \geq \alpha\, 2^{2j}\right) + \eta \\
\leq & \quad \sum_{\{j:2^j>M\}} \frac{E\left[\sup_{H_{n,j}} |W_n(\boldsymbol{\theta}) - W_n(\boldsymbol{\theta}_0)|\right]}{\alpha n^{-1/2} 2^{2j}} + \eta \\
\leq & \quad \sum_{\{j:2^j>M\}} \frac{\kappa}{\alpha 2^{j-1}} + \eta,
\end{aligned}
$$

Desde que Lema 3, Lema 4 mantienen mediante el uso de la desigualdad de Markov. Esto prueba la afirmación. □

Lema 4. *Bajo las condiciones C1 y C4, para todo $\epsilon > 0$ existe una constante $k > 0$ de modo que $E\left[\sup_{\boldsymbol{\theta}\in V_\epsilon(\boldsymbol{\theta}_0)} |W_n(\boldsymbol{\theta}) - W_n(\boldsymbol{\theta}_0)|\right] \leq k\epsilon$, para todo $n \in \mathbb{N}$.*

Demostración. Sea $\overline{\boldsymbol{\beta}} = \left(\boldsymbol{\beta}_1', \boldsymbol{\beta}_2'\right)'$ y $\overline{S}(t, \boldsymbol{\theta}) := \frac{1}{n}\sum_{i=1}^n R_i(t)\exp(\boldsymbol{\beta}'\widetilde{\boldsymbol{X}}_i(t))$. Se reescribe $W_n(\boldsymbol{\theta}) - W_n(\boldsymbol{\theta}_0) := W_{1n}(\boldsymbol{\theta}) - W_{2n}(\boldsymbol{\theta})$.

$$
\begin{aligned}
W_{1n}(\boldsymbol{\theta}) &= n^{-1/2}(\overline{\boldsymbol{\beta}}-\overline{\boldsymbol{\beta}}_0)' \sum_{i=1}^{n}\left[\int_0^\tau E\begin{pmatrix}\boldsymbol{X}_{1i}(t)\\ \boldsymbol{X}_{2i}(t)\end{pmatrix} dN_i(t) - \int_0^\tau E\begin{pmatrix}\boldsymbol{X}_{1i}(t)\\ \boldsymbol{X}_{2i}(t)\end{pmatrix}\lambda(t,\boldsymbol{\theta}_0)dt\right] \\
&+ n^{-1/2}\boldsymbol{\beta}_3' \sum_{i=1}^{n}\left[\int_0^\tau (\boldsymbol{X}_{2i}(t)-\boldsymbol{\omega})dN_i(t) - \int_0^\tau E(\boldsymbol{X}_{2i}(t)-\boldsymbol{\omega})\lambda(t,\boldsymbol{\omega}_0 dt)\right] \\
&- n^{-1/2}\boldsymbol{\beta}_{30}' \sum_{i=1}^{n}\left[\int_0^\tau (\boldsymbol{X}_{2i}(t)-\boldsymbol{\omega}_0)dN_i(t) - \int_0^\tau E(\boldsymbol{X}_{2i}(t)-\boldsymbol{\omega}_0)\lambda(t,\boldsymbol{\omega}_0 dt)\right] \\
&= n^{-1/2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)' \sum_{i=1}^{n}\left[\int_0^\tau \widetilde{\boldsymbol{X}}_i(t,\boldsymbol{\omega}_0)dN_i(t) - \int_0^\tau E\left[\widetilde{\boldsymbol{X}}_i(t,\boldsymbol{\omega}_0)\right]\lambda(t,\boldsymbol{\theta}_0)dt\right] \\
&+ n^{-1/2}\boldsymbol{\beta}_3' \left[\sum_{i=1}^{n}\int_0^\tau (\boldsymbol{X}_{2i}(t)-\boldsymbol{\omega}) - (\boldsymbol{X}_{2i}(t)-\boldsymbol{\omega}_0)dN_i(t)\right. \\
&- \left.\int_0^\tau \left[E(\boldsymbol{X}_{2i}(t)-\boldsymbol{\omega}) - E(\boldsymbol{X}_{2i}(t)-\boldsymbol{\omega}_0)\right]\lambda(t,\boldsymbol{\theta}_0)dt\right]
\end{aligned}
$$

y

$$
\begin{aligned}
W_{2n}(\boldsymbol{\theta}) &= \sqrt{n}\left(\int_0^\tau \log\left(\bar{S}(t,\boldsymbol{\theta})\right)d\bar{N}(t) - \int_0^\tau \log\left(s(t,\boldsymbol{\theta})\right)s(t,\theta_0)d\bar{\Lambda}_0(t)\right. \\
&- \left.\int_0^\tau \log\left(\bar{S}(t,\boldsymbol{\theta}_0)\right)d\bar{N}(t) + \int_0^\tau \log\left(s(t,\boldsymbol{\theta}_0)\right)s(t,\theta_0)d\bar{\Lambda}_0(t)\right) \\
&= n^{-1/2}\sum_{i=1}^{n}\left[\int_0^\tau \log\left(\frac{s(t,\boldsymbol{\theta})}{s(t,\boldsymbol{\theta}_0)}\right)dN_i(t) - \int_0^\tau \log\left(\frac{s(t,\boldsymbol{\theta})}{s(t,\boldsymbol{\theta}_0)}\right)s(t,\boldsymbol{\theta}_0)d\Lambda_0(t)\right] \\
&+ n^{-1/2}\sum_{i=1}^{n}\left[\int_0^\tau \log\left(\frac{\bar{S}(t,\boldsymbol{\theta})}{s(t,\boldsymbol{\theta})}\right) - \log\left(\frac{\bar{S}(t,\boldsymbol{\theta}_0)}{s(t,\boldsymbol{\theta}_0)}\right)dN_i(t)\right]
\end{aligned}
$$

Primero se contempla $W_{1n}(\boldsymbol{\theta})$. La expectativa del supremo de valor absoluto del primer término es $O(\in)$, desde $\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|<\epsilon$. En el segundo término se considera los vectores componente a componente. Los conjuntos de las funciones $\{f_\omega : \omega \in [\omega_1,\omega_2]\}$ y $\{g_\omega : \omega \in [\omega_1,\omega_2]\}$ con $f_\omega(a,b) = abI\{b>\omega\}$ y $g_\omega(a,b) = a\omega I[b>\omega]$ forma la clase Vapnik - Cervonenkis. Desde $\int_0^\tau (\boldsymbol{X}_{2i}(t)-(\boldsymbol{\omega}_0-\epsilon))dN_i(t)$ es una función envolvente para $\int_0^\tau (\boldsymbol{X}_{2i}(t)-\omega)dN_i(t)$ en $V_\epsilon(\boldsymbol{\theta}_0)$. Por lo tanto, la norma $L_2(P)$ para la función envolvente está limitado por

$$
\begin{aligned}
&E\sup_{\boldsymbol{\omega}\in V_\epsilon(\omega_0)}\left\|\int_0^\tau \left[(\boldsymbol{X}_{2i}(t)-\boldsymbol{\omega}) - (\boldsymbol{X}_{2i}(t)-\boldsymbol{\omega}_0)\right]dN_i(t)\right\| \\
&\leq \left\{E\int_0^\tau \left\|(\boldsymbol{X}_{2i}(t)-(\boldsymbol{\omega}-\epsilon)) - (\boldsymbol{X}_{2i}(t)-\boldsymbol{\omega}_0)\right\|^2 dN_i(t)\right\}^{1/2} = O_(\epsilon)
\end{aligned}
$$

La acotación de

$$E \sup_{\omega \in V_\epsilon(\omega_0)} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^\tau \left[(\boldsymbol{X}_{2i}(t) - \boldsymbol{\omega}) - (\boldsymbol{X}_{2i}(t) - \boldsymbol{\omega}_0) \right] dN_i(t) \right.$$
$$\left. - \int_0^\tau \left[E(\boldsymbol{X}_{2i}(t) - \boldsymbol{\omega}) - E(\boldsymbol{X}_{2i}(t) - \boldsymbol{\omega}_0) \right] \lambda(t, \boldsymbol{\theta}_0) dt \right\|$$

