

TESIS PARA OPTAR POR EL TÍTULO DE DOCTOR EN ESTADÍSTICA

**Análisis de Sobrevida Multivariado: Una
Comparación de Modelos Alternativos**

por

Mirta Sara Santana

Comisión de Supervisión

Prof. Dr. Sergio Muñoz Navarro (Ph.D.), Director

Prof. Dr. Raúl P. Mentz (Ph.D.), Codirector

Prof. Dr. Aldo J. Viollaz (Ph.D.)

Prof. C.P.N. Santiago Di Lullo

*Facultad de Ciencias Económicas
Universidad Nacional de Tucumán*

2000

RESUMEN

El Análisis de Sobrevida estudia el *tiempo* que transcurre hasta que se observa un evento o *falla* en presencia de *censuras*. Si k fallas se pueden observar en cada sujeto bajo estudio, se presenta entonces el problema del *Análisis de Sobrevida Multivariado*. En este contexto se deben distinguir dos tipos de situaciones: los sujetos experimentan sucesivas fallas del mismo tipo, (*eventos ordenados o recurrentes*); los individuos experimentan eventos de naturaleza diferente, (*eventos no ordenados o diferentes*). Varios modelos de regresión han sido propuestos en la literatura, en el caso de datos de sobrevida multivariados, usando extensiones del modelo de "hazard" proporcional debido a Cox. La diferencia fundamental entre las propuestas radica, principalmente, en como intenta manejar la estructura de correlación de los tiempos de falla dentro de los sujetos. Así, algunos modelan el "hazard" condicional (*modelos condicionales*) mientras que otros modelan el "hazard" marginal de cada falla (*modelos marginales*) y la estructura de correlación se considera en la etapa de estimación. Sin embargo, no se estudió hasta ahora cómo afecta a la precisión de los estimadores de los modelos la falta de cumplimiento de los supuestos bajo los cuales se derivaron las propiedades de los estimadores. A partir de este problema se plantea como objetivos de esta tesis estudiar comparativamente: el efecto sobre la precisión de los estimadores de los modelos de regresión para datos de sobrevida multivariados, la falta de cumplimiento de los supuestos (*independencia y falta de memoria*) y el efecto del tamaño de muestra y del porcentaje de censuras. Además, se quiere esclarecer la controversia sobre la bondad de los diferentes modelos para las dos situaciones (eventos ordenados y no ordenados), aún en el caso de tiempos de falla independientes y con distribuciones sin memoria. Para cada uno de los problemas, en una primera etapa, se llevan a cabo estudios de Monte Carlo, luego se derivan cotas para los sesgos del estimador y se compara con los resultados alcanzados en los estudios por simulación. A partir de esto se arriba a las siguientes conclusiones: i) el problema más importante que presentan los estimadores de los parámetros cuando no se cumplen los supuestos del modelo es la introducción de sesgos; ii) la precisión de los estimadores se ve más afectada por la memoria de los procesos que por la falta de correlación entre los tiempos de falla; iii) en la situación de eventos ordenados los modelos debido a Andersen y Gill y a Prentice y Cai aparecen como los más recomendables, a excepción que se tenga evidencia que el proceso tiene memoria marginal y conjunta, y la correlación entre los tiempos de falla sea pequeña o nula, en tal caso se sugiere usar el modelo debido a Wei, Lin y Weissfeld; iv) en la situación de eventos no ordenados, los modelos marginales, considerando un proceso Markov para el modelo de Prentice y Cai, generan los menores sesgos en los estimadores si la distribución no tiene memoria marginal, en caso contrario se debería usar cualquiera de los modelos condicionales.

AGRADECIMIENTOS

Deseo expresar el más profundo agradecimiento a mi director, Prof. Dr. Sergio Muñoz, cuya paciencia y estímulo constante fueron los mejores incentivos durante el desarrollo de esta tesis. Mayor aún es la gratitud por la confianza que depositó en mí y por la generosidad con que me brindó los medios que permitieron alcanzar los objetivos planteados. Hago extensivo este agradecimiento a mi codirector, Prof. Dr. Raúl P. Mentz, quien fue mi "maestro" que me guió desde los comienzos de mi carrera y generosamente contribuyó a mi enriquecimiento profesional.

Mi gratitud y reconocimiento va también para:

Los miembros de la Comisión de Supervisión, Prof. Dr. Aldo Viollaz y Prof. Santiago Di Lullo, por las importantes sugerencias y los aportes bibliográficos que contribuyeron, de manera muy significativa, a mejorar los resultados de este trabajo.

Al personal de la Facultad de Medicina de la Universidad de la Frontera, Temuco, Chile, que generosamente me brindaron los medios para llevar a cabo las Pasantías por esa Casa de Estudios.

A mis amigas y colegas, cuyo afecto y amistad me alentaron en los momentos difíciles y, en especial, a Cristina Martínez cuya desinteresada ayuda contribuyó a mejorar la redacción de este documento.

A mi querida familia, en especial mis hijos, sin cuyo amor y comprensión no hubiera podido llevar a cabo los estudios.

Y a todos los que de una u otra manera me apoyaron en esta tarea.

INDICE

RESUMEN	I
AGRADECIMEINTOS	II
CAPITULO I – INTRODUCCIÓN GENERAL	1
1. Introducción	1
1.1. Aspectos especiales de los datos de sobrevida	1
1.2. Datos censurados	3
1.3. Condicionamiento	4
1.4. Elección de los modelos	7
2. Modelos de regresión	8
3. Datos de sobrevida Multivariados	10
3.1. Modelos de regresión para datos de sobrevida multivariados ..	11
CAPITULO II - MODELO PROBABILÍSTICO	16
1. Introducción	16
2. Modelo probabilístico de los datos de sobrevida	16
3. Mecanismos de Censuras	24
4. Función de Verosimilitud	29
CAPITULO III - MODELOS DE "HAZARD" PROPORCIONAL ..	35
1. Introducción	35
2. Modelos de regresión para datos censurados	36
3. Regresión para datos multivariados con fallas correlacionadas ...	43
3.1. Modelo de Andersen y Gill (AG)	45
3.2. Modelo de Oakes (OK)	47
3.3. Modelo de Wei, Lin y Weissfeld (WLW)	50
3.4. Modelo de Prentice y Cai (PC)	51
4. Implementación computacional	53

CAPITULO IV - CORRELACIÓN ENTRE LOS TIEMPOS DE FALLA	59
1. Introducción	59
2. Estudio de Monte Carlo	60
3. Análisis del Comportamiento del Sesgo Relativo	72
4. Cotas para el Sesgo Relativo	75
4.1. Sesgo en la Situación de Eventos Recurrentes	76
4.2. Sesgo en la Situación de Eventos Diferentes	83
5. Conclusiones	88
CAPITULO V - DISTRIBUCIONES CON MEMORIA	91
1. Introducción	91
2. Propiedades de Falta de Memoria	92
3. Modelos de Distribución Bivariada	94
3.1. Distribución Block-Basu (ACBVE)	94
3.2. Distribución Sarkar (ACBVE₂)	96
3.3. Distribución Ryu	97
4. Algoritmos Computacionales	100
5. Estudio de Monte Carlo	112
6. Cotas para el Sesgo de los Estimadores en el Modelo de Regresión	117
6.1. Sesgo en la Situación de Eventos Recurrentes	118
6.2. Sesgo en la Situación de Eventos Diferentes	128
7. Conclusiones	138
CAPITULO VI - TAMAÑO DE MUESTRAS Y CENSURAS	140
1. Introducción	140
2. Efecto del Tamaño de Muestras	141
3. Efecto del Porcentaje de Censuras	148
4. Conclusiones	154

CAPITULO VII – ANÁLISIS DE LOS RESULTADOS Y CONCLUSIONES.....	157
ANEXO 1 – CONCEPTOS Y RESULTADOS ADICIONALES DE LA TEORÍA DE PROCESOS DE CONTEO APLICADOS AL ANÁLISIS DE SOBREVIDA.....	163
BIBLIOGRAFÍA.....	171

CAPITULO I

INTRODUCCIÓN GENERAL

En los últimos años del siglo pasado ha habido un considerable interés en los métodos para analizar datos sobre eventos observados a través del tiempo (*datos de sobrevida o datos de historia de vida*), y en los estudios de factores asociados con la tasa de ocurrencia de aquellos eventos.

En la Sección 1 de este capítulo se hace una caracterización de los datos de sobrevida, cada uno de los cuales mide el tiempo que transcurre desde que comienza la observación hasta que se produce un evento previamente definido (*tiempo de falla*), y se define las funciones que caracterizan la distribución del tiempo de falla. En la Sección 2 se describen los modelos de regresión más usados para estudiar el efecto de ciertas covariables sobre los tiempos de falla. Por último, en la Sección 3 se describe la situación de eventos múltiples, los problemas que surgen para la generalización del modelo de "hazard" proporcional al caso multivariado y a partir de ellos se plantea los objetivos de esta tesis.

1. Características de los Datos de Sobrevida

En estudios donde el objetivo es evaluar el efecto de algunas intervenciones del investigador (tratamientos), el análisis resulta relativamente simple para un estadístico si se tiene una sola variable respuesta bien definida para medir la diferencia de tratamientos (ANOVA). Cuando existe más de una variable respuesta igualmente importante involucrada en la evaluación, los aspectos sobre diseño del estudio, análisis de los datos e interpretación de los resultados llega a ser más complicado (MANOVA). Estos problemas de respuestas múltiples han sido extensamente investigados y son usados en forma constante, en especial en la industria farmacéutica y entes reguladores de medicamentos.

Aun en el caso de tener una sola variable respuesta, si ella es el *tiempo que transcurre hasta que se produce un evento*, por ejemplo cura de una enfermedad,

los datos que resultan presentan algunas características particulares, razón por la cual han recibido en la literatura el nombre especial de *datos de sobrevida* o *datos de historia de vida* en el caso de estudios de confiabilidad.

Estos datos pueden provenir de investigaciones en diferentes campos de la ciencia y tecnología, como ser, medicina, demografía, ingeniería, economía, educación, sociología, entre otros.

En el caso más simple, y de ahí el nombre de sobrevida, el evento es la muerte de una persona o rotura de un dispositivo, pero el término puede cubrir otros tipos de eventos como ser la ocurrencia, complicación o cura de una enfermedad. En aplicaciones industriales, el evento es típicamente la falla de una unidad o algún componente de una unidad (estudios de confiabilidad). En economía, los datos de sobrevida pueden obtenerse a partir de un estudio sobre el tiempo que demora una persona desempleada en conseguir un trabajo. En psicología, el evento puede ser crisis psicótica; mientras que en educación puede ser la deserción, el egreso o la aprobación de un examen.

Desde un enfoque de procesos, el evento es la transición de un estado a otro, por ejemplo, la ocurrencia de una enfermedad define el evento: "transición del estado de salud al estado de enfermedad". En el ejemplo de psicología la crisis psicótica marca el evento: "transición del estado normal al estado psicótico". Así, dependiendo del contexto, se usan palabras como muerte, evento, falla o transición para referirse al mismo concepto.

Por último, para el tratamiento de los datos de sobrevida se debe definir un punto, tiempo cero, a partir del cual se mide el tiempo hasta la falla (*escala de tiempo*) y conocer en todo momento del estudio en qué estado se encuentra cada individuo bajo observación.

1.1 Aspectos especiales de los datos de sobrevida

En un sentido amplio, lo que hace especial a este tipo de datos es que la variable respuesta es el *tiempo*, que presenta algunas características especiales que la hacen diferente de cualquier otra variable; entre las que se puede mencionar el

hecho que es observada secuencialmente, lo cual trae varias consecuencias. Entre las consecuencias más importante figura la presencia de *datos incompletos*, generados por la imposibilidad de conocer el tiempo de falla de todos los individuos bajo estudio, es decir, para algunos individuos bajo estudio sólo se conoce que el tiempo de falla excede a un cierto valor. Sin embargo, tanto o más importante que el problema de datos incompletos, es el hecho que la probabilidad de sobrevida a la falla cambia según el tiempo de referencia, aspecto conocido como *condicionamiento*, lo que conduce a tener que evaluar la probabilidad condicional de falla en un instante t , dado que sobrevivió a la falla hasta ese instante. Así por ejemplo, la esperanza de vida de una persona a los 20 años es menor que la esperanza de vida de una persona a los 60 años, ya que ella debe calcularse condicional al tiempo de sobrevida. Otras importantes consideraciones incluyen *la elección del modelo paramétrico* adecuado a esta estructura de datos y la posibilidad de tener *eventos múltiples* o *covariables dependientes del tiempo*.

1.2 Datos Censurados

Como se mencionó previamente, una de las características particulares de los datos de sobrevida es la presencia de datos incompletos. La forma más común de observaciones incompletas son las denominadas *Censuras a derecha* que son las que surgen cuando en un conjunto de individuos hay algunos de ellos a los cuales no puede observarse los valores del tiempo de sobrevida, sino que para ellos sólo se conoce que son mayores que un cierto valor.

Una razón común para tener datos censurados es que el individuo observado sobreviva a la falla cuando el estudio termina, y así todo lo que se conoce sobre su tiempo de vida es que excede a un cierto valor lo que puede ser raramente evitado en datos de historia de vida. La presencia de censuras trae aparejado una complicación en el manejo de la distribución teórica de los estimadores aún cuando el mecanismo de censura sea simple y bien entendido.

Para usar la metodología estadística clásica en este tipo de datos se debe eliminar del análisis aquellas observaciones incompletas, lo que produciría una

pérdida de eficiencia. Se hace necesario, entonces, recurrir a una metodología que use toda la información disponible. Tal metodología se dio en llamar *Análisis de Sobrevida*.

El uso de la mayoría de los modelos y métodos para analizar datos censurados implican importantes suposiciones sobre la naturaleza de las censuras y su relación a los procesos de falla, ya que su presencia puede alterar la estimación de la probabilidad de ocurrencia del evento de interés. Así, una primera condición para evitar que se introduzcan sesgos en la estimación de las probabilidades, es que la ocurrencia de censuras no debe estar relacionada al tiempo de falla futura, o lo que es lo mismo, que el tiempo de censura de un individuo (tiempo de observación libre de falla) dependa sólo de la historia del proceso observado (tiempo de falla). Esto es equivalente a formular que la probabilidad del tiempo residual de sobrevida a la falla para aquellos individuos censurados debe ser igual al correspondiente a los individuos no censurados (siguen bajo observación), lo cual implica que el tiempo de censura no da ninguna información acerca de la distribución del tiempo de falla. Si el esquema de censuras responde a estas suposiciones se habla de *Censuras no Informativas*.

1.3 Condicionamiento

Este aspecto es más fácil introducirlo a través de un ejemplo. Supongamos que se quiere estudiar la mortalidad en una dada ciudad. Con algunas aproximaciones tales como suponer densidad constante en cada año y extrapolando más allá de 100 años, se obtiene que el tiempo mediana de vida de un hombre, es de 75 años y 80 días. Así, para un hombre que cumple 75 años, la mediana sugiere que tiene sólo 80 días de vida, sin embargo esto no es correcto, ya que para evaluar su tiempo residual de vida se debe tener en cuenta que él ya vivió hasta los 75 años y así la mediana condicional dada la sobrevida hasta los 75 años resulta ser de 82 años y 294 días. Por lo tanto, lo relevante es la distribución de sobrevida truncada a izquierda de 75 años, aspecto que tiene una enorme consecuencia para los procedimientos que tratan con datos de sobrevida. Una lógica consecuencia es que

es más conveniente manejar la distribución del *tiempo de sobrevida a la falla* (T) por medio de la *función "hazard"* (h(t)) definida como la probabilidad de falla dentro de un intervalo corto de tiempo, dado que el individuo no experimentó la falla al comienzo de dicho intervalo. Si la distribución de T es absolutamente continua, la definición formal de h(t) está dada por

$$h(t) = \lim_{\Delta t \downarrow 0} \frac{\Pr(t < T < t + \Delta t \mid T > t)}{\Delta t} \quad (1.3.1)$$

Las relaciones entre la función de densidad f(t) y la función "hazard" son las siguientes:

$$h(t) = \frac{f(t)}{1 - F(t)} \quad (1.3.2)$$

$$f(t) = h(t) \exp\left\{-\int_0^t h(t)\right\} \quad (1.3.3)$$

Para datos de sobrevida, más conveniente que usar la función de distribución ordinaria F(t) es usar la llamada *función de sobrevida* (S(t)) que representa la probabilidad de que el tiempo libre de falla T sea mayor que t; que se expresa como

$$S(t) = \Pr[T > t] = 1 - F(t), \quad (1.3.4)$$

de donde se sigue que

$$f(t) = -\frac{dS(t)}{dt} \quad (1.3.5)$$

De esta manera la relación entre las funciones de densidad, "hazard" y de sobrevida está dada por:

$$h(t) = \frac{f(t)}{S(t)} = - \frac{d \log S(t)}{dt} \quad (1.3.6)$$

La siguiente tabla muestra cómo las funciones de densidad, "hazard" y de sobrevida cambian al ser truncadas en el tiempo v , es decir, condicionadas para $T > v$.

Tabla 1.3.1. Funciones de Sobrevida, densidad y "hazard" completas y truncadas para $T > v$.

Funciones	Distribución Completa	Distribución Truncada para $T > v$
Sobrevida	$S(t)$	$S(t) / S(v)$
Densidad	$f(t)$	$f(t) / S(v)$
"Hazard"	$h(t)$	$h(t)$

La tabla muestra la ventaja de la función "hazard", la cual, a la inversa de las otras funciones, condiciona la sobrevida hasta t ; y, por lo tanto, no cambia si se condiciona hasta un tiempo $v < t$. Esta función es también conocida como *tasa de falla instantánea* ya que mide el riesgo de falla por unidad de tiempo durante el proceso de envejecimiento. Por otra parte, la función "hazard" puede crecer, decrecer, permanecer constante o indicar un proceso más complicado. Por ejemplo, los pacientes con leucemia aguda que no responden al tratamiento, tienen una función "hazard" creciente.

En estudios de confiabilidad la función "hazard" es más conocida como *tasa condicional de falla* y la función de densidad recibe el nombre de *tasa incondicional de falla*; mientras que en demografía es conocida como *fuerza de mortalidad*.

Ahora, el concepto de condicionamiento corresponde al concepto de *exposición al riesgo*, es decir, antes de la falla los individuos tienen que encontrarse en un estado desde el cual es posible la transición. Por ejemplo, un individuo solamente está al riesgo de morir si en el momento considerado está vivo.

Desde el punto de vista matemático, este fenómeno es naturalmente analizado como un proceso aleatorio desarrollándose en el tiempo, es decir, un proceso estocástico con una realización del proceso para cada individuo.

Puesto que a cada evento la variable respuesta es dicotómica, el proceso estocástico a elegir es un proceso de conteo que toma el valor cero antes del evento y uno después del evento.

Esta manera de describir el fenómeno, aunque aparentemente más complicada, simplifica el tratamiento matemático de los métodos de estimación, y permite incorporar en los modelos los aspectos especiales de este tipo de datos.

1.4 Elección de modelos

Las distribuciones estándares que se usan generalmente en otro tipo de datos, en la mayoría de los casos no son relevantes para los datos de sobrevida. En general, para datos de sobrevida se necesitan distribuciones que, además de estar definidas sólo para valores positivos, sean asimétricas y, quizás la propiedad más importante, es que sean *fácilmente adaptables a truncamiento*, en el sentido que el hazard correspondiente sea una función simple de los parámetros de la distribución. Ejemplos de ellas son las distribuciones exponencial, Weibull y Gompertz. Otras distribuciones como la log-normal, inversa Gausiana y Gamma son menos convenientes computacionalmente pero todavía frecuentemente usadas.

Por ejemplo, la Weibull con $h(t) = \mu \gamma t^{\gamma-1}$, $\mu > 0, \gamma > 0$ y la distribución Gompertz con $h(t) = \mu \varphi^t$, $\mu > 0, \varphi \geq 1$ incluyen como caso particular a la distribución exponencial ($\gamma=1$ y $\varphi=1$ respectivamente) que tiene "hazard" constante. En el caso de la Gompertz, el tiempo de sobrevida a la falla después del truncamiento es también una distribución Gompertz con el mismo valor de φ .

Los *métodos no paramétricos* han llegado a ser muy populares dentro del análisis de sobrevida, por varias razones. Una de ella es que los datos tienen, a menudo, algunos aspectos que no son fácilmente descriptos por modelos paramétricos. Por ejemplo, los tiempos de vida en los humanos muestran un

“hazard” decreciente los primeros 5 años de vida. Este aspecto es relevante en algunos casos y en otros no, pero suponiendo una distribución no paramétrica, se evita la preocupación por este problema.

2. Modelos de Regresión

En muchos casos el objetivo del estudio es establecer la dependencia del tiempo de falla con ciertos factores o covariables. En efecto, cuando hay covariables, se está típicamente más interesado en el efecto de aquellas sobre el tiempo de sobrevida o el “hazard” que, en cómo cambia este último con el tiempo.

Las covariables pueden ser factores que son de importancia intrínseca, como tratamientos en ensayo de drogas, o factores que se conoce o se sospecha que tienen influencia sobre el riesgo de falla, pero que no son determinantes por sí mismas del tiempo de falla.

Si $\mathbf{Z} = (z_1, \dots, z_p)$ es el vector de covariables de dimensión p para un individuo, el modelo de regresión más común con datos de sobrevida es el modelo de “hazard” proporcional, el cual especifica que el “hazard” $h(t)$ para un sujeto con vector de covariables \mathbf{Z} está dado por:

$$h(t) = h_0(t) \exp(\beta' \mathbf{Z}) \quad (2.1)$$

donde $h_0(t)$ representa el “hazard” para un individuo con vector de covariables $\mathbf{Z}=\mathbf{0}$ y recibe el nombre de “hazard” basal. Así, el “hazard” es el producto del “hazard” basal por un factor dependiente de las covariables. Las componentes del vector de parámetro β describen la importancia de la covariable correspondiente.

Cox (1972) sugirió un procedimiento de estimación que elimina el efecto de $h_0(t)$ dejándolo completamente inespecificado; de esta manera, el análisis se concentra en el efecto de las covariables.

Otro modelo de regresión que aparece en la literatura es el llamado modelo de *tiempo de falla acelerado*. Este modelo más que describir el “hazard” como función de las covariables, se formula a partir del supuesto que éstas actúan

directamente en el tiempo de falla vía un factor de escala. Así, existe una función de sobrevida basal, $S_0(t)$, que representa la función de sobrevida para un individuo con covariables iguales a cero y una función de sobrevida para un individuo con covariables \mathbf{Z} , ligadas a través de la siguiente expresión:

$$S(t) = S_0 \left(\frac{t}{\exp(\eta' \mathbf{Z})} \right) \quad (2.2)$$

donde las componentes del vector de parámetros η describe la importancia de las covariables correspondientes. De aquí se sigue que el modelo en términos del "hazard" y del "hazard" basal puede ser escrito como:

$$h(t) = h_0(t \exp(-\eta' \mathbf{Z})) \exp(-\eta' \mathbf{Z}) \quad (2.3)$$

Si $S_0(t)$ es la función de sobrevida correspondiente a una distribución Weibull con parámetro de forma γ , el modelo (2.3) coincide con el modelo de "hazard" proporcional pero con diferente parametrización; y la relación entre los parámetros viene dada por $\eta_j = -\beta_j / \gamma_j$. No existe otra situación donde ambos modelos coincidan.

La manera estándar de describir un modelo de tiempo de falla acelerado es:

$$\log T_i = \eta' \mathbf{Z}_i + \varepsilon_i \quad (2.4)$$

donde las ε_i son independientes e idénticamente distribuidas. De esta manera, (2.4) describe un modelo lineal en el logaritmo de T . La distribución del error puede ser Weibull o cualquier otra; pero, en todos los casos es paramétrica.

La cuestión que surge naturalmente en la práctica es, entonces, si el modelo de tiempo de falla acelerado tiene más sentido que el modelo de "hazard" proporcional. El primero parece más sensible a procesos relacionados con el uso de alguna maquinaria o dispositivo y su nombre deriva precisamente del hecho que es más adecuado para medir el efecto de un factor que incrementa el riesgo de falla. Mientras que el modelo basado en la expresión (2.1) es mejor tanto para el manejo

de las covariables dependientes del tiempo, como en la descripción de los cambios en el riesgo correspondiente a puntos fijos en el tiempo. Pero, la principal ventaja del modelo de "hazard" proporcional es que permite estimar los parámetros del modelo sin suposiciones sobre el tipo de distribución del tiempo de sobrevida; a la inversa del modelo de tiempo de falla acelerado, el cual se basa en modelos paramétricos.

Para comprender mejor las diferencias y similitudes entre los modelos descriptos se puede expresar ambos en función del "hazard" acumulado, $H(t) = \int_0^t h(u) du$. Así, el modelo de "hazard" proporcional dice que, si se tiene 2 grupos se verifica que $H_1(t) = c H_2(t)$ para algún c , mientras que el modelo de tiempo de falla acelerado establece que $H_2(t) = H_1(t/c)$ para algún c .

3. Datos de Sobrevida Multivariados

Los datos de sobrevida multivariados surgen cuando *cada sujeto* bajo estudio puede experimentar *varios eventos, fallas o transiciones*, o cuando existe algún agrupamiento natural o artificial de sujetos (pequeños cluster) lo cual induce dependencia entre los tiempos de falla de los individuos del mismo grupo. Ejemplos en investigaciones médicas son: la recurrencia de tumores o episodios infecciosos; el desarrollo de síntomas o enfermedades en distintos sistemas de órganos; la ocurrencia de ceguera en el ojo izquierdo y derecho; la aparición de una enfermedad genética entre los miembros de una familia; etc. Ejemplos en otras áreas incluyen la rotura repetida de cierto tipo de maquinaria en la industria; la repetición de actos delictivos por parte de un individuo en estudios sociales; la elección de productos por cada consumidor en investigaciones de mercado.

Por lo tanto, si k eventos o fallas se pueden producir en cada sujeto o cluster bajo estudio, se define entonces un vector aleatorio de tiempos de falla $\mathbf{T}=(T_1, \dots, T_k)$, lo que da origen al problema del *Análisis de sobrevida multivariado*.

La generalización de la metodología para estimar la función de sobrevida al caso multivariado ha sido extensamente estudiada y, sin embargo, un estimador

óptimo de $S(t)$ todavía es desconocido ya que en muchas situaciones, con tiempos de sobrevida multivariado, no se satisface la suposición de censuras independientes.

3.1. Modelos de regresión para datos de sobrevida multivariados.

El estudio del efecto de covariables sobre el riesgo de fallas, en el caso de tener datos de sobrevida multivariados ha cobrado gran interés en los últimos años, fundamentalmente para la industria farmacéutica. Los ensayos clínicos diseñados para estudiar el efecto de nuevas drogas en enfermedades crónicas como asma y artritis con sucesivas crisis; o para el tratamiento de enfermedades como lupus, cáncer y SIDA que producen afecciones en diferentes órganos, requieren de metodología apropiada que contemple la posibilidad de manejar los tiempos de sobrevida a múltiples fallas.

Como ya se dijo en la sección anterior, el modelo de "hazard" proporcional de Cox ha sido extensamente usado para analizar la influencia de covariables en datos de sobrevida univariados, debido a su flexibilidad y al hecho que las propiedades asintóticas de sus estimadores ha sido elegantemente justificada usando la teoría de martingalas (Andersen *et.al.* (1992) y Fleming-Harrington (1991)).

Varios modelos de regresión han sido propuestos en la literatura para manejar el problema, en el caso de datos de sobrevida multivariados, usando extensiones del modelo de "hazard" proporcional de Cox. Sin embargo, hay algunos aspectos que surgen de la extensión del modelo original que deben ser considerados. Por ejemplo, cómo manejar la correlación entre los tiempos de falla dentro de un mismo individuo; cómo definir la escala de tiempo y los conjuntos bajo riesgo en cada instante de tiempo, ya que ellos cambian con el tiempo, y todo esto agravado por la presencia de censuras.

Por otra parte, en ese contexto se hace necesario distinguir dos tipos de situaciones marcadamente diferentes:

1. Los sujetos bajo estudio pueden experimentar sucesivas fallas del mismo tipo, las que reciben las denominaciones de *eventos recurrentes* o *eventos ordenados*. Ejemplo de ellas son crisis asmáticas o repetidas infecciones de un mismo órgano; roturas sucesivas del mismo circuito, en la industria; sucesivas deserciones, en ciencias de la educación; repetidos hechos delictivos, en ciencias sociales; etc.
2. Los sujetos bajo estudio pueden experimentar eventos de naturaleza enteramente diferente, y a las cuales se les dio la denominación de *eventos diferentes* o *eventos no ordenados*. Ejemplos de este tipo de eventos son metástasis en diferentes órganos; roturas de diferentes partes en un proceso industrial; elección de diferentes productos en un estudio de mercado.

En la situación de eventos no ordenados se puede considerar que los sujetos están "expuestos al riesgo" de cualquiera de las fallas desde el comienzo del período de observación; así, la escala de tiempo en este caso corresponde a los *tiempos totales*, es decir, tiempo desde que comienza el estudio.

Por otra parte, para los eventos ordenados la principal dificultad para el análisis es que los sujetos no pueden ser considerados como "expuestos al riesgo" para el segundo evento hasta que ellos no han experimentado el primero. Así, en este caso la escala de tiempo corresponde a los *tiempos intereventos*, es decir, el tiempo desde el evento anterior.

La diferencia fundamental entre las propuestas radica, principalmente, en cómo se maneja la estructura de correlación de los tiempos de falla dentro del modelo y en la etapa de la estimación, y cómo se definen las escalas de tiempo y los conjuntos bajo riesgo a la falla en cada momento. Teniendo en cuenta estas diferencias, los modelos se clasificaron en: *modelos marginales* y *modelos condicionales*. Los primeros consideran que el "hazard" marginal para cada falla sigue un modelo de "hazard" proporcional, como el dado en (2.1); luego, maximizando por separado la verosimilitud parcial, se propone tener en cuenta la

estructura de dependencia entre las distintas fallas dentro de cada sujeto, en el estimador de la matriz de covarianza de los parámetros del modelo.

Los modelos condicionales establecen que es el "hazard" condicional a la ocurrencia de los eventos anteriores el que sigue un modelo de "hazard" proporcional similar a (2.1), es decir, ellos contemplan la estructura de dependencia de los tiempos de falla en cada sujeto, dentro del mismo modelo, y no en la etapa de la estimación como los modelos marginales. Las propiedades de los estimadores de los parámetros, en los diferentes modelos, especialmente de los modelos condicionales, fue ampliamente estudiada por los autores. Sin embargo, las fuertes suposiciones que cada uno de ellos hacen sobre la estructura probabilística que deben tener los datos para que tales propiedades se cumplan, todavía no tiene una solución acabada.

Actualmente no existe acuerdo, entre los autores de las diferentes propuestas, sobre la utilidad de los distintos modelos en eventos ordenados o no ordenados, aún en el caso de tiempos de falla independientes (Oakes (1997), Wei y Gliddeb (1997)). Tampoco se encontró en la bibliografía, un estudio de cómo afecta a la precisión de los estimadores de los parámetros del modelo el aumento del grado de correlación entre los tiempos de falla.

En las últimas décadas, apareció en la literatura varias propuestas de distribuciones bivariadas de tipo exponencial que responden a la estructura de datos de sobrevida [Marshall y Olkin (1967), Block y Basu (1974), Sarkar(1987), Ryu (1993)]. En la propuesta de Ryu él establece la diferencia entre su distribución y las previas, basado en las propiedades de *falta de memoria marginal* y *falta de memoria conjunta*.

Puesto que en los modelos condicionales subyace el supuesto de que procesos no tiene memoria, surgió en este trabajo un nuevo interrogante, no planteado hasta ahora, y es el referido a cómo afecta a los estimadores de estos modelos si la distribución subyacente de los datos no cumple con esta propiedad.

En síntesis, los problemas que todavía no tienen una solución consensuada son aquellos referidos a cómo afecta a la eficiencia de los modelos: a) las diferentes

consideraciones que deben hacerse respecto a la escala de tiempos a utilizar en la situación de eventos ordenados; b) cuando aumenta el grado de correlación entre los tiempos de falla. Y, a partir de la publicación de Ryu, se plantea el problema adicional de conocer el comportamiento de los modelos cuando la distribución subyacente de los datos no cumple con la propiedad de falta de memoria.

Por otra parte, aunque estos modelos son especialmente adecuados para datos censurados, a excepción de Prentice y Cai que afirma que los sesgos en los estimadores se hacen importantes si el porcentaje de censuras es superior al 25%, ninguna otra propuesta analiza el efecto de las censuras en los estimadores de los modelos propuestos.

La importancia de resolver problemas puramente prácticos ya fue puntualizada por Cramer (1960) cuando expresa

“La teoría matemática pertenece por completo a la esfera conceptual y trata con entes puramente abstractos. Sin embargo, el designio de la teoría es construir un modelo de un cierto número de fenómenos del mundo físico, y las proposiciones y entes abstractos de la teoría tienen sus contrapartes en ciertas relaciones entre objetos y ciertos objetos observables. Si el modelo *ha de tener utilidad práctica*, debe haber un cierto tipo general de concordancia entre las proposiciones teóricas y sus contrapartes empíricas. Cuando una cierta proposición tenga como contraparte alguna relación directamente observable, debemos exigir que nuestras observaciones muestren que dicha relación se cumple efectivamente. Si en pruebas repetidas se ha encontrado una concordancia de este tipo, y consideramos que esta concordancia es suficientemente precisa y permanente, la teoría podrá considerarse aceptable en la práctica ...Los resultados de la teoría pueden aplicarse, como instrumentos en el *análisis científico* de los fenómenos observados...El principio general en que tales aplicaciones se basan puede expresarse así: *DEBE MODIFICARSE TODA TEORÍA QUE NO SE AJUSTE A LOS HECHOS*”.