□

Es una consecuencia del teorema 2.14.1 de Vaart y Wellner (1996). Ahora considerando $\boldsymbol{W}_{2n}(\boldsymbol{\theta})$. Para la clase de funciones $\left\{ \log \left(\frac{s(t,\boldsymbol{\theta})}{s(t,\boldsymbol{\theta})_0} \right) : \boldsymbol{\theta} \in V_\epsilon(\boldsymbol{\theta}_0) \right\}$ se puede demostrar que tiene una función sobre la norma $L_2(P)$ de orden $O_{(\epsilon)}$ y su integral agrupada $L_2(P)$ es finita. Por lo tanto como consecuencia del teorema 2.14.2 de Vaart y Wellner (1996) la cota de

$$E \left[\sup_{\boldsymbol{\theta} \in V_\epsilon(\boldsymbol{\theta}_0)} \left| n^{-1/2} \sum_{i=1}^{n} \left[\int_0^\tau \log \left(\frac{s(t,\boldsymbol{\theta})}{s(t,\boldsymbol{\theta}_0)} \right) dN_i(t) \quad - \int_0^\tau \log \left(\frac{s(t,\boldsymbol{\theta})}{s(t,\boldsymbol{\theta}_0)} \right) s(t,\boldsymbol{\theta}_0) d\Lambda_0(t) \right] \right| \right]$$

Es de orden $O(\epsilon)$. Después de usar la expansión de Taylor del logaritmo en 1 , el segundo término puede ser tratado de manera similar. Mas detalles de estos se encuentran en Pons (2003) quien comprobó esto en el caso de que $N_i(t)$ saltara solo una vez.

B.5. Normalidad Asintótica

Usamos el teorema 3 que establece la normalidad asintótica de M-estimadores en el caso la función de criterio de Lipshitz y sus funciones límites admiten la expansión de Taylor de segundo orden. Considerar la función de criterio

$$m_{\boldsymbol{\theta}} = m_{\boldsymbol{\theta}}(x) = \int_0^\tau \left[\left(\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \boldsymbol{\beta}_3' \right) \begin{pmatrix} x_1(t) \\ x_2(t) \\ (x_2(t) - \omega)^+ \end{pmatrix} - \log(s(t,\boldsymbol{\theta})) \right] dN(t)$$

Y la matriz $\Delta(\boldsymbol{\theta}_0) = E\dot{m}_{\boldsymbol{\theta}_0} \dot{m}_{\boldsymbol{\theta}_0}'$, donde $\dot{m}_{\boldsymbol{\theta}_0}$ está dado por

$$\dot{m}_{\boldsymbol{\theta}_0} = \begin{pmatrix} -\int_0^\tau \left(\boldsymbol{\beta}_{30} I\left\{ x_2(t) > \omega_0 \right\} + \frac{1}{s(t,\boldsymbol{\theta}_0)} \frac{\partial}{\partial \omega} s(t,\boldsymbol{\theta}_0) \right) dN(t) \\ \int_0^\tau \left(x_1(t) - \frac{1}{s(t,\boldsymbol{\theta}_0)} \frac{\partial}{\partial \boldsymbol{\beta}_1} s(t,\boldsymbol{\theta}_0) \right) dN(t) \\ \int_0^\tau \left(x_2(t) - \frac{1}{s(t,\boldsymbol{\theta}_0)} \frac{\partial}{\partial \boldsymbol{\beta}_2} s(t,\boldsymbol{\theta}_0) \right) dN(t) \\ \int_0^\tau \left((x_2(t) - \omega_0)^+ - \frac{1}{s(t,\boldsymbol{\theta}_0)} \frac{\partial}{\partial \boldsymbol{\beta}_3} s(t,\boldsymbol{\theta}_0) \right) dN(t) \end{pmatrix}$$

Desde $H(\boldsymbol{\theta})$ y $\Delta(\boldsymbol{\theta})$ son continuas en $\boldsymbol{\theta}_0$, ellas pueden ser estimadas consistentemente por $H(\widehat{\boldsymbol{\theta}})$ y $\Delta(\widehat{\boldsymbol{\theta}})$.

Teorema 3. *Bajo las condiciones C1 y C4, y bajo la asunción de que $\widehat{\boldsymbol{\theta}}_n$ es un estimador consistente de $\boldsymbol{\theta}_0$*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \longrightarrow N(0, H(\boldsymbol{\theta}_0)^{-1} \Delta(\boldsymbol{\theta}_0) H(\boldsymbol{\theta}_0)^{-1}) \; n \to \infty$$

,

Donde $H(\boldsymbol{\theta}_0)$ es dado en Lema 2.

Demostración. Reescribir la función objetivo como sigue

$$m_{\boldsymbol{\theta}} = (\boldsymbol{\beta}_1^{'}, \boldsymbol{\beta}_1^{'}) \int_0^{\tau} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} dN_t + \boldsymbol{\beta}_3^{'} \int_0^{\tau} (x_2(t) - \omega) dN(t) - \int_0^{\tau} \log(s(t, \boldsymbol{\theta})) dN(t)$$

La función $x \mapsto m_{\boldsymbol{\theta}}(x)$ es una función medible de modo que $\boldsymbol{\theta} \mapsto m_{\boldsymbol{\theta}}(x)$ es diferenciable en $\boldsymbol{\theta}_0$ para $P - almost$ casi todo x porque la condición $C1$. Esto puede ser fácilmente visto, que el primer término de $m_{\boldsymbol{\theta}}$ es Lipshitz en $\boldsymbol{\beta}_1$ y $\boldsymbol{\beta}_2$, desde que es linear en $\boldsymbol{\beta}_1$ y $\boldsymbol{\beta}_2$. El segundo término es Lipshitz en una vecindad de $\boldsymbol{\theta}_0$, desde

$$\begin{aligned} &\left\| \int_0^{\tau} \widetilde{\boldsymbol{\beta}}_3^{'} (x_2(t) - \widetilde{\omega})^+ dN(t) - \int_0^{\tau} \widetilde{\boldsymbol{\beta}}_3^{'} (x_2(t) - \omega)^+ dN(t) \right\| \\ &\leq \|\omega - \widetilde{\omega}\| \, N(\tau) \left\| \boldsymbol{\beta}_3^{'} \right\| + \left\| \widetilde{\boldsymbol{\beta}}_3^{'} - \boldsymbol{\beta}_3^{'} \right\| \int_0^{\tau} \|(x_2(t) - \widetilde{\omega})\| \, dN(t) \end{aligned}$$

Ahora para el tercer término por la expansión de Taylor en $\boldsymbol{\theta}$

$$log(s(t, \widetilde{\boldsymbol{\theta}})) - log(s(t, \boldsymbol{\theta})) = \frac{\frac{\partial}{\partial \boldsymbol{\beta}} s(t, \boldsymbol{\theta}')}{s(t, \boldsymbol{\theta}')} (\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) + \frac{\frac{\partial}{\partial \omega} s(t, \boldsymbol{\theta}')}{s(t, \boldsymbol{\theta}')} (\widetilde{\omega} - \omega)$$

Donde $\boldsymbol{\theta}^{'}$ esta en el segmento de línea entre $\boldsymbol{\theta}$ y $\boldsymbol{\theta}^{'}$. Las derivadas parciales son uniformes delimitadas lejos de cero por las condiciones $C2$ y $C3$.Por lo tanto el último término es Lipschitz en $\boldsymbol{\theta}$.

Ademas el mapa $\boldsymbol{\theta} \mapsto Em_{\boldsymbol{\theta}} = z(\boldsymbol{\theta})$ admite la expansión de Taylor de segundo orden en $\boldsymbol{\theta}_0$, con la matriz simétrica hessiana no singular $H(\boldsymbol{\theta}_0)$ dada en Lema 2.

Finalmente, desde que $\widehat{\boldsymbol{\theta}}_n$ es consistente para $\boldsymbol{\theta}_0$ en una vecindad de $\boldsymbol{\theta}_0$, esto sigue que $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ es asintóticamente normal con matriz de covarianza $H(\boldsymbol{\theta}_0)^{-1} \Delta(\boldsymbol{\theta}_0) H(\boldsymbol{\theta}_0)^{-1}$ por el teorema 5.23 en der Vaart (1998).

□

Teorema 4. *Bajo las condiciones $C1$ y $C4$, el proceso*

$$\sqrt{n}(\widehat{\Lambda}_n(t) - \Lambda_0(t)) + \sqrt{n}(\widehat{\boldsymbol{\beta}}_n(t) - \boldsymbol{\beta}_0(t))' \int_0^{'} \frac{E[R(s) \widetilde{X}(\mu, \omega_0) \exp \boldsymbol{\beta}_0^{'} \widetilde{X}(\mu, \omega_0)]}{s(\mu, \boldsymbol{\theta}_0)}$$

Converge débilmente a la media cero del proceso gausiano con covarianza $\int_0^{s \wedge t} \frac{1}{s(\mu, \boldsymbol{\theta}_0)} d\Lambda_0(\mu)$, s , $t \in [0, \tau]$ y $\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$ los procesos anteriores son asintóticamente independientes.

Demostración.

$$\begin{aligned}
&\sqrt{n} \quad (\widehat{\Lambda}_n(t) - \Lambda_0(t)) \\
&= \quad \sqrt{n}\left\{\int_0^{'} \frac{d(n\overline{N})(\mu)}{S(\mu,\widehat{\boldsymbol{\theta}}_n)} - \Lambda_0(t)\right\} \\
&= \quad \sqrt{n}\left\{\int_0^{'} \frac{d(n\overline{N})(\mu)}{S(\mu,\widehat{\boldsymbol{\theta}}_n)} - \int_0^{'} \frac{S(\mu,\boldsymbol{\theta}_0}{S(\mu,\widehat{\boldsymbol{\theta}}_n)} d\Lambda_0(\mu)\right. \\
&+ \quad \left.\int_0^{'} \frac{S(\mu,\boldsymbol{\theta}_0}{S(\mu,\widehat{\boldsymbol{\theta}}_n)} d\Lambda_0(\mu) - \int_0^{'} \frac{S(\mu,\widehat{\boldsymbol{\theta}}_n}{S(\mu,\widehat{\boldsymbol{\theta}}_n)} d\Lambda_0(\mu)\right\} \\
&= \quad \sqrt{n}\left\{\int_0^{'} \frac{d(n\overline{N})(\mu) - S(\mu,\boldsymbol{\theta}_n) d\Lambda_0(\mu)}{S(\mu,\widehat{\boldsymbol{\theta}}_n)} - \int_0^{'} \frac{S(\mu,\widehat{\boldsymbol{\theta}}_n) - S(\mu,\boldsymbol{\theta}_0)}{S(\mu,\widehat{\boldsymbol{\theta}}_n)} d\Lambda_0(\mu)\right\} \\
&= \quad \int_0^{'} \frac{d[n^{1/2}\overline{M}(\mu)]}{n^{-1}S(\mu,\widehat{\boldsymbol{\theta}}_n)} - \int_0^{'} \frac{n^{-1/2}[S(\mu,\widehat{\boldsymbol{\theta}}_n) - S(\mu,\boldsymbol{\theta}_0)]}{n^{-1}S(\mu,\widehat{\boldsymbol{\theta}}_n)} d\Lambda_0(\mu)
\end{aligned}$$

Donde $\overline{M}(\mu) = \frac{1}{n}\sum_{i=1}^{n} M_i(\mu)$. El primer término en la última expresión converge a un proceso gausiano centrado con covarianza $\int_0^{s\wedge t} \frac{1}{s(\mu,\boldsymbol{\theta}_0)} d\Lambda_0(\mu)$ por el teorema de Rebolledos.

El segundo término puede ser manejado como sigue: Una expansión de Taylor en $\boldsymbol{\beta}_0$ rendimientos

$$n^{-1/2}\left(S(\mu,\widehat{\boldsymbol{\theta}}_n) - S(\mu,\widehat{\boldsymbol{\theta}}_0)\right) = n^{-1/2}\left(S(\mu,\boldsymbol{\beta}_0,\widehat{\omega}) - S(\mu,\boldsymbol{\beta}_0,\omega_0)\right)$$

$$+n^{-1/2}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)^{'}\left(\sum_{i=1}^{n} R_i(\mu)\widetilde{X}_i(\mu,\widehat{\omega})\exp\left\{\boldsymbol{\beta}_*^{'}\widetilde{X}_i(\mu,\widehat{\omega})\right\}\right)$$

,

Donde $\boldsymbol{\beta}_*$ esta en la línea de segmento entre $\boldsymbol{\beta}_0$ y $\widehat{\boldsymbol{\beta}}_0$. El primer término de la expansión de Taylor converge uniformemente en el $\in [0,\tau]$ a cero en probabilidad usando el teorema de mapeo continuo, desde S es una función continua en ω. Por la ley de grandes números dado por Andersen y Gill (1982)

$$\sup_{\mu\in[0,\tau]} \sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}} \left|\frac{1}{n}\sum_{i=1}^{n} R_i(\mu)\widetilde{X}_i(\mu,\omega)\exp\boldsymbol{\beta}^{'}\widetilde{X}_i(\mu,\omega) - E[R_i(\mu)\widetilde{X}_i(\mu,\omega)\exp\boldsymbol{\beta}_T\widetilde{X}_i(\mu,\omega)]\right| \xrightarrow{P} 0$$

y

$$sup_{\mu\in[0,\tau]} \sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}} \left|n^{-1}S(\mu,\boldsymbol{\theta}) - s(\mu,\boldsymbol{\theta})\right| \xrightarrow{P} 0$$

La independencia asintótica sigue de la aproximación

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) = I^{-1}(\boldsymbol{\theta}_0).n^{1/2}\frac{\partial}{\partial\boldsymbol{\beta}}\log L(\boldsymbol{\theta}_0) + Op(1)$$

,

Donde $n^{-1}\frac{\partial^2}{(\partial\boldsymbol{\beta})^2}\log L(\boldsymbol{\theta}_0) \xrightarrow{P} I(\boldsymbol{\theta}_0)$ y

$$\frac{\partial}{\partial \boldsymbol{\beta}} \log L(\boldsymbol{\theta}_0) = \sum_{i=1}^{n} \int_0^{\tau} \widetilde{X}_i(\mu, \omega_0) dM_i(\mu)$$

$$- \int_0^{\tau} \frac{E[R_i(u)\widetilde{X}_i(\mu, \omega_0) \exp \boldsymbol{\beta}_0 \widetilde{X}_i(\mu, \omega_0)]}{S(\mu, \boldsymbol{\theta}_0)} d\overline{M}(\mu)$$

Desde $n^{-1/2}\frac{\partial}{\partial \boldsymbol{\beta}} \log L(\boldsymbol{\theta}_0)$ y $W(t) = \int_0^{'} \frac{n^{1/2} d\overline{M}(\mu)}{s(\mu, \boldsymbol{\theta}_0)}$ son asintóticamente gausiana con media cero y $E[W(t).n^{-1/2}\frac{\partial}{\partial \boldsymbol{\beta}} \log L(\boldsymbol{\theta}_0)] = 0$ para todo $t \in [0, \tau]$. □

B.6. Bondad de ajuste

Si el modelo con punto de cambio es adecuado o no debe ser examinado por medio de pruebas de bondad de ajuste. Se usan pruebas que fueron desarrolladas por Gandy y Jensen (2006) para una versión extendida de un tipo del modelo de regresión de Cox $\lambda_i(t) = \lambda_0(t)\rho_i(\boldsymbol{\beta}, t)$, donde $\lambda_0(t)$ es la función de riesgo base y $\rho_i(\boldsymbol{\beta}, t)$ es un proceso estocástico observable que puede depender de un vector de dimensión finita $\boldsymbol{\beta}$. En este caso tenemos solo el modelo básico de regresión de cox que es independiente del tiempo como hipótesis nula de modo que $\rho_i(\boldsymbol{\beta}, t) = R_i(t) \exp \boldsymbol{\beta}^{'} \boldsymbol{X}_i$. Donde el R_i toma valores 1 o 0 para indicar si el individuo está en riesgo o no. El test esta basado en sumas ponderadas de los residuos martingalas y el test estadístico esta dado por

$$T(c(\widehat{\boldsymbol{\vartheta}}, .)) = n^{-1/2} \sum_{i=1}^{n} \int_0^{\tau} c_i(\widehat{\boldsymbol{\vartheta}}, s) dN(s)$$

.