En base a todo lo expuesto, en esta tesis se plantean los siguientes objetivos:

- A. Estudiar comparativamente el comportamiento de las diferentes propuestas de modelos de regresión tipo Cox para datos de sobrevida multivariados respecto a la precisión en la estimación de los parámetros, cuando no se cumplen las

condiciones bajo las cuales se derivaron las propiedades de los estimadores, a saber:

- i. Cuando los tiempos de falla para los distintos eventos definidos en cada individuo no son independientes.
 - ii. Cuando distribución conjunta subyacente no cumple con la condición de falta de memoria marginal, conjunta o ambas.
- B.** Estudiar, por medio de simulaciones, el efecto del tamaño de muestra y del porcentaje de censuras sobre cada uno de los modelos considerados.
- C.** Esclarecer la controversia sobre la bondad de los diferentes modelos para la situación de eventos ordenados y no ordenados, aún en el caso de tiempos de falla independientes y con distribuciones que gozan de la propiedad de falta de memoria marginal.

En el presente trabajo se considera de manera especial, el estudio de datos de sobrevida en ciencias biomédicas, aunque todos los métodos son adecuados para aplicaciones en confiabilidad industrial (tiempo libre de rotura), criminología (tiempo para volver a delinquir), ciencias sociales (duración del matrimonio), marketing (tiempo de suscripción a revistas), etc.

CAPITULO II

MODELO PROBABILÍSTICO

1. Introducción

Los procesos de conteo y martingalas proveen una manera directa de estudiar las propiedades de los estimadores y de los tests de hipótesis estadística para datos de tiempo de falla censurados a derecha, y en especial permiten una generalización directa de los métodos y modelos al caso multivariado.

En la Sección 2 de este capítulo, se presenta el modelo probabilístico de los datos de sobrevida con un enfoque de proceso de conteo. Se define la función "hazard", que, como ya se dijo, es una de las funciones que caracteriza la distribución de los tiempos de falla, de una manera que pueda ser fácilmente relacionada con los procesos estocásticos asociados a los datos de sobrevida.

Como se mencionó en el capítulo anterior, una importante razón para plantear modelos y métodos especiales para los datos de tiempo de falla es la necesidad de tratar con las censuras; de allí que en la Sección 3 se hace una breve análisis de posibles estructuras y mecanismos de censuras, así como los problemas adicionales que surgen en el análisis de sobrevida multivariado, en presencia de ellas. Por último, en la Sección 4, se define la función de verosimilitud en términos del compensador del proceso de conteo multivariado autogenerado y a partir de ella se deduce la *verosimilitud parcial*.

2. Modelo Probabilístico de los Datos de Sobrevida

Típicamente, en el análisis de sobrevida, las observaciones *completas* de una muestra de tiempos X_1, \dots, X_n no son posibles. Así, se observa el par (T_i, D_i) , $i=1, \dots, n$, donde D_i es un indicador de censura, es decir, una variable aleatoria dicotómica describiendo si T_i representa el tiempo de falla X_i o es sólo una cota inferior de él, es decir

$$\begin{aligned} X_i = T_i & \quad \text{si} \quad D_i = 1 \\ X_i > T_i & \quad \text{si} \quad D_i = 0 \end{aligned} \quad (2.2.1)$$

Como ya se dijo en el capítulo anterior, el análisis estadístico de este tipo de datos no es posible sin hacer suposiciones sobre las censuras. Por el momento se considera la suposición más general: *censuras independientes*, lo cual significa que, en cualquier tiempo t , la experiencia de sobrevida en el futuro no es estadísticamente alterada por las censuras y la experiencia de sobrevida en el pasado. Matemáticamente la noción de pasado y futuro se define a través del concepto de *filtración* o *historia del proceso* $(\mathfrak{F}_t)_{t \geq 0}$, donde \mathfrak{F}_t representa los datos disponibles hasta el tiempo observado t . La expresión \mathfrak{F}_t corresponde a los datos disponibles hasta justo antes de t . La definición formal de filtración se presenta en el Anexo 1, aquí simplemente tomamos \mathfrak{F}_t para significar los valores de (T_i, D_i) para $i=1, \dots, n$, tales que $T_i \leq t$ y \mathfrak{F}_t los valores de (T_i, D_i) tal que $T_i < t$.

Sea X la variable aleatoria "tiempo de sobrevida a la falla" y sea U la variable aleatoria "tiempo de sobrevida censurado". Por la suposición de censuras independientes, X y U resultan ser variables aleatorias independientes, luego el par de variables observadas (T, D) queda definido como $T = \min(X, U)$ y $D = I\{X \leq U\}$.

La independencia entre X y U lleva a que la función "hazard" puede expresarse como

$$\begin{aligned} h(t) &= \lim_{\Delta \downarrow 0} \frac{\Pr(t < T < t + \Delta t \mid T \geq t)}{\Delta t} \\ &= \lim_{\Delta \downarrow 0} \frac{\Pr(t < T < t + \Delta t \mid T \geq t, U \geq t)}{\Delta t} \end{aligned} \quad (2.2.2)$$

Así, de manera todavía muy informal se puede escribir

$$\Pr[T_i \in [t, t + dt), D = 1 \mid \mathfrak{F}_{t-}] = \begin{cases} h(t) dt & \text{si } T_i \geq t \\ 0 & \text{si } T_i < t \end{cases} \quad (2.2.3)$$

Reemplazando la probabilidad del primer miembro por la esperanza de una variable indicadora se tiene

$$\begin{aligned} E[I\{T_i \in [t, t+dt), D=1\} | \mathfrak{F}_{t-}] &= \#\{i : T_i \in [t, t+dt), D=1\} h(t) dt \\ &= Y(t) h(t) dt \\ &= \lambda(t) dt \end{aligned} \quad (2.2.4)$$

donde $Y(t) = \#\{T \in [t, t+dt), D=1\}$ es el número de individuos al riesgo de falla en $[t, t+dt)$, o sea, el tamaño del conjunto bajo riesgo en ese intervalo de tiempo, y $\lambda(t) = Y(t) h(t)$.

Si se define una familia de variables aleatorias $N = \{N(t), t \in [0, \infty)\}$, donde $N(t) = \#\{i : T_i \leq t, D_i = 1\}$ representa el número de eventos que ocurrieron hasta t , entonces la familia de variables aleatorias N es un *proceso estocástico de conteo*.

El proceso N se dice *adaptado a la filtración* \mathfrak{F}_t si $N(t)$ es \mathfrak{F}_t -medible para todo t .

Por lo tanto, si se denota $dN(t)$ o $N(dt)$ al incremento de N sobre el intervalo infinitesimal $[t, t+dt)$, es decir, $dN(t) = N(t+dt) - N(t)$; dicho incremento es la expresión que aparece dentro del operador esperanza dado en (2.2.4), así la ecuación se puede reescribir como

$$\lambda(t) dt = E[dN(t) | \mathfrak{F}_t] \quad (2.2.5)$$

Se debe notar que $\lambda(t)$ es una variable aleatoria, ya que ella depende de las variables condicionantes en \mathfrak{F}_t , y, en este contexto, toma el nombre de *intensidad del proceso* $N(t)$.

A partir de (2.2.5) se puede inferir que $\lambda(t)$ representa la tasa de cambio instantáneo en N en $[t, t+dt)$ dado \mathfrak{F}_t , o lo que es lo mismo, dado que el tiempo de falla y el tiempo de censura son ambos mayores que t .

De una manera general, un proceso de conteo $\{N(t), t \in [0, \infty)\}$ es un proceso estocástico que toma valores $0, 1, 2, \dots$, que tiene un *proceso intensidad* definido a través de (2.2.5).

El proceso $\Lambda(t)$ dado por

$$\begin{aligned}\Lambda(t) &= \int_0^t \lambda(s) ds & t \geq 0 \\ &= \int_0^t Y(s)h(s) ds\end{aligned}\tag{2.2.6}$$

es llamado *proceso de intensidad acumulado*, el cual en cada tiempo t es una variable aleatoria que representa el número de saltos de N en $(0, t]$. En efecto, ya que $N((t + dt) -) - N(t -)$ es una variable aleatoria 0-1 en t , y si $C(u) = \Pr[U > u]$, se tiene que

$$\begin{aligned}E[N(t)] &= \Pr[T \leq t, D=1] = \Pr[T \leq t, T \leq U] \\ &= \int_0^t C(s-) dF(s) \\ &= \int_0^t C(s-) S(s) \frac{dF(s)}{S(s)} \\ &= \int_0^t \Pr[T \geq s] h(s) ds \\ &= E \left[\int_0^t I\{T \geq s\} h(s) ds \right] \\ &= E \left[\int_0^t \lambda(s) ds \right] = E[\Lambda(t)]\end{aligned}\tag{2.2.7}$$

Supongamos por el momento que el proceso $N(t)$ no es censurado, entonces, si se define el proceso $M(t)$ como

$$M(t) = N(t) - \Lambda(t),\tag{2.2.8}$$

por (2.2.7) se sigue que $M(t)$ tiene media cero. El proceso $M(t)$ recibe el nombre de *proceso de conteo compensado* o *proceso de conteo martingala*.

Tomando la esperanza condicional del incremento de $M(t)$ sobre el intervalo $[t, t + dt)$, dada la información del pasado \mathfrak{F}_t , se tiene

$$E[dM(t) | \mathfrak{F}_t] = E[dN(t) - d\Lambda(t) | \mathfrak{F}_t].$$

Por (2.6) esta expresión se puede escribir como

$$E[dM(t) | \mathfrak{F}_t] = E[dN(t) - \lambda(t)dt | \mathfrak{F}_t].$$

Se puede notar, a partir de la definición dada en (2.2.5), que $\lambda(t)$ es no aleatoria dado \mathfrak{F}_t , y entonces,

$$E[dM(t) | \mathfrak{F}_t] = E[dN(t) | \mathfrak{F}_t] - \lambda(t)dt = 0. \quad (2.2.9)$$

Por las relaciones (2.2.8) y (2.2.9), $\Lambda(t)$ representa el *compensador de N(t)*, ya que $M(t)$ es *una martingala*, considerando que (2.2.9) es equivalente, al menos heurísticamente, a la definición de proceso martingala (Ver Apéndice 1)

$$E[dM(t) | \mathfrak{F}_s] = M(s), \quad \forall s < t. \quad (2.2.10)$$

En efecto, ya que (2.2.10) se debe cumplir para todos los intervalos $(s, t]$, sumando los incrementos de M sobre pequeños subintervalos $[u, u + du)$ que se obtienen particionando $[s + ds, t + dt) = (s, t]$ se encuentra que

$$\begin{aligned} E[M(t) | \mathfrak{F}_s] - M(s) &= E[M(t) - M(s) | \mathfrak{F}_s] \\ &= E\left[\int_{s < u \leq t} dM(u) | \mathfrak{F}_s\right] \\ &= \int_{s < u \leq t} E[dM(u) | \mathfrak{F}_s] \\ &= \int_{s < u \leq t} E[E(dM(u) | \mathfrak{F}_{u-}) | \mathfrak{F}_s] \\ &= 0 \end{aligned} \quad (2.2.11)$$

La primera igualdad se cumple por la medibilidad de $M(s)$ respecto a \mathfrak{F}_s y la última igualdad vale ya que $\mathfrak{F}_s \subset \mathfrak{F}_{u-}$.

El proceso $N(t)$ en este contexto, también recibe el nombre de *proceso de conteo compensado*.

Puesto que, $\lambda(t)$ y como consecuencia $\Lambda(t)$, son conocidas dado \mathfrak{F}_t , ellas son funciones no aleatorias justo antes de t , y por ello se dice que son *procesos predictibles* (Ver Anexo 1). Luego, a partir de (2.2.8) se podría considerar al proceso martingala como que es un *proceso puro ruido* siendo la parte sistemática del proceso de conteo $N(t)$ su compensador $\Lambda(t)$: un proceso predecible variando suavemente, el cual cuando se sustrae del proceso de conteo $N(t)$ deja un proceso de ruido, predecible de media cero.

Un proceso de conteo multivariado es un proceso estocástico, el cual en este contexto se le puede considerar que registra la ocurrencia en el tiempo de varios tipos de eventos disjuntos.

Aquí se supone que la filtración a la cual el proceso es adaptado ya está dada o bien se construye a partir de él, es decir, es generado por el mismo proceso y entonces recibe el nombre de *filtración autogenerada* y que se denotará con \mathfrak{N}_t y se define como

$$\mathfrak{N}_t = \sigma(N(s): s \leq t) \quad \forall t. \quad (2.2.13)$$

Así, se supone una filtración $(\mathfrak{F}_t, t \in \tau)$ sobre un espacio de probabilidad $(\Omega, \mathfrak{F}, P)$ la cual satisface las condiciones usuales (ver Anexo 1), excepto posiblemente el de completitud. Un *proceso de conteo multivariado*

$$N(t) = (N_1(t), \dots, N_k(t))' \quad (2.2.12)$$

es un vector de k procesos *cadlag* (ver definición en Anexo 1), todos ceros al tiempo cero, con trayectorias las cuales son constantes a trozos y no decrecientes, teniendo saltos de tamaño $+1$ solamente y, además, dos componentes cualesquiera no pueden saltar simultáneamente.

Se supone que $N_k(t)$ es finito excepto en un conjunto de medida nula para cada k y para todo $t \in \tau$. El proceso es llamado *autogenerado* si la filtración asociada es aquella generada por el mismo proceso.

Para un proceso de conteo N se puede asociar una sucesión de *tiempos de saltos* y de *marcas de saltos*, los cuales dan el tiempo de falla y tipo de cada evento. Debido a que dos componentes no saltan simultáneamente, la suma de ellas es también un proceso de conteo, definido por

$$N_{\bullet} = \sum_{h=1}^k N_h \quad (2.2.14)$$

El compensador del proceso N es un vector Λ cuyas componentes son los compensadores Λ_k de las componentes N_k de N .

Es claro que, debido a la existencia y unicidad del compensador, existe y es único el proceso de intensidad integrado o acumulado Λ_{\bullet} . Por otra parte, si λ existe y es continuo a izquierda y con límites a derecha, entonces él es único.

Lima y Sen (1998) expresan las componentes del vector intensidad λ_k , en el caso bivariado, en términos de los "hazard" condicionales y de los procesos \mathcal{N}_t -predictibles $Y_k(t) = I\{T_k \geq t\}$, con $t \geq 0$ y $k = 1, 2$, notando que en este caso \mathcal{N}_t contiene información sobre si una o ambas fallas tuvieron lugar justo antes de t . Se necesita entonces considerar las siguientes situaciones:

- i) Ningún evento tuvo lugar antes de t , es decir, $Y_1(t) = Y_2(t) = 1$.
- ii) Se presentó el primer evento antes de t pero el segundo no tuvo lugar, es decir, $Y_1(t) = 0, Y_2(t) = 1$.
- iii) Solamente se presentó el segundo evento antes de t , esto significa que, $Y_1(t) = 1, Y_2(t) = 0$.
- iv) Ambos eventos se presentaron antes de t , es decir, $Y_1(t) = Y_2(t) = 0$.

De esta manera, si se quiere escribir la primera componente de λ , sólo habrá que considerar el caso en que $Y_1(t) = 1$, y así, de acuerdo a la definición de λ_k se puede escribir que

$$\lambda_1(t) = E\left\{dN_1(t) | \mathcal{N}_{t^-}\right\} = p_1^{(1)}(t)Y_1(t)[1 - Y_2(t)] + p_2^{(1)}(t)Y_1(t)Y_2(t) \quad (2.2.15)$$

donde

$$p_1^{(1)}(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P\{T_1 \in [t, t + \Delta t) \mid T_1 \geq t, T_2 < t\}$$

(2.2.16)

y

$$p_2^{(1)}(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P\{T_1 \in [t, t + \Delta t) \mid T_1 \geq t, T_2 \geq t\}$$

Pueden ser interpretadas como las funciones "hazard" condicional dado T_2 . Similarmente, para la componente 2,

$$\lambda_2(t) = E\left\{dN_2(t) | \mathcal{N}_{t^-}\right\} = p_2^{(2)}(t)Y_2(t)[1 - Y_1(t)] + p_1^{(2)}(t)Y_1(t)Y_2(t) \quad (2.2.17)$$

para

$$p_2^{(2)}(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P\{T_2 \in [t, t + \Delta t) \mid T_2 \geq t, T_1 < t\}$$

(2.2.18)

y

$$p_1^{(2)}(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P\{T_2 \in [t, t + \Delta t) \mid T_2 \geq t, T_1 \geq t\}$$

Basándose en (2.2.15) y (2.2.17) se puede representar el vector de proceso intensidad como un producto matricial

$$\lambda(t) = \begin{pmatrix} \lambda_1(t) \\ \lambda_2(t) \end{pmatrix} = \begin{pmatrix} Y_1(t) & 0 \\ 0 & Y_2(t) \end{pmatrix} \begin{pmatrix} \alpha_{11}(t) & \alpha_{12}(t) \\ \alpha_{21}(t) & \alpha_{22}(t) \end{pmatrix} \begin{pmatrix} Y_1(t) \\ Y_2(t) \end{pmatrix} \quad (2.2.19)$$

donde $\alpha_{11} = p_1^{(1)}$, $\alpha_{12} = p_2^{(1)} - p_1^{(1)}$, $\alpha_{21} = p_1^{(2)} - p_2^{(2)}$ y $\alpha_{22} = p_2^{(2)}$.

Esta formulación permite ver claramente el rol de los "hazard" condicionales en el vector intensidad del proceso multivariado y como se verá más adelante, facilitará la interpretación estadística de la formulación de los modelos semiparamétricos tipo Cox.

3. Mecanismos de censuras

Se comienza esta sección con una introducción a las censuras a derecha y ejemplos de modelos de mecanismos de censuras a derecha. Se introduce luego el concepto de *censuras a derecha independientes*, y en la Sección 4 se discute el concepto de *censuras a derecha no informativas*.

Si se considera un estudio de sobrevida con una muestra de tamaño n , entonces se tiene una matriz de procesos de conteo que se puede expresar en término de los vectores N_1, \dots, N_n , que son n copias del proceso $N(t)$ definido en (2.2.12), luego la matriz de procesos de conteo estará dada por

$$N(t) = (N_1(t), \dots, N_n(t))' \quad (2.3.1)$$

Ahora se verá cómo la forma más común de observaciones incompletas, las *censuras a derecha*, pueden ser introducida dentro del modelo probabilístico.

El proceso de conteo multivariado dado en (2.3.1) se puede expresar también como

$$N = (N_{hi}; i = 1, \dots, n; h = 1, \dots, k)$$

Donde i representa al i -ésimo *individuo* y h el h -ésimo *tipo de evento* que los individuos pueden experimentar. Por ejemplo, $h = 1, \dots, k$ puede indicar las sucesivas crisis asmáticas.

Una censura a derecha de N es la situación donde la observación de $N_{hi}(\cdot)$, con $h = 1, \dots, k$, finaliza después de algún tiempo U_i (posiblemente aleatorio), es decir, N_{hi} es observado solamente en un conjunto aleatorio $A_i = \{t \leq U_i\} \subseteq \mathfrak{T}$ o de manera equivalente, cuando el proceso

$$C_i(t) = I(t \in A_i) = I(t \leq U_i) = 1 \quad (2.3.2)$$

La introducción de censuras a derecha dentro del modelo a menudo producirá una variación aleatoria extra, es decir, el proceso de censuras

$$C(t) = \{C_i(t), i = 1, \dots, n\},$$

el cual no necesariamente es adaptado a la filtración original (\mathfrak{F}_t) . Así, se debe trabajar con una filtración *agrandada* $\mathfrak{F}_t \supseteq \mathfrak{F}_t$ con la suposición que $C(\cdot)$ es \mathfrak{F}_t -predictible y esto se puede conseguir tomando

$$\mathfrak{F}_t = \mathfrak{F}_t \vee \sigma(C(u), u \leq t) \quad (2.3.3)$$

donde $\sigma(C(u), u \leq t)$ es la filtración generada por C . De este modo el proceso de censura depende sólo del pasado y no del futuro.

Después de tener la filtración *agrandada* \mathfrak{F}_t para incluir cualquier variación aleatoria adicional de los tiempos censurados, ella puede reducirse especificando cuales son los datos disponibles para el investigador en cualquier tiempo t después de la censura. Así, la parte observable de N o sea el *proceso de conteo censurado a derecha* $N^c = (N_{hi}^c)$ está dado por

$$N_{hi}^c = \int_0^t C_i(s) dN_{hi}(s) \quad (2.3.4)$$

y entonces (\mathfrak{F}_t^c) será la filtración generada por los datos observados.

Para dar algunos ejemplos de modelos para mecanismos censurados a derecha primero consideramos el caso especial donde el proceso de conteo multivariado N es definido a partir de n tiempos de sobrevida independientes X_1, \dots, X_n . En este contexto las observaciones consisten típicamente de $(T_i, D_i, i = 1, \dots, n)$, donde $T_i = X_i \wedge U_i$ y $D_i = I(T_i = X_i)$. Así, para cada individuo está disponible el valor de T_i y la información de si se trata de un tiempo de falla o un tiempo de censura. Si $D_i = 1$, entonces no se observa el tiempo de censura U_i .

Los datos disponibles al tiempo t están dados por el proceso

$$(N^c(u), Y^c(u); 0 \leq u \leq t)$$

donde $N^c = (N_i^c, i = 1, \dots, n)$ e $Y^c = (Y_i^c, i = 1, \dots, n)$. Aquí,

$$N_i^c(t) = I(T_i \leq t, D_i = 1)$$

es el proceso de conteo censurado a derecha para el individuo i , y el proceso

$$Y_i^c(t) = I(T_i \geq t)$$

indica si el individuo i está al riesgo de falla justo antes de t .

En muchas situaciones, con datos de sobrevida se tiene un proceso de conteo definido por $N_i(t) = I(X_i \leq t)$ y su compensador se estudia con respecto a la filtración generada por dicho proceso (\mathfrak{N}_i), y donde el proceso de censura $C_i(t) = I(t \leq U_i)$ es predecible con respecto a dicha filtración, es decir, que la filtración original \mathfrak{N}_i no necesita ser agrandada para incluir las censuras dentro del modelo.

En la bibliografía se considera los siguientes tipos de censuras a derecha:

➤ Censuras Tipo I simples

Aquí las observaciones de cada individuo cesan en un tiempo *determinístico* u_0 común a todos los sujetos bajo estudio, de esta manera $C_i(t) = I(t \leq u_0)$ no es aleatorio y trivialmente predecible con respecto a \mathfrak{N}_i . Este esquema de censura es más común en la industria donde n ítems idénticos se ponen a prueba simultáneamente y son observados en un intervalo fijo $[0, u_0]$.

➤ Censuras Tipo II simples

En este caso el estudio termina en el tiempo de la r -ésima falla, $r < n$, es decir, $U_i = X_{(r)}$, $i = 1, \dots, n$. Luego $C_i = I(t \leq X_{(r)})$ es predecible con respecto a \mathfrak{N}_i . Este tipo de censura es también común en experimentos industriales. Se debe notar que en este ejemplo las observaciones T_1, \dots, T_n son dependientes.

En los siguientes mecanismos de censura, \mathfrak{F}_i es estrictamente más grande que \mathfrak{N}_i .

➤ Censuras de Tipo I progresivas

Se presenta cuando el estudio termina en una fecha predeterminada, pero los individuos entran al mismo consecutivamente, este es el caso de muchos ensayos clínicos en medicina.

Cuando el interés se centra en el tiempo de vida *a partir* de la entrada al estudio, el máximo tiempo de observación para el individuo i será el tiempo U_i desde la entrada hasta la fecha de cierre del experimento.

Si los tiempos de sobrevida en el período de observación son *independientes* del tiempo de entrada (X_o), y se designa con $\varphi_o = \sigma(X_o)$, entonces, el proceso de censura C con componentes $C_i(t) = I(t \leq U_i)$ es predecible con respecto a $\varphi_t = \varphi_o \vee N_t$.

Otra manera de establecer la φ_t - predictibilidad de C , es diciendo que dada φ_o , los tiempos de censura $U_1 = u_1, \dots, U_n = u_n$ son *determinísticos* y así tenemos que las censuras Tipo I progresivas son la generalización de las censuras Tipo I simples.

Aquí, todos los tiempos de censura son observables ya que se conoce si el individuo tiene la falla o no antes de la fecha de cierre. Esto, por supuesto, es una consecuencia de la suposición que las censuras se producen porque los individuos no presentan la falla hasta la finalización del estudio.

➤ Censuras aleatorias

El modelo de censuras aleatorias es una generalización de (a) y (c), donde $U = (U_1, \dots, U_n)$ es independiente de $X = (X_1, \dots, X_n)$, pero donde U puede tener una distribución arbitraria.

El modelo de censuras *clásico* en el cual los U_1, \dots, U_n son independientes e idénticamente distribuidos, es matemáticamente el modelo más manejable para los mecanismos de censuras y es el que se supone válido en la mayoría de las publicaciones sobre datos de sobrevida. En este caso, se supone que φ_t está definida por la expresión (2.3.3). Una manera alternativa de definir φ_t sería $\varphi_t = \varphi_o \vee N_t$, donde $\varphi_o = \sigma(U)$, pero, excepto para los modelos de censuras Tipo I progresivas, es muy antinatural modelar los tiempos de censuras como realizándose al tiempo 0, y también contradice la convención de que los datos están disponibles en el tiempo t .

El hecho que se considere otra filtración puede tener la consecuencia que el compensador de N cambie; en efecto, un problema fundamental que hay que enfrentar en presencia de censuras a derecha es que ellas *pueden alterar las intensidades del evento* de interés. Si por ejemplo, en un ensayo clínico se excluye del estudio aquellos pacientes que presentan características particulares de la enfermedad, entonces los restantes que quedan al riesgo no son ya representativos de la muestra con que se inició el estudio. Los mecanismos de censura a derecha que evitan éste y similares problemas, permitiendo mantener la representatividad de los conjuntos bajo riesgo son generalmente llamados *independientes*.

La formalización matemática para modelos de procesos de conteo generales no es directa, sino que se realiza a través de la siguiente definición:

Definición: Sea N un proceso de conteo multivariado con compensador Λ^0 respecto a una dada filtración (\mathfrak{F}_t) y medida de probabilidad $P_{\theta\phi}$, y, sea C el proceso de censuras a derecha predecible con respecto a la filtración $\mathcal{G}_t \supseteq \mathfrak{F}_t$. Entonces, se dice que las censuras de N generadas por C son independientes si el $P_{\theta\phi}$ -compensador de N con respecto a \mathcal{G}_t es también Λ^0 .

En otras palabras: el conocimiento adicional de los tiempos censurados a derecha hasta justo antes de t no debería alterar el proceso intensidad para N al tiempo t . Por lo tanto, si, como en el ejemplo dado anteriormente, los pacientes que mejoran (empeoran) notablemente son excluidos del estudio, entonces el conocimiento a cerca de un paciente que todavía está en el estudio (no censurado) indicaría que tiene una intensidad de muerte mayor (menor) comparada con la situación sin censura. Entonces, tal esquema de censuras no será independiente a menos que las características de los pacientes excluidos se introduzcan en el modelo como *covariables*.

Las censuras Tipo I y II simples son obviamente independientes ya que $(\mathcal{G}_t) = (\mathfrak{F}_t)$.

A causa de que $N_{hi}(t) = \Lambda_{hi}(t, \theta) + M_{hi}(t)$ donde M_{hi} es una martingala con respecto a (\mathcal{G}_t) , la parte observable de N_{hi} de acuerdo a (2.3.3) está dada por

$$\begin{aligned}
 N_{hi}^c &= \int_0^t C_i(s) dN_{hi}(s) \\
 &= \int_0^t C_i(s) d\Lambda(s, \theta) + \int_0^t C_i(s) dM_{hi}(s) \quad (2.3.5) \\
 &= \Lambda_{hi}^c(t, \theta) + M_{hi}^c(t)
 \end{aligned}$$

Ya que M_{hi}^c es la integral estocástica de un proceso predecible y acotado $C_i(\cdot)$ con respecto a una martingala M_{hi} , ella es también una martingala. Entonces, bajo censuras independientes N_{hi}^c tiene un $(P_{0\phi} - (\phi_t))$ - compensador para todo ϕ dado por

$$\Lambda_{hi}^c = \int_0^t C_i(s) d\Lambda_{hi}(s) \quad (2.3.6)$$

4. Función de Verosimilitud

Previo a la formulación de la función de verosimilitud, es necesario introducir el concepto de *producto integral* y la noción intuitiva de cómo se debe considerar los procesos de conteo bajo diferentes medidas de probabilidad.

La notación usada aquí, \mathcal{P} , para el *producto integral* intenta sugerir que se trata de una versión continua del producto ordinario, Π ; así como la notación \int generaliza la suma \sum . Para una definición más formal ver Gill y Johansen (1990).

Se puede pensar que una medida de probabilidad \mathbf{P} sobre \mathfrak{T} se puede construir a través de los siguientes pasos:

1. Especificar \mathbf{P} sobre \mathfrak{T}_0 .
2. Especificar la distribución condicional, dada \mathfrak{T}_0 , del tiempo T_1 perteneciente a $(0, \tau]$, del primer salto de \mathbf{N} .

3. Especificar la distribución condicional del *tipo* del primer salto (evento) J_1 perteneciente al conjunto $\{0, 1, \dots, k\}$ dado \mathfrak{F}_0 y T_1 donde $J_1 = 0$ sólo se permite si $T_1 = \tau$ y significa que no hay ningún primer salto de N .
4. Especificar la distribución condicional de T_2 perteneciente a $(T_1, \tau]$ dada \mathfrak{F}_0 , T_1 y J_1 y así sucesivamente.

Bajo ciertas condiciones suficientes se puede estar seguro de que no ocurran infinitos saltos con probabilidad positiva en el intervalo considerado y así la construcción no tendría problemas.

Otra manera alternativa de describir la distribución de un proceso de conteo es pensar que la distribución condicional de los T_h , dado el pasado, puede también especificarse a través de sus funciones "hazard" acumuladas. Multiplicando la tasa "hazard" al tiempo t por la probabilidad de un salto de tipo h al tiempo t , dado que ocurrió un salto en ese tiempo, se obtiene el "hazard" específico del tipo de evento h ; inversamente, a partir del "hazard" específico para (T_h, J_h) dado el pasado, se puede reconstruir el "hazard" total y los "hazard" condicionales. Estos "hazard" condicionales son la probabilidad, dependiente del tiempo, de que un evento de cada tipo ocurra, dado que ocurrió uno y en términos de la intensidad del proceso N ellos representan las componentes del vector intensidad asociado al vector de proceso de conteo N . Así, este argumento informal sugiere que bajo las condiciones de adaptabilidad del proceso a las filtraciones, el proceso de intensidad acumulado de N está determinado por las distribuciones de probabilidad condicionales e, inversamente, a partir del proceso intensidad, se puede obtener las distribuciones condicionales y entonces, junto con el conocimiento de \mathbf{P} sobre \mathfrak{F}_0 , construir \mathbf{P} sobre \mathfrak{F} .

Ahora, en lugar de avanzar de salto en salto de N , se podría ir también de intervalos infinitesimales a intervalos infinitesimales, con probabilidad de salto de tipo h , $d\Lambda_h(t)$, en el intervalo $[t, t + dt)$, dado el pasado (\mathfrak{F}_0 y la trayectoria de N hasta el tiempo t). Esto da una manera alternativa de escribir la densidad de probabilidad o la razón de verosimilitud, que puede ser interpretada más fácilmente

y conduce a la noción de *verosimilitud parcial* de una manera natural. En efecto, si P_h representa la distribución conjunta de (T_h, J_h) dado $\mathfrak{F}_{T_{h-1}} = \mathfrak{F}_0 \vee \sigma\{T_1, J_1, \dots, T_{h-1}, J_{h-1}\}$ y $A = \{1, \dots, k\}$, entonces sobre el intervalo (T_{h-1}, T_h) se tiene que

$$P_h[(t, \tau], A) = \rho_{T_{h-1}}^t \left(1 - d \Lambda_{\bullet}(s) \right) \quad (2.4.1)$$

donde \bullet denota la suma del "hazard" acumulado sobre todos los k eventos. Por otra parte si $P_h(h | t)$ denota la distribución condicional de J_h dado $T_h = t$, entonces,

$$P_h(h | t) = \frac{d\Lambda_h(t)}{d\Lambda_{\bullet}(t)} \quad (2.4.2)$$

Las expresiones (2.4.1) y (2.4.2) son parte de las llamadas fórmulas de Jacod para el proceso intensidad.