Los pesos $c_i(., .)$ son escogidos de modo que una simple distribución asintótica puede ser derivada y ademas que el test es potente contra algunas alternativas, los cuales son llamados modelos alternativos. Por lo tento el test estadístico no solo contiene el parámetro del modelo nulo sino también del modelo alternativo. El vector parámetro $\boldsymbol{\vartheta} = (\boldsymbol{\beta}^{'}, \boldsymbol{\gamma}^{'})^{'}$, donde $\boldsymbol{\beta}$ es un vector de parámetros del modelo nulo $\boldsymbol{\gamma}$ es un vector parámetros del modelo alternativo y es estimado por el estimador de máxima verosimilitud $\widehat{\boldsymbol{\vartheta}} = (\widehat{\boldsymbol{\beta}}^{'}, \widehat{\boldsymbol{\gamma}}^{'})^{'}$. En nuestro caso consideramos la hipótesis nula

$$H_0 : \lambda_i(t) = \lambda_0(t)\rho_i(\boldsymbol{\beta}, t)$$

.

Los modelos alternativos están dados

$$\lambda_i(t) = a(t)h_i(\gamma, t)$$

,

Donde $a(t)$ es una función de riesgo base sin especificar $\boldsymbol{\gamma}$ es un vector parámetro desconocido y los procesos estocásticos $h_i(\boldsymbol{\gamma}, t), i = 1, ..., n$ son observables. En nuestro caso tenemos el modelo básico de cox $\rho_i(\boldsymbol{\beta}, t) = R_i(t) \exp \boldsymbol{\beta}^{'} \boldsymbol{X}_i$ como hipótesis nula y el modelo de punto de cambio $h_i(\gamma, t) = R_i(t) \exp \gamma^{'} \widetilde{\boldsymbol{X}}_i(t, \omega_0)$ como un modelo competitivo. Bajo la hipótesis nula (y algunas leves decisiones técnicas) el test estadístico es asintóticamente normal:

$$T(c(\widehat{\boldsymbol{\vartheta}}, t)) \xrightarrow{d} N(0, \sigma^2) \tag{B.2}$$

Donde σ^2 puede ser estimado consistentemente por $\widehat{\sigma}^2(c) = n^{-1} \sum_{i=1}^{n} \int_0^\tau c_i^2(\widehat{\vartheta}, s) dN_i(s)$. Estudios de simulación muestran que el test se realiza bien inclusive para muestras de tamaño moderado. Para la elección explícita de tamaños y todos los otros detalles referirse a Gandy y Jensen (2006)

Apéndice C

Implementación de los programas

C.1. Cálculo de los estadísticos en R

Se muestra el código desarrollado en R (2012) para el cálculo de los estadísticos que se encuentran en la sección 4.2

```
################################### VARIABLES ######################################
#ORDEN
#MES
#START
#STOP
#EVENT
#MSISDN
#MES_RECIBO
#EDAD
#MONTO_RECIBO
#TRAF_VOZ
#MIN_VOZ
#TRAF_DATOS
#MB_DATOS
#CARGO_FIJO
#FEC_NACIMIENTO
#CUSTOMER_ID
#CO_ID
#FEC_ACTIVACION
#CH_STATUS
#FEC_ESTADO
#DES_MOTIVO_EST
#SEXO
#TIEMPO_VIDA_MESES
################################### FIN VARIABLES ##################################

library(survival)
library(plotrix)
###################### ANÁLISIS DESCRIPTIVO MUESTRA VS REAL ##########################
#Carga de información
data_poblacion="C://Users//LUCIA//Google Drive//Tesis Final//Codigo//BASE_R_FINAL_25.csv"
data_muestra="C://Users//LUCIA//Google Drive//Tesis Final//Codigo//BASE_R_FINAL_25_MUESTRA.csv"

lineas_poblacion=read.table(data_poblacion,header=T,sep=",")
lineas_muestra=read.table(data_muestra,header=T,sep=",")
```

```
casos_poblacion = diff(c(0,lineas_poblacion$ORDEN)) != 0
casos_muestra = diff(c(0,lineas_muestra$ORDEN_2)) != 0

#gráfico de histogramas
op = par(mfrow = c(4, 2))

hist(lineas_poblacion$EDAD[casos_poblacion],
main = paste("Población: Histograma de","Edad") ,
xlab = "Edad",ylab="Frecuencia",breaks=c(10,20,30,40,50,60,70,80,90))

hist(lineas_muestra$EDAD[casos_muestra],main = paste("Muestra: Histograma de","Edad"),
xlab = "Edad",ylab="Frecuencia",breaks=c(10,20,30,40,50,60,70,80,90),col="gray")


hist(lineas_poblacion$CARGO_FIJO[casos_poblacion],
main = paste("Población: Histograma de","Cargo Fijo") ,
xlab = "Cargo Fijo",ylab="Frecuencia",breaks=c(10,30,50,70,90,110,130,150,170))

hist(lineas_muestra$CARGO_FIJO[casos_muestra],
main = paste("Muestra: Histograma de","Cargo Fijo"),
xlab = "Cargo Fijo",ylab="Frecuencia",breaks=c(10,30,50,70,90,110,130,150,170),col="gray")


hist(lineas_poblacion$MIN_VOZ[casos_poblacion],
main = paste("Población: Histograma de","Tráfico de voz") ,
xlab = "Tráfico de Voz",ylab="Frecuencia")

hist(lineas_muestra$MIN_VOZ[casos_muestra],
main = paste("Muestra: Histograma de","Tráfico de voz") ,
xlab = "Tráfico de Voz",ylab="Frecuencia",col="gray")


hist(lineas_poblacion$MB_DATOS[casos_poblacion],
main = paste("Población: Histograma de","Tráfico de datos") ,
xlab = "Tráfico de datos",ylab="Frecuencia")

hist(lineas_muestra$MB_DATOS[casos_muestra],
main = paste("Muestra: Histograma de","Tráfico de datos") ,
xlab = "Tráfico de datos",ylab="Frecuencia",col="gray")


#cálculo de estadísticos
#Edad
IQR(lineas_muestra$EDAD[casos_muestra]) #el rango intercuartil
mean(lineas_muestra$EDAD[casos_muestra]) # media
summary(lineas_muestra$EDAD[casos_muestra]) # min y max
median(lineas_muestra$EDAD[casos_muestra]) # min y max
edad = table(lineas_muestra$EDAD[casos_muestra])
edad[which(edad == max(edad))[1]] # moda
sd(lineas_muestra$EDAD[casos_muestra]) #desviacion estandar
var(lineas_muestra$EDAD[casos_muestra]) #varianza
```

```
std.error(lineas_muestra$EDAD[casos_muestra]) #standart error o error standar
skewness=function(x){ m3=mean((x-mean(x))^3)
skew=m3/(sd(x)^3)
skew }
skewness(lineas_muestra$EDAD[casos_muestra]) # asimetria
kurtosis=function(x) {
m4=mean((x-mean(x))^4)
kurt=m4/(sd(x)^4)-3
kurt}
kurtosis(lineas_muestra$EDAD[casos_muestra])

#Monto Facturado
IQR(lineas_muestra$CARGO_FIJO[casos_muestra]) #el rango intercuartil
mean(lineas_muestra$CARGO_FIJO[casos_muestra]) # media
summary(lineas_muestra$CARGO_FIJO[casos_muestra]) # min y max
median(lineas_muestra$CARGO_FIJO[casos_muestra]) # min y max
cargo = table(lineas_muestra$CARGO_FIJO[casos_muestra])
cargo[which(cargo== max(cargo))[1]] # moda
sd(lineas_muestra$CARGO_FIJO[casos_muestra]) #desviacion estandar
var(lineas_muestra$CARGO_FIJO[casos_muestra]) #varianza
std.error(lineas_muestra$CARGO_FIJO[casos_muestra]) #standart error o error standar
skewness(lineas_muestra$CARGO_FIJO[casos_muestra]) # asimetria
kurtosis(lineas_muestra$CARGO_FIJO[casos_muestra]) # kurtosis


# tráfico de datos
IQR(lineas_muestra$MB_DATOS[casos_muestra]) #el rango intercuartil
mean(lineas_muestra$MB_DATOS[casos_muestra]) # media
summary(lineas_muestra$MB_DATOS[casos_muestra]) # min y max
median(lineas_muestra$MB_DATOS[casos_muestra]) # min y max
datos = table(lineas_muestra$MB_DATOS[casos_muestra])
datos[which(datos== max(datos))[1]] # moda
sd(lineas_muestra$MB_DATOS[casos_muestra]) #desviacion estandar
var(lineas_muestra$MB_DATOS[casos_muestra]) #varianza
std.error(lineas_muestra$MB_DATOS[casos_muestra]) #standart error o error standar
skewness(lineas_muestra$MB_DATOS[casos_muestra]) # asimetria
kurtosis(lineas_muestra$MB_DATOS[casos_muestra]) # kurtosis

# tráfico de voz
IQR(lineas_muestra$MIN_VOZ[casos_muestra]) #el rango intercuartil
mean(lineas_muestra$MIN_VOZ[casos_muestra]) # media
summary(lineas_muestra$MIN_VOZ[casos_muestra]) # min y max
median(lineas_muestra$MIN_VOZ[casos_muestra]) # min y max
voz = table(lineas_muestra$MIN_VOZ[casos_muestra])
voz[which(voz== max(voz))[1]] # moda
sd(lineas_muestra$MIN_VOZ[casos_muestra]) #desviacion estandar
var(lineas_muestra$MIN_VOZ[casos_muestra]) #varianza
std.error(lineas_muestra$MIN_VOZ[casos_muestra]) #standart error o error standar
skewness(lineas_muestra$MIN_VOZ[casos_muestra]) # asimetria
kurtosis(lineas_muestra$MIN_VOZ[casos_muestra]) # kurtosis

#################### FIN ANÁLISIS DESCRIPTIVO MUESTRA VS REAL ############################
```