Puesto que, para cada t , N_h es una variable dicotómica para la cual, a partir de (2.2.5), se verifica que $P[dN_h(t) | \mathfrak{F}_t] = d\Lambda_h(t)$, de una manera intuitiva, la verosimilitud del proceso total podría escribirse como productos de experimentos multinomiales condicionales e infinitesimales ($n=1$; $k+1$ celdas). En efecto, Jacod (1975) expresa la razón de verosimilitudes en función de la siguiente expresión para la función de verosimilitud,

$$dP |_{\mathfrak{F}_t} = dP |_{\mathfrak{F}_0} \rho_{t \in [0, \tau]} \left[\prod_h d\Lambda_h(t)^{\Delta N_h(t)} (1 - d\Lambda_{\bullet}(t))^{1 - \Delta N_{\bullet}(t)} \right] \quad (2.4.3)$$

La interpretación heurística de la expresión (2.4.3) conecta naturalmente a la noción de *verosimilitud parcial* dada por Cox (1975); para ver esto, se supone, por el momento, que $k=1$, entonces, la densidad de la muestra \mathbf{X} puede ser factorizada como el producto de densidades condicionales $p(x_i | x_0, x_1, \dots, x_{i-1})$ de cada X_i dados sus predecesores X_0, \dots, X_{i-1} . Como una función de un parámetro desconocido θ de la distribución de \mathbf{X} , esto es la verosimilitud para θ basada en \mathbf{X} . Cox (1975) afirma que si se extrae algunos factores, por ejemplo todos los pares, el resto todavía puede usarse como base para la inferencia estadística. En particular,

siguen siendo válidas las propiedades para muestras grandes de los estimadores por máxima verosimilitud, y también se podrían aplicar a esta "verosimilitud parcial" los tests de razón de verosimilitud y test "score". Lo que justifica realmente esta afirmación es la teoría de martingala ya que las componentes del "score" parcial $(\partial/\partial\theta)\ln p(X_i | X_0, X_1, \dots, X_{i-1}; \theta)$, $i = 1, 3, \dots$ forman la sucesión de diferencias martingalas de tiempo discreto y de aquí se deriva que la verosimilitud parcial satisface la usual relación entre las esperanzas de la primera y segunda derivadas. Más aún, si el número de componentes es grande y su influencia individual es pequeña, el teorema central del límite para las martingalas podría aplicarse para obtener la normalidad asintótica del "score" parcial.

Todo lo expuesto sugiere, entonces, que de la expresión (2.4.3) se podría extraer las contribuciones para el intervalo de tiempo dt de acuerdo al valor de un proceso indicador predecible C , y en efecto, el resultado es llamado *la verosimilitud parcial para el proceso de conteo filtrado* $\int C dN$, definido en la expresión (2.3.4).

En el caso en que $k > 1$, para construir la verosimilitud para N^c bajo censuras independientes, se puede factorizar cada verosimilitud multinomial en (2.4.3) como el producto de una verosimilitud marginal correspondiente al evento del primer tipo y la verosimilitud condicional del segundo dado el primero, luego se podría formar una verosimilitud parcial de cada uno de estos factores.

Sea N el proceso de conteo básico, es decir completamente observado, luego la (\mathfrak{F}_t) -verosimilitud basada en las observaciones de N y X_0 , donde la distribución de X_0 puede depender también de θ y de ϕ , entonces la expresión (2.4.3) puede escribirse como

$$L(\theta, \phi) = L_0(\theta, \phi) \mathcal{P}_t \left[\prod_{i=1}^n \prod_{h=1}^k d\Lambda_{hi}^\theta(t)^{\Delta N_{hi}(t)} (1 - d\Lambda_{..}^\theta(t))^{1 - \Delta N_{..}(t)} \right] \quad (2.4.4)$$

donde $N_{..} = \sum_{i=1}^n \sum_{h=1}^k N_{hi}$ y $\Lambda_{..}^\theta$ es el $P_{\theta\phi}$ -compensador de $N_{..}$ con respecto a (\mathfrak{F}_t) . De

manera abreviada (2.4.4) se puede escribir como,

$$L(\theta, \phi) = L_0(\theta, \phi) L_t(\theta)$$

Si \mathbf{X}_0 no depende de θ , entonces $L_t(\theta)$ es la verosimilitud completa para θ basada en las observaciones de \mathbf{N} ; de otra manera, $L_t(\theta)$ es la verosimilitud parcial.

Sea el proceso de conteo censurado a derecha,

$$\mathbf{N}^c = (N_{hi}^c, i = 1, \dots, n; h = 1, \dots, k),$$

se supone que las observaciones al tiempo t incluyen a \mathbf{X}_0 y $(\mathbf{N}^c(u), 0 \leq u \leq t)$ junto con los tiempos de censura a derecha $U_i \leq t$ de los individuos para los cuales no experimentaron ningún evento. Luego, las observaciones pueden formalizarse como la terna $(\mathbf{X}_0, \mathbf{N}^c, \mathfrak{N}^c(t))$, donde $\mathfrak{N}^c(t) = (\Lambda^c(u, \theta), \theta \in \Theta, 0 \leq u \leq t)$ es la familia de $\mathbf{P}_{\theta\phi}$ -compensadores de \mathbf{N}^c con respecto a $(\mathcal{F}_t) = (\mathfrak{F}_t^c)$.

Por los argumentos dado anteriormente la verosimilitud puede reescribirse como

$$L_\tau(\theta, \phi) = L_0(\theta, \phi) \rho_t \left[\mathbf{P}_{\theta\phi}(d\mathbf{N}(t)) \middle| \mathfrak{F}_{t-}^c \right], \quad (2.4.5)$$

y de aquí se sigue que

$$\begin{aligned} L_\tau(\theta, \phi) &= \rho_t \left[\mathbf{P}_{\theta\phi}(d\mathbf{N}^c(t)) \middle| \mathfrak{F}_{t-}^c \right] L_0(\theta, \phi) \rho_t \left[\mathbf{P}_{\theta\phi}(d\mathbf{N}(t)) \middle| d\mathbf{N}^c(t), \mathfrak{F}_{t-}^c \right] \\ &= L_\tau^c(\theta) L_\tau''(\theta, \phi) \end{aligned} \quad (2.4.6)$$

donde la contribución de $L_0(\theta, \phi)$ a partir de \mathbf{X}_0 ha sido incluida en el segundo factor, $L_\tau''(\theta, \phi)$.

En (2.4.6), el primer factor, correspondiente al segundo factor de (2.4.3) es igual a

$$L_\tau^c(\theta) = \rho_t \left[\prod_{i,h} d\Lambda_{hi}^c(t, \theta)^{\Delta N_{hi}^c(t)} (1 - d\Lambda_{..}^c(t, \theta))^{1 - \Delta N_{..}^c(t)} \right] \quad (2.4.7)$$

donde $N_{..}^c = \sum_{i=1}^n \sum_{h=1}^k N_{hi}^c$ y $\Lambda_{..}^c(\cdot, \theta)$ es el $P_{\theta\phi, (\mathfrak{T}_t^c)}$ -compensador de $N_{..}^c$. De esta manera, la función de *verosimilitud parcial* tiene la misma forma que la verosimilitud parcial $L_t(\theta)$ dada en (2.4.4) basada en el proceso no censurado N . Entonces, *las censuras a derecha independientes preservan la forma de la verosimilitud parcial*; lo que trae como consecuencia que, por ejemplo, el proceso "score" $(\partial/\partial\theta)\ln L_t^c(\theta)$ es una $P_{\theta\phi, (\mathfrak{T}_t^c)}$ -martingala así como $(\partial/\partial\theta)\ln L_t^c(\theta)$ es una $P_{\theta\phi, (\mathfrak{T}_t)}$ -martingala en el modelo sin censura.

Una importante cuestión es ahora si $L_t^c(\theta)$ es la verosimilitud completa para θ para cada ϕ fijo basada en las observaciones de N , la respuesta a esto lo da la siguiente definición de *censura no informativa*.

Definición 1.1 Si, para cada ϕ fijo, la verosimilitud (2.4.6) puede ser factorizada como $L_t(\theta, \phi) = L_t^c(\theta)L_t^*(\phi)$, es decir, de manera que L no dependa de θ , entonces el mecanismo de censuras a derecha independientes C es llamado no informativo para θ .

Esta definición se puede interpretar como que C es no informativo para θ si para cada ϕ fijo, la distribución condicional de $dN(t)$, dada \mathfrak{T}_t^c y $dN^c(t)$, no depende de θ ; o en otras palabras, si para cada ϕ fijo y para cada t , la intensidad condicional de ciertos individuos censurados al tiempo t , no depende de θ , dado el pasado hasta antes de t y dada una posible falla al tiempo t .

Todo lo expuesto lleva a que la inferencia estadística para muestras grandes, de datos con censuras independientes o censuras no informativas, basada en la verosimilitud parcial será la misma que para datos no censurados basados en la verosimilitud completa ya que la estructura martingala juega un rol central en la teoría asintótica de procesos de conteo.

CAPITULO III

MODELOS DE "HAZARD" PROPORCIONAL

1. Introducción

Como ya se mencionó en el Capítulo I, los modelos de regresión proporcional brindan un método para explorar la asociación de covariables con la tasa de fallas y la distribución de sobrevida, y para estudiar el efecto de covariables primarias, tales como tratamiento, mientras se ajusta por otras covariables.

Estos modelos han sido marcadamente exitosos para analizar datos de tiempos de falla por una variedad de razones. Empíricamente, ellos parecen dar una razonable aproximación a la asociación entre las covariables y las variables respuestas tiempos de falla, que surgen a menudo en estudios médicos y de ingeniería. Con apropiadas restricciones, las covariables pueden ser determinísticas o funciones aleatorias dependientes del tiempo, dando al modelo mayor flexibilidad que los modelos lineales tradicionales para analizar datos de seguimiento. Los parámetros en el modelo tienen una interpretación simple y directa y, aunque las ecuaciones de estimación para los parámetros deben ser resueltas iterativamente, los métodos numéricos estándar funcionan bien.

En la Sección 1 se analiza en detalle el modelo de "hazard" proporcional, la *verosimilitud parcial* propuesta por Cox como herramienta para obtener los estimadores de los parámetros del modelo y su generalización directa a procesos de conteo. En la Sección 2 se describe aquellas propuestas de generalización del modelo de Cox para el caso multivariado, para las cuales se encuentra disponible el software que permite su implementación computacional. Por último, en la Sección 3 se discute la implementación computacional de cada uno de los cuatro modelos de "hazard" proporcional para datos de sobrevida multivariados, que se consideran en este trabajo.

2. Modelos de Regresión para Datos Censurados

Los datos disponibles en el problema de regresión con censuras a derecha son observaciones independientes de la terna (T, D, \mathbf{Z}) , donde $T = \min(X, U)$, siendo X es el tiempo de falla y U el tiempo de censura; $D = I_{\{X \leq U\}}$ es el indicador de que el evento que falló ha sido observado y, $\mathbf{Z} = (Z_1, \dots, Z_p)'$ es un vector columna p -dimensional de covariables. La importante cuestión de la inferencia es sobre la distribución condicional de falla, dadas las covariables.

La información contenida en el par (T, D) es equivalente a aquella contenida en los procesos N e Y , ya definidos en el Capítulo I. Esto conduce a dos enfoques de los modelos de regresión para datos censurados, uno basado en el empleo tradicional de las distribuciones condicionales y funciones "hazard" usado por Cox (1972); y, otro basado en el proceso estocástico N compensado por el proceso de intensidad acumulado Λ y el proceso martingala M asociado a ellos. En la mayoría de las aplicaciones los dos métodos son equivalentes.

Para presentar brevemente el enfoque tradicional y por simplicidad, se supone que ninguna de las covariables varía con el tiempo. Sea $S(t | \mathbf{Z}) = P\{T > t | \mathbf{Z}\}$ la función de sobrevida condicional. La función "hazard" condicional está definida por

$$h(t) = \lim_{\Delta t \downarrow 0} \frac{\Pr(t < T < t + \Delta t | T \geq t, \mathbf{Z})}{\Delta t}$$

Sea \mathbf{z}_1 y \mathbf{z}_2 dos valores distintos de la covariable, si se supone que la tasa de falla condicional asociada a ellos tiene una razón fija en el tiempo, entonces, $h(t | \mathbf{z}_1) = k(\mathbf{z}_1, \mathbf{z}_2) h(t | \mathbf{z}_2)$ donde k es una función de valores reales no negativa dependiente sólo de las covariables y no del tiempo. Si $h_0(t, \gamma)$ denota la tasa "hazard" condicional dada $\mathbf{Z} = \mathbf{0}$, llamada la tasa "hazard" basal, y $g(\mathbf{Z}) = k(\mathbf{Z}, \mathbf{0})$, entonces

$$h(t | \mathbf{Z}) = h_0(t, \gamma)g(\mathbf{Z})$$

Tomando $g(x) = e^x$, asegura que $g \geq 0$ y conduce a la forma más común del modelo de regresión de "hazard" proporcional,

$$h(t | \mathbf{Z}) = h_0(t, \gamma) \exp(\mathbf{b}'\mathbf{Z}) \quad (3.2.1)$$

Ya que la función de sobrevida $S(t | \mathbf{Z}) = \exp\left\{-\int_0^t h(u | \mathbf{Z}) du\right\}$, una forma alternativa del modelo de "hazard" proporcional es

$$S(t | \mathbf{Z}) = \{S_0(t)\}^{\exp(\mathbf{b}'\mathbf{Z})} \quad (3.2.2)$$

Las censuras juegan aquí el mismo rol que en el caso de muestras homogéneas y de esta manera, requieren la misma condición de censuras independientes.

Un modelo de "hazard" proporcional más general se basa en los procesos de conteo N , donde no se agrega ninguna complejidad conceptual si las covariables dependen del tiempo.

En el enfoque de procesos de conteo para datos de tiempo censurados, se modela la intensidad del proceso más que la tasa de "hazard" condicional para T . Y , se supone que el modelo está caracterizado por una estructura multiplicativa de los procesos de intensidad individuales, es decir,

$$\lambda_i^0(t) = Y_i(t) h_i^0(t; \mathbf{Z}_i(t)), \quad i = 1, \dots, n. \quad (3.2.3)$$

De esta manera, el modelo de regresión proporcional se expresa como

$$\lambda^0(t) = Y(t) h_0(t, \gamma) \exp(\mathbf{b}'\mathbf{Z}(t)). \quad (3.2.4)$$

En el caso de que \mathbf{Z} no varíe con el tiempo, la única información adicional acumulada a través del tiempo, está contenida en N e Y , y entonces el modelo de "hazard" proporcional (3.2.1) y el modelo de intensidad multiplicativa (3.2.4) coincide bajo censuras independientes. Si \mathbf{Z} varía con el tiempo, entonces, se hace necesario imponerle condiciones adicionales al modelo a fin de que $\lambda^0(t)$ represente la intensidad del proceso N . De esta manera, una suposición fundamental es que el

proceso $Z_i(t)$ es (\mathfrak{F}_t) -predictible, significando que los valores de las covariables al tiempo t deben ser conocidos justo antes de t ; y, se supone además que $Z_i(\cdot)$ es *localmente acotado*.

Una clara interpretación del modelo generalizado, en el contexto de los datos que se quieren analizar, llega a través de los resultados presentados en el Capítulo II, los cuales muestran cómo se encuentran relacionados el compensador Λ^0 del proceso de conteo, el proceso intensidad asociado, dado que existe, y la función "hazard". En efecto, cuando $\mathfrak{F}_t = \sigma\{N(s), Y(s+), Z(s+) : 0 \leq s \leq t\}$, tanto el modelo de "hazard" proporcional (3.2.1) como el modelo de intensidad multiplicativa (3.2.4) tiene ambos la interpretación

$$P\{N(t+dt) - N(t) = 1 \mid \mathfrak{F}_t\} = Y(t) h_0(t, \gamma) \exp(\mathbf{b}'Z(t)) dt, \quad (3.2.4)$$

bajo las condiciones ya mencionadas.

Muchos métodos paramétricos están disponibles para analizar la regresión de tiempos de falla con censuras a derecha. Estos modelos especifican tanto el efecto de las covariables como la forma de la tasa "hazard" basal dependiente del parámetro γ desconocidos.

Cuando sólo se quiere estimar los coeficientes de regresión, la función h_0 es esencialmente un molesto parámetro infinito dimensional; por esta razón, y debido a su directa generalización al caso multivariado, el estudio se concentra en *modelos semiparamétricos*, en los cuales la forma del "hazard" basal no es especificada.

Desde un enfoque tradicional, la dimensionalidad de h_0 dificulta la maximización conjunta de la verosimilitud con respecto a h_0 y \mathbf{b} ; entonces, la verosimilitud estándar no se puede usar para estimar \mathbf{b} cuando no se especifica la forma del "hazard" basal, es decir, cuando se usa un modelo no paramétrico para h_0 . Una manera de resolver este problema es recurrir a la descomposición de la verosimilitud total $f(\mathbf{t}, \beta)$ con $\beta = (h_0(t), \mathbf{b})$ como el producto de una verosimilitud marginal y una condicional, es decir

$$f(\mathbf{t}, \beta) = f_{w|v}(\mathbf{w}|\mathbf{v}, \beta) f_v(\mathbf{v}, \beta)$$

donde $\mathbf{t}' = (\mathbf{w}', \mathbf{v}')$. Si sólo uno de estos factores, verosimilitud condicional o marginal, depende del parámetro \mathbf{b} , entonces sus componentes se pueden estimar maximizando esta función. Basado en una idea similar, Cox deriva la expresión que denominó *función de verosimilitud parcial* con respecto a \mathbf{b} , y a partir de ella se estima el parámetro del modelo de regresión (3.2.1), que viene dado por

$$L(\mathbf{b}) = \prod_{i=1}^n \left[\frac{\exp\{\mathbf{b}' \mathbf{Z}_i\}}{\sum_{l \in \mathfrak{R}} \exp\{\mathbf{b}' \mathbf{Z}_l\}} \right]^{\delta_i} \quad (3.2.5)$$

donde $\mathfrak{R} = \{l : T_l \geq X_i\}$ es el conjunto de individuos bajo el riesgo de tener la falla l al tiempo X_i , y δ_i es una variable indicador que toma el valor cero si la observación es una censura. La ecuación (3.2.5) es una consecuencia directa del hecho que

$$P[\text{individuo } (i) \text{ falle en } X_i \mid \text{una falla en } X_i, \mathfrak{R}] = \frac{\exp\{\mathbf{b}' \mathbf{Z}_i\}}{\sum_{l \in \mathfrak{R}} \exp\{\mathbf{b}' \mathbf{Z}_l\}}$$

Para la construcción de una verosimilitud parcial general para \mathbf{b} , usando el enfoque de procesos de conteo, se recurre a otra herramienta estadística frecuentemente usada, que consiste en maximizar la verosimilitud condicional, dado un estadístico suficiente para h_0 . En efecto, bajo el modelo de intensidad multiplicativa, cuando en cada individuo se pueden observar k eventos, el modelo de intensidad proporcional se puede escribir como,

$$\lambda_{hi}^0(\mathbf{t}, \mathbf{Z}_i(\mathbf{t})) = Y_{hi}(\mathbf{t}) h_{h0}(\mathbf{t}, \gamma) \exp(\mathbf{b}' \mathbf{Z}_{hi}(\mathbf{t})), \quad i = 1, \dots, n; h = 1, \dots, k \quad (3.2.6)$$

donde las covariables de tipo específico $\mathbf{Z}_{hi}(\cdot)$, que son procesos predictibles localmente acotados, se definen como

$$\mathbf{Z}_{hi}(\mathbf{t}) = (Z_{hi1}(\mathbf{t}), \dots, Z_{hip}(\mathbf{t}))', \quad i = 1, \dots, n; h = 1, \dots, k$$

y son calculados a partir del vector de covariables básico $\mathbf{Z}_i(\mathbf{t})$ y de la historia \mathfrak{Z}_{i-} . Por otra parte, se supone que \mathbf{b} tiene dimensión p , mientras que γ es infinito

dimensional; en efecto, como se mencionó antes, las funciones "hazard" subyacentes son completamente inespecificadas, y funciones no negativas con

$$\int_0^t h_{h0}(u) du < \infty, \quad \forall t, \quad h = 1, \dots, k \quad (3.2.7)$$

Si se supone que tenemos censuras no informativas, de acuerdo a lo discutido en el Capítulo II Sección 4, la verosimilitud parcial (2.4.7) es proporcional al producto integral

$$\mathcal{P}_{t \in \tau} \left[\prod_{h,i} d\Lambda_{h0}(t) \exp(\mathbf{b}' \mathbf{Z}_{hi}(t))^{\Delta N_{hi}(t)} \left(1 - \sum_{h=1}^k d\Lambda_{h0}(t) S_h^{(0)}(\mathbf{b}, t) \right)^{1 - \Delta N_{..}(t)} \right], \quad (3.2.8)$$

donde

$$S_h^{(0)} = \sum_{i=1}^n \exp(\mathbf{b}' \mathbf{Z}_{hi}(t)) Y_{hi}(t), \quad h = 1, \dots, k; \quad (3.2.9)$$

$$\Lambda_{h0} = \int_0^t h_{h0}(u) du, \quad h = 1, \dots, k; \quad (3.2.10)$$

y $Y_{hi}(t)$ es el indicador de que el individuo i está bajo el riesgo del evento tipo h en el tiempo t . Para $\Lambda_{h0}(t)$ continua la verosimilitud multinomial (3.2.8) se puede escribir en la forma Poisson como

$$\mathcal{P}_{t \in \tau} \left[\prod_{h,i} d\Lambda_{h0}(t) \exp(\mathbf{b}' \mathbf{Z}_{hi}(t))^{\Delta N_{hi}(t)} \right] \exp \left[- \sum_{h=1}^k \int_0^{\tau} S_h^{(0)}(\mathbf{b}, u) d\Lambda_{h0}(u) \right]. \quad (3.2.11)$$

Para usar la metodología antes mencionada para determinar el estimador de \mathbf{b} , se debe encontrar previamente un estimador por máxima verosimilitud de $\Lambda_{h0}(t)$. Karr(1987) estudió la estimación por máxima verosimilitud de h_0 en un modelo de intensidad multiplicativa usando el método de tamizado, que consiste en maximizar la función de verosimilitud sobre h_0 en un subespacio restringido del espacio

paramétrico de dimensión infinita, de manera tal que en el espacio restringido la derivada logarítmica se encuentre acotada por una constante M . La solución \hat{h}_M , llamada un *tamiz*, genera una función de "hazard" acumulada $\hat{\Lambda}_M$. Si cuando M tiende a infinito, $\hat{\Lambda}_M$ converge (en el sentido de convergencia débil o en distribución), entonces $\hat{\Lambda}_M$ converge al estimador de Nelson-Aalen $\hat{\Lambda}$. Karr mostró que $\hat{\Lambda}$ es un estimador consistente.

La estimación de la función "hazard" basal acumulada $\Lambda_{h0}(t)$ dado \mathbf{b} puede ser encontrada a partir de (3.2.8) siguiendo la solución dada por Karr. Para valores fijos de \mathbf{b} la maximización de (3.2.8) y (3.2.11) conducen al estimador de Nelson-Aalen dado por

$$\hat{\Lambda}_{h0}(t, \mathbf{b}) = \int_0^t \frac{J_h(u)}{S_h^{(0)}(\mathbf{b}, u)} dN_h(u) \quad h = 1, \dots, k; \quad (3.2.12)$$

donde $J_h(u) = I\{Y_h(u) > 0\}$, $Y_h = Y_{h1} + \dots + Y_{hn}$, y $N_h = N_{h1} + \dots + N_{hn}$. Insertando (3.2.12) en (3.2.8) se obtiene la siguiente verosimilitud parcialmente maximizada dependientes solamente de \mathbf{b} :

$$L(\mathbf{b}) \mathcal{P}_t \left[\prod_h dN_h(t)^{\Delta N_h(t)} (1 - dN_{\bullet\bullet}(t))^{1 - \Delta N_{\bullet\bullet}(t)} \right] \quad (3.2.12)$$

donde

$$L(\mathbf{b}) = \mathcal{P}_{t \in \tau} \left[\prod_{h,i} \left(\frac{\exp(\mathbf{b}' \mathbf{Z}_{hi}(t))}{S_h^{(0)}(\mathbf{b}, t)} \right)^{\Delta N_{hi}(t)} \right], \quad (3.2.13)$$

que, sin mucha dificultad, se puede observar que para el caso en que \mathbf{Z} no depende de t y $k=1$, es la verosimilitud parcial derivada por Cox (1972, 1975) para el caso de datos censurados. La estimación de \mathbf{b} se puede basar en (3.2.13) o en su logaritmo, es decir,

$$C_{\tau}(\mathbf{b}) = \log L(\mathbf{b}) \\ = \sum_{h=1}^k \left[\sum_{i=1}^n \int_0^{\tau} \mathbf{b}' \mathbf{Z}_{hi}(t) dN_{hi}(t) - \int_0^{\tau} \ln S_h^{(0)}(\mathbf{b}, t) dN_h(t) \right]. \quad (3.2.14)$$

Luego, el valor $\hat{\mathbf{b}}$ que maximiza $C_{\tau}(\mathbf{b})$ será el estimador del vector de parámetros de

la regresión. El vector de "score" estadístico $\mathbf{U}_{\tau}(\mathbf{b}) = \left(\frac{\partial}{\partial b_j} C_{\tau}(\mathbf{b}) \right)$, $j = 1, \dots, p$

puede escribirse, conforme a (3.2.14), como

$$\mathbf{U}_{\tau}(\mathbf{b}) = \sum_{h=1}^k \left[\sum_{i=1}^n \int_0^{\tau} \mathbf{Z}_{hi}(t) dN_{hi}(t) - \int_0^{\tau} \mathbf{E}_h(\mathbf{b}, t) dN_h(t) \right], \quad (3.2.15)$$

donde

$$\mathbf{E}_h(\mathbf{b}, t) = (E_{hj}(\mathbf{b}, t), j = 1, \dots, p) = \frac{\sum_{i=1}^n \mathbf{Z}_{hi}(t) \exp(\mathbf{b}' \mathbf{Z}_{hi}(t)) Y_{hi}(t)}{S_h^{(0)}(\mathbf{b}, t)} \quad h=1, \dots, k. \quad (3.2.16)$$

La matriz de derivadas parciales de segundo orden de $C_{\tau}(\mathbf{b})$, denotada con $-\mathbf{I}_{\tau}(\mathbf{b})$, se expresa a través de

$$\mathbf{I}_{\tau}(\mathbf{b}) = \sum_{h=1}^k \int_0^{\tau} \mathbf{V}_h(\mathbf{b}, t) dN_h(t), \quad (3.2.17)$$

donde

$$\mathbf{V}_h(\mathbf{b}, t) = (V_{hj}(\mathbf{b}, t), j = 1, \dots, p) \\ = \frac{\sum_{i=1}^n \mathbf{Z}_{hi}(t) \mathbf{Z}'_{hi}(t) \exp(\mathbf{b}' \mathbf{Z}_{hi}(t)) Y_{hi}(t)}{S_h^{(0)}(\mathbf{b}, t)} - \mathbf{E}_h(\mathbf{b}, t) \mathbf{E}'_h(\mathbf{b}, t) \quad (3.2.18)$$

Bajo algunas condiciones de regularidad, se verifica que para n grande, $\sqrt{n}(\hat{\mathbf{b}} - \mathbf{b}_0)$ tiene aproximadamente una distribución multinormal con media cero y matriz de

covarianza que puede estimarse consistentemente por $(n^{-1}I_{\tau}(\hat{\mathbf{b}}))^{-1}$, donde \mathbf{b}_0 es el verdadero valor del parámetro. A partir de este resultado, se sigue inmediatamente que la distribución asintótica del *estadístico del test de Wald*, dado por $(\hat{\mathbf{b}} - \mathbf{b}_0)' I_{\tau}(\hat{\mathbf{b}})(\hat{\mathbf{b}} - \mathbf{b}_0)$ para la hipótesis simple de que $\mathbf{b} = \mathbf{b}_0$, es χ^2 con p grados de libertad. La teoría de verosimilitud estándar sugiere también que, tanto el test estadístico *del test de razón de verosimilitud (parcial)*

$$-2 \ln Q = 2(C_{\tau}(\hat{\mathbf{b}}) - C_{\tau}(\mathbf{b}))$$

como el estadístico del *test de "score"*

$$U_{\tau}(\mathbf{b}_0)' I_{\tau}(\mathbf{b}_0)^{-1} U_{\tau}(\mathbf{b}_0)$$

tienen la misma distribución asintótica que el estadístico del test de Wald, bajo las mismas condiciones de regularidad.

3. Regresión para Datos Multivariados con Fallas Correlacionadas

El modelo de "hazard" proporcional de Cox ha sido ampliamente usado para analizar datos de tiempos de falla univariado, a causa de su natural flexibilidad y de que las propiedades de sus procedimientos de inferencia para muestras grandes, han sido justificadas usando teoría de martingalas.

La generalización del modelo de Cox al modelo de intensidad proporcional (3.2.6), provee de una herramienta para el estudio de problemas más complejos, en los cuales se puede observar un número importante de eventos involucrados, es decir, los sujetos pueden experimentar varias *fallas*, las cuales corresponden a ocurrencias repetidas del mismo tipo de evento (*Eventos Recurrentes u Ordenados*) o a la ocurrencia de eventos de naturaleza enteramente diferente (*Eventos Diferentes o No Ordenados*). Sin embargo, implícitamente, este modelo *supone que los eventos son independientes*, lo cual resulta muy poco realista en muchas situaciones, principalmente en el caso de eventos recurrentes.

Por otra parte, las escalas que se usan para medir los tiempos de falla deben responder a la naturaleza del problema, es decir, cuando se está frente a una situación de Eventos No Ordenados (ver Fig. 4.1, Capítulo IV), es claro que los individuos están bajo el riesgo de cualquiera de las fallas desde el comienzo de la observación, las escalas de tiempo que interesan, entonces, son las denominadas escalas de *tiempos totales* (T_{ih} , $i=1, \dots, n$; $h=1, \dots, k$), los cuales se definen como el tiempo desde el comienzo de la observación hasta la h -ésima falla para el i -ésimo individuo. Sin embargo, en la situación de Eventos Recurrentes u Ordenados, el individuo que no ha experimentado la h -ésima falla, definitivamente no está bajo el riesgo de la $(h+1)$ -ésima recurrencia, y así, el individuo que no ha experimentado la h -ésima falla tiene que ser excluido del análisis para la $(h+1)$ -ésima recurrencia, lo cual viola la suposición de "missing" completamente al azar, por esta razón, las escalas de tiempo que interesan aquí son las denominadas escalas de *tiempos intereventos* (ver Fig. 4.1, Capítulo IV).

Varias propuestas para modelar datos con fallas correlacionadas, usando una extensión del modelo de regresión proporcional de Cox, se pueden encontrar en la literatura, y la diferencia entre las distintas propuestas radica principalmente en los siguientes aspectos: el manejo de la correlación dentro de los sujetos y la secuencia de ocurrencia de los eventos.

Sobre la base de estas consideraciones, las propuestas se clasifican en los llamados: *modelos marginales* y *modelos condicionales*. Los primeros proponen modelar la distribución marginal de cada tiempo de falla con un modelo de "hazard" proporcional, donde no se impone ninguna estructura particular de dependencia entre los distintos tiempos de falla en cada individuo, mientras que si se tiene en cuenta la correlación en la etapa de estimación de la matriz de covarianza de los estimadores. Los segundos, expresan el modelo de "hazard" proporcional en función del "hazard" condicional, de esta manera, ellos introducen la estructura de correlación entre los tiempos de falla en el modelo.

Therneau [1995] desarrolló un programar para la aplicación de cuatro de las propuestas de modelos de regresión para datos de sobrevida multivariados que aparecen en la literatura:

- *modelo de Andersen y Gill (AG),*
- *modelo de Oakes (OK),*
- *modelo de Wei, Lin y Weissfeld (WLW),*
- *modelo de Prentice y Cai (PC).*

Los dos primeros corresponden a los clasificados como modelos condicionales y los dos últimos a los clasificados como modelos marginales.

3.1. Modelo de Andersen y Gill (AG)

La formulación de Andersen y Gill (1982) para el modelo de "hazard" proporcional como un proceso de conteo multivariado, ha sido de gran utilidad en el desarrollo teórico de la solución del problema de la extensión del modelo de regresión proporcional de Cox. Como ya se explicitó en el Capítulo II, el enfoque de procesos de conteo conduce a que cada sujeto i sea representado como un proceso de conteo $N_{hi}(t)$ que indica el número de eventos del tipo h que el individuo i experimentó hasta t .

En esta formulación, la suposición básica es que el proceso de conteo multivariado $\mathbf{N}(t) = (N_{hi}; h=1, \dots, k; i=1, \dots, n, t \in [0, \tau])$ tiene un proceso de intensidad aleatorio $\lambda(t) = (\lambda_{hi}(t); h=1, \dots, k; i=1, \dots, n, t \in [0, \tau])$ tal que λ_{hi} cumple con el modelo de intensidad proporcional dado en (3.2.6), es decir,

$$\lambda_{hi}(t, \mathbf{Z}_i(t)) = Y_{hi}(t) h_{h0}(t) \exp(\mathbf{b}' \mathbf{Z}_{hi}(t)), \quad i = 1, \dots, n; h = 1, \dots, k$$

donde \mathbf{b} es un vector de p parámetros, λ_{h0} es la función "hazard" subyacente e, Y_{hi} es un proceso predecible que toma valores $\{0,1\}$ indicando con el valor 1 que el individuo i está bajo el riesgo de tener el evento h . Finalmente, $\mathbf{Z}_{hi} = (Z_{hi1}, \dots, Z_{hip})$ es un vector columna de p procesos covariantes para el i -ésimo individuo y el h -ésimo evento. Se supone además, que los $Z_{hi}(\cdot)$ son localmente acotados y $\mathfrak{I}_t(h)$ -

predictibles, significando que el valor de las covariables al tiempo t debería conocerse antes de t , que es el caso, por ejemplo, cuando Z_{hi} es continua a izquierda y con límite por la derecha.

La suposición que N_{hi} tiene proceso de intensidad λ_{hi} para todo $i = 1, \dots, n$ y para todo $h = 1, \dots, k$, asegura que el proceso M_{hi} definido por

$$M_{hi}(t) = N_{hi}(t) - \int_0^t \lambda_{hi}(u) du \quad i=1, \dots, n; h=1, \dots, k; t \in [0, \tau] \quad (3.3.1.1)$$

es una martingala local en $[0, \tau]$, y como consecuencia es cuadrado integrable con proceso de variación predecible $\langle M_{hi} \rangle$ (ver Definición 20 del Apéndice 1), dado por

$$\langle M_{hi}, M_{hi} \rangle(t) = \int_0^t \lambda_{hi}(u) du \quad \text{y} \quad \langle M_{hi}, M_{hj} \rangle = 0, \quad i \neq j. \quad (3.3.1.2)$$

Esto permite la aplicación del teorema central del límite para las martingalas debida a Rebolledo (1980) y con ello demostrar la normalidad asintótica de los estimadores de \mathbf{b} .

El segundo término de (3.3.1.1.), que no es otra cosa que el proceso de intensidad acumulado, $\Lambda_{hi}(t)$, resulta ser el compensador del proceso de conteo N_{hi} .

La estimación de \mathbf{b} se puede basar en (3.2.13) o en su logaritmo, (3.2.14)

Cuando no hay covariables, el modelo AG se reduce a un simple proceso Poisson no estacionario en el cual la función intensidad λ_{hi} para el h -ésimo evento en el i -ésimo sujeto, condicionado sobre los eventos previos que sucedieron en el mismo sujeto antes de t , es

$$\lambda_{hi}(t) = \lim_{\delta \rightarrow 0} \frac{1}{\delta} P \left[N_{hi}(t+\delta) - N_{hi}(t) = 1 \mid \mathfrak{F}_t \right] = \lambda_{hi}(t+). \quad (3.3.1.3)$$

De esta manera, el número total de eventos del tipo h para el individuo i sobre cualquier intervalo $(0, t)$ sigue entonces una distribución Poisson con media $\Lambda_{hi}(t)$.

3.2. Modelo de Oakes (OK)

La propuesta de Oakes (1992) para el análisis de sobrevida multivariado usa el concepto de "frailty" introducido por Vaupel, Manton y Stallard [1979]. El "frailty" (W) es un factor aleatorio inobservable, el cual induce dependencia entre los eventos múltiples afectando proporcionalmente la función de intensidad para cada individuo

Los efectos aleatorios W sobre los diferentes individuos se suponen independientes e idénticamente distribuidas de acuerdo a una distribución F con transformada de Laplace

$$L(s) = E(e^{-sW}) = \int e^{-sW} dF(w).$$

Si la función de intensidad o "hazard" basal para un sujeto cuyo $W=1$ esta dada por

$$\lambda^0(t) = \lim_{\delta \rightarrow 0} \frac{1}{\delta} P[T \leq t + \delta | T > t, W = 1], \quad (3.3.2.1)$$

entonces, la misma función para un sujeto cuyo efecto aleatorio es $W = w$ será

$$\lambda(t|w) = w\lambda^0(t). \quad (3.3.2.2)$$

Introduciendo dentro del modelo AG efectos aleatorios específicos para cada sujeto, la función intensidad, condicionada sobre los valores w_i , viene dada por

$$\lambda_{hi}(t | w_i) = w_i \lambda_{hi}(t) \quad (3.3.2.3)$$

De esta manera, los llamados modelo "frailty", como los descriptos por Oakes, son modelos como los AG pero con efectos aleatorios. Condicionada sobre estos efectos aleatorios, para cada individuo, la función de intensidad de riesgo para los eventos sigue un modelo AG.

A partir de (3.3.2.3) se puede mostrar que, todavía condicionando sobre los w_i , el número total de eventos del individuo i en el intervalo $(0, t)$ sigue una distribución Poisson con media $D_i(t) = w_i \Lambda_i(t)$.

Integrando sobre la distribución de los efectos aleatorios y de acuerdo a Karr (1991, pag. 274), dado el número de eventos $D_i(t) = m$ en $(0, t)$, los tiempos de ocurrencia T_1, \dots, T_m son independiente de W . Y de esta manera, $D_i(t)$ lleva toda la información del pasado del proceso sobre su futuro. Esto implica que la función de intensidad del modelo "frailty" observable en t , todavía condicionada sobre los tiempos t_1, \dots, t_m de los m eventos previos en $(0, t)$, pero integrada sobre la distribución de los "frailty" inobservables W , que se expresa como

$$\lambda_m(t | T_1, \dots, T_m) = E(W \lambda_m(t) | T_1, \dots, T_m)$$

es función sólo de t , lo que concuerda con lo formulado por AG.

Oakes muestra que el problema de identificabilidad en el modelo, producida por la introducción de efectos de heterogeneidad vía los "frailty" cuando se observa un solo evento en cada sujeto, notado por Aalen [1988], desaparece en el caso bivariado cuando los datos están disponibles sobre pares de individuos con el mismo valor de W y cuando los tiempos de sobrevida T_1 y T_2 son condicionalmente independientes dado el valor de W .

Otro problema que surge en este modelo es respecto a cómo se especifica la acción de los efectos fijos (Z_{ih}), condicional sobre los W o incondicionalmente, en términos de la distribución de los tiempos de sobrevida observables. En los modelos lineales estándares, como por ejemplo el ANOVA mixto, ambas especificaciones son equivalentes, pero esto no ocurre en este caso, ya que si la acción condicional sigue un modelo de "hazard" proporcional, la acción incondicional en general no responde a él. Hougaard [1986] mostró que este problema puede solucionarse si y solo si los "frailty" siguen una distribución positiva estable con una transformada de Laplace $L(s) = \exp(-\alpha s)$ con $0 < \alpha \leq 1$. Entonces, si β es la acción incondicional de una covariable entre sujetos que tiene el mismo valor de W , la acción incondicional de la misma covariable entre sujetos

que tienen valores distintos de W será $\alpha\beta$. De esta manera, el efecto de los términos aleatorios es hacer que los coeficientes de regresión disminuyan pero el modelo se preserva.

Oakes propone usar W con distribución $\Gamma(1, \kappa)$, donde las covariables actúan sobre las intensidades marginales a través de un modelo de "hazard" proporcional ya que, como él mismo demuestra, un modelo "frailty" gamma asegura que las $\lambda_{hi}(t)$ ($h=1, \dots, k$) son proporcionales y el factor de proporcionalidad es independiente de h . De esta manera, el modelo modificado por los efectos aleatorios tiene validez y los programas de computación estándares, que trabajan con el producto de la verosimilitud parcial sobre estratos dependientes del tiempo, pueden usarse para estimar los efectos de las covariables.

Sin embargo, Oakes propone formar una sola verosimilitud parcial que permita la estimación conjunta del parámetro κ y de los efectos de las covariables, a partir de la siguiente expresión para la función de intensidad

$$\lambda_{hi}(t) = (1+h\kappa)Y_{ih}(t)\lambda_{h0}(t)\exp(\mathbf{b}'\mathbf{Z}_{ih}(t)) \quad (3.3.2.4)$$

donde $\kappa = \text{var}(W)$.

La verosimilitud parcial, propuesta por Oakes, para datos de tiempos de falla censurados, en los parámetros del modelo de regresión \mathbf{b} y el parámetro κ , se expresa como

$$L(\mathbf{b}, \kappa) = \prod_{i=1}^n \prod_{j=1}^{d(i)} \left[\frac{(1+(j-1)\kappa)\exp(\mathbf{b}'\mathbf{Z}_i)}{\sum_{l \in \mathfrak{R}(i,l)} (1+j'(l)\kappa)\exp(\mathbf{b}'\mathbf{Z}_l)} \right] \quad (3.3.2.5)$$

donde $d(i)$ es el número observado de fallas experimentadas por el i -ésimo individuo, y $\mathfrak{R}(i,l) = \{l: \exists j'(l); T_{ij'(l)} < T_{ij} < T_{ij'(l)+1}\}$, en donde $T_{ij'(l)+1}$ son los tiempos de censuras de los sujetos.

3.3. Modelo de Wei, Lin y Weissfeld (WLW)

Wie, Lin y Weissfeld (1989), proponen modelar la distribución marginal de cada una de las variables tiempo de falla T_j ($j = 1, \dots, k$) con un modelo de regresión proporcional de Cox, sin imponer ninguna estructura particular de dependencia entre los tiempos de falla sobre un mismo individuo. De esta manera, para la h -ésima falla del i -ésimo individuo, la función "hazard" se supone que adopta la forma

$$h_{ji}(t) = h_{j0}(t) \exp\{\mathbf{b}_j' \mathbf{Z}_{ji}(t)\}, \quad t \geq 0 \quad (3.3.3.1)$$

donde $\mathbf{b}_j = (b_{j1}, \dots, b_{jp})'$ es el vector de parámetros de regresión específico para cada falla.

Los parámetros en el modelo (3.3.3.1) son estimados maximizando la verosimilitud parcial específica de la falla dada por

$$L_j(\mathbf{b}) = \prod_{i=1}^n \left[\frac{\exp(\mathbf{b}_j' \mathbf{Z}_{ji}(T_{ji}))}{\sum_{l \in \mathfrak{R}_j(T_{ji})} \exp(\mathbf{b}_j' \mathbf{Z}_{jl}(T_{jl}))} \right]^{\delta_{ji}} \quad (3.3.3.2)$$

donde $\mathfrak{R}_j(t) = \{l : T_{jl} \geq t\}$ es el conjunto de individuos bajo el riesgo de presentar la falla j justo antes de t , y δ_{ji} es una variable indicador que toma el valor cero si la observación es una censura.

Para simplificar la notación, se supone que $p = 1$, es decir, se tiene una covariable para cada falla. Bajo la suposición de independencia de los tiempos de falla, las ecuaciones basadas en los "score" de la verosimilitud parcial para el vector de parámetros $\mathbf{b} = (b_1, \dots, b_k)'$, de acuerdo a lo discutido en la Sección 1 de este capítulo, puede escribirse como,

$$\sum_{j=1}^k \sum_{i=1}^n \int_0^{T_{ji}} Z_{ji}(u) \hat{M}_{ji}(du) = 0 \quad (3.3.3.3)$$

donde

$$\hat{M}_{ji}(t) = N_{ji}(t) - \int_0^{T_{ji}} \exp(\mathbf{b}_j Z_{ji}(u)) \hat{\lambda}_{0j}(u) du \quad (3.3.3.4)$$

siendo

$$\hat{\lambda}_{0j}(u) = \left[\sum_{l \in \mathfrak{R}_j(u)} \exp(\mathbf{b}_j Z_{lj}(u)) \right]^{-1}$$

el estimador propuesto por Breslow (1974) para el "hazard" basal correspondiente a la falla j y con \mathfrak{R}_j definido como antes.

Los autores muestran que, a pesar del supuesto de independencia para los T_{ji} , para n grande, el vector de los estimadores $(\hat{\mathbf{b}}_1, \dots, \hat{\mathbf{b}}_k)$ es consistente y aproximadamente normal con media $(\mathbf{b}_1, \dots, \mathbf{b}_k)$ y matriz de covarianza \mathbf{Q} . Por otra parte, proponen un estimador consistente para \mathbf{Q} al cual intentan adaptarlo a la estructura de dependencia de los T_{ji} en un dado individuo, y con esto proveen la base para la inferencia simultánea sobre los \mathbf{b}_j .

Mediante un estudio de Monte Carlo y en la situación de eventos recurrentes, los autores verifican la precisión de la aproximación normal y afirman que este método provee una interesante herramienta de análisis. Por otra parte, también muestran que los modelos que suponen estructuras de dependencia entre los eventos son más sensibles al apartamiento del supuesto de proporcionalidad de los "hazard".

3.4. Modelo de Prentice y Cai (PC)

Prentice y Cai (1992) consideran, en primer lugar, diferentes modelos especializados para el proceso de intensidad basal, correspondiente al i -ésimo individuo en la k -ésima falla en ausencia de covariables, a saber:

- $\lambda_0(t)$ especifica un proceso Poisson no homogéneo
- $\lambda_{0q}(t)$ especifica un proceso Markov, donde q es una función del tipo de evento y del tiempo, es decir, $q = q(j, t)$.

- $\lambda_0(t - t_{q-1})$ especifica un proceso de renovación ordinaria.
- $\lambda_{0q}(t - t_{q-1})$ especifica un proceso semi Markov.

Luego, los modelos de regresión correspondientes a un proceso Markov y a un proceso semi Markov pueden escribirse respectivamente como

$$\lambda_{qi}(t | \mathbf{Z}_{ji}) = \lambda_{q0}(t) \exp\{\mathbf{b}_j' \mathbf{Z}_{ji}(t)\}, \quad t \geq 0 \quad (3.3.4.1)$$

$$\lambda_{qi}(t | \mathbf{Z}_{ji}) = \lambda_{q0}(t - t_{q-1}) \exp\{\mathbf{b}_j' \mathbf{Z}_{ji}(t)\}, \quad t \geq 0 \quad (3.3.4.2)$$

El primero, enfatiza el tiempo total de seguimiento, mientras que el segundo incorpora el tiempo inter-evento.

Prentice y Cai proponen modelar la distribución marginal de los tiempos de falla T_{ji} ($j=1, \dots, k; i=1, \dots, n$), a través de los modelos dados en (3.3.4.1) o (3.3.4.2) según corresponda.

A fin de obtener un estimador más eficiente que los propuestos por Wei, Lin y Weissfeld, bajo suposición de independencia entre los tiempos de falla, Prentice y Cai modifican las ecuaciones de "score" de la verosimilitud parcial dada en (3.3.3.3), ponderando la contribución de un dado individuo por un factor relacionado con la dependencia entre sus tiempos de falla, y entonces, en lugar de estimar los parámetros del modelo a través de la expresión (3.3.3.3), ellos proponen usar la siguiente ecuación de estimación

$$\sum_{j=1}^k \int_0^{T^*} \mathbf{Z}_j(u) \hat{\Omega}_j^{-1} \hat{M}_j'(du) = 0 \quad (3.3.4.3)$$

donde T^* es el máximo tiempo de seguimiento en la muestra, $\mathbf{Z}_j(u) = (Z_{j1}, \dots, Z_{jn})$, $\hat{M}_j'(u) = (\hat{M}_{j1}'(u), \dots, \hat{M}_{jn}'(u))$ y $\hat{\Omega}_j$ es un estimador de la matriz de correlación de los M_j . En efecto, bajo la suposición de que $k=2$, es decir que, en cada individuo sólo se puede observar dos tipos de fallas, los autores muestran que si

$$C(t_1, t_2) = E[M_1(t_1), M_2(t_2)] \quad (3.3.4.4)$$

representa la función de covarianza entre las martingalas, entonces se verifica que

$$S(t_1, t_2) = S_1(t_1)S_2(t_2) \left[1 + \int_0^{t_1} \int_0^{t_2} \{S_1(u_1)S_2(u_2)\}^{-1} C(du_1, du_2) \right] \quad (3.3.4.5)$$

donde $S(t_1, t_2)$ es la función de sobrevida bivariada, y entonces, a través de la expresión (3.3.4.5) se puede ver que $C(t_1, t_2)$ caracteriza completamente la naturaleza de la dependencia entre T_1 y T_2 dadas las marginales.

Por otra parte, se verifica que

$$\begin{aligned} C(t_1, t_2) &= \Lambda_1(t_1)\Lambda_2(t_2)S(t_1, t_2) + \Lambda_1(t_1) \int_0^{t_2} [1 - \Lambda_2(u_2)] S(t_1, du_2) \\ &\quad + \Lambda_2(t_2) \int_0^{t_1} [1 - \Lambda_1(u_1)] S(ds_1, t_2) \\ &\quad + \int_0^{t_1} \int_0^{t_2} [1 - \Lambda_1(u_1)][1 - \Lambda_2(u_2)] S(du_1, du_2) \end{aligned} \quad (3.3.4.6)$$

Para estimar $S(t_1, t_2)$ y $C(t_1, t_2)$, Prentice y Cai proponen usar estimadores de Kaplan-Meier para $S_j(t_j)$, $j = 1, 2$, y el estimador de Nelson-Aalen para la $\Lambda_j(t_j)$ y muestran que, de esta manera, se obtienen estimadores consistentes y aproximadamente insesgados bajo un modelo de censuras aleatorias, cuando la probabilidad de que una observación sea censurada no supera el valor 0.25.

4. Implementación Computacional

Therneau (1995) publica las bases teóricas y ventajas del software que desarrolló para la implementación computacional, en S-Plus y más recientemente en SAS, de los modelos de regresión proporcional de Cox con un enfoque de procesos de conteo. Entre las ventajas computacionales de este procedimiento se enumeran las siguientes:

- La facilidad para compilar la rutina.
- El proceso de conteo puede manejarse de manera más eficiente y por lo tanto es más fácil de usar.

- Permite una fácil extensión del modelo de Cox a datos de sobrevida multivariados, tanto con modelos condicionales como marginales.
- Proporciona gran flexibilidad en la formación de estratos y conjuntos de riesgo y en la manipulación de las escalas de tiempo.
- Provee de un estimador de la varianza bien desarrollado.

Sin embargo, la implementación computacional, de cada tipo de modelo de regresión para datos de sobrevida multivariados, requiere la toma de decisiones respecto a los siguientes aspectos:

- La inclusión o no de un término de interacción entre las covariables y la variable que Therneau denomina "Strata", que, en este caso corresponde a las diferentes fallas del individuo.
- La elección de las escalas de tiempo (tiempos totales o tiempos inter-
eventos).
- La organización de los datos.

Respecto a este último punto, cualquiera sea el modelo de regresión multivariado elegido, la matriz de datos se organiza de manera tal que cada individuo es representado como un conjunto de filas ("Cluster"), cada una de las cuales representa una falla de ese individuo.

Para los modelos condicionales (AG o OK), y bajo el supuesto de que el proceso subyacente es de incrementos independientes, es decir, un proceso sin memoria, las filas correspondientes a cada individuo (fallas) contienen datos sobre el período de tiempo que el individuo se encuentra bajo el riesgo de la falla correspondiente, es decir, la información sobre el tiempo de falla o censura. Estos datos se organizan en dos columnas que registran el comienzo ("start") y el final ("stop") del período de exposición al riesgo de la falla asociada a la fila considerada. Por ejemplo, los intervalos de tiempo correspondientes a las $(j + 1)$ fallas de un individuo serían:

fila	“start”	“stop”
1ra. fila	(comienzo de la observación,	primer evento]
2da. fila	(primer evento,	segundo evento]
.....
última fila	(j-ésimo evento,	final del seguimiento]

De esta manera, un sujeto con 0 eventos estaría representado por una sola fila, otro sujeto con un evento podría estar representado por una o dos filas dependiendo de la realización de seguimiento adicional o no después del primer evento. Si se piensa que el proceso subyacente es de renovación, la manera adecuada de organizar los datos sería en intervalos como: $(0, t_1]$, $(0, t_2 - t_1]$,

Para los modelos marginales, cada evento o tipo de evento es modelado como un estrato separado, y dentro de cada estrato se usan los datos marginales, es decir, aquellos que resultan de ignorar toda información excepto la producida para ese tipo de evento. Para lograrlo, cada individuo debe aparecer en todos los estratos y todos los intervalos de tiempo deben comenzar en cero, de esta manera, se necesita sólo una columna de tiempos de falla, y entonces, se debe introducir en el modelo la información adicional sobre cuál evento o tipo de evento está bajo consideración (“Strata”). Computacionalmente, esto equivale a considerar los términos de interacción entre la variable “strata” y las covariables bajo estudio. El vector de coeficientes combinado, es la concatenación de los coeficientes que se obtienen a partir de los ajustes para cada estrato y esto es equivalente a tener coeficientes individuales ajustando cada tipo de falla (estrato) como un conjunto de datos separado. Therneau afirma que usar un ajuste global más que un ajuste por separado para cada evento tiene algunas ventajas prácticas, a saber:

- Son más fáciles de compilar, sobre todo cuando el número de eventos por sujeto es grande.
- Permite considerar otros modelos, y en particular, no se necesita incluir todos los términos de interacción entre las covariables y los estratos.

La diferencia entre los modelos condicionales (AG y OK) radica en la asignación o no de cada evento o tipo de evento a diferentes estratos (Tabla 3.4.1), puesto que, Therneau introduce computacionalmente el efecto aleatorio en el modelo AG por consideración de términos de interacción entre las covariables y el tipo de falla ("strata").

Therneau propone el uso del modelo (3.3.4.2) para PC, de esta manera, la diferencia entre los modelos marginales (WLW y PC), se establece a través de la indicación de usar el estimador corregido y la consideración del tiempo de falla, de la siguiente manera:

Modelos	Tiempo ("Stop") de la j-ésima falla
WLW	T_j
PC	$T_j - T_{j-1}$

En la Tabla 3.4.1 se presenta un ejemplo de cómo deben ordenarse los datos para cada uno de los cuatro modelos, en el caso en que $k = 2$, es decir cuando los individuos pueden tener a lo sumo dos eventos o tipos de eventos.

Tabla 3.4.1: Ordenamiento de los tiempos de falla para la aplicación de los diferentes modelos de regresión.

Modelos	Intervalo	Estrato
Andersen - Gill	$(0, t_1]$	1
(AG)	$(t_1, t_2]$	1
Oakes	$(0, t_1]$	1
(OK)	$(t_1, t_2]$	2
Wei-Lin-Weissfeld	$(0, t_1]$	1
(WLW)	$(0, t_2]$	2
Prentice - Cai	$(0, t_1]$	1
(PC)	$(0, t_1 - t_2]$	2

Si se designa con T_1 al extremo izquierdo del intervalo ("Start"), con T_2 el extremo derecho ("Stop"), como ID a la columna que indica el individuo ("Cluster"), como Z a la única covariable bajo estudio, con CEN la columna que registra si es una falla o una censura ("Status") y con E la columna que registra el tipo de evento ("Strata"), entonces las sentencias que requiere el S-Plus para la aplicación de modelos condicionales vienen dadas por:

- **AG** <- coxph(Surv(T_1 , T_2 , CEN) ~ Z + cluster(ID), data = nombre archivo de datos)
- **OK** <- coxph(Surv(T_1 , T_2 , CEN) ~ Z * strata(E) + cluster(ID), data = nombre archivo de datos).

Las sentencias correspondientes a los modelos marginales son:

- **WLW** <- coxph(Surv(X_1 , CEN) ~ Z * strata(E) + cluster(ID), data = nombre archivo de datos)
- **PC** <- coxph(Surv(X_2 , CEN) ~ Z * strata(E) + cluster(ID), data = nombre archivo de datos)

donde X_1 es el tiempo de falla total de los eventos y X_2 es el tiempo de falla interevento.

La comparación de la precisión de los estimadores de cada modelo, en la situación de eventos recurrentes y de eventos diferentes, mediante simulaciones, se lleva a cabo usando el paquete de programa S-Plus, con las recomendaciones proporcionadas por Therneau para la aplicación de cada modelo.

CAPITULO IV

CORRELACIÓN ENTRE LOS TIEMPOS DE FALLA

1. Introducción

Como se puntualizó en el Capítulo III, uno de los problemas más importantes que surgen en la extensión del modelo de Cox a datos de sobrevida multivariados, es el manejo de la correlación entre los tiempos de falla dentro de los sujetos. Por este motivo, la diferencia entre las distintas propuestas de extensión del modelo de "hazard" proporcional al caso multivariado, radica principalmente en cómo se tiene en cuenta dicha correlación.

En la mayoría de los problemas prácticos es muy difícil suponer que los tiempos de falla dentro de un mismo individuo sean independientes, en especial si se trata de recurrencias del mismo evento. Si bien todas las propuestas intentan el manejo de esta correlación, dentro del mismo modelo o a través de la estimación de la matriz de covarianza de los estimadores de los parámetros, como lo puntualiza Oakes (1998); no hay un estudio sobre la eficiencia de los diferentes modelos cuando aumenta el grado de correlación entre los tiempos de falla.

En este capítulo se estudia cómo afecta el grado de correlación entre los tiempos de falla, a la precisión de los estimadores de los parámetros de los diferentes modelos de regresión tipo Cox.

En la Sección 2, se detallan las condiciones bajo las cuales se lleva a cabo el estudio de Monte Carlo y se presentan los valores estimados del *Sesgo Relativo*, *Error Estándar* y *Error Cuadrático Medio* de los estimadores del parámetro de los diferentes modelos de "hazard" proporcional considerados. En la Sección 3 se hace un análisis de los resultados encontrados en la Sección 2, usando la analogía con los estudios de confiabilidad de sistemas en serie y paralelo en el área de ingeniería. En la Sección 4 se encuentran cotas para el sesgo, y se analiza el grado de dependencia que ellas presentan con el parámetro de correlación de la distribución subyacente de

los datos. Por último, en la Sección 5, se presentan las conclusiones de los resultados alcanzados y se propone las estrategias a seguir para la elección del modelo en situaciones prácticas.

2. Estudio de Monte Carlo

Como se verá en el capítulo siguiente, la familia de distribuciones bivariadas propuesta por Sarkar (1987), es la que tiene un comportamiento más estable en todos los modelos, cuando se estudia el problema de la memoria del proceso de conteo. De esta manera, a fin de que el sesgo producido por la falta de correlación entre los tiempos de falla se pueda visualizar con mayor claridad, sin la contaminación del sesgo por memoria de la distribución, las muestras en las diferentes situaciones se obtienen a partir de la distribución Sarkar, cuya definición y propiedades se presentan en el Capítulo V.

Para llevar a cabo el estudio, se simulan datos como proveniente de la aplicación de dos tratamientos hipotéticos, reflejados en la covariable Z , definida más adelante. Se considera el problema de dos fallas, es decir $k=2$. Se simulan 100 muestras de tamaño $n = 500$, para cada una de las siguientes situaciones:

I) *eventos recurrentes* en los cuales los tiempos de falla intereventos tienen conjuntamente la distribución Sarkar, así, los tiempos totales T_1 y T_2 se obtiene como $T_1 = X_1$ y $T_2 = X_1 + X_2$, (Ver Fig. 4.2.1).

II) *eventos diferentes* donde los tiempos de falla totales se distribuyen conjuntamente como una distribución Sarkar, lo cual corresponden a $T_1 = X_1$ y $T_2 = X_2$, (Ver fig. 4.2.1).

Los datos con distribución Sarkar son generados por el algoritmo que se presenta en la Sección 4 del Capítulo V.

El parámetro b_k ($k=1, 2$) del modelo es en todos los casos igual a -2 y para cada tipo de evento se consideran distintos grados de correlación entre los tiempos

de falla. Los valores de los parámetros de la distribución y su correspondiente correlación entre los tiempos de falla se muestran en la Tabla 4.2.1.

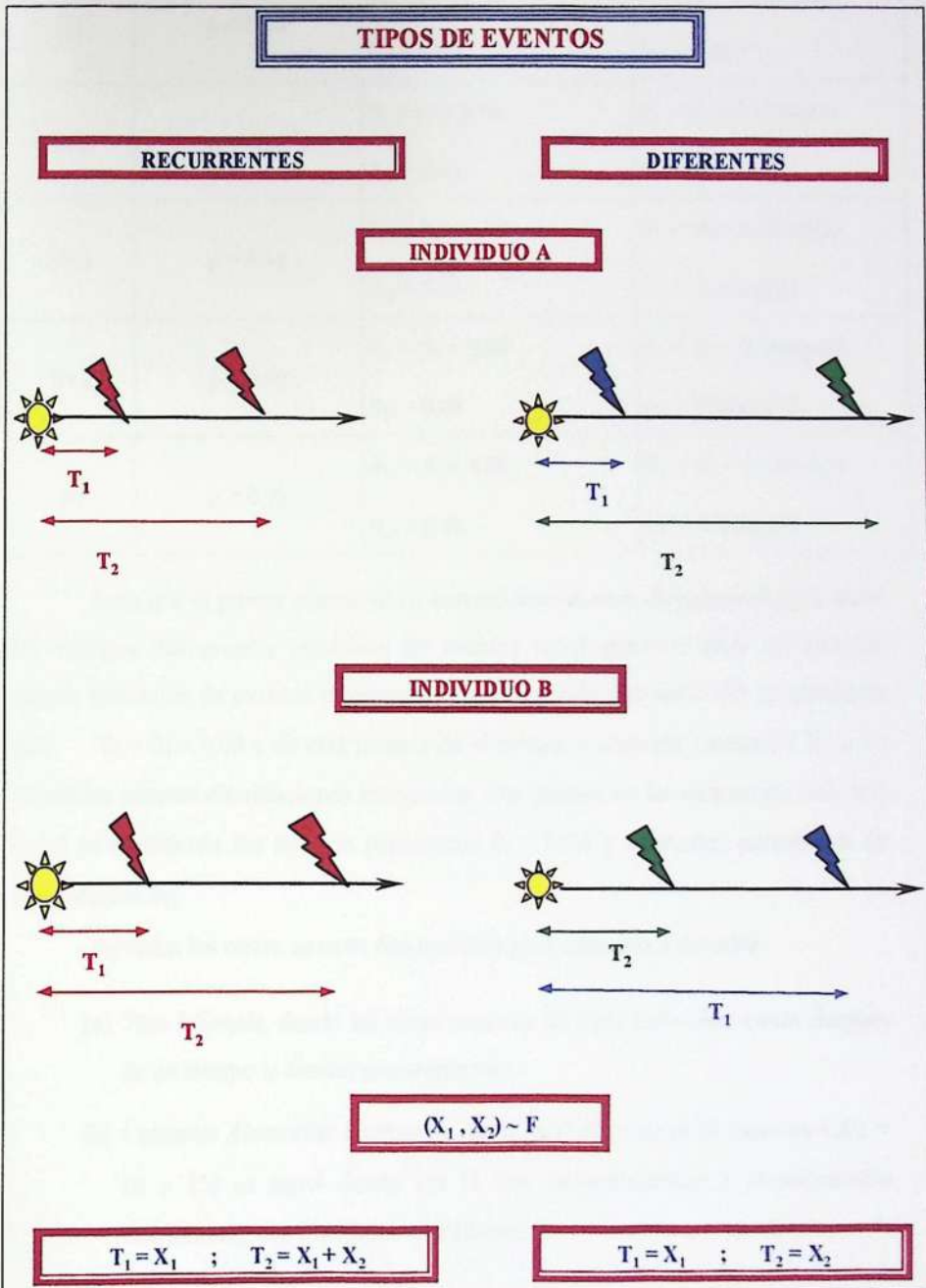


Fig. 4.2.1.: Esquema de los diferentes tipos de eventos

Tabla 4.2.1: Valores de los parámetros para cada grado de correlación y para cada valor de Z

Escenarios	Correlación (Y_1, Y_2)	Parámetros si $Z=0$	Parámetros si $Z=1$
(i)	$\rho = 0.00$	$\theta_1 = \theta_2 = 0.05$ $\theta_{12} = 0.00$	$\theta_1 = \theta_2 = 0.05\exp(2)$ $\theta_{12} = 0.00$
(ii)	$\rho = 0.13$	$\theta_1 = \theta_2 = 0.04$ $\theta_{12} = 0.01$	$\theta_1 = \theta_2 = 0.04\exp(2)$ $\theta_{12} = 0.01\exp(2)$
(iii)	$\rho = 0.34$	$\theta_1 = \theta_2 = 0.04$ $\theta_{12} = 0.03$	$\theta_1 = \theta_2 = 0.04\exp(2)$ $\theta_{12} = 0.03\exp(2)$
(iv)	$\rho = 0.60$	$\theta_1 = \theta_2 = 0.04$ $\theta_{12} = 0.08$	$\theta_1 = \theta_2 = 0.04\exp(2)$ $\theta_{12} = 0.08\exp(2)$
(v)	$\rho = 0.95$	$\theta_1 = \theta_2 = 0.04$ $\theta_{12} = 0.40$	$\theta_1 = \theta_2 = 0.04\exp(2)$ $\theta_{12} = 0.40\exp(2)$

Note que el primer escenario (i) corresponde al caso de independencia entre los tiempos intereventos (situación de eventos recurrentes) o entre los tiempos totales (situación de eventos diferentes). En el segundo escenario (ii) se considera que $\theta_1 = \theta_2 = 0.04$ y de esta manera en el primer y segundo escenario, Y_1 e Y_2 tienen las mismas distribuciones marginales. Por último, en los escenarios (iii), (iv) y (v) se consideran los mismos parámetros $\theta_i = 0.04$ y diferentes parámetros de dependencia θ_{12} .

En todas los casos, se usan dos modelos para censuras a derecha:

- Tipo I Simple*, donde las observaciones de cada individuo cesan después de un tiempo t_0 común preestablecido.
- Censuras Aleatorias Uniformes*, en la cual el proceso de censura $C_i(t) = I(t \leq U_i)$ es aquel donde los U_i son independientes e idénticamente distribuidos, con distribución uniforme.

En todos los casos se considera que aproximadamente el 25% de observaciones son censuradas. En la situación de eventos ordenados, si un individuo tiene la primera recurrencia censurada, entonces tiene también censurada la segunda recurrencia.

A cada conjunto de datos simulados se aplica los cuatro modelos de regresión usando el software desarrollado por Therneau (1995) para S-Plus. En cada modelo se requiere un ordenamiento especial de los datos que se presentan en la Tabla 3.4.1 del Capítulo III.