C.2. Identificación del punto de cambio en R

Se adjunta el código realizado en R para la identificación del rango en donde se encuentra al punto de cambio a través de los residuos de martingala.

```
###################### IDENTIFICACIÓN DEL PUNTO DE CAMBIO ################################
#ALGORITMO 2
#Se busca la covariable que tenga puntos de cambio,
#para esto se trabaja con el modelo de cox clásico
#se arma una base para que sea carga a R
summary(lineas_muestra)
#Se crea el objeto de sobrevivencia
lineas.surv = Surv(lineas_muestra$START, lineas_muestra$STOP, lineas_muestra$EVENT)
lineas1.cox = coxph(lineas.surv ~(MONTO_RECIBO+EDAD),data=lineas_muestra)
lineas2.cox = coxph(lineas.surv ~(MB_DATOS+EDAD),data=lineas_muestra)
lineas3.cox = coxph(lineas.surv ~(MIN_VOZ+EDAD),data=lineas_muestra)

#Estimación de la función de supervivencia
par(mfrow=c(1,2))

#S(t)
plot(survfit(lineas1.cox),conf.int=F,lty=c(1,3),main="Función de supervivencia"
,xlab= "Tiempo ",ylab=" Probabilidad de supervivencia",lwd=2,col ="blue")
axis(2, seq(0,1,.1))
#legend("bottomleft",c("Curva de supervivencia"),lty=1,col="blue")

plot(survfit(lineas1.cox),conf.int=F,fun="cumhaz",main="Función de riesgo acumulado"
,xlab= "Tiempo ",ylab=" Riesgo acumulado",lwd=2,col ="blue")
#legend("upleft",c("Curva de riesgo"),lty=1,col="blue")
#plot(basehaz(lineas1.cox,centered = TRUE))
#plot(Ft)

#S(t)= probabilidad que un individuo sobreviva mas que el tiempo determinado a priori
#f(t) = P[T > t] = 1 - F(t)
st=survfit(lineas1.cox)$surv
tt=survfit(lineas1.cox)$time
t=25
Ft = 1 -  st
ft = rep(0, len = t)
i=1
while(i <= t){
if(i==1){ft[i]=Ft[i]}
else{ft[i]=Ft[i]-Ft[i-1]}
i=i+1
}

par(mfrow=c(1,2))
plot(Ft,main="Función de dist acumulada F(t)"
,xlab= "Tiempo ",ylab="F(t)",lwd=2,col ="blue")
lines(lowess(tt,Ft),col='red')
plot(ft,main="Función de dist f(t)"
```

```
,xlab= "Tiempo ",ylab="f(t)",lwd=2,col ="blue")
lines(lowess(tt,ft),col='red')


#RESIDUOS DE MARTINGALA
#Esto será utilizado para determinar la forma funcional de la covariable.
par(mfrow=c(2,2))
lineas1.res = residuals(lineas1.cox,type=c("martingale"),collapse=lineas_muestra$ORDEN_2)
lineas2.res = residuals(lineas2.cox,type=c("martingale"),collapse=lineas_muestra$ORDEN_2)
lineas3.res = residuals(lineas3.cox,type=c("martingale"),collapse=lineas_muestra$ORDEN_2)

cases = diff(c(0,lineas_muestra$ORDEN_2)) != 0
plot(lineas_muestra$EDAD[cases],lineas1.res,xlab = "Edad" ,ylab="Residuos de Martingala")
lines(lowess(lineas_muestra$EDAD[cases],lineas1.res),col='red')
plot(lineas_muestra$MONTO_RECIBO[cases],lineas1.res,xlab="Recibo",ylab="Residuos de Martingala")
lines(lowess(lineas_muestra$MONTO_RECIBO[cases],lineas1.res),col='red')
plot(lineas_muestra$MB_DATOS[cases],lineas2.res,xlab="Datos (Mb)",ylab="Residuos de Martingala")
lines(lowess(lineas_muestra$MB_DATOS[cases],lineas2.res),col='red')
plot(lineas_muestra$MIN_VOZ[cases],lineas3.res,xlab="Voz (Min)",ylab="Residuos de Martingala")
lines(lowess(lineas_muestra$MIN_VOZ[cases],lineas3.res),col='red')

### SE DEFINE RANGO PARA QUE LA CURVA SE GRAFIQUE MEJOR EN EL PUNTO DE CAMBIO #########
summary(survfit(lineas1.cox))
summary(lineas1.cox)
H0 = basehaz(fit, centered=TRUE)
plot(lineas_muestra$MONTO_RECIBO[cases],lineas1.res,xlab="Recibo"
,ylab="Residuos de Martingala", xlim=c(50,150))
lines(lowess(lineas_muestra$MONTO_RECIBO[cases],lineas1.res),col='red')

######################## FIN DE IDENTIFICACIÓN DEL PUNTO DE CAMBIO ############################
```

C.3. Calculo de los parámetros de las covariables y punto de cambio en Mathematica