Para cada situación, cada escenario y cada tipo de modelo de regresión se estima el Error Cuadrático Medio (ECM), el Error Estándar (SE) y, por último, el Sesgo Relativo (SR) del estimador \hat{b} del parámetro del modelo por la expresión

$$SR = \frac{\left(\hat{E}(\hat{b}) - b \right) * 100}{b} \quad (4.2.1)$$

Los valores del ECM obtenidos a partir de las simulaciones para cada situación se muestran en la Tabla 4.2.2 y Tabla 4.2.3, y se representan gráficamente en las figuras 4.2.2 a 4.2.5; los correspondientes valores del SR se presentan en las Tablas 4.2.4 y 4.2.5, y en las figuras 4.2.6 a 4.2.9. El SE se muestra sólo en el caso de censuras Tipo I, en la Tabla 4.2.6, y Figuras 4.2.10 y 4.2.11, ya que sus valores no se modifican sustancialmente por el tipo de censura. Se calcula además, para cada valor de correlación ρ_K , la tasa de cambio del ECM definida como

$$TC_{ECM} = ECM(\rho_K) - ECM(0),$$

donde $ECM(0)$ es el valor del ECM cuando los tiempos de falla son independientes. De manera similar se calcula la tasa de cambio para el SR. Los valores de las tasas de cambio se muestran en las Tablas 4.2.7 a 4.2.8 y Figuras 4.2.12 a 4.2.15.

A partir de estos resultados, se puede observar que la correlación entre los tiempos de falla afecta la precisión del estimador del parámetro del modelo en forma diferente, según se trate de eventos diferentes o de eventos recurrentes, como era de esperar. El modelo WLW en el caso de eventos recurrentes y los modelos

AG, OK y PC para el caso de eventos diferentes, muestran un comportamiento atípico, ya que su ECM disminuye a medida que aumenta el grado de correlación entre los tiempos de falla. Este comportamiento, como se verá en la Sección 4 resulta en algunos casos de un problema del modelo y en otros de su implementación computacional. El SE muestra también un comportamiento diferente según el tipo de situación, pero en todos los casos sus valores son pequeños y de esta manera, no tienen demasiada influencia en la variación de ECM con el grado de correlación entre los tiempos de falla.

Tabla 4.2.2. Error cuadrático Medio (ECM) de los estimadores del parámetro $b = -2$ para las diferentes situaciones y según el grado de correlación entre los tiempos de falla cuando se usa censuras Tipo I simples.

Correlación (T_1, T_2)	Error Cuadrático Medio			
	Eventos Ordenados		Eventos no Ordenados	
	AG	OK	PC	WLW
0.00	0.02	0.02	0.02	0.35
0.13	0.025	0.043	0.018	0.099
0.34	0.063	0.159	0.058	0.042
0.60	0.108	0.313	0.056	0.027
0.95	0.157	0.488	0.089	0.021

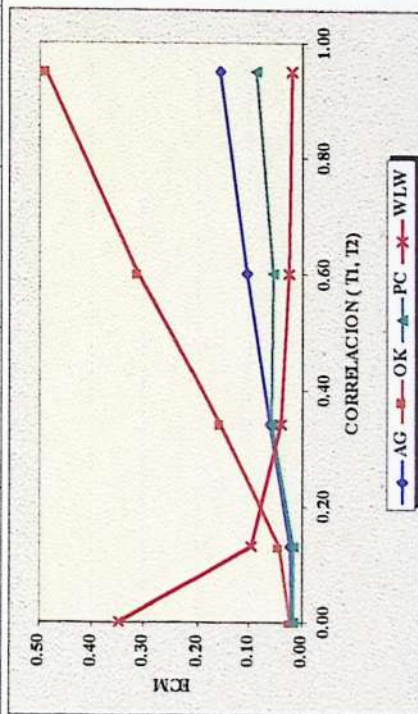


Figura 4.2.2.: ECM del estimador de b para cada modelo según el grado de correlación de los tiempos de falla en eventos recurrentes y censuras Tipo I.

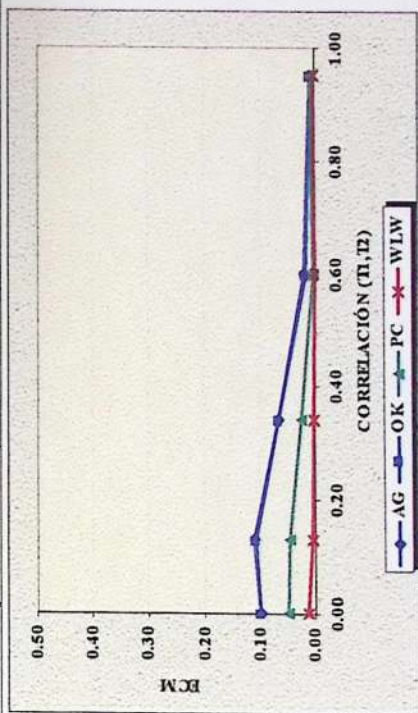


Figura 4.2.3.: ECM del estimador de b para cada modelo según el grado de correlación de los tiempos de falla en eventos diferentes y censuras Tipo I.

Tabla 4.2.3. Error cuadrático Medio (ECM) de los estimadores del parámetro $b = -2$ para las diferentes situaciones y según el grado de correlación entre los tiempos de falla cuando se usa censuras aleatorias Uniformes.

Correlación (T_1, T_2)	Error Cuadrático Medio						
	Eventos Ordenados			Eventos no Ordenados			
	AG	OK	PC	AG	OK	PC	WLW
0.00	0.024	0.021	0.012	0.411	0.102	0.028	0.002
0.13	0.014	0.026	0.003	0.170	0.132	0.053	0.008
0.34	0.026	0.103	0.010	0.112	0.089	0.052	0.004
0.60	0.058	0.248	0.019	0.063	0.056	0.093	0.011
0.95	0.102	0.409	0.033	0.037	0.021	0.078	0.021

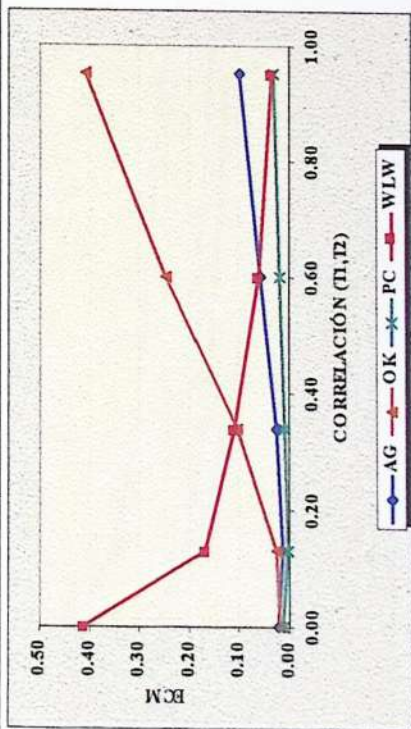


Figura 4.2.4. ECM del estimador de b para cada modelo según el grado de correlación de los tiempos de falla en eventos recurrentes y censuras aleatorias.

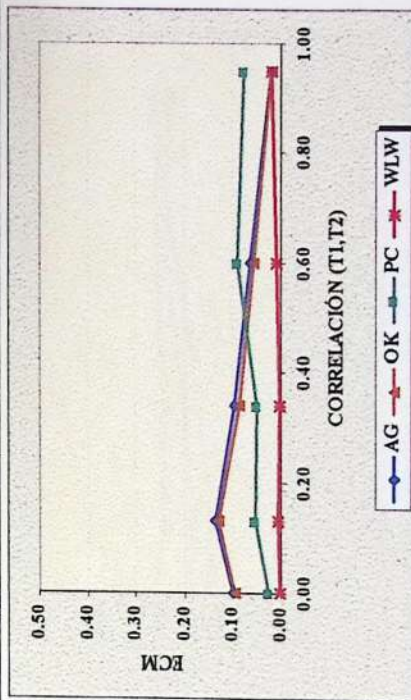


Figura 4.2.5. ECM del estimador de b para cada modelo según el grado de correlación de los tiempos de falla en eventos diferentes y censuras aleatorias.

Tabla 4.2.4. Sesgo Relativo (SR) de los estimadores del parámetro $b = -2$ para las diferentes situaciones y según el grado de correlación entre los tiempos de falla cuando se usa censuras Tipo I Simples.

Correlación (T_1, T_2)	Sesgo Relativo (SR)									
	Eventos Ordenados					Eventos no Ordenados				
	AG	OK	PC	WLW	WLR	AG	OK	PC	WLW	WLR
0.00	2%	2%	2%	28%	2%	-15%	-15%	-10%	1%	1%
0.13	-7%	-10%	-7%	16%	16%	-16%	-16%	-10%	-2%	-2%
0.34	-12%	-20%	-11%	9%	9%	-13%	-13%	-8%	-1%	-1%
0.60	-16%	-28%	-10%	6%	6%	-7%	-6%	-2%	1%	1%
0.95	-19%	-35%	-14%	2%	2%	4%	4%	-3%	2%	2%

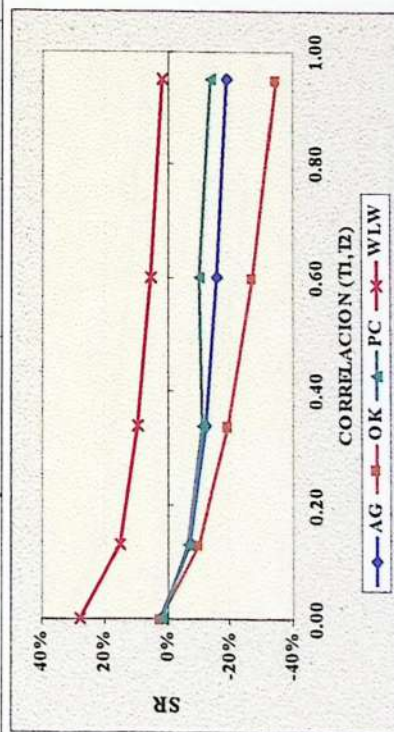


Figura 4.2.6.: SR del estimador de b para cada modelo según el grado de correlación de los tiempos de falla en eventos recurrentes y censuras Tipo I.

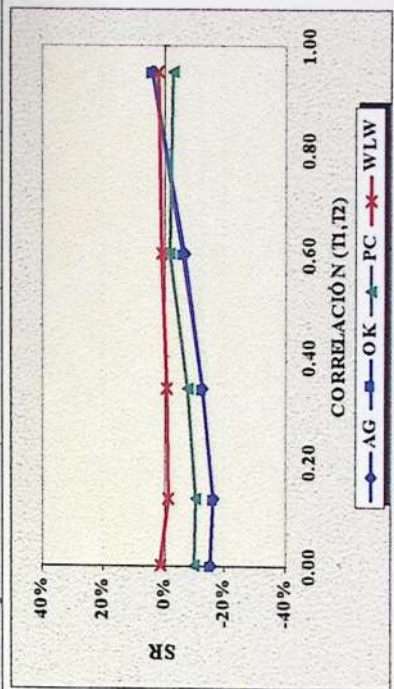


Figura 4.2.7.: SR del estimador de b para cada modelo según el grado de correlación de los tiempos de falla en eventos diferentes y censuras Tipo I.

Tabla 4.2.5. Sesgo Relativo (SR) de los estimadores del parámetro $b=-2$ para las diferentes situaciones y según el grado de correlación entre los tiempos de falla cuando se usa censuras aleatorias Uniformes.

Correlación (T_1, T_2)	Sesgo Relativo (SR)									
	Eventos Ordenados					Eventos no Ordenados				
	AG	OK	PC	WLW	WLV	AG	OK	PC	WLV	WLV
0.00	5%	5%	5%	31%	31%	-16%	-16%	-8%	-8%	0%
0.13	-4%	-8%	1%	20%	20%	-19%	-18%	-11%	-11%	-4%
0.34	-8%	-16%	3%	16%	16%	-15%	-15%	-11%	-11%	-1%
0.60	-12%	-25%	5%	11%	11%	-11%	-11%	-14%	-14%	1%
0.95	-15%	-32%	5%	7%	7%	-3%	-2%	-13%	-13%	3%

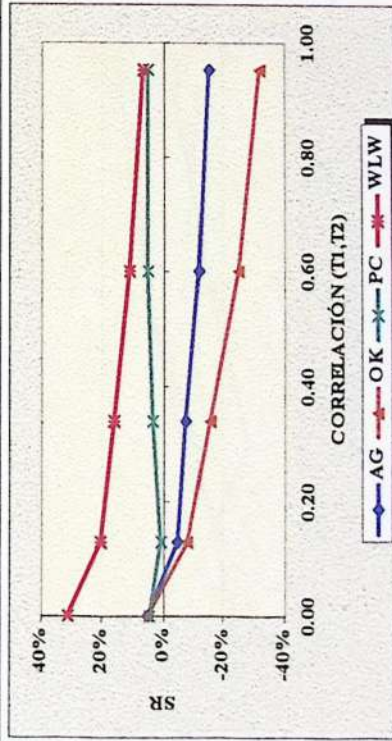


Figura 4.2.8.: SR del estimador de b para cada modelo según el grado de correlación de los tiempos de falla en eventos recurrentes y censuras aleatorias.

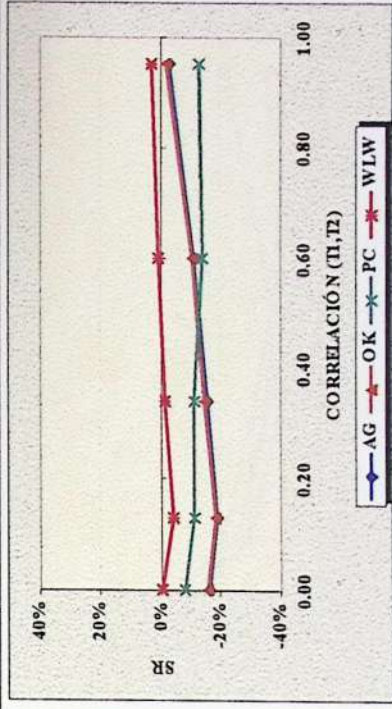


Figura 4.2.9.: SR del estimador de b para cada modelo según el grado de correlación de los tiempos de falla en eventos diferentes y censuras aleatorias.

Tabla 4.2.6. Error Estándar (SE) de los estimadores del parámetro $b = -2$ para las diferentes situaciones y según el grado de correlación entre los tiempos de falla cuando se usa censuras Tipo I Simples.

Correlación (T_1, T_2)	Error Estándar (SE)							
	Eventos Ordenados			Eventos no Ordenados				
	AG	OK	PC	WLW	AG	OK	PC	WLW
0.00	0.02	0.06	0.05	0.05	0.10	0.10	0.10	0.11
0.13	0.08	0.03	0.05	0.06	0.06	0.05	0.05	0.07
0.34	0.04	0.07	0.09	0.08	0.04	0.05	0.03	0.05
0.60	0.07	0.09	0.12	0.12	0.03	0.03	0.04	0.03
0.95	0.10	0.08	0.13	0.14	0.03	0.04	0.07	0.04

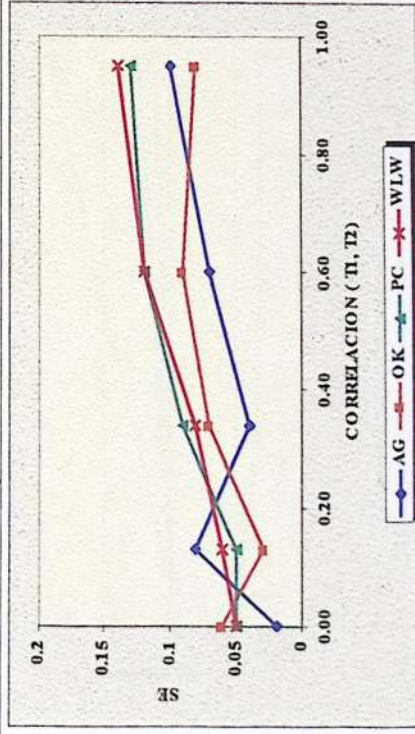


Figura 4.2.10.: SE del estimador de b para cada modelo según el grado de correlación de los tiempos de falla en eventos recurrentes y censuras Tipo I.

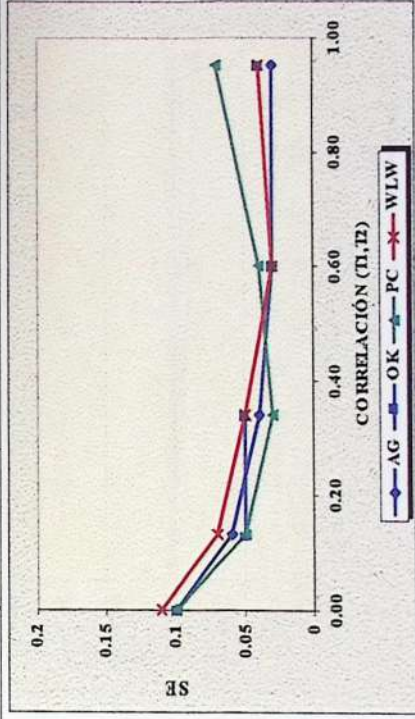


Figura 4.2.11.: SE del estimador de b para cada modelo según el grado de correlación de los tiempos de falla en eventos diferentes censuras Tipo I.

Tabla 4.2.7. Tasa de cambio del ECM de los estimadores del parámetro $b = -2$ para las diferentes situaciones y según el grado de correlación entre los tiempos de falla cuando se usa censuras Tipo I Simples.

Correlación (T_1, T_2)	Tasa de Cambio del ECM							
	Eventos Ordenados			Eventos no Ordenados				
	AG	OK	PC	WLW	AG	OK	PC	WLW
0.13	0.005	0.019	0.002	-0.249	0.009	0.008	-0.004	-0.007
0.34	0.044	0.134	0.042	-0.305	-0.034	-0.035	-0.024	-0.010
0.60	0.088	0.289	0.039	-0.320	-0.082	-0.083	-0.047	-0.011
0.95	0.137	0.464	0.073	-0.327	-0.094	-0.092	-0.043	-0.010

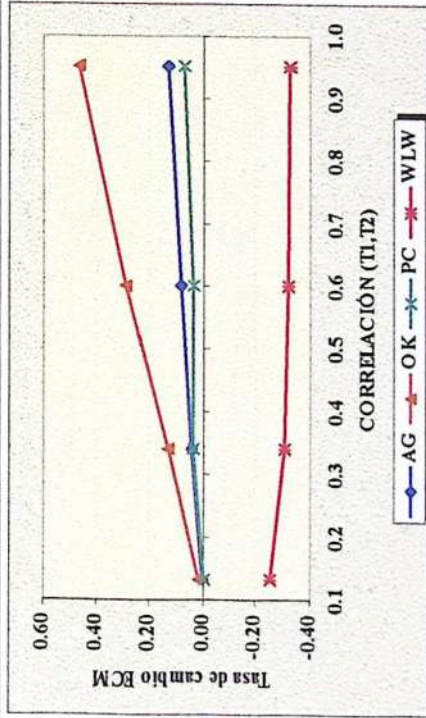


Figura 4.2.12.: Tasa de cambio del ECM del estimador de b para cada modelo según el grado de correlación de los tiempos de falla en eventos recurrentes y censuras Tipo I.

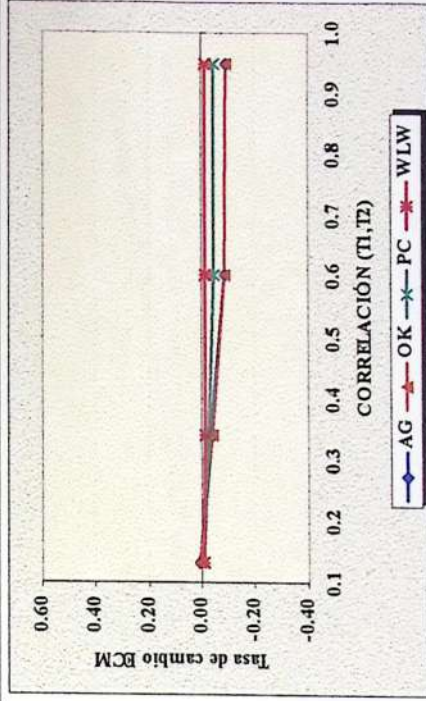


Figura 4.2.13.: Tasa de cambio del ECM del estimador de b para cada modelo según el grado de correlación de los tiempos de falla en eventos diferentes y con censuras Tipo I.

Tabla 4.2.8. Tasa de cambio del SR de los estimadores del parámetro $b = -2$ para las diferentes situaciones y según el grado de correlación entre los tiempos de falla cuando se usa censuras Tipo I Simples.

Correlación (T_1, T_2)	Tasa de Cambio del SR									
	Eventos Ordenados					Eventos no Ordenados				
	AG	OK	PC	WLW	AG	OK	PC	WLW		
0.13	-9	-13	-8	-13	-1	-1	0	-3		
0.34	-15	-22	-13	-19	2	2	2	-2		
0.60	-18	-30	-12	-23	8	9	8	0		
0.95	-21	-37	-16	-26	19	19	7	1		

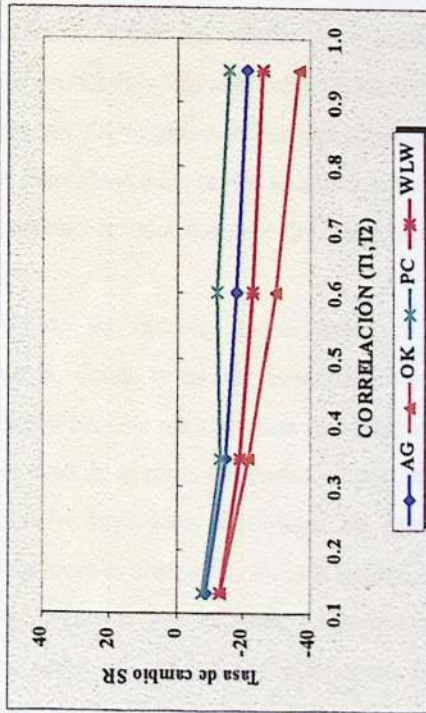


Figura 4.2.14.: Tasa de cambio del SR del estimador de b para cada modelo según el grado de correlación de los tiempos de falla en eventos recurrentes y censuras Tipo I.

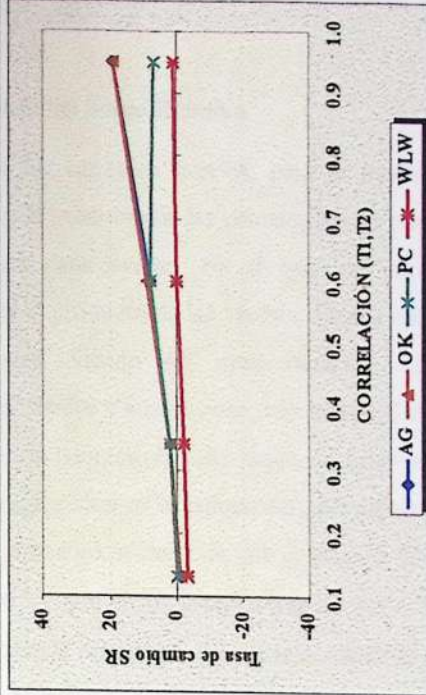


Figura 4.2.15.: Tasa de cambio del SR del estimador de b para cada modelo según el grado de correlación de los tiempos de falla en eventos diferentes y con censuras Tipo I.

3. Análisis del Comportamiento del Sesgo Relativo

Como se puntualiza en los capítulos previos, para el modelo WLW, la verosimilitud parcial conjunta es el producto de las verosimilitudes parciales de los *tiempos totales* marginales para cada evento. En el caso del modelo PC, la verosimilitud parcial conjunta es el producto de las verosimilitudes parciales de los *tiempos inter-eventos* marginales. Siendo así, estos modelos, implícitamente suponen que los tiempos de falla, totales e intereventos, son independientes. De esta manera, si la correlación entre los tiempos de falla crece, se debería esperar una pérdida de eficiencia estadística (precisión en la estimación), principalmente cuando los eventos son ordenados. Sin embargo, a causa de que el modelo WLW presenta un sesgo positivo cuando $\theta_{12} = 0$ y que el SR decrece cuando $\rho(Y_1, Y_2)$ crece, se produce una ganancia en la eficiencia estadística del modelo a medida que aumenta la correlación entre los tiempos de falla.

Por otra parte, por la definición de SR según la expresión (4.2.1), un valor positivo del SR indica una sobrestimación del efecto de tratamiento, y una consecuente sobrestimación de la probabilidad de sobrevida de los individuos bajo tratamiento. De la misma manera, un SR negativo indica una subestimación de la probabilidad de sobrevida de los individuos a los cuales se aplicó el tratamiento. De aquí la importancia de estudiar no sólo el valor absoluto del sesgo sino también su signo.

Respecto a la aplicación de los modelos mencionados en estudios de confiabilidad del área de ingeniería, donde la ésta se examina sobre sistemas en serie o en paralelo, se puede establecer una analogía con la situación de eventos ordenados y no ordenados. En el caso de eventos ordenados, el segundo evento no puede ocurrir hasta que el primero haya sido observado. De esta manera, el tratamiento será efectivo, al menos hasta el tiempo t , cuando el tiempo de la primera y segunda falla sean ambos mayores a t . Por lo tanto, se podría considerar la situación de eventos ordenados como un sistema en serie, como lo puntualiza Turnbull (1997). Por otra parte, en la situación de eventos no ordenados o diferentes, el tratamiento es efectivo, al menos hasta el tiempo t , cuando cualquier

falla no ocurre hasta t , de este modo, esta situación se puede considerar como un sistema en paralelo.

Puesto que en nuestro caso, los tiempos de falla intereventos (X_1, X_2) para el caso de eventos recurrentes tiene distribución Sarkar Bivariada $S(\theta_1, \theta_2, \theta_{12})$ con $\theta_1 = \theta_2 = \theta_c$ y $\theta_{12} > 0$, considerando la expresión de la función de sobrevida conjunta de la distribución Sarkar dada en (5.2.2.1) cuando $\theta_{12} > 0$, la confiabilidad de los sistemas en serie está dada por

$$C^D(t) = S(t, t) = P[X_1 \geq t, X_2 \geq t \mid \theta_c, \theta_{12}] = \exp[-2\theta_c - \theta_{12}]t, \quad (4.3.1)$$

donde $C^D(t)$ es la confiabilidad bajo una estructura dependiente de (X_1, X_2) .

Las marginales de una distribución Sarkar son exponenciales con parámetro $\theta_c + \theta_{12}$; por lo tanto si $\theta_{12} = 0$, la confiabilidad de un sistema en serie con tiempos de vida inter-eventos independientes estará dada por

$$C^I(t) = S_1(t) S_2(t) = P[X_1 \geq t, X_2 \geq t \mid \theta_c, \theta_{12}] = \exp[-2\theta_c - 2\theta_{12}]t. \quad (4.3.2)$$

Entonces, la razón

$$\frac{C^D(t)}{C^I(t)} = \exp[\theta_{12}t] \quad (4.3.3)$$

es mayor que 1, para todo $t > 0$.

A partir de la expresión (4.3.3) se observa que tratar el problema como si los tiempos intereventos fueran independientes produce una subestimación de la función de sobrevida, y el grado de ésta aumenta con el valor del parámetro de correlación de la distribución θ_{12} . Es decir, el valor del sesgo del estimador del modelo decrece cuando aumenta la correlación entre X_1 y X_2 . Esto explica el comportamiento del SR en la situación de eventos recurrentes, que se muestra en las figuras (4.2.6) y (4.2.8).

Respecto a la confiabilidad de los sistemas en paralelo, cuando la distribución conjunta de los tiempos de falla es Sarkar y el parámetro de correlación de dicha distribución es $\theta_{12} > 0$, está dada por

$$C^D(t) = 1 - P[Y_1 \leq t, Y_2 \leq t | \theta_c, \theta_{12}] = \exp[-(\theta_c + \theta_{12})t][2 - \exp(-\theta_c t)]. \quad (4.3.4)$$

Sin embargo, si se trata el problema como si X_1 y X_2 fueran independientes, entonces la confiabilidad sería

$$C^I(t) = 1 - F_1(t) F_2(t) = \exp[-(\theta_c + \theta_{12})t] \{2 - \exp[-(\theta_c + \theta_{12})t]\}. \quad (4.3.5)$$

de donde el cociente

$$\frac{C^D(t)}{C^I(t)} = \frac{2 - \exp(-\theta_c t)}{2 - \exp[-(\theta_c + \theta_{12})t]} \quad (4.3.6)$$

el cual es menor que 1, para todo $t > 0$.

De esta última expresión, se infiere que en el caso de eventos diferentes, la consideración de independencia entre X_1 y X_2 , cuando $\theta_{12} > 0$, produce una sobrestimación de la confiabilidad del sistema, que aumenta con el valor del parámetro de correlación θ_{12} . Ahora bien, una sobrestimación creciente de la sobrevida a cualquiera de las fallas tiene como consecuencia un crecimiento del sesgo cuando aumenta el valor de θ_{12} . En el caso de los modelos AG, OK y PC, para los cuales el SR es negativo si $\theta_{12} = 0$, su crecimiento produce estimadores menos sesgados cuando aumenta la correlación de los tiempos de falla. Esto explicaría el comportamiento del SR en el caso de eventos diferentes, y del ECM, sobre todo en los modelos condicionales. Sin embargo, si el tiempo de sobrevida de los sistemas es muy grande la correlación no afecta de manera importante la estimación de la sobrevida en esta situación.

4. Cotas para el Sesgo Relativo

Para evaluar el sesgo de los estimadores del parámetro de los modelos de regresión, se debe tener en cuenta que ellos son un promedio de los estimadores correspondientes a cada evento, o sea que

$$\hat{b} = \frac{b_1^* + b_2^*}{2} = \frac{b + k_1 + b + k_2}{2} = b + \frac{k_1 + k_2}{2} \quad (4.4.1)$$

donde k_i ($i=1, 2$) es el sesgo del estimador del parámetro cuando se considera la i -ésima falla. Por lo tanto, se debe calcular por separado el sesgo introducido cuando se estima el parámetro para cada evento y luego, el sesgo total será el promedio de los sesgos para cada falla.

Por otra parte, tanto en la situación de eventos recurrentes como en la de eventos diferentes, se debe considerar por separado los modelos condicionales de los modelos marginales. Esto debido a que, como se puntualizó en capítulos previos, los modelos de "hazard" proporcional de tipo condicional se expresan en función de las componentes del vector intensidad $\lambda(\mathbf{t})$, mientras que los modelos marginales se expresan en función de los "hazard" correspondientes a las distribuciones marginales de los tiempos de falla (totales o intereventos).

Teniendo en cuenta lo referido, *para cada evento* y cuando $Z=1$, se considera el cociente

$$\frac{e^{b_i^*}}{e^b} = \frac{h_i^*(\theta)}{h_i(\theta)}, \quad i = 1, 2, \quad (4.4.2)$$

donde θ es el vector de parámetros de la distribución; b es el verdadero valor del parámetro; b_i^* es el estimador del parámetro del modelo, cuando no se considera la correlación entre los tiempos de falla dentro del modelo; y por último, h y h_i^* son los "hazard" asociados a los respectivos modelos. Cabe destacar que b representa el

valor del parámetro solamente si se introducen sesgos por correlación, en caso contrario se debe reemplazar b por \hat{b} , que corresponde al *estimador del parámetro*. Entonces, a partir de la expresión (4.4.2) se obtiene

$$E(b_i^*) - b = E \left\{ \ln \left[\frac{h_i^*(\theta)}{h_i(\theta)} \right] \right\}, \quad i = 1, 2, \quad (4.4.3)$$

que es la expresión del sesgo del estimador en el caso que b sea el verdadero valor del parámetro, o bien

$$E(b_i^*) - E(\hat{b}) = E \left\{ \ln \left[\frac{h_i^*(\theta)}{h_i(\theta)} \right] \right\}, \quad i = 1, 2, \quad (4.4.4)$$

en el caso que se introduzcan otros tipos de sesgos en el modelo.

4.1. Sesgos en la Situación de Eventos Recurrentes

Modelos condicionales

Como ya se mencionó, los modelos de "hazard" proporcional de tipo condicional se expresan en función de las componentes del vector intensidad $\lambda(t)$. Entonces, si se considera, por ejemplo, el vector intensidad $\lambda_1(t)$ correspondiente al primer evento, se debe tener en cuenta que es igual a la suma de dos términos,

$$\lambda_1(t) = p_1^{(1)}(t)Y_1(t)[1-Y_2(t)] + p_2^{(1)}(t)Y_1(t)Y_2(t)$$

el primero de los cuales está asociado al "hazard" condicional si $T_2 < t$ y el segundo asociado al "hazard" condicional cuando $T_2 \geq t$. Por lo tanto, la correlación entre los tiempos de falla se toma en cuenta dentro de los modelos condicionales, por esto se espera que en ellos no se introduzcan sesgos por correlación. Sin embargo, en el estudio de Monte Carlo el sesgo aumenta con el grado de correlación, pero como se verá en detalle en el próximo capítulo, el sesgo observado se introduce por causa de

la memoria conjunta de la distribución Sarkar. En efecto, usando la expresión (5.5.1.5) y tomando como referencia valores de los parámetros de la distribución Block-Basu, que producen similares grados de correlación a los usados en el estudio por simulación, se obtiene el siguiente cuadro comparativo

Tabla 4.4.1: Comparación de la cota del SR con los valores estimados en el estudio de Monte Carlo en modelos condicionales en la situación de eventos ordenados o recurrentes

Correlación	Parámetros de la distribución Block-Basu	Cota para el SR por memoria de la distribución	SR estimado por simulación del modelo AG
$\rho = 0.13$	$\theta_1 = \theta_2 = 0.015$ $\theta_{12} = 0.035$	$\leq - 6 \%$	-7%
$\rho = 0.34$	$\theta_1 = \theta_2 = 0.015$ $\theta_{12} = 0.04$	$\leq - 11 \%$	-12%
$\rho = 0.60$	$\theta_1 = \theta_2 = 0.015$ $\theta_{12} = 0.06$	$\leq - 16 \%$	-16%

Las diferencias observadas entre las cotas y los valores estimados del sesgo en el modelo OK (ver Tabla 4.2.4), se deben a que este modelo se ve afectado en mayor grado por la memoria del proceso como se discute en detalle en el capítulo siguiente.

Modelos Marginales

El hecho que la verosimilitud parcial conjunta de los modelos marginales se exprese como el producto de las verosimilitudes parciales de los tiempos de falla marginales, implícitamente supone independencia entre los tiempos de falla; tiempos de falla totales en el modelo WLW y tiempos de falla intereventos en el modelo PC. Osea que, a diferencia de los modelos condicionales, en los modelos marginales no se modela cada una de las componentes del vector intensidad $\lambda(\mathbf{t})$, sino los "hazard" correspondientes a las distribuciones marginales de los tiempos

de falla (totales o intereventos). Por esta razón, la función $h^*(\theta)$ de la ecuación (4.4.2) representa en este caso el "hazard" marginal, mientras que $h(\theta)$ es el correspondiente "hazard" condicional.

Por el mismo argumento empleado para obtener la expresión (4.4.1), el sesgo que se introduce al estimar el parámetro b del modelo, sólo para el primer evento, será el promedio de los sesgos introducidos en la estimación de b cuando $h(\theta)$ sea el "hazard" condicional dado $T_2 < t$ y $T_2 \geq t$. Entonces, el sesgo para los modelos marginales de cada uno de los eventos se puede expresar como

$$E(b_i^*) - b = \frac{1}{2} E \left\{ \ln \left[\frac{p_1^{(i)*}}{p_1^{(i)}} \right] + \ln \left[\frac{p_2^{(i)*}}{p_2^{(i)}} \right] \right\}, \quad i = 1, 2, \quad (4.4.1.1)$$

donde $p_i^{(j)*}$ ($i, j=1, 2$) es el "hazard" marginal de los tiempos de falla, y $p_i^{(j)}$ ($i, j=1, 2$) es el correspondiente "hazard" condicional.

Debido a la diferencia que existe entre los modelos WLW y PC, respecto a cómo se consideran los tiempos de falla de la segunda recurrencia (X_1+X_2 para WLW y X_2 para PC), a fin de obtener una cota para el sesgo se debe tener en cuenta dos tipos:

- (a). El sesgo que se genera por considerar que los individuos que no han experimentado la primera falla están bajo el riesgo de sufrir la segunda falla; que es equivalente a considerar como tiempo de falla para la segunda recurrencia, en el modelo WLW, al tiempo total, es decir, $T_2 = X_1 + X_2$.
- (b). El sesgo que se genera si los tiempos de falla son correlacionados.

Por esta razón, para evaluar el sesgo por correlación cuando se considera la segunda recurrencia, en el modelo WLW, se usa la expresión (4.4.4), debido a que, el estimador del parámetro en el denominador se ve afectado por el sesgo de tipo

- (a).

Evaluación del sesgo tipo (a)

Considerar que el tiempo de falla del segundo evento en el modelo WLW es $T_2 = X_1 + X_2$, equivale a pensar que los datos provienen de la distribución conjunta de las variables (X_1, T_2) y no de las variables originales (X_1, X_2) .

Por razones de simplicidad y para explicar el comportamiento atípico del modelo WLW en la situación de eventos recurrentes cuando los tiempos de falla son independientes, para evaluar el sesgo tipo (a) se parte de la consideración que X_1 y X_2 son independientes e idénticamente distribuidas en forma exponencial con parámetro θ_c . Entonces $T_2 = X_1 + X_2$ se distribuye como una Gamma, $\Gamma(2, \theta_c)$, por lo tanto

$$P[X_2 \geq t] = e^{-\theta_c t}$$

y

$$P[T_2 \geq t] = e^{-\theta_c t} (\theta_c t + 1),$$

de la última igualdad se sigue que la función de densidad de T_2 está dada por

$$f_{T_2} = \theta_c^2 t e^{-\theta_c t},$$

y por lo tanto las funciones "hazard" asociadas vienen dada por

$$h_{T_2} = \frac{\theta_c^2 t}{\theta_c t + 1}$$

$$h_{X_2} = \theta_c.$$

Luego, reemplazando estas expresiones en (4.4.2) se obtiene

$$\frac{e^b}{e^b} = \frac{h_{T_2}}{h_{X_2}} = \frac{\theta_c t}{\theta_c t + 1},$$

de donde se sigue que

$$\hat{b} - b = \ln(\theta_c t) - \ln(\theta_c t + 1). \quad (4.4.1.2)$$

Pero, por ser $\ln(x)$ una función cóncava se verifica que

$$\ln\left(\frac{x+x'}{2}\right) \geq \frac{1}{2}\ln x + \frac{1}{2}\ln x'.$$

Y, usando esta propiedad en la expresión (4.4.1.2), se obtiene que

$$\hat{b} - b \leq \ln(\theta_c t) - \ln 2 - \frac{1}{2}\ln(\theta_c t),$$

de donde se sigue que la expresión para el sesgo está dada por

$$E(\hat{b}) - b \leq \frac{1}{2}E[\ln(\theta_c t)] - \ln 2. \quad (4.4.1.3)$$

Entonces, por la desigualdad de Jensen y la concavidad de la función $\ln(x)$ se verifica que

$$E(\ln x) \leq \ln E(x).$$

Usando esta desigualdad en (4.4.1.3) y considerando que la $E(T_2) = 2/\theta_c$, se obtiene que

$$E(\hat{b}) - b \leq -\frac{1}{2}\ln 2. \quad (4.4.1.4)$$

En el caso en que $b = -2$, como el considerado en el estudio de Monte Carlo, se tiene que la cota para el SR es

$$SR = \frac{E(\hat{b}) - b}{b} \geq 17\%, \quad (4.4.1.5)$$

resultado que concuerda con lo obtenido por simulación.

Evaluación del sesgo tipo (b)

El sesgo tipo (b) se introduce tanto en el modelo PC como en el modelo WLW para ambos eventos, pero en este último, al estimar el parámetro correspondiente al segundo evento, se genera además el sesgo tipo (a). Por todo lo mencionado, para obtener una cota para este tipo de sesgo correspondiente al primer y segundo evento se usa la expresión (4.4.1.1) para el modelo PC; mientras que, para el modelo WLW al evaluar el sesgo del estimador cuando se considera el segundo evento se debe reemplazar en la expresión antes mencionada, b por \hat{b} , que corresponde a un estimador sesgado del modelo, con sesgo tipo (a).

Como se verá en detalle en el capítulo siguiente, las expresiones para los $p_i^{(i)}$ correspondiente a la distribución Sarkar, cuando $\theta_1 = \theta_2 = \theta_c$, son las siguientes

$$\begin{aligned} p_2^{(2)} = p_1^{(1)} &= \theta_c + \frac{1}{2} \theta_{12} + \frac{1}{2} \theta_{12} [1 - \exp(-\theta_c t)]^{-1} \\ p_1^{(2)} = p_2^{(1)} &= \theta_c + \frac{1}{2} \theta_{12}; \end{aligned} \quad (4.4.1.6)$$

mientras que los "hazard" marginales de los tiempos de falla, son los correspondientes a la distribución exponencial con parámetro $\theta_c + \theta_{12}$, por lo que $p_i^{(i)*} = \theta_c + \theta_{12}$, para todo $i, j = 1, 2$. Reemplazando en (4.4.1.1) se tiene que, para el modelo PC

$$E(\hat{b}_{PC}^*) - b = \frac{1}{2} E \left[\ln \left(\frac{\theta_c + \theta_{12}}{p_1^{(1)}} \right) + \ln \left(\frac{\theta_c + \theta_{12}}{p_2^{(1)}} \right) \right]. \quad (4.4.1.7)$$

Puesto que en el modelo WLW se introduce el sesgo tipo (a), se tiene que

$$E(\hat{b}_{WLW}^*) - E(\hat{b}) = \frac{1}{2} E \left[\ln \left(\frac{\theta_c + \theta_{12}}{p_1^{(1)}} \right) + \ln \left(\frac{\theta_c + \theta_{12}}{p_2^{(1)}} \right) \right].$$

Combinando esta expresión con (4.4.1.4) se obtiene

$$E(b_{WLW}^*) - b \leq \frac{1}{2} E \left[\ln \left(\frac{\theta_c + \theta_{12}}{p_1^{(1)}} \right) + \ln \left(\frac{\theta_c + \theta_{12}}{p_2^{(1)}} \right) \right] - \frac{1}{2} \ln 2 \quad (4.4.1.8)$$

Sin embargo, el "hazard" condicional $p_1^{(1)}$ de la distribución Sarkar depende de t , por lo tanto, para obtener una expresión independiente del tiempo de falla, a partir de las expresiones (4.4.1.7) y (4.4.1.8), que simplifique el cálculo de la esperanza, se usan cotas para $p_1^{(1)}$ que se obtienen a partir de su definición dada en (4.4.1.6), es decir,

$$\theta_c + \frac{1}{2} \theta_{12} \leq p_1^{(1)} \leq \theta_c + \frac{3}{2} \theta_{12} \quad (4.4.1.9)$$

Usando esta desigualdad en (4.4.1.7) se tiene que la expresión para el sesgo del estimador del modelo PC toma la forma

$$\frac{1}{2} \left[\ln \left(\frac{\theta_c + \theta_{12}}{\theta_c + \frac{3}{2} \theta_{12}} \right) + \ln \left(\frac{\theta_c + \theta_{12}}{\theta_c + \frac{1}{2} \theta_{12}} \right) \right] \leq E(b_{PC}^*) - b \leq \left[\ln \left(\frac{\theta_c + \theta_{12}}{\theta_c + \frac{1}{2} \theta_{12}} \right) \right] \quad (4.4.1.10)$$

Reemplazando el lado izquierdo de la desigualdad (4.4.1.9) en (4.4.1.8) se tiene que

$$E(b_{WLW}^*) - b \leq \left[\ln \left(\frac{\theta_c + \theta_{12}}{\theta_c + \frac{1}{2} \theta_{12}} \right) \right] - \frac{1}{2} \ln 2 \quad (4.4.1.11)$$

Usando los valores de los parámetros de la distribución Sarkar en los diferentes escenarios y del parámetro de los modelos de regresión PC y WLW del estudio de Monte Carlo, se encuentran las respectivas cotas del SR, y como se puede ver a través de la Tabla 4.4.2, ellas concuerdan con los valores estimados del SR en el estudio por simulación.

Estos resultados permiten explicar el aumento de la eficiencia estadística del modelo WLW cuando aumenta el grado de correlación entre los tiempos de falla, en contra de lo esperado por la suposición implícita de independencia que plantea la

propuesta de estimación. Es claro, a partir del análisis de cómo se generan los sesgos de estos modelos, que si en lugar de considerar los *tiempos totales* se consideran los *tiempos intereventos*, la situación sería similar al del modelo PC; es decir, el sesgo aumentaría con el grado de correlación entre los tiempos de recurrencia de los eventos.

Tabla 4.4.2: Comparación de la cota del SR con los valores estimados en el estudio de Monte Carlo en modelos marginales en la situación de eventos ordenados o recurrentes

Modelo	Correlación	Cota para el SR	Valor estimado del SR
PC	$\rho = 0.13$	[- 5 %, - 0.2 %]	-7%
	$\rho = 0.34$	[- 12 %, - 1 %]	-11%
	$\rho = 0.60$	[- 20 %, - 3 %]	-10%
	$\rho = 0.95$	[- 30 %, - 6 %]	-14%
WLW	$\rho = 0.13$	≥ 12 %	16%
	$\rho = 0.34$	≥ 5 %	9%
	$\rho = 0.60$	≥ -3 %	6%
	$\rho = 0.95$	≥ -13 %	2%

La falta de concordancia entre el SR calculado con la expresión (4.4.1.11) y el valor estimado para una correlación de 0.95, en el modelo WLW, se puede explicar si se considera que cuando la correlación es cercana a uno el efecto por ordenamiento es casi nulo, debido a que los valores de X_1 y X_2 tienden a coincidir en cada individuo.

4.2. Sesgos en la Situación de Eventos Diferentes

En la situación de eventos diferentes se debe tener en cuenta dos aspectos fundamentales:

- i) Los individuos están expuestos al riesgo de cualquiera de las fallas desde el comienzo del estudio y entonces, como ya se puntualiza en la figura 4.2.1, se verifica que $T_1 = X_1$ y $T_2 = X_2$.
- ii) Las fallas no se presentan en un orden preestablecido, puesto que se trata de eventos diferentes, por lo tanto, puede ser $T_1 > T_2$ o viceversa.

Cualquier modelo que no tenga en cuenta estas características introducirá sesgos en la estimación del parámetro de regresión.

Sin embargo, de los cuatro modelos que se consideran, el modelo WLW es el único que tiene en cuenta ambas características, como se verá en detalle más adelante. Por otra parte, puesto que las distribuciones marginales de X_1 y X_2 involucran el parámetro de correlación de la distribución bivariada, al estimar el parámetro de regresión usando tales marginales, implícitamente se considera el grado de correlación entre ellas. De esta manera, se espera que el sesgo por correlación del modelo WLW sea próximo a cero. En efecto, el estudio de Monte Carlo muestra que el sesgo para el modelo WLW es comparativamente pequeño, cualquiera sea el grado de correlación entre las variables y en ambos tipos de censuras.

Por todo lo expuesto, para evaluar el sesgo en la situación de eventos diferentes se considera como modelo de referencia a WLW, por lo que en el denominador de la expresión (4.4.2) se usa el "hazard" marginal de la distribución de X_1 y X_2 respectivamente.

Modelos condicionales

Para evaluar el sesgo de los modelos condicionales, se debe tener en cuenta que el software desarrollado por Therneau (1995) requiere, en esta situación, ordenar los datos de manera que el primer tiempo de falla T_1 sea el $\min(X_1, X_2)$, mientras que T_2 sea la diferencia entre el $\min(X_1, X_2)$ y el $\max(X_1, X_2)$. El uso de los modelos condicionales con los datos así ordenados, es equivalente, entonces, a aplicar tales modelos a muestras provenientes de la distribución conjunta de las

variables $W = \min(X_1, X_2)$ y $Z = |X_1 - X_2|$, y no de las variables originales (X_1, X_2) . Por lo tanto, el sesgo de los modelos condicionales en la situación de eventos diferentes, se genera principalmente por el ordenamiento de los datos para el cálculo computacional y no por la estructura del modelo. A partir de estas consideraciones se evalúa el SR para estos casos.

➤ Evaluación del Sesgo por Ordenamiento

Por las propiedades de la distribución Sarkar las variables W y Z son independientes y sus respectivas funciones "hazard", como se verá en detalle en la Sección 5 del Capítulo 5, vienen dadas por: $h_w = 2\theta_c + \theta_{12}$ y $h_z = \theta_c + \theta_{12}$. Luego, a partir de la expresión (4.4.3) y para cada falla, el sesgo por ordenamiento de estos modelos se puede expresar como

$$E(\hat{\beta}_1) - b = \ln \left[\frac{h_w}{h_{X_1}} \right], \quad (4.4.2.1)$$

$$E(\hat{\beta}_2) - b = \ln \left[\frac{h_z}{h_{X_2}} \right].$$

Luego, reemplazando las expresiones de las funciones "hazard" de W , Z , X_1 y X_2 , y considerando que las funciones "hazard" de Z y X_2 son iguales, se obtiene que el promedio de estas igualdades da como resultado que el sesgo introducido en los modelos condicionales, por el ordenamiento de los datos, se exprese como

$$E(\hat{\beta}) - b = \frac{1}{2} \ln \left[\frac{2\theta_c + \theta_{12}}{\theta_c + \theta_{12}} \right] \quad (4.4.2.2)$$

➤ Evaluación del Sesgo por Correlación

Puesto que los modelos condicionales usan el "hazard" condicional en la ecuación de verosimilitud parcial, la expresión para el sesgo en esta situación se expresa como

$$E(\hat{b}^*) - E(\hat{b}) = \frac{1}{2} E \left\{ \ln \left[\frac{p_1^{(1)}}{h_{X_1}} \right] + \ln \left[\frac{p_2^{(1)}}{h_{X_1}} \right] \right\}$$

Reemplazando los $p_i^{(1)}$ ($i=1, 2$) se tiene que

$$E(\hat{b}^*) - E(\hat{b}) \geq \frac{1}{2} \ln \left[\frac{\theta_c + \frac{1}{2}\theta_{12}}{\theta_c + \theta_{12}} \right]$$

Luego, se obtiene $E(\hat{b})$ a partir de la expresión (4.4.2.2) y reemplazándolo en esta última desigualdad se llega a

$$E(\hat{b}^*) - b \geq \frac{1}{2} \ln \left[\frac{\theta_c + \frac{1}{2}\theta_{12}}{\theta_c + \theta_{12}} \right] + \frac{1}{2} \ln \left[\frac{2\theta_c + \theta_{12}}{\theta_c + \theta_{12}} \right] \quad (4.4.2.3)$$

Por último, reemplazando los valores de los parámetros para los distintos escenarios, se obtiene el sesgo en cada una de las situaciones consideradas en el estudio de Monte Carlo. La Tabla 4.4.3 muestra comparativamente los valores del SR obtenidos a través de la expresión (4.4.2.3) y los valores estimados por simulación para cada grado de correlación considerado en el caso de censuras Tipo I. También se puede observar que a medida que aumenta la correlación el valor estimado se aleja más de la cota superior; lo que se puede explicar por el hecho que cuando la correlación se aproxima a uno el efecto por ordenamiento es casi nulo, ya que los valores de X_1 y X_2 tienden a coincidir en cada individuo.

Tabla 4.4.3: Comparación de la cota del SR con los valores estimados en el estudio de Monte Carlo en modelos condicionales en la situación de eventos no ordenados o diferentes

Correlación	Cota para el SR	Valor estimado del SR para el modelo AG	Valor estimado del SR para el modelo OK
$\rho = 0.13$	$\leq -12\%$	-16%	-16%
$\rho = 0.34$	$\leq -5\%$	-13%	-13%
$\rho = 0.60$	$\leq 3\%$	-7%	-6%
$\rho = 0.95$	$\leq 13\%$	4%	4%

Modelos Marginales

Como se analizó al comienzo de esta sección, el modelo WLW no genera sesgos en esta situación. Sin embargo, basado en los mismos argumentos, el empleo del modelo PC tampoco debería generar sesgos en la situación de eventos diferentes; pero, de acuerdo a las recomendaciones dadas por Therneau, para usar el modelo PC se deben ordenar los datos de una manera similar a la requerida para los modelos condicionales (ver Tabla 4.2.2). Por lo tanto, el sesgo que se introduce en el modelo PC es el mismo que el generado por ordenamiento en los modelos condicionales, es decir,

$$E(\hat{b}_{PC}) - b = \frac{1}{2} \ln \left[\frac{2\theta_c + \theta_{12}}{\theta_c + \theta_{12}} \right] \quad (4.4.2.4)$$

A través de esta expresión y con los valores de los parámetros utilizados en el estudio de Monte Carlo se obtienen los valores del SR para cada uno de los diferentes escenarios considerados. La Tabla 4.4.4 muestra comparativamente los valores calculados con la expresión (4.4.2.4) y los estimados por simulación para el caso de censuras Tipo I.

Tabla 4.4.4: Comparación de la cota del SR con los valores estimados en el estudio de Monte Carlo en modelos condicionales en la situación de eventos ordenados o recurrentes

Correlación	Cota para el SR	Valor estimado del SR para el modelo PC
$\rho = 0.13$	-15 %	-10%
$\rho = 0.34$	-11 %	- 8%
$\rho = 0.60$	- 7 %	- 2%
$\rho = 0.95$	- 2 %	-3%

5. Conclusiones

Antes de analizar el efecto de la correlación en los estimadores del parámetro del modelo de regresión en las diferentes situaciones consideradas, se hace necesario hacer algunas observaciones sobre los sesgos adicionales que se introducen en el modelo, aún en el caso de tiempos de fallas independientes.

Para comenzar, se deben distinguir dos casos, uno que corresponde al modelo WLW en la situación de eventos recurrentes y la otro a los modelos AG, OK y PC en la situación de eventos diferentes. El primer caso surge de la consideración de los individuos bajo el riesgo de la segunda falla, y responde a la estructura del modelo; mientras que en el segundo caso, el sesgo adicional se genera por un problema computacional que surge por el requerimiento del ordenamiento de los datos para el uso del software.

La consideración de los tiempos de falla totales en el caso del modelo WLW, que conduce a considerar a los sujetos, bajo el riesgo de la segunda falla antes que ellos hayan experimentado la primera, ha sido cuestionada por muchos autores, entre ellos Oakes (1992), Pepe y Cai (1993). Oakes (1997) en su comentario sobre este problema afirma que *“puede haber alguna pérdida de eficiencia estadística (poder de los tests o precisión de los estimadores) si se ignora la información que está disponible -por ejemplo que ciertos individuos están o no definitivamente bajo el riesgo de un evento. Sorpresivamente, el grado de esta pérdida de eficiencia no parece haber sido estudiado”*. Efectivamente, esta consideración produce un SR mayor del 17%, si los tiempos de falla intereventos son independientes. Sin embargo este sesgo se hace despreciable si hay una fuerte correlación entre los tiempos de falla, lo que conduce, en la situación de eventos recurrentes, a obtener mejores estimadores si hay una fuerte correlación entre los tiempos de falla. Esto explica los resultados del estudio de Monte Carlo, obtenidos por los autores de esta propuesta, sobre la eficiencia de los estimadores.

Sin embargo, para la situación de eventos recurrentes si en el modelo WLW se usan los tiempos intereventos en lugar de los tiempos totales, entonces, los

sesgos serían equivalentes al del modelo PC, como se puede ver a través de la formulación de dichos modelos y del estudio para obtener las cotas.

Del análisis de los resultados de la evaluación del sesgo, se deduce que los sesgos por ordenamiento del modelo PC, en la situación de eventos diferentes, se pueden evitar si se considera como T_2 (tiempo de falla del segundo evento) a $X_1 + X_2$. Esto corresponde a considerar al proceso no como un proceso semi-Marcov, según lo sugiere Therneau, sino como un proceso Marcov en esta situación, que es la otra propuesta de Prentice y Cai. De esta manera, los modelos PC y WLW serían equivalentes en esta situación.

Por otra parte, el efecto del ordenamiento en los modelos condicionales, como se puede observar en los valores de la tasa de cambio del SR (Tabla 4.2.8 y Fig. 4.2.15), es similar al de los modelos PC si la correlación es menor de 0.6. Sin embargo, y como se infiere a partir de la expresión del sesgo (ecuación 4.4.2.3) para correlaciones superiores a 0.6, el valor del sesgo generado por usar el "hazard" condicional en la ecuación de verosimilitud parcial, empieza a ser importante frente al sesgo por ordenamiento, si el parámetro θ_c es pequeño; lo que se advierte claramente en la Fig. 4.2.15.

Por otra parte, un análisis global del efecto de la correlación entre los tiempos de falla, permite afirmar que en general los cuatro modelos estudiados son más robustos a la falta de cumplimiento del supuesto de independencia en el caso de eventos no ordenados que en el de eventos ordenados. Sin embargo, en ambos casos, la variabilidad de los estimadores es pequeña ya que se consideró tamaños de muestras grandes a fin de, en lo posible, aislar el efecto de la correlación entre los tiempos de falla.

Ahora, debido a las diferencias de comportamiento de los modelos para las diferentes situaciones, se discute por separado cada una de ellas.

A. En la situación de *eventos recurrentes*:

La tasa de cambio del SR, con respecto al sesgo bajo independencia (Fig. 4.2.14), muestra que no hay grandes diferencias entre los modelos con respecto a la

velocidad de crecimiento de la subestimación del parámetro, a excepción del modelo OK que aparece como más sensible a los efectos del aumento de la correlación. Pero, como se verá en el próximo capítulo, este modelo es más sensible a la memoria conjunta del proceso que a la falta de independencia.

Sin embargo, las diferencias se ponen en evidencia si se considera el valor absoluto de los sesgos y, entonces, los modelos AG y PC aparecen como los más recomendables para cualquier grado de correlación. En cambio, si la correlación es muy grande la recomendación sería usar el modelo WLW.

Otro aspecto que se debe tener en cuenta, a la hora de elegir el modelo, es su error estándar y, en esta situación, los modelos condicionales parecen tener ventaja sobre los marginales.

B. En la situación de *eventos diferentes*:

Si se considera tanto el cambio del SR como del ECM con respecto al modelo bajo independencia, el modelo con mejor comportamiento es WLW, aunque con la corrección antes discutida, el modelo PC sería equivalente y se conseguiría además una disminución del SE del estimador.

CAPITULO V

DISTRIBUCIONES CON MEMORIA

1. Introducción

Como ya se mencionó en los capítulos previos, los modelos de regresión tipo Cox para datos de sobrevida multivariados, si bien no requieren supuestos a cerca de la forma paramétrica que debe tener la función "hazard", si lo hacen respecto al proceso estocástico subyacente. Las propiedades de los estimadores se obtienen a partir de la suposición que los datos responden a un proceso Poisson no homogéneo y por lo tanto a un *proceso sin memoria*, de lo cual se deriva que su distribución debe cumplir con algunas propiedades, entre las cuales se encuentran la *continuidad absoluta* y la *falta de memoria*.

En la mayoría de los problemas prácticos, resulta razonable suponer que los tiempos de falla de los eventos no pueden coincidir. Sin embargo, la propiedad falta de memoria (LMP) puede no ser siempre real y, muchas veces, hay fuertes razones para suponer que ella no es válida. Por este motivo, en este capítulo se estudia cómo afecta a la precisión de los estimadores el hecho que la estructura subyacente de los datos no cumple con esta propiedad pero, sí con la continuidad absoluta.

En la Sección 2, se define dos tipos de propiedades de falta de memoria a la luz de las diferentes propuestas de distribuciones bivariadas adecuadas a datos de sobrevida. En la Sección 3, se presentan los modelos probabilísticos de las tres distribuciones bivariadas absolutamente continuas que se consideran, para comparar el comportamiento de distintos modelos de regresión proporcional multivariados, en la situación de eventos recurrentes y eventos diferentes. En la Sección 4, se definen las propiedades de cada distribución, a partir de las cuales se construyen los algoritmos que se usan para generar las muestras aleatorias de tiempos de sobrevida. En la Sección 5, se muestra y analiza los resultados de la aplicación de los modelos de regresión a los datos simulados. En la Sección 6, se obtienen analíticamente, cotas para el sesgo de los estimadores en cada una de las situaciones

planteadas y se compara con los valores obtenidos por simulación. Y, por último, en la Sección 7 se presentan las conclusiones de los resultados alcanzados y se propone algunas estrategias a seguir para la elección del modelo en situaciones prácticas.

2. Propiedad de Falta de Memoria

Marshall y Olkin (1967) derivaron una distribución exponencial bivariada (BVE) a partir del planteo de modelar el riesgo de rotura de dos componentes de una planta fabril. Las suposiciones que hacen sobre el proceso son:

- i) la planta falla si una o ambas componentes falla;
- ii) las fallas para las diferentes componentes ocurren independientemente;
- iii) una vez que una planta “sobrevive” a la falla de una de las componentes, entonces su condición funcional es naturalmente la misma que la de una planta que no ha sufrido ninguna falla;
- iv) los tiempos de falla de cada componente siguen una distribución exponencial.

Estas suposiciones conducen a una distribución que no es absolutamente continua pero tiene la propiedad de falta de memoria.

Ryu (1993) puntualiza que en una distribución bivariada se debe distinguir dos tipos de propiedades de falta de memoria:

- a) La *propiedad de falta de memoria marginal* (MLMP), la cual implica que los tiempos de fallas marginales tienen una distribución exponencial, de lo cual se deriva que el proceso marginal no tiene memoria.
- b) La *propiedad de falta de memoria conjunta* (JLMP), la cual significa que dado que ambas componentes no fallan hasta un tiempo t , la distribución del tiempo de vida conjunta remanente es exactamente la misma que la de una nueva planta. Matemáticamente la propiedad de falta de memoria se define como:

Definición 1.1: Una distribución bivariada se dice que tiene la propiedad de falta de memoria conjunta si satisface la siguiente condición:

$$\bar{F}(s_1 + t, s_2 + t) = \bar{F}(s_1, s_2)\bar{F}(t, t) \quad \text{para } s_1, s_2, t \geq 0$$

$$\text{donde } \bar{F}(x_1, x_2) = P[X_1 > x_1, X_2 > x_2].$$

Block y Basu (1974) demostraron que la única distribución bivariada absolutamente continua y que mantiene las propiedades de falta de memoria marginal y conjunta, es aquella con marginales exponenciales independientes; lo que implica que se debe sacrificar ya sea la MLMP o la JLMP para tener una distribución bivariada absolutamente continua que permita dependencia entre las variables (tiempos de falla).

Motivados por esto Block y Basu derivaron una distribución exponencial bivariada absolutamente continua (ACBVE), pero que no goza de la propiedad de falta de memoria marginal, teniendo marginalmente cada una de las variables una distribución que es un promedio ponderado de dos distribuciones exponenciales. Sin embargo, ACBVE todavía goza de la propiedad de falta de memoria conjunta.

Sarkar (1987) con la misma motivación propone una distribución exponencial bivariada absolutamente continua (ACBVE₂) pero que no goza de la JLMP. En esta distribución, las marginales son exponenciales y como consecuencia satisface la MLMP.

Para derivar otra extensión de la distribución (BVE) de Marshall y Olkin, Ryu toma una situación más general que ellos, permitiendo que la falla de una de las componentes de la planta incremente la tasa de falla de la otra componente; la distribución conjunta exhibe entonces una estructura de envejecimiento, en el sentido que la vida remanente, libre de falla, de la planta es menor que la de otra similar sin falla. En muchas situaciones de la vida real, estas condiciones son más razonables que aquellas supuestas por Marshall y Olkin para derivar su distribución. La extensión de Ryu genera una distribución que es absolutamente continua pero que no satisface la MLMP, ni la JLMP; por esta razón, como los mismos autores lo mencionan, esta distribución no es de tipo exponencial en el sentido estricto.

3. Modelos de las Distribuciones Bivariadas

En esta Sección, se presentan los modelos probabilísticos y las propiedades de las tres distribuciones bivariadas absolutamente continuas antes mencionadas.

3.1. Distribución de Block-Basu (ACBVE)

La distribución ACBVE representa la parte absolutamente continua de la distribución exponencial bivariada de Marshall y Olkin y es tal que, para $(\theta_1, \theta_2, \theta_{12}) \in [0, \infty) \times [0, \infty) \times [0, \infty)$, la función de sobrevida bivariada está dada por

$$S(x_1, x_2) = P[X_1 \geq x_1, X_2 \geq x_2] = \frac{\theta}{\theta_1 + \theta_2} \exp[-\theta_1 x_1 - \theta_2 x_2 - \theta_{12} \max(x_1, x_2)] - \frac{\theta_{12}}{\theta_1 + \theta_2} \exp[-\theta \max(x_1, x_2)] \quad (5.3.1.1)$$

para $x_1, x_2 > 0$

donde $\theta = \theta_1 + \theta_2 + \theta_{12}$.

Sin dificultad se puede observar que esta distribución goza de la propiedad de falta de memoria conjunta, en efecto, puesto que

$$\max(x_1 + t, x_2 + t) = \max(x_1, x_2) + t,$$

entonces

$$S(x_1 + t, x_2 + t) = S(x_1, x_2) S(t, t)$$

Las distribuciones marginales de la distribución ACBVE están dadas por:

$$S(x_1) = P[X_1 \geq x_1] = \frac{\theta}{\theta_1 + \theta_2} \exp[-(\theta_1 + \theta_{12})x_1] - \frac{\theta_{12}}{\theta_1 + \theta_2} \exp[-\theta x_1] \quad \text{para } x_1 > 0$$
(5.3.1.2)

$$S(x_2) = P[X_2 \geq x_2] = \frac{\theta}{\theta_1 + \theta_2} \exp[-(\theta_2 + \theta_{12})x_2] - \frac{\theta_{12}}{\theta_1 + \theta_2} \exp[-\theta x_2] \quad \text{para } x_2 > 0$$

Es decir, ambas son promedios ponderados de exponenciales, y en consecuencia, la distribución ACBVE no cumple con la MLMP.

Otras propiedades de la distribución de Block y Basu son:

i. $\min(X_1, X_2)$ tiene distribución exponencial con parámetro θ , $e(\theta)$.