Se desarrolla el algoritmo para el cálculo de los parámetros del modelo y también el valor del punto de cambio en Mathematica.

```
Lectura de los datos de la muestra y construcción de la base que se trabajará

Clear [data, X, x, x1, x2, beta, b1, b2, b3, R, i, g, j, k, theta, T, n, NN,
  a, lambda0, Lambda0] ;

Timing[data =
   Import["C:/Lucía/2014/BASE_MATHEMATICA_FINAL_25_MUESTRA_N.txt", "Table"];]

Lambda0 = Import["C:/Lucía/2014/riesgo_base_acum_lamda_0.txt", "Table"];

lambda0 = Import["C:/Lucía/2014/riesgo_base_lamda_0.txt", "Table"];

beta = {b1, b2,
   b3}; (* parametros de la regresion de cox con ptos de cambio en las \
```

```
covariables *)

T = 25;(* periodo de tiempo de evaluación *);
n = Length[data]/T ;(* se calcula el tamaño de la muestra *);

(* Estos son las covariables disponibles y el orden
1- RECIBO
  2- MIN_TOTAL
  3- KB_TOTAL
  4- SMS_TOTAL
  5- EDAD
  6- ESTADO_RIESGO
  7- CH_STATUS
  8- FEC_ESTADO
  9- SEXO
  10-FEC_ACTIVACION
  11-DES_MOTIVO_EST
*)

Clear[i]; (* limpia la variable que incrementa*)
(* Este código pasa los datos a un vector de matrices las covariables para \
cada individuo *)
Do[a[i] = Table[{Abs[data[[j, 1]]], data[[j, 5]], data[[j, 3]],
     data[[j, 6]] }, {j, 25*i - 24, 25*i}], {i, 1, n}];

aa = Table[{a[i] [[All, 2]], a[i] [[All, 1]]}, {i, 1, n}];

(* Se separa el vector R que es el vector de Riesgo *)
Do [R[i] = a[i][[All, 4]] , {i, 1, n}];

(* Construccion del vector de riesgo para todos los individuos en el instante \
j *)
r[j_] := Table[R[i][[j]], {i, 1, n}]

(*Se crea un arreglo con la cantidad de eventos de la muestra*)
Array[NN, n];

Table[NN[i] = 1 - R[i], {i, 1, n}];

(*Se crea un arreglo con las covariables que se estudiarán*)
Array[X, n];

Clear[i];

(*Función que crea el arreglo de covaribles en base a los que se van a \
considerar en el modelo de Jensen y Lutkebohmert*)

X[i_, g_] := {a[i][[All, 2]], a[i][[All, 1]], a[i][[All, 1]] - g};

(*Función que crea el arreglo de covaribles en base a los que se van a \
considerar en el modelo de cox clásico, sin considerar el punto de cambio*)
```

```
X0[i_] := {a[i][[All, 2]], a[i][[All, 1]]};

X00 = Table[{a[i][[All, 2]], a[i][[All, 1]]}, {i, 1, n}];

MatrixForm[X00[[All, 2, All]]];

Clear[g];

(* Vector que une los betas con el punto de cambio *)
theta = {beta, g}; (*g=gamma*)

(* 1- RECIBO
 5- EDAD
 3- KB_TOTAL
 6- ESTADO_RIESGO *)

(*x={x1,x2,x2-g}*)

b1 = -0.011779796363351181; b2 = -0.0092819937155466; b3 = \
-0.00828281989721975; g = 105.58;

Clear[x]

x = {{x1}, {x2}, {-g + x2}}

{b1, b2, b3}.{x1, x2, x2 - g}

Construcción del modelo de la función de verosimilitud para el modelo de Uwe Jensen y
Constanze Lutkebohmert

Clear[beta, b1, b2, b3, i, j];

L[b1_, b2_, b3_, g_] := \!\(
\*UnderoverscriptBox[\(\[Sum]\), \(i = 1\), \(n\)]\((
\*UnderoverscriptBox[\(\[Sum]\), \(j = 0\), \(T - 1\)]
\*SubsuperscriptBox[\(\[Integral]\), \(j\), \(j +
         1\)]{b1, b2, b3} . \(X[i, g]\)[\([\)\(All, j + 1\)\(]\)]*\(NN[
          i]\)[\([\)\(j + 1\)\(]\)] \[DifferentialD]t)\)\) -

    \!\(
\*UnderoverscriptBox[\(\[Sum]\), \(i = 1\), \(n\)]\((
\*UnderoverscriptBox[\(\[Sum]\), \(k = 0\), \(T - 1\)]\
\*SubsuperscriptBox[\(\[Integral]\), \(k\), \(k + 1\)]Log[\
\*UnderoverscriptBox[\(\[Sum]\), \(j =
            1\), \(n\)]\((\(R[j]\)[\([\)\(k + 1\)\(]\)]*
            Exp[{b1, b2, b3} . \(X[j, g]\)[\([\)\(All,
               k + 1\)\(]\)]])\)]\ *\ \(NN[i]\)[\([\)\(k +
          1\)\(]\)] \[DifferentialD]t)\)\);

Cálculo de los Betas y Omega, según el modelo con punto de cambio

Primera Fase del Algoritmo: Fijar "g", para elegir valores de los beta's.
```

```
Los intervalos para los betas están aproximadamente entre -1.5 y 0;
para el valor del g que corresponde a los recibos están entre el 90 y el 110, aproximadamente.

Timing[beta =
  NArgMax[{L[b1, b2, b3, 100], -1.0 <= b1 <= 0, -1.0 <= b2 <= 0, -1.0 <= b3 <=
      0}, {b1, b2, b3}]]

(* Se evalúa la función creada ingresando como parámetro el punto donde se \
observa el cambio, calculado para el modelo de regresión de Cox clásico con \
los residuos de martingala, que fué realizada en R, la función Timing nos \
dirá el tiempo en Segundos que demora en ejecutarse la función*)

Segunda Fase del Algoritmo: Dejar libre a "g", y Fijar los valores de los beta's

(* Se evalúa la función creada ingresando como parámetro los betas calculados \
con el modelo de regresión de Cox clásico que fué realizada en R la función \
Timing nos dirá el tiempo en Segundos que demora en ejecutarse la función*)

Timing[ArgMax[{L[-0.008747994425385194`, -0.012172015706168125`, \
-0.027136960600184043`, g], 100 <= g <= 110}, g]]


Iteraciones para hallar el valor de g que converge
1- En esta iteracción los dos valores van cambiando en principio solo fijamos g

Clear[vector1, vector2, p, m, g, b1, b2, b3];
m = 105.26614071250273`;

Timing[Do[vector1[p] =
   ArgMax[{L[b1, b2, b3, m], -1 <= b1 <= 0, -1 <= b2 <= 0, -1 <= b3 <=
      0}, {b1, b2, b3}]; Print[vector1[p]];
  vector2[p] =
   ArgMax[{L[vector1[p][[1]], vector1[p][[2]], vector1[p][[3]], g],
     100 <= g <= 110}, g]; m = vector2[p]; Print[vector2[p]];
  pp = p, {p, 1, 10}]]

2- En esta iteracción fijamos b1,b2 y b3, ya que convergen siempre al mismo valor

Clear[vector2, p, g]
Timing[Do[
  vector2[p] =
   ArgMax[{L[-0.6430759116735106, 0.959708209491969, -0.9979207852426352`, g],
      100 <= g <= 110}, g]; vector2[p]; Print[vector2[p]], {p, 1, 10}]]



Gráficos

(*Se grafica la función de riesgo de cox con puntos de cambio*)

hgraf[x1_, x2_] := Exp[{b1, b2, b3}.{x1, x2, x2 - g}]