(5.3.1.3)

ii. $\min(X_1, X_2)$ es independiente de $X_1 - X_2$ y también de $|X_1 - X_2|$.

iii. El coeficiente de correlación está dado por

$$\rho_{BB}(X_1, X_2) = \frac{\theta_{12} [(\theta_1^2 + \theta_2^2)\theta + \theta_1 \theta_2 \theta_{12}]}{[(\theta_1 + \theta_2)^2 (\theta_1 + \theta_{12})^2 + (\theta_2 + \theta_{12})^2] [(\theta_1 + \theta_2)^2 (\theta_2 + \theta_{12})^2 + (\theta_1 + \theta_2)^2 \theta^2]}$$
(5.3.1.4)

De la expresión (5.3.1.4) se sigue que $\rho_{BB} = 0$ si $\theta_{12} = 0$ ó $\theta_1 = \theta_2 = 0$.

Como se expuso en el Capítulo II, Pedroso de Lima y Sen (1998) muestran que el proceso de conteo bivariado $N(t) = (N_1(t), N_2(t))'$ se encuentra asociado al vector de intensidades $\lambda(t) = (\lambda_1(t), \lambda_2(t))'$ cuyas componentes se expresan como:

$$\lambda_1(t) = p_1^{(1)}(t)Y_1(t)[1 - Y_2(t)] + p_2^{(1)}(t)Y_1(t)Y_2(t)$$
(5.3.1.5)

$$\lambda_2(t) = p_2^{(2)}(t)Y_2(t)[1 - Y_1(t)] + p_1^{(2)}(t)Y_1(t)Y_2(t)$$

donde $Y_i(t) = I\{T_i \geq t\}$. Puesto que las $p_j^{(i)}(t)$ son las funciones "hazard" condicionales, ellas dependen de los parámetros de la distribución y, como consecuencia, las componentes de $\lambda(t)$ se pueden expresar en función de los parámetros de la distribución bivariada de T .

En el caso particular de la distribución de Block y Basu ellos encuentran que:

$$p_i^{(i)} = \theta_i + \theta_{12} \quad , \quad \text{con } i=1,2 \quad (5.3.1.6)$$

$$p_j^{(i)} = \theta_i \theta \quad (\theta_1 + \theta_2)^{-1} \quad , \quad \text{con } i \neq j \quad ; \quad i,j=1,2 \quad (5.3.1.7)$$

3.2. Distribución Sarkar (ACBVE₂)

La función de sobrevida bivariada de la distribución derivada por Sarkar (ACBVE₂) toma la forma siguiente:

$$S(x_1, x_2) = \Pr(X_1 \geq x_1; X_2 \geq x_2) = \begin{cases} e^{-(\theta_2 + \theta_{12})x_2} \left\{ 1 - [A(\theta_1 x_2)]^{-\gamma} [A(\theta_1 x_1)]^{1+\gamma} \right\} & \text{para } 0 < x_1 \leq x_2 \\ e^{-(\theta_1 + \theta_{12})x_1} \left\{ 1 - [A(\theta_2 x_1)]^{-\gamma} [A(\theta_2 x_2)]^{1+\gamma} \right\} & \text{para } 0 < x_2 \leq x_1 \end{cases} \quad (5.3.2.1)$$

donde $\theta_1 > 0$, $\theta_2 > 0$, $\theta_{12} \geq 0$, $\gamma = \theta_{12}/(\theta_1 + \theta_2)$ y $A(z) = 1 - e^{-z}$.

Su derivación se basa en modificaciones de las propiedades que caracterizan a las propuestas previas de Marshall y Olkin, y de Block y Basu, requiriendo que ella sea absolutamente continua y que el $\min(X_1, X_2)$ sea independiente de $g(X_1, X_2)$ para algún $g \in \mathcal{L}$ (funciones continuas).

Entonces, estableciendo que el $\min(X_1, X_2)$ es independiente de la función $X_1 - X_2 + k(X_1, X_2)$ para algún $k(X_1, X_2)$; Sarkar demostró que (ACBVE₂) goza de las siguientes propiedades:

- i. X_1 e X_2 son marginalmente exponenciales
- ii. El $\min(X_1, X_2)$ tiene distribución exponencial con parámetro θ , siendo $\theta = \theta_1 + \theta_2 + \theta_{12}$. (5.3.2.2)

iii. El coeficiente de correlación entre X_1 e X_2 está dado por

$$\rho_S(X_1, X_2) = \frac{\theta_{12}}{\theta} + \frac{\theta_{12}}{\theta} \sum_{j=1}^{\infty} \frac{j!}{\theta + (\theta_1 + \theta_2)j} \left\{ \theta_1^j \prod_{k=1}^j (\theta + k\theta_1)^{-1} + \theta_2^j \prod_{k=1}^j (\theta + k\theta_2)^{-1} \right\} \quad (5.3.2.3)$$

De la expresión (5.3.2.3) se sigue que $\rho_S = 0$ si y sólo si $\theta_{12} = 0$; más aún, en este caso las variables son independientes.

Además, por la propiedad i) es claro que la distribución Sarkar cumple con la MLMP. Por otra parte, de la expresión (5.3.2.1) y del hecho que las variables no son independientes, se sigue la distribución ACBVE₂ no cumple la JMLP.

Pedroso de Lima y Sen muestran que las componentes del vector intensidad $\lambda(t)$ se pueden expresar en función de los parámetros de (ACBVE₂), a través de las los "hazard" condicionales dados por

$$p_1^{(1)} = \theta_1 (1 + \gamma) + \gamma \theta_2 [A(\theta_2 t)]^{-1} \quad (5.3.2.4)$$

$$p_2^{(2)} = \theta_2 (1 + \gamma) + \gamma \theta_1 [A(\theta_1 t)]^{-1}$$

$$p_2^{(1)} = \theta_1 (1 + \gamma) \quad (5.3.2.5)$$

$$p_1^{(2)} = \theta_2 (1 + \gamma)$$

3.3. Distribución Ryu

La distribución propuesta por Ryu surge a partir de modificaciones en las condiciones de falla de las componentes del sistema, permitiendo la existencia de una componente común cuya falla produce una disminución de la probabilidad de sobrevida a la falla del sistema. Para esto, modela el entorno físico de un individuo expuesto al riesgo de dos eventos (fallas) a través de tres procesos Poisson independientes $\{N_1(t), t \geq 0\}$, $\{N_2(t), t \geq 0\}$ y $\{N_{12}(t), t \geq 0\}$, con intensidades λ_1 , λ_2 y λ_{12} respectivamente, donde N_i ($i=1,2$) gobierna la tasa de falla de cada

componente, mientras que N_{12} determina la chance de falla de la componente común que causa un daño potencial a ambas componentes.

Para derivar la distribución propuesta, Ryu supone que la tasa "hazard" al tiempo t para la falla de la i -ésima componente es $d_i N_i(t) + s_i N_{12}(t)$, $i=1,2$, dando una realización del proceso estocástico bivariado. Dicha tasa "hazard" especificada como un proceso estocástico, recibe el nombre de *entorno doblemente estocástico o función "hazard" aleatoria*. Además, se supone que los d_i son infinitos, dando como resultado que la función de distribución sea absolutamente continua y que cualquier realización del proceso Poisson N_i produce una falla en la i -ésima componente. Pero, una realización de N_{12} sólo incrementa la tasa "hazard" para la falla de la i -ésima componente por una cantidad s_i , $i=1,2$, actuando de una manera acumulada. Condicional a la realización de este proceso, los tiempos de falla de cada componente se suponen independientes. Por lo tanto, lo que conduce a la dependencia es la existencia del entorno común especificado por N_{12} .

Si Y_i es el tiempo hasta la realización del proceso N_i , y Z_i es la variable para la cual el "hazard" condicional al tiempo t es $s_i N_{12}(t)$, dada una realización de N_{12} ; entonces X_i , el tiempo de falla de la i -ésima componente, será el $\min(Y_i, Z_i)$, es decir, que a través del proceso estocástico común, N_{12} , se establece la dependencia entre los dos tiempos de falla, X_1 y X_2 .

La función de sobrevida bivariada en este caso es:

$$S(x_1, x_2) = \begin{cases} \exp\left\{-\left(\theta_1 + \theta_{12}\right)x_1 - \theta_2 x_2 + \frac{\theta_{12}}{s_1} \left[1 - e^{-s_1(x_1 - x_2)}\right] + \frac{\theta_{12}}{s_1 + s_2} \left[e^{-s_1(x_1 - x_2)} - e^{-s_1 x_1 - s_2 x_2}\right]\right\} \\ \quad \text{para } x_1 > x_2 \\ \exp\left\{-\left(\theta_2 + \theta_{12}\right)x_2 - \theta_1 x_1 + \frac{\theta_{12}}{s_2} \left[1 - e^{-s_2(x_2 - x_1)}\right] + \frac{\theta_{12}}{s_1 + s_2} \left[e^{-s_2(x_2 - x_1)} - e^{-s_1 x_1 - s_2 x_2}\right]\right\} \\ \quad \text{para } x_1 \leq x_2 \end{cases} \quad (5.3.3.1)$$

Cabe destacar que si s_1 y $s_2 \rightarrow \infty$ esta función de sobrevida bivariada se reduce a la propuesta por Marshall y Olkin.

Las distribuciones marginales correspondientes toman la forma:

$$S(x_1) = P[X_1 \geq x_1] = \exp\left[-(\theta_1 + \theta_{12})x_1 - \frac{\theta_{12}}{s_1}(1 - e^{-s_1 x_1})\right] \quad \text{para } x_1 > 0 \quad (5.3.3.2)$$

$$S(x_2) = P[X_2 \geq x_2] = \exp\left[-(\theta_2 + \theta_{12})x_2 - \frac{\theta_{12}}{s_2}(1 - e^{-s_2 x_2})\right] \quad \text{para } x_2 > 0$$

A partir de la expresión (5.3.3.2) se puede observar fácilmente que la distribución de Ryu no cumple con la MLMP, y más aún, sus marginales tienen una tasa de falla creciente dadas por:

$$h_i(x_i) = \theta_i + \theta_{12}(1 - e^{-s_i x_i}) \quad \text{para } i=1,2 \text{ e } x_i > 0 \quad (5.3.3.3)$$

Además, a partir de la expresión (5.3.3.1) se puede demostrar que:

$$\frac{S(x_1, x_2)S(t, t)}{S(x_1 + t, x_2 + t)} = \exp\left\{\frac{\theta_{12}}{s_1 + s_2} [1 - e^{-(s_1 + s_2)t}] [1 - e^{-s_1 x_1 - s_2 x_2}]\right\} \geq 1 \quad (5.3.3.4)$$

lo cual implica que

$$\frac{S(y_1 + t, y_2 + t)}{S(t, t)} \leq S(y_1, y_2) \quad \forall y_1, y_2, t \geq 0 \quad (5.3.3.5)$$

Siendo así, la distribución conjunta muestra un envejecimiento en el sentido que la probabilidad de sobrevida conjunta para el tiempo de vida remanente es menor o igual que la probabilidad de sobrevida de un nuevo sistema, dado que no se producen ninguna falla hasta un cierto tiempo t .

Por otra parte, se puede demostrar que en esta distribución el $\min(X_1, X_2)$ no es independiente de $X_1 - X_2$, a menos que $\theta_{12} = 0$ ó $s_1 + s_2 = \infty$; es decir, que no exista la componente común o que la falla de la componente común produzca una falla en al menos una de las componentes ($s_1 = \infty$ ó $s_2 = \infty$).

Para el caso de la distribución de Ryu, obtener las expresiones de las funciones "hazard" condicionales requiere la integración de la función de densidad bivariada, que, dada su complejidad, sólo se puede llevar a cabo usando un

desarrollo en serie, y por lo tanto se llega a una formulación aproximada de las mismas.

4. Algoritmos Computacionales

En esta sección se definen algunas propiedades de las distribuciones descritas en la sección anterior, sobre la base de las cuales es posible generar las muestras aleatorias, y además el algoritmo que se usa en cada caso.

⇒ *Distribución de Block y Basu*

Si $Z = X_1 - X_2$, Block y Basu muestran que la distribución de Z está dada por

$$F(z) = \begin{cases} \frac{\theta_1}{\theta_1 + \theta_2} \exp[(\theta_2 + \theta_{12})z] & \text{para } Z \leq 0 \\ 1 - \frac{\theta_2}{\theta_1 + \theta_2} \exp[-(\theta_1 + \theta_{12})z] & \text{para } Z > 0 \end{cases}$$

Por simplicidad, aquí se considera $\theta_1 = \theta_2$, en tal caso, la distribución de Z toma la forma

$$F(z) = \begin{cases} \frac{1}{2} \exp[(\theta_1 + \theta_{12})z] & \text{para } Z \leq 0 \\ 1 - \frac{1}{2} \exp[-(\theta_1 + \theta_{12})z] & \text{para } Z > 0 \end{cases} \quad (5.4.1)$$

Por las propiedades (5.3.1.3) se cumple que $W = \min(X_1, X_2)$ se distribuye como $e(\theta)$ y es independiente de Z , de lo cual se sigue que:

- ◆ Si $Z \leq 0 \Rightarrow X_1 = W \Rightarrow X_2 = X_1 - Z$
- ◆ Si $Z > 0 \Rightarrow X_2 = W \Rightarrow X_1 = X_2 + Z$.

A partir de estas expresiones y usando la distribución de Z dada en (5.4.1) se generan las muestras aleatorias de (X_1, X_2) con distribución ACBVE empleando el siguiente algoritmo:

- 1° Generar la variable $V \sim U_{[0,1]}$
- 2° Generar la variable Z , usando la transformada integral de (5.4.1), o sea

$$Z = \frac{\ln V}{\theta_1 + \theta_{12}}$$

- 3° Generar la variable $W \sim e(2\theta_1 + \theta_{12})$
- 4° Si $Z \leq 0$ hacer $X_1 = W$ e $X_2 = X_1 - Z$ en caso contrario hacer $X_2 = W$ e $X_1 = X_2 + Z$.

⇒ *Distribución de Sarkar*

Sarkar demuestra el siguiente teorema que permite generar muestras de (X_1, X_2) con distribución ACBVE₂.

Teorema: Sean Z_1, Z_2, Z_3 y V variables aleatorias independientes con distribuciones $e(\theta_1), e(\theta_2), e(\theta_{12})$ y $U_{[0,1]}$ respectivamente, se define

$$\begin{aligned} X_1 &= W_1 + \{1 - I(Z_1 > Z_2)\} \left\{ \frac{1}{\theta_1} A^{-1} \left[V^{1/(1+\gamma)} A(\theta_1 W_2) \right] - W_1 \right\} \\ X_2 &= W_2 + \{I(Z_1 > Z_2)\} \left\{ \frac{1}{\theta_2} A^{-1} \left[V^{1/(1+\gamma)} A(\theta_2 W_1) \right] - W_2 \right\} \end{aligned} \quad (5.4.1)$$

donde $W_1 = \min(Z_1, Z_3)$ y $W_2 = \min(Z_2, Z_3)$, entonces (X_1, X_2) tienen distribución ACBVE₂ $(\theta_1, \theta_2, \theta_{12})$.

El algoritmo para generar muestras de la distribución de Sarkar queda definido entonces de la siguiente manera:

- 1° Generar la variable aleatoria $Z_1 \sim e(\theta_1)$
- 2° Generar la variable aleatoria $Z_2 \sim e(\theta_2)$

- 3° Generar la variable aleatoria $Z_3 \sim e(\theta_{12})$
- 4° Generar la variable aleatoria $V \sim U_{[0,1]}$
- 5° Generar la variable aleatoria $W_1 = \min(Z_1, Z_3)$
- 6° Generar la variable aleatoria $W_2 = \min(Z_2, Z_3)$
- 7° Generar la variable aleatoria $I = 1$ si $Z_1 > Z_2$ en caso contrario, $I = 0$
- 8° Calcular $\gamma = \theta_{12}/(\theta_1 + \theta_2)$
- 9° Generar la variable aleatoria $S = V^{1/(1+\gamma)}$
- 10° Generar la variable aleatoria $A_1 = 1 - \exp(-\theta_1 \cdot W_2)$
- 11° Generar la variable aleatoria $A_2 = 1 - \exp(-\theta_2 \cdot W_1)$
- 12° Generar la variable aleatoria $A_{11} = 1/\theta_1 \cdot \ln[1 - S \cdot A_1]$
- 13° Generar la variable aleatoria $A_{22} = 1/\theta_2 \cdot \ln[1 - S \cdot A_2]$
- 14° Generar las variables aleatorias X_1 y X_2 , a partir de las expresiones (5.4.1), es decir,

$$X_1 = W_1 + (1 - I) \cdot (A_{11} - W_1)$$

$$X_2 = W_2 + I \cdot (A_{22} - W_2)$$

⇒ Distribución de Ryu

Como ya se dijo en la Sección 2.3, Ryu derivó su modelo de distribución a partir de tres procesos Poisson independientes, para ello, define la variable Y_i como el tiempo hasta la realización del proceso $\{N_i(t)\}$ con intensidad θ_i , $i=1,2$; y, la variable Z_i para la cual el "hazard" al tiempo t es $s_i N_{12}(t)$, dada una realización del proceso $\{N_{12}(t)\}$ con intensidad θ_{12} . El autor demuestra que si se producen k realizaciones de N_{12} hasta t en los tiempos τ_1, \dots, τ_k respectivamente, entonces

$$E \left\{ \exp \left[-s_i \int_0^t N_{12}(u) du \right] \right\} = \exp \left[-\theta_{12} t + \frac{\theta_{12}}{s_i} (1 - e^{-s_i t}) \right]$$

Este resultado se obtiene considerando que $t - \tau_1, \dots, t - \tau_k$ son variables aleatorias independientes con distribución $U_{[0,1]}$, si ellas no están ordenadas.

Para obtener muestras aleatorias de la distribución se parte de: w_1, w_2, \dots realizaciones de una variable aleatoria exponencial con parámetro θ_{12} ; luego, se define $\tau_0 = 0$ y $\tau_j = w_1 + \dots + w_j$, con $j=1,2,\dots$. Obtenidos los τ_j , el incremento del "hazard" de la i -ésima componente del sistema al tiempo t , está dado por js_i , para $t \in [\tau_j, \tau_{j+1})$. De esta manera, para conseguir una realización de Z_i se necesita obtener un valor de la distribución que tiene dicho "hazard". En efecto, la distribución de Z_i es una mezcla de variables aleatorias $(\tau_j + v_j)$ con $j=1,2,\dots$, donde v_j es una variable aleatoria exponencial con parámetro js_i y truncada en w_{j+1} , donde las probabilidades de mezcla están dadas por: $P_1 = 1 - \exp(-s_1 w_2)$ para $j = 1$, y para $j=2,3,\dots$ por $P_j = \exp[-(s_1 w_2 + 2s_1 w_3 + (j-1)s_1 w_j)][1 - \exp(-js_i w_{j+1})]$. Entonces, dados w_1, w_2, \dots ; Z_1 y Z_2 son generadas independientemente y en ese caso $X_i = \min(Y_i, Z_i)$.

Basado en estos resultados el algoritmo para generar las muestras aleatorias con distribución Ryu resulta ser el siguiente:

- 1° Generar la variable aleatoria $Y_1 \sim e(\theta_1)$
- 2° Generar la variable aleatoria $Y_2 \sim e(\theta_2)$
- 3° Hacer $j = 1$
- 4° Hacer $\tau_1 = 0$
- 5° Hacer $Z_1 = 0$
- 6° Hacer $Z_2 = 0$
- 7° Si $(Z_1 \neq 0$ ó $Z_2 \neq 0)$ ir a 15°
- 8° Generar la variable aleatoria $W \sim e(\theta_{12})$
- 9° Hacer $\tau_2 = \tau_1 + W$
- 10° Generar la variable aleatoria $V_1 \sim e(s_1 \cdot j)$
- 11° Generar la variable aleatoria $V_2 \sim e(s_2 \cdot j)$
- 12° Si $(V_1 \geq \tau_1$ y $V_1 < \tau_2)$, hacer $Z_1 = V_1$
- 13° Si $(V_2 \geq \tau_1$ y $V_2 < \tau_2)$, hacer $Z_2 = V_2$
- 14° Hacer $\tau_1 = \tau_2$, $j = j+1$ y retornar al 7°
- 15° Hacer $X_1 = \min(Z_1, Y_1)$
- 16° Hacer $X_2 = \min(Z_2, Y_2)$.

La Figura 5.4.5 muestra la representación gráfica de la distribución Block-Basu y las Figuras 5.4.1 a 5.4.4 y 5.4.6 a 5.4.10 presentan histogramas de muestras de tamaño 500 de cada una de las distribuciones con parámetros como los dados en la Tabla 5.5.1 para el grupo control, generadas a partir de los algoritmos antes descritos, y de las respectivas marginales.

Para cada una de las muestras se aplica el test de Komogorv-Simirov bajo la hipótesis nula que las marginales tienen distribución exponencial. Para el caso de la distribución de Block - Basu y de la distribución de Ryu se rechaza la hipótesis nula al nivel del 5%; sin embargo, no hubo suficiente evidencia para rechazar H_0 en el caso de la distribución Sarkar.

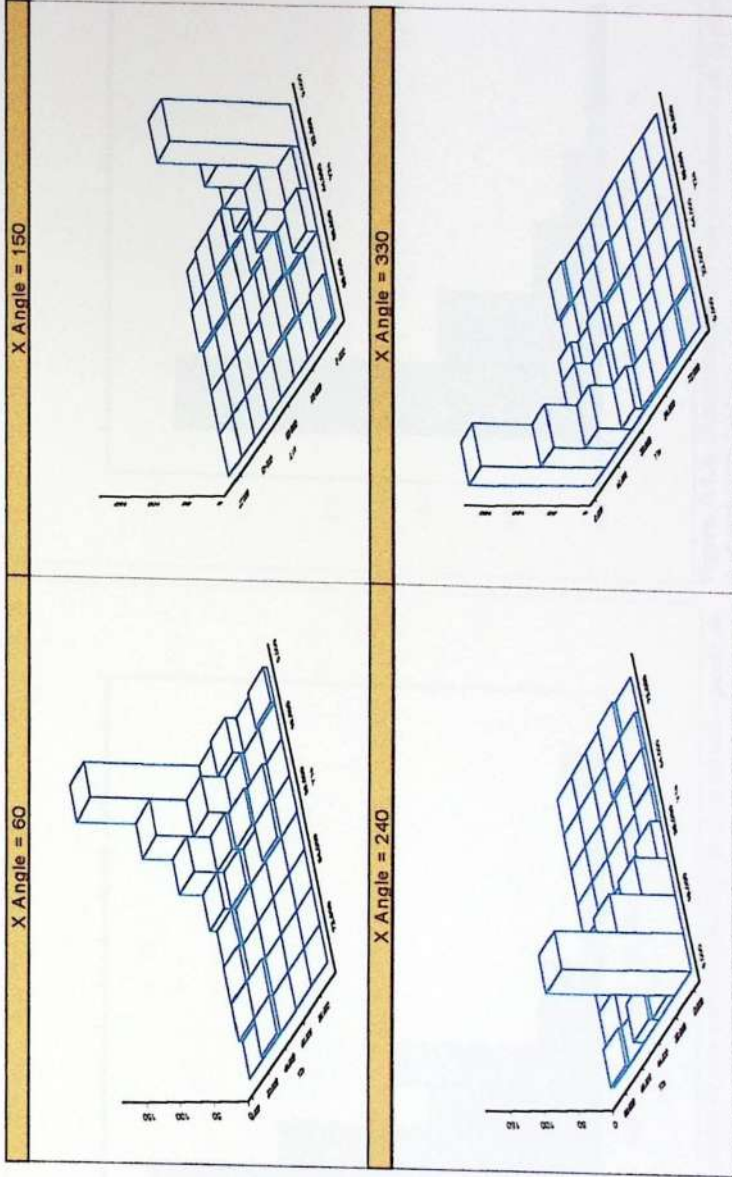


Figura. 5.4.1: Histogramas de una muestra aleatoria de (Y_1, Y_2) con distribución Sarkar.

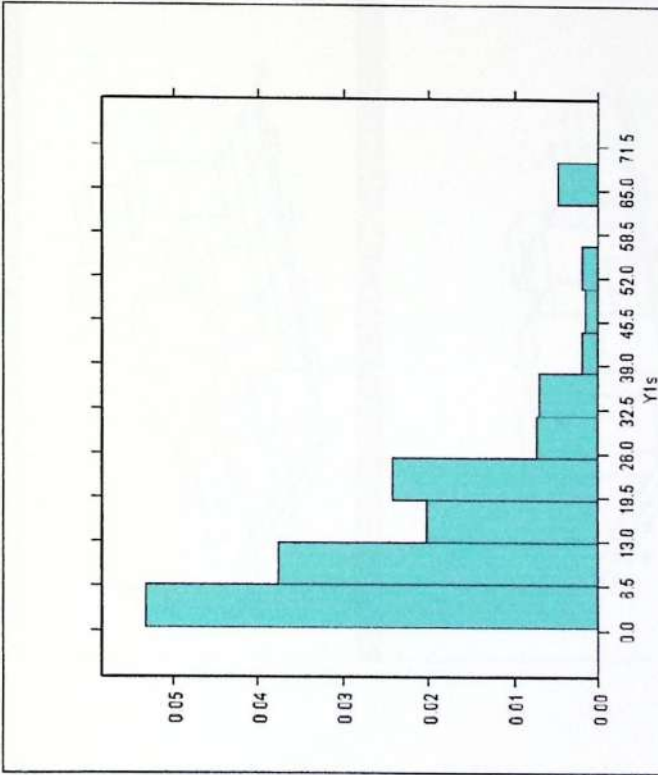


Figura. 5.4.2: Histograma de una muestra aleatoria de Y_1 generada a partir de la distribución Sarkar.

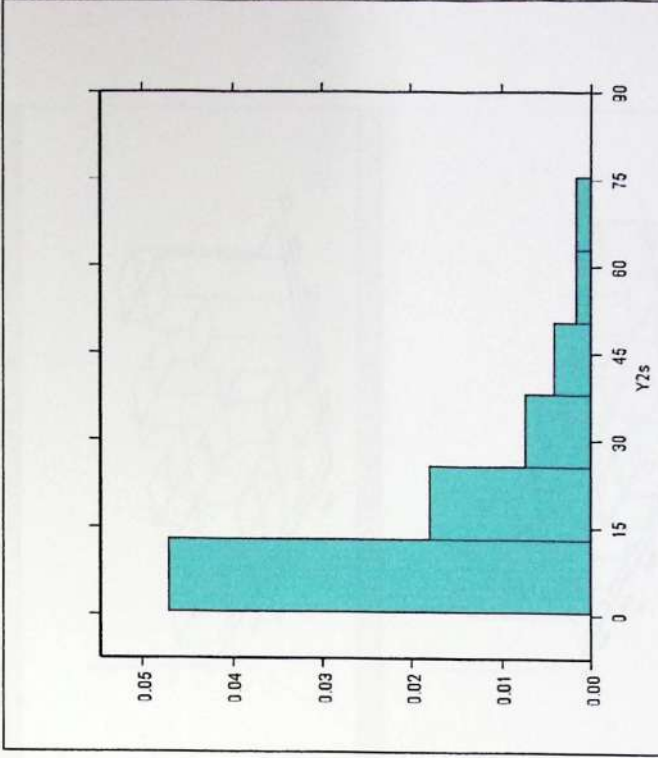


Figura. 5.4.3: Histograma de una muestra aleatoria de Y_2 generada a partir de la distribución Sarkar.

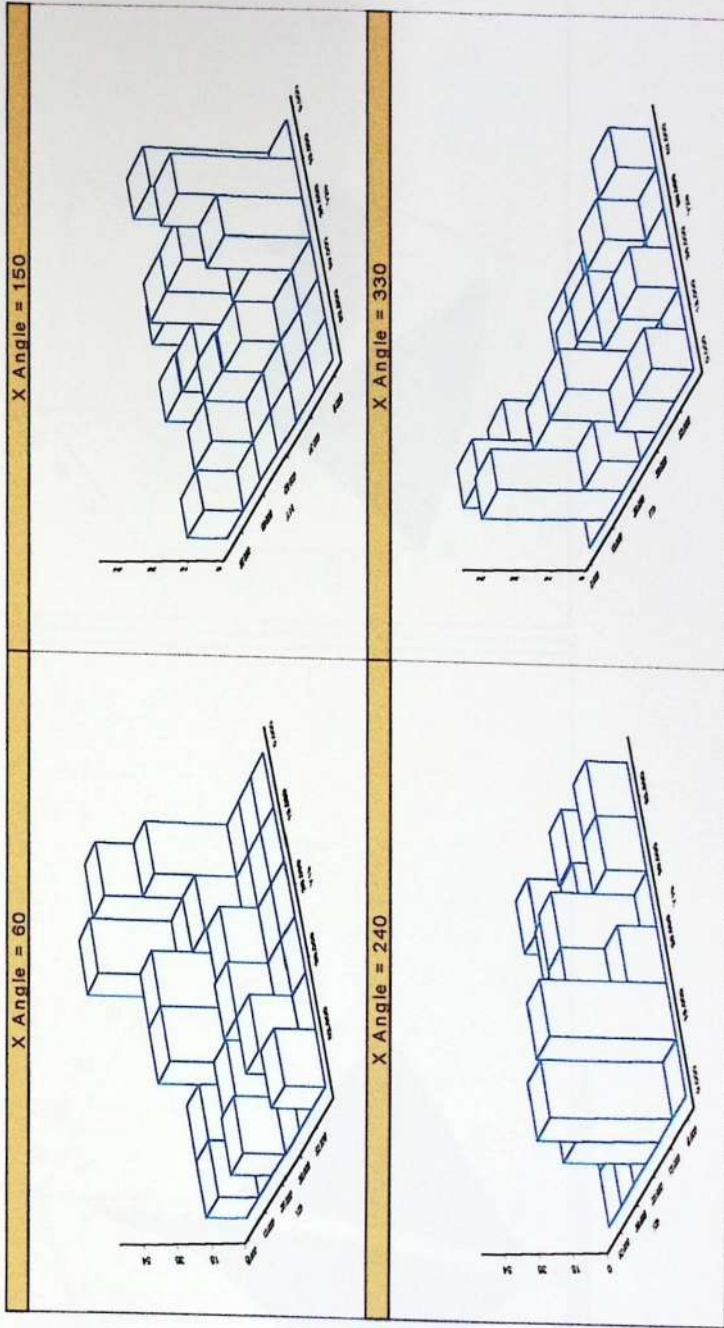


Figura 5.4.4: Histogramas de una muestra aleatoria de (Y_1, Y_2) con distribución Block-Basu.

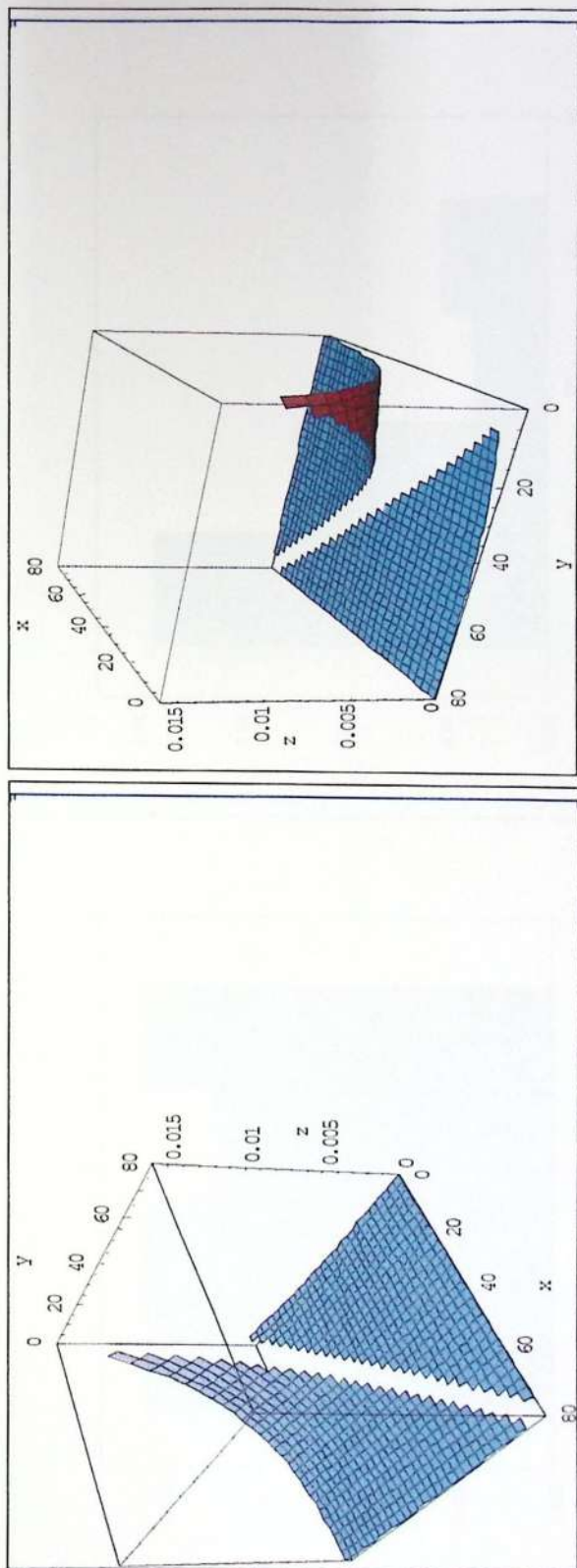


Figura. 5.4.5: Distribución Block-Basu

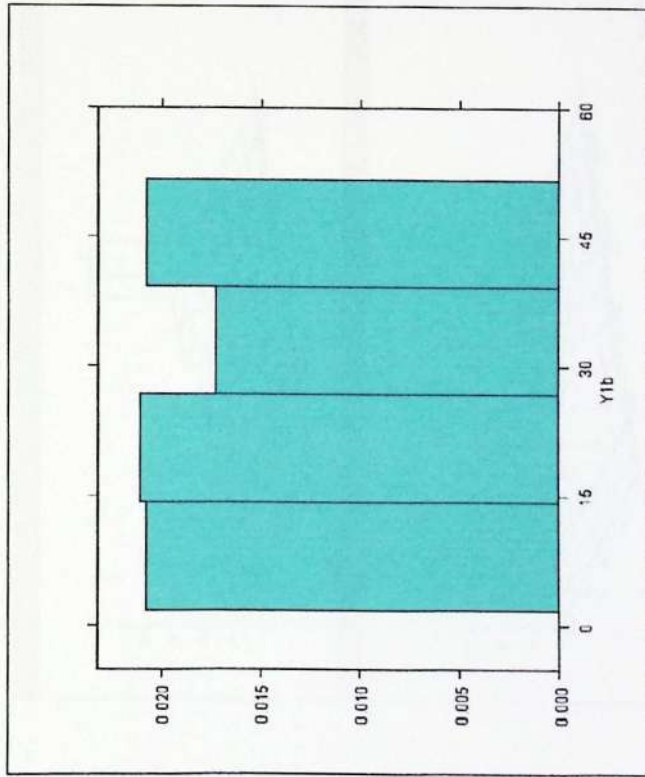


Figura. 5.4.6: Histograma de una muestra aleatoria de Y_1 generada a partir de la distribución Block-Basu.