```

```
lambdaaPC[x1_, x2_, j_] := \!\(
\*UnderoverscriptBox[\(\[Sum]\), \(k = 1\), \(j\)]\(lambda0[\([\)\(k,
     1\)\(]\)]*hgraf[x1, x2]\)\);

(* Para un tiempo fijo t, se evalua la función de riesgo calculado haciendo \
continua el monto_recibo y la edad*)

Plot3D[lambdaaPC[x1, x2, 5], {x1, 70, 70.01}, {x2, 50, 500}]

Plot3D[lambdaaPC[x1, x2, 5], {x1, 20, 20.01}, {x2, 50, 500}]

Plot3D[lambdaaPC[x1, x2, 5], {x1, 20, 76}, {x2, 200, 200.01}]

Plot3D[lambdaaPC[x1, x2, 5], {x1, 20, 76}, {x2, 100, 100.01}]

Clear[b11, b22, b33, g, slambdaCoxPC, slambdaPC]
b11 = -0.011779796363351181; b22 = -0.0092819937155466; b33 = \
-0.00828281989721975; g = 105.58;
hh[b1_, b2_, b3_, g_, i_, j_] := Exp[{b1, b2, b3}.X[i, g]][[j]];

lambdaCoxPC[b1_, b2_, b3_, g_, i_, j_] :=
  lambda0[[j, 1]]*R[i][[j]]*hh[b1, b2, b3, g, i, j];

slambdaCoxPC1[i_, j_] := \!\(
\*UnderoverscriptBox[\(\[Sum]\), \(k = 1\), \(j\)]\(lambda0[\([\)\(k,
     1\)\(]\)]*\(R[i]\)[\([\)\(k\)\(]\)]*hh[b11, b22, b33, g, i, k]\)\);

Timing[slambdaCoxPC = Table[\!\(
\*UnderoverscriptBox[\(\[Sum]\), \(k = 1\), \(j\)]\(lambda0[\([\)\(k,
       1\)\(]\)]*\(R[i]\)[\([\)\(k\)\(]\)]*hh[b11, b22, b33, g, i, k]\)\), {j,
      1, T}, {i, 1, n}];]

MatrixForm[slambdaCoxPC];

Timing[slambdaPC = Table[slambdaCoxPC[[j, i]], {j, 1, T}, {i, 1, n}];]

Timing[fslambdaPC = Table[(1/n)*\!\(
\*UnderoverscriptBox[\(\[Sum]\), \(i = 1\), \(n\)]\(slambdaCoxPC[\([\)\(j,
       i\)\(]\)]\)\), {j, 1, T}];]

paresPago[i_, j_] := {X00[[i]][[2, j]],
  lambdaaPC[X00[[i]][[1, j]], X00[[i]][[2, j]], j]}

paresEdad[i_, j_] := {X00[[i]][[1, j]], slambdaCoxPC[[j, i]]}

Timing[ternasEdad3D =
   Table[{j, X00[[i]][[1, j]], fslambdaPC[[j]]}, {j, 1, T}, {i, 1, n}];]

ternasEdad3D[[1]]; ternasEdad3D[[2]];

tEdad3D = Union[ternasEdad3D[[1]], ternasEdad3D[[2]], ternasEdad3D[[3]],
```

```
   ternasEdad3D[[4]], ternasEdad3D[[5]], ternasEdad3D[[6]], ternasEdad3D[[7]],
    ternasEdad3D[[8]], ternasEdad3D[[9]], ternasEdad3D[[10]],
   ternasEdad3D[[11]], ternasEdad3D[[12]], ternasEdad3D[[13]],
   ternasEdad3D[[14]], ternasEdad3D[[15]], ternasEdad3D[[16]],
   ternasEdad3D[[17]], ternasEdad3D[[18]], ternasEdad3D[[19]],
   ternasEdad3D[[20]], ternasEdad3D[[21]], ternasEdad3D[[22]],
   ternasEdad3D[[23]], ternasEdad3D[[24]], ternasEdad3D[[25]]];

ListPlot3D[tEdad3D, PlotRange -> {{0, 25}, {20, 76}}]

Timing[ternasPago3D =
   Table[{j, X00[[i]][[2, j]], fslambdaPC[[j]]}, {j, 1, T}, {i, 1, n}];]

tPago3D = Union[ternasPago3D[[1]], ternasPago3D[[2]], ternasPago3D[[3]],
   ternasPago3D[[4]], ternasPago3D[[5]], ternasPago3D[[6]], ternasPago3D[[7]],
    ternasPago3D[[8]], ternasPago3D[[9]], ternasPago3D[[10]],
   ternasPago3D[[11]], ternasPago3D[[12]], ternasPago3D[[13]],
   ternasPago3D[[14]], ternasPago3D[[15]], ternasPago3D[[16]],
   ternasPago3D[[17]], ternasPago3D[[18]], ternasPago3D[[19]],
   ternasPago3D[[20]], ternasPago3D[[21]], ternasPago3D[[22]],
   ternasPago3D[[23]], ternasPago3D[[24]], ternasPago3D[[25]]];

Construcción de Función de Riesgo: Lambda de Cox Clásico y Con punto de Cambio

Funciones de regresión de Cox Clásico

(* Valores de beta para el modelo de Cox Clásico, calculados en R *)
b1 = -0.00890;
b2 = -0.00746;
rho[b1_, b2_, i_, j_] := Exp[{b1, b2}.X0[i]][[j]];
lambdaCox[b1_, b2_, i_, j_] := lambda0[[j, 1]]*R[i][[j]]*rho[b1, b2, i, j];

slambdaCox[i_, j_] := \!\(
\*UnderoverscriptBox[\(\[Sum]\), \(k = 1\), \(j\)]\(
\*SubsuperscriptBox[\(\[Integral]\), \(k - 1\), \(k\)]lambda0[\([\)\(k,
      1\)\(]\)]*\(R[i]\)[\([\)\(k\)\(]\)]*
     rho[b1, b2, i, k] \[DifferentialD]t\)\);

slambda[j_] := Table[slambdaCox[i, j], {i, 1, n}]

fslambda[j_] := (1/n)*\!\(
\*UnderoverscriptBox[\(\[Sum]\), \(i = 1\), \(n\)]\(slambdaCox[i, j]\)\);

mlambda[j_] := Median[slambda[j]];

Timing[vectorslambdaCox = Table[fslambda[j], {j, 1, T}];]

Timing[paresm = Table[{j, mlambda[j]}, {j, 1, T}];]

ListLinePlot[paresm]

Timing[paresmS = Table[{j, Exp[-paresm[[j, 2]]]}, {j, 1, T}];]
```

```
ListLinePlot[paresmS, AxesOrigin -> {0.5, .89}, PlotRange -> All]

Timing[vectorlambdaCox = Table[flambda[j, b1, b2], {j, 1, T}];]

Timing[pares5 = Table[vectorslambdaCox[[j]], {j, 1, T}];]

Timing[paresfS = Table[{j, Exp[-vectorslambdaCox[[j]]]}, {j, 1, T}];]


Funciones de regresión de Cox con puntos de cambio en las covariables

Clear[b11, b22, b33, g];

b11 = -0.011779796363351181; b22 = -0.0092819937155466; b33 = \
-0.00828281989721975; g = 105.58;
hh[b1_, b2_, b3_, g_, i_, j_] := Exp[{b1, b2, b3}.X[i, g]][[j]];

lambdaCoxPC[b1_, b2_, b3_, g_, i_, j_] :=
  lambda0[[j, 1]]*R[i][[j]]*hh[b1, b2, b3, g, i, j];

slambCoxPC[i_, j_] := \!\(
\*UnderoverscriptBox[\(\[Sum]\), \(k = 1\), \(j\)]\(
\*SubsuperscriptBox[\(\[Integral]\), \(k - 1\), \(k\)]lambda0[\([\)\(k,
      1\)\(]\)]*\(R[i]\)[\([\)\(k\)\(]\)]*
     hh[b11, b22, b33, g, i, k] \[DifferentialD]t\)\);

slambPC[j_] := Table[slambCoxPC[i, j], {i, 1, n}]

slambPC[2];

fslambPC[j_] := (1/n)*\!\(
\*UnderoverscriptBox[\(\[Sum]\), \(i = 1\), \(n\)]\(slambCoxPC[i, j]\)\);

Timing[vectorslambdaCoxPC = Table[fslambPC[j], {j, 1, T}];]

Timing[paressPC = Table[{j, vectorslambdaCoxPC[[j]]}, {j, 1, T}];]

Timing[paresfSPC = Table[{j, Exp[-vectorslambdaCoxPC[[j]]]}, {j, 1, T}];]

mslambPC[j_] := Median[slambPC[j]];

Timing[vectormslambdaCoxPC = Table[mslambPC[j], {j, 1, T}];]

paresmsPC = Table[{j, vectormslambdaCoxPC[[j]]}, {j, 1, T}];

ListLinePlot[paresmsPC]

paresmfSPC = Table[{j, Exp[-vectormslambdaCoxPC[[j]]]}, {j, 1, T}];

ListLinePlot[paresmfSPC, AxesOrigin -> {0.3, .82}, PlotRange -> All]
```

C.4. Prueba de bondad de ajuste en Mathematica

Se desarrolla en Mathematica la prueba de bondad de ajuste entre el modelo propuesto por Jensen y Lutkebohmert (2008) y el modelo Cox (1972)

```
Prueba de bondad de Ajuste

Tercera Fase del Algoritmo: Realizar prueba de bondad de Ajuste
Construir \!\(\(
\*SubscriptBox[\(h\), \(i\)](t)\)\ , \
\*SubscriptBox[\(\[Del]\), \(1\)]\)\[Rho](t), Subscript[\[Del], 2]\[Rho](t)

rho[b1_, b2_, i_] := Exp[{b1, b2}.X0[i]];

derho1[b1_, b2_, i_] := X0[i][[1]]*Exp[{b1, b2}.X0[i]];

derho2[b1_, b2_, i_] := X0[i][[2]]*Exp[{b1, b2}.X0[i]];

h[b1_, b2_, b3_, g_, i_, j_] := Exp[{b1, b2, b3}.X[i, g]][[j]];

Construcción de Cocientes:
(rho(beta,i)/h(gamma,i), derho2(beta,i)/h(gamma,i),rho(beta,i)/h(gamma,i))

ra[i_, j_] := \!\(
\*UnderoverscriptBox[\(\[Sum]\), \(k = 1\), \(j\)]\(lambda0[\([\)\(k,
     1\)\(]\)]*\(R[i]\)[\([\)\(k\)\(]\)]*rho[b1, b2, i, k]\)\);

ha[i_, j_] := \!\(
\*UnderoverscriptBox[\(\[Sum]\), \(k = 1\), \(j\)]\(lambda0[\([\)\(k,
     1\)\(]\)]*\(R[i]\)[\([\)\(k\)\(]\)]*h[b11, b22, b33, g, i, k]\)\);

drho1a[i_, j_] := \!\(
\*UnderoverscriptBox[\(\[Sum]\), \(k = 1\), \(j\)]\(lambda0[\([\)\(k,
     1\)\(]\)]*\(R[i]\)[\([\)\(k\)\(]\)]*\(derho1[b1, b2,
      i]\)[\([\)\(k\)\(]\)]\)\);

drho2a[i_, j_] := \!\(
\*UnderoverscriptBox[\(\[Sum]\), \(k = 1\), \(j\)]\(lambda0[\([\)\(k,
     1\)\(]\)]*\(R[i]\)[\([\)\(k\)\(]\)]*\(derho2[b1, b2,
      i]\)[\([\)\(k\)\(]\)]\)\);

co1a[i_, j_] := ra[i, j]/ha[i, j]

co2a[i_, j_] := drho1a[i, j]/ha[i, j];

co3a[i_, j_] := drho2a[i, j]/ha[i, j];

c1a = Table[co1a[i, j], {i, 1, n}, {j, 1, T}];

c2a = Table[co2a[i, j], {i, 1, n}, {j, 1, T}];

MatrixForm[c2a];
```

```
c3a = Table[co3a[i, j], {i, 1, n}, {j, 1, T}];

qa[i_, j_] := {co1a[i, j], co2a[i, j], co3a[i, j]}

QQa = Table[qa[i, j], {i, 1, n}, {j, 1, T}];

Construyendo la matriz que corresponde a los n individuos para el instante j; con j=1, 2, ...,25

Timing[qqa[j_] := Table[qa[i, j], {i, 1, n}]]

Construcción de matriz B, para hallar valor c para cada tiempo-j, del conjunto total
de los n individuos.

Clear[c, cc, t]

vh[j_] := Table[ h[b11, b22, b33, g, i, j], {i, 1, n}]

DiagonalMatrix[vh[1]];

Inverse[Transpose[qqa[1]].DiagonalMatrix[vh[1]].qqa[1]];

c[j_] := r[j] -
  qqa[j].Inverse[Transpose[qqa[j]].DiagonalMatrix[vh[j]].qqa[j]].Transpose[
    qqa[j]].DiagonalMatrix[vh[j]].r[j]

Timing[cc = Table[c[j], {j, 1, T}];]

Timing[sigma2 = (1/n)*\!\(
\*UnderoverscriptBox[\(\[Sum]\), \(i = 1\), \(n\)]\((
\*UnderoverscriptBox[\(\[Sum]\), \(j = 0\), \(T - 1\)]
\*SubsuperscriptBox[\(\[Integral]\), \(j\), \(j +
         1\)]\(((\(cc[\([\)\(j +
            1\)\(]\)]\)[\([\)\(i\)\(]\)])\)*\(((\(cc[\([\)\(j +
            1\)\(]\)]\)[\([\)\(i\)\(]\)])\)*\(NN[i]\)[\([\)\(j +
          1\)\(]\)] \[DifferentialD]t)\)\)]

Timing[cmuestral = Sqrt[(1/n)]*\!\(
\*UnderoverscriptBox[\(\[Sum]\), \(i = 1\), \(n\)]\((
\*UnderoverscriptBox[\(\[Sum]\), \(j = 0\), \(T - 1\)]
\*SubsuperscriptBox[\(\[Integral]\), \(j\), \(j +
         1\)]\(((\(cc[\([\)\(j + 1\)\(]\)]\)[\([\)\(i\)\(]\)])\)*\(NN[
          i]\)[\([\)\(j + 1\)\(]\)] \[DifferentialD]t)\)\)]

 Timing[Sqrt[(1/n)]*(\!\(
\*UnderoverscriptBox[\(\[Sum]\), \(j = 0\), \(T - 1\)]\(
\*SubsuperscriptBox[\(\[Integral]\), \(j\), \(j +
       1\)]\(((\(cc[\([\)\(j + 1\)\(]\)]\)[\([\)\(1\)\(]\)])\)*\(NN[
        1]\)[\([\)\(j + 1\)\(]\)] \[DifferentialD]t\)\))]

tt = cmuestral/Sqrt[sigma2]
```

Bibliografía

Aalen, O. (1978). Nonparametric inference for a family of counting processes, *The Annals of Statistics* **6**: 701–726.

Andersen, P. K., Borgan, O., Gill, R. D. y Keiding, N. (1993). *Statistical Models Based on Counting Processes*, Springer.

Andersen, P. K. y Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study, *Annals of Statistics* **10**: 1100–1120.

Breslow, N. (1974). Covariance analysis of censored survival data, *Biometrics* **30**: 89–99.

Chappell, R. (1989). Fitting bent lines to data with applications to allometry, *Journal of Theoretical Biology* **138**: 235–256.

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

Cox, D. R. (1975). Partial likehood, *Biometrika* **62**: 269–276.

Cox, D. R. y Oakes, D. (1984). *Analysis of Survival Data*, Chapman and Hall.

der Vaart, A. V. (1998). *Asymptotic statistics*, Cambridge University Press.

Fleming, T. R. y Harrington, D. P. (1991). *Counting processes and survival analysis*, Wiley Series in Probability and Statistics.

Gandy, A. y Jensen, U. (2005). On goodness of fit tests for aalen's additive risk model, *Scandinavian Journal of Statistics* **32**: 425–445.

Gandy, A. y Jensen, U. (2006). Model checks for cox-type regression models based on optimally weighted martingale residuals, *Lifetime data analysis* **15**: 534–557.

Gandy, A., Jensen, U. y Lutkebohmert, C. (2005). A cox model with a change-point applied to an actuarial problem, *Brazilian Journal of Probability and Statistic* **19**: 93–109.

Ibragimov, I., Has'minskii, R. y Kotz, S. (1981). *Statistical estimation. Asymptotic theory*, Springer.

Jensen, U. y Lutkebohmert, C. (2008). A cox-type regression model with change-points in the covariates, *Lifetime data analysis* **14**: 267–285.

K. Liang, S. S. y Liu, X. (1990). The cox proportional hazard model with change-point: an epidemioliogie application, *Biometrics* **46**: 783–793.

Kosorok, M. y Song, R. (2007). Inference under right censoring for transformations models with a change-point based on a covariate threshold, *Annals of Statistics* **35**: 957–989.

Lehmann, E. L. y Casella, G. (1998). *Theory of Point Estimation*, Springer.

Lohr, S. L. (1999). *Muestreo: Diseño y análisis*, International Thomson Editores.

Luo, X. y Boyett, J. (1997). Estimations of a threshold parameter in cox regression., *Communications in Statistics - Theory and Methods* **26**: 2329–2346.

Mathematica (2012). Versión 9.0, *Wolfrang Research* .

Pons, O. (2003). Estimation in a cox regression model with a change-point according to a threshold in a covariate, *Annals of Statistics* **31**: 442–463.

R (2012). Versión 2.14.2, *The R foundation for Statistical Computing* .

Survival (2013). Versión 2.37-4, *Therneau, T. - Survival Analysis Library in R* .

Vaart, A. V. y Wellner, J. (1996). *Weak Convergence and empirical processes*, Springer.