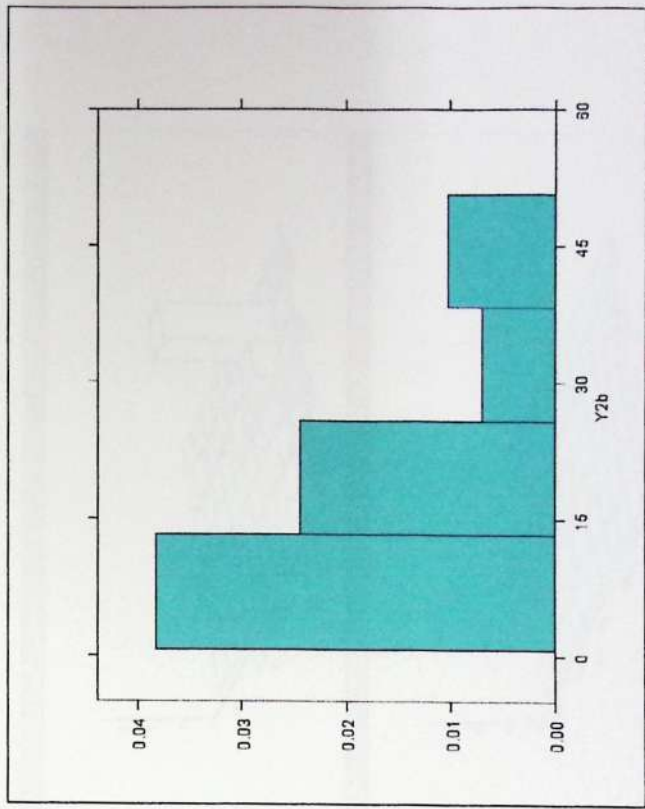


Figura. 5.4.7: Histograma de una muestra aleatoria de Y_2 generada a partir de la distribución Block-Basu.

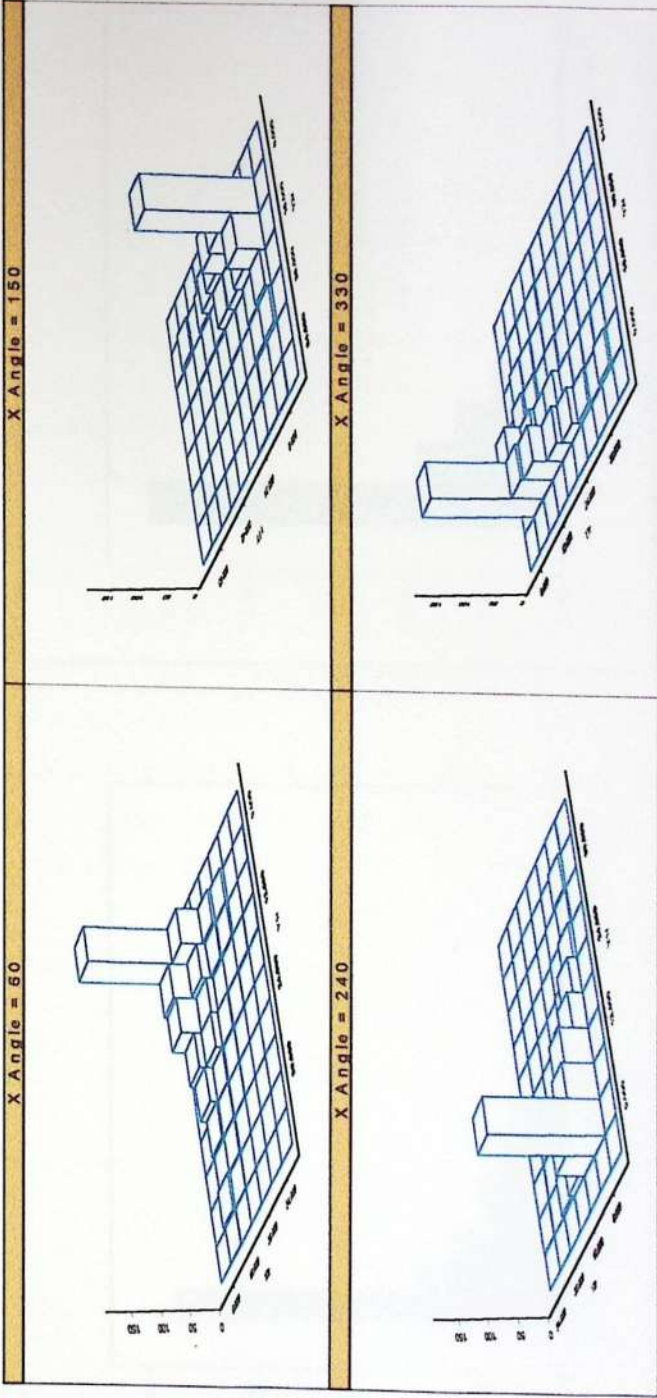


Figura. 5.4.8: Histogramas de una muestra aleatoria de (Y_1, Y_2) con distribución Ryy.

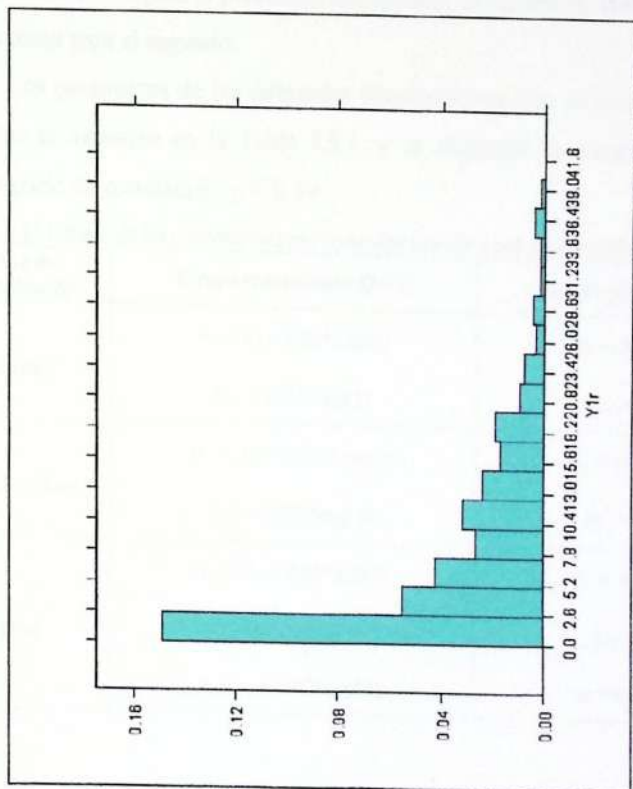


Figura. 5.4.9: Histograma de una muestra aleatoria de Y_1 , generada a partir de la distribución Ryt.

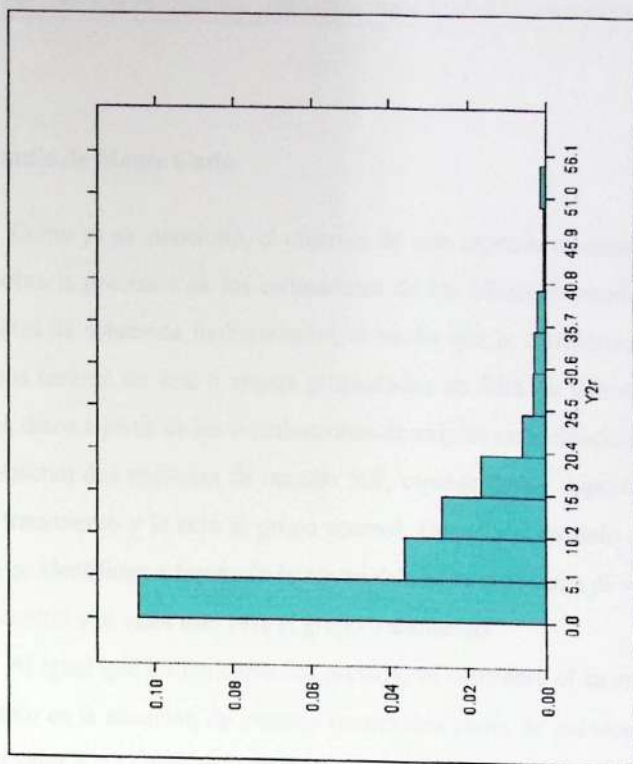


Figura. 5.4.10: Histograma de una muestra aleatoria de Y_2 , generada a partir de la distribución Ryt.

5. Estudio de Monte Carlo

Como ya se mencionó, el objetivo de este capítulo es estudiar el efecto que tiene sobre la precisión de los estimadores de los diferentes modelos de regresión para datos de sobrevida multivariados, el hecho que la distribución subyacente de los datos carezca de una o ambas propiedades de falta de memoria. Para ello se simulan datos a partir de las distribuciones descritas en la Sección 3. En cada caso se consideran dos muestras de tamaño 500, representando hipotéticamente una al grupo tratamiento y la otra al grupo control. Dentro del modelo de regresión, los grupos se identifican a través de la covariable Z , la cual toma el valor cero para el grupo control y el valor uno para el grupo tratamiento.

Al igual que en los capítulos previos, se considera el caso de dos eventos, $k=2$, tanto en la situación de eventos recurrentes como de eventos diferentes, y, se toma el valor $b = -2$ para el parámetro del modelo de regresión, tanto para el primer evento como para el segundo.

Los parámetros de las diferentes distribuciones que se usaron para generar los datos se muestran en la Tabla 5.5.1, y se eligieron de manera de obtener el mismo grado de correlación, $\rho = 0.19$.

Tabla 5.5.1: Valores de los parámetros para cada distribución y para cada grupo.

Tipo de Distribución	Grupo tratamiento ($Z=1$)	Grupo placebo ($Z=0$)
Sarkar	$\theta_1 = \theta_2 = 0.04 * \exp(2)$ $\theta_{12} = 0.02 * \exp(2)$	$\theta_1 = \theta_2 = 0.04$ $\theta_{12} = 0.02$
Block y Basu	$\theta_1 = \theta_2 = 0.015 * \exp(2)$ $\theta_{12} = 0.035 * \exp(2)$	$\theta_1 = \theta_2 = 0.015$ $\theta_{12} = 0.035$
Ryu	$\theta_1 = \theta_2 = 0.05 * \exp(2)$ $\theta_{12} = 0.01 * \exp(2)$ $s_1 = s_2 = 0.05 * \exp(2)$	$\theta_1 = \theta_2 = 0.05$ $\theta_{12} = 0.01$ $s_1 = s_2 = 0.05$

En todas las situaciones se usa censuras aleatorias uniformes, es decir, aquellas donde el proceso de censura es $C_i(t) = I(t \leq U_i)$ y donde U_i es una variable aleatoria con distribución uniforme y con un 10% de observaciones censuradas. En el caso de eventos ordenados, si un sujeto tiene el primer evento censurado, entonces tiene también censurado el segundo evento.

Para comparar la precisión de los estimadores de los diferentes modelos se estima en cada situación el *Error Cuadrático Medio* (ECM), el Error Estándar (SE) y el *Sesgo Relativo* (SR) del estimador \hat{b} , el cual está dado por

$$SR = \frac{\left(\hat{E}(\hat{b}) - b \right) * 100}{b} \quad (5.5.1)$$

Los valores del ECM obtenidos a partir de las 100 simulaciones para cada situación se muestran en la Tabla 5.5.2 y Fig.5.5.1 y 5.5.2, los correspondientes valores del SR y SE se presentan en la Tabla 5.5.3 y 5.5.4 respectivamente y se representan gráficamente en las Figuras 5.5.3 a 5.5.6.

Tabla 5.5.2. Error cuadrático Medio (ECM) de los estimadores del parámetro $b = -2$ según las diferentes situaciones y según el tipo de distribución subyacente de los datos

Tipo de Distribución	Error Cuadrático Medio								
	Eventos Ordenados			Eventos no Ordenados					
	AG	OK	PC	WLW	AG	OK	PC	WLW	
Sarkar	0.016	0.034	0.004	0.146	0.095	0.092	0.025	0.002	
Block y Basu	0.002	0.005	0.205	0.638	0.001	0.002	0.300	0.134	
Ryu	0.273	0.341	0.221	0.052	0.541	0.542	0.352	0.254	

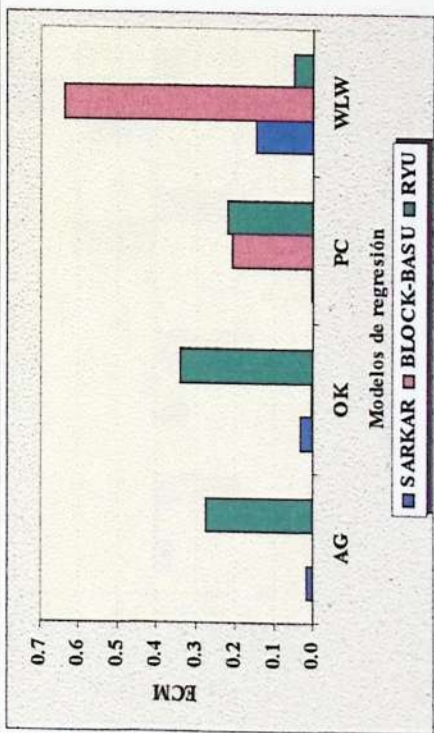


Figura. 5.5.1: Error cuadrático medio para eventos ordenados o recurrentes

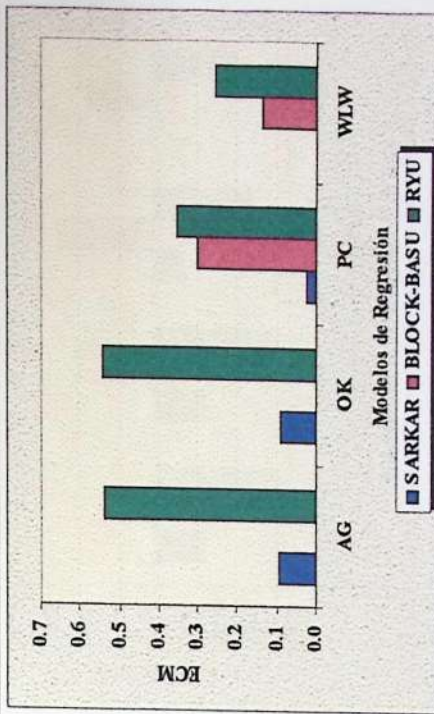


Figura. 5.5.2: Error cuadrático medio para eventos no ordenados o diferentes

Tabla 5.5.3. Sesgo relativo (SR) de los estimadores del parámetro $b = -2$ según las diferentes situaciones y según el tipo de distribución subyacente de los datos

Tipo de Distribución	Sesgo Relativo								
	Eventos Ordenados			Eventos no Ordenados					
	AG	OK	PC	WLW	AG	OK	PC	WLW	
Sarkar	-6 %	-9 %	2 %	19 %	-15 %	-15 %	-8 %	0 %	
Block y Basu	-2 %	3 %	23 %	40 %	1 %	2 %	27 %	18 %	
Ryu	-25 %	-29 %	-23 %	-9 %	-36 %	-36 %	-29 %	-25 %	

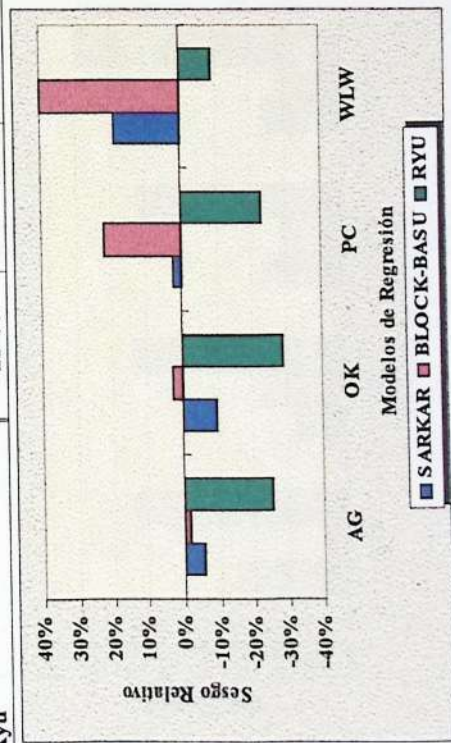


Figura 5.5.3: Sesgo relativo para eventos ordenados o recurrentes

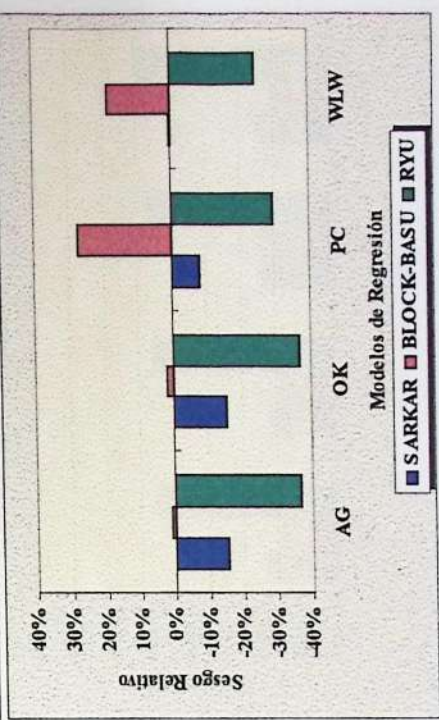


Figura 5.5.4: Sesgo relativo para eventos no ordenados o diferentes

Tabla 5.5.4. Error estándar de los estimadores del parámetro $b=-2$ según las diferentes situaciones y según el tipo de distribución subyacente de los datos

Tipo de Distribución	Error Estándar							
	Eventos Ordenados			Eventos no Ordenados				
	AG	OK	PC	WLW	AG	OK	PC	WLW
Sarkar	0.060	0.026	0.045	0.044	0.039	0.036	0.037	0.040
Block y Basu	0.013	0.048	0.033	0.035	0.029	0.028	0.038	0.032
Ryu	0.113	0.116	0.113	0.144	0.095	0.097	0.100	0.105

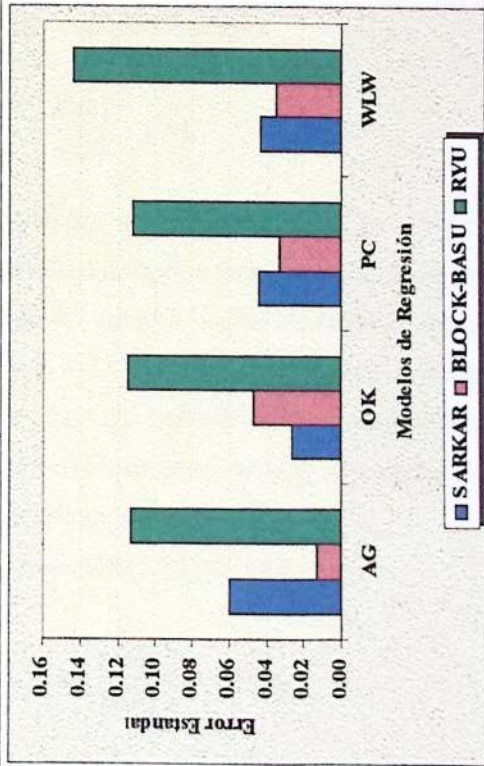


Figura. 5.5.5: Error Estándar para eventos ordenados o recurrentes

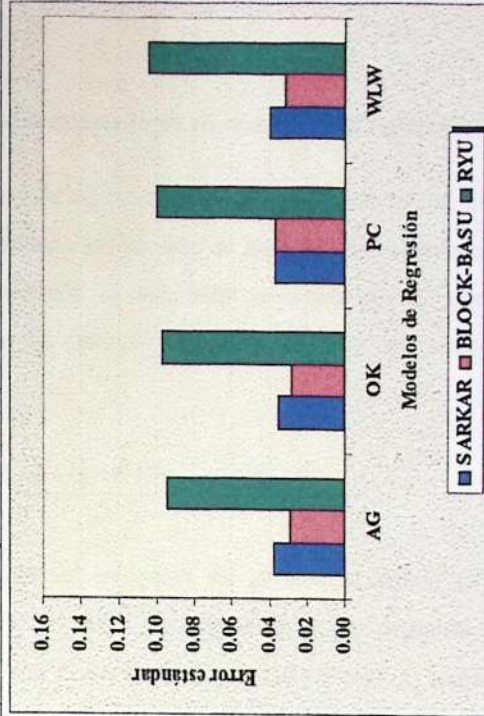


Figura. 5.5.6: Error Estándar para eventos no ordenados o diferentes

6. Cota para el sesgo de los estimadores en el modelo de regresión.

Para evaluar el sesgo de los estimadores del parámetro de los modelos de regresión cuando la distribución subyacente de los datos no cumple con uno o ambos tipos de falta de memoria, se debe tener en cuenta que el estimador del parámetro es un promedio del estimador correspondiente a cada evento, o sea que

$$\hat{b} = \frac{b_1^* + b_2^*}{2} = \frac{b + k_1 + b + k_2}{2} = b + \frac{k_1 + k_2}{2} \quad (5.6.1)$$

donde k_i ($i=1,2$) es el sesgo del estimador del parámetro cuando se tiene en cuenta la i -ésima falla. Por lo tanto, se debe evaluar por separado el sesgo introducido cuando se estima el parámetro para cada evento y , siendo así, el sesgo total será el promedio de los sesgos para cada falla. De esta forma, *para cada evento* y cuando $Z=1$, se considera el cociente

$$\frac{e^{b_i^*}}{e^b} = \frac{h_i^*(\theta)}{h_i(\theta)}, \quad i = 1, 2, \quad (5.6.2)$$

donde θ es el vector de parámetro de la distribución; b_i^* es el estimador del parámetro del modelo cuando la distribución tiene memoria; h_i^* es el "hazard" condicional o marginal, según se evalúe un modelo condicional o marginal de la distribución con memoria; y, h_i es el "hazard" correspondiente de la distribución sin memoria. Cabe destacar que b representa el valor del parámetro si sólo se introducen sesgos por memoria, en caso contrario se debe reemplazar b por \tilde{b} , que corresponde al *estimador del parámetro*.

Luego, a partir de la expresión (5.6.2) se obtiene

$$E(b_i^*) - b = E \left\{ \ln \left[\frac{h_i^*(\theta)}{h_i(\theta)} \right] \right\}, \quad i = 1, 2, \quad (5.6.3)$$

que da la expresión del sesgo del estimador en el caso que b sea el verdadero valor del parámetro, o bien

$$E(\hat{b}_i^*) - E(\tilde{b}) = E \left\{ \ln \left[\frac{h_i^*(\theta)}{h_i(\theta)} \right] \right\}, \quad i = 1, 2, \quad (5.6.4)$$

en el caso que se introduzcan otros tipos de sesgos en el modelo.

6.1. Sesgos en la Situación de Eventos Recurrentes

Para la evaluación del sesgo en la situación de eventos recurrentes se debe considerar por separado los modelos condicionales de los modelos marginales.

Modelos condicionales

Como ya se vio en los capítulos previos, los modelos de "hazard" proporcional de tipo condicional se expresan en función de las componentes del vector intensidad $\lambda(t)$. Si se considera, por ejemplo $\lambda_1(t)$, se debe tener en cuenta que ella es la suma de dos términos (ver ecuación (5.3.1.5)), uno de los cuales corresponde al "hazard" condicional si $T_2 < t$ y el otro si $T_2 \geq t$. Por lo tanto, siguiendo el argumento dado para obtener (5.6.1), el sesgo que se introduce al estimar el parámetro b del modelo sólo para el primer evento, será el promedio de los sesgos en la estimación de b cuando los $h(\theta)$ sean los "hazard" condicionales dado $T_2 < t$ y $T_2 \geq t$. Entonces, el sesgo para los modelos condicionales de cada uno de los eventos se puede expresar como

$$E(\hat{b}_i^*) - b = \frac{1}{2} E \left\{ \ln \left[\frac{p_1^{(i)*}}{p_1^{(i)}} \right] + \ln \left[\frac{p_2^{(i)*}}{p_2^{(i)}} \right] \right\}, \quad i = 1, 2, \quad (5.6.1.1)$$

donde $p_i^{(j)}$ ($i, j=1,2$) son los "hazard" condicionales de la distribución sin memoria y $p_i^{(j)*}$ ($i, j=1,2$) son los correspondientes a la distribución con memoria.

De las tres distribuciones consideradas la única que goza de la JLMP es la debida a Block y Basu, por tal motivo en los denominadores se usa las expresiones de los "hazard" condicionales dadas en (5.3.1.6) y (5.3.1.7).

⇒ *Sesgo para la Distribución Sarkar*

Por simplicidad se considera $\theta_1 = \theta_2 = \theta_c$, luego por la simetría de las funciones que definen las distribuciones de Sarkar y Block-Basu se tiene que $p_1^{(1)} = p_2^{(2)}$ y $p_2^{(1)} = p_1^{(2)}$ y en consecuencia $\lambda_1(t) = \lambda_2(t)$. Es suficiente, entonces, evaluar el sesgo para una sola de las componentes del vector intensidad $\lambda(t)$.

De las expresiones (5.3.1.6) y (5.3.1.7) se obtiene que la expresión de los "hazard" condicionales de la distribución Block-Basu toman la forma

$$\begin{aligned} p_1^{(1)} &= \theta_c + \theta_{12} \\ p_2^{(1)} &= \theta_c + \frac{1}{2} \theta_{12} \end{aligned} \quad (5.6.1.2)$$

A partir de (5.3.2.4) y (5.3.2.5) se tiene que los "hazard" condicionales correspondientes a la distribución Sarkar serán

$$\begin{aligned} p_1^{(1)*} &= \theta_c^* + \frac{1}{2} \theta_{12}^* + \frac{1}{2} \theta_{12}^* [1 - \exp(-\theta_c^* t)]^{-1} \\ p_2^{(1)*} &= \theta_c^* + \frac{1}{2} \theta_{12}^* \end{aligned}$$

En estas últimas igualdades se puede ver que el "hazard" condicional cuando $T_2 < t$ depende de t ; luego, acotando $p_1^{(1)*}$ para obtener una expresión independiente del t se obtiene

$$\begin{aligned} p_1^{(1)*} &\geq \theta_c^* + \theta_{12}^* \geq \theta_c^* + \frac{1}{2} \theta_{12}^* \\ p_2^{(1)*} &= \theta_c^* + \frac{1}{2} \theta_{12}^* \end{aligned} \quad (5.6.1.3)$$

En consecuencia, se logra que el primer término de la expresión (5.6.1.1) se pueda escribir como

$$\ln \left[\frac{p_1^{(1)*}}{p_1^{(1)}} \right] \geq \ln \left[\frac{\theta_c^* + \frac{1}{2}\theta_{12}^*}{\theta_c + \theta_{12}} \right], \quad (5.6.1.4)$$

ya que, como se muestra en el capítulo anterior, en los eventos recurrentes, el sesgo de los estimadores de los modelos condicionales no es significativo si la correlación entre los tiempos de falla es muy pequeña. Usando, en la expresión (5.6.1.1), la desigualdad (5.6.1.4) y las correspondientes igualdades dadas en (5.6.1.2) y (5.6.1.3), se obtiene que, el sesgo que se introduce cuando la distribución subyacente de los datos es $ACBVE_2(\theta_c, \theta_{12})$ está dado por:

$$E(b^*) - b \geq \frac{1}{2} \left\{ \ln \left[\frac{\theta_c^* + \frac{1}{2}\theta_{12}^*}{\theta_c + \theta_{12}} \right] + \ln \left[\frac{\theta_c^* + \frac{1}{2}\theta_{12}^*}{\theta_c + \frac{1}{2}\theta_{12}} \right] \right\} \quad (5.6.1.5)$$

Reemplazando los valores de los parámetros usados en las simulaciones se obtiene que

$$E(b^*) - b \geq 0.11,$$

de aquí que la cota para el SR, en el caso $b = -2$, será

$$SR \leq -6\%.$$

Es decir, se subestima el efecto de tratamiento por lo menos en un 6 %, valor que concuerda con el estudio de Monte Carlo.

⇒ *Sesgo para la Distribución Ryu*

En el caso de la distribución Ryu, dada la complejidad de la expresión para la función de densidad, surgen problemas para obtener una expresión exacta de los "hazard" condicionales, razón por la cual la cota para el sesgo en este caso se obtiene de acuerdo a la expresión (5.6.1.1) sin calcular los $p_i^{(i)*}$.

Por simplicidad, nuevamente, se considera $\theta_1 = \theta_2 = \theta_c$ y $s_1 = s_2 = s$; por lo tanto, sólo se evalúa el sesgo para una de las componentes del vector $\lambda(t)$. Pero, por ser la función logaritmo una función cóncava se verifica que

$$\frac{1}{2} \left\{ \ln \left(\frac{p_1^{(1)*}}{p_1^{(1)}} \right) + \ln \left(\frac{p_2^{(1)*}}{p_2^{(1)}} \right) \right\} \leq \ln \left\{ \frac{1}{2} \left(\frac{p_1^{(1)*}}{p_1^{(1)}} + \frac{p_2^{(1)*}}{p_2^{(1)}} \right) \right\} = \ln \left[\frac{p_2^{(1)} p_1^{(1)*} + p_1^{(1)} p_2^{(1)*}}{2 p_1^{(1)} p_2^{(1)}} \right].$$

Puesto que, si la distribución es Block-Basu, se verifica que $p_1^{(1)} > p_2^{(1)}$, entonces,

$$\frac{1}{2} \left\{ \ln \left(\frac{p_1^{(1)*}}{p_1^{(1)}} \right) + \ln \left(\frac{p_2^{(1)*}}{p_2^{(1)}} \right) \right\} < \ln \left[\frac{p_1^{(1)} (p_1^{(1)*} + p_2^{(1)*})}{2 p_1^{(1)} p_2^{(1)}} \right] < \ln \left[\frac{h^*(t)}{p_2^{(1)}} \right] \quad (5.6.1.6)$$

donde en el tercer término de la desigualdad, $h^*(t)$ representa el "hazard" marginal correspondiente a la primera recurrencia, la cual surge de la definición de los "hazard" condicionales $p_j^{(1)}$ dados en el Capítulo II.

La función "hazard" marginal de la distribución Ryu dada en la expresión (5.3.3.3), para $\theta_1 = \theta_2 = \theta_c$ y $s_1 = s_2 = s$, toma la forma:

$$h^*(t) = \theta_c^* + \theta_{12}^* (1 - e^{-st}).$$

Por otra parte, una cota para $h^*(t)$ independiente de t se puede obtener considerando que t varía entre 0 e ∞ , así si $t \rightarrow \infty$, $h^*(t) \rightarrow \theta_c^* + \theta_{12}^*$, mientras que si $t \rightarrow 0$, $h^*(t) \rightarrow \theta_c^*$; en consecuencia,

$$\theta_c^* \leq h^*(t) \leq \theta_c^* + \theta_{12}^*$$

Cabe destacar que, si tiene en cuenta que la memoria de la distribución Ryu se introduce a través del factor e^{-st} , entonces, el "hazard" marginal que involucre el parámetro s se puede obtener evaluando $E[h^*(\theta)]$. En efecto, sin mayores dificultades de cálculo, se obtiene que

$$E[h^*(t)] < \theta_c^* + \theta_{12}^* + \frac{1}{2} s^* \quad (5.6.1.7)$$

Reemplazando esta desigualdad en (5.6.1.6) se obtiene la cota para el sesgo del estimador del parámetro de los modelos de regresión tipo condicional si la distribución subyacente de los datos es Ryu, es decir,

$$E(b^*) - b < \ln \left[\frac{\theta_c^* + \theta_{12}^* + \frac{1}{2}s}{\theta_c + \frac{1}{2}\theta_{12}} \right] \quad (5.6.1.8)$$

Ahora, reemplazando los valores de los parámetros usados en las simulaciones se obtiene que

$$E(b^*) - b < 0.96.$$

De aquí que la cota para el SR, en el caso $b = -2$, será

$$SR > -48 \%$$

Es decir, se subestimaría el efecto de tratamiento en menos de un 48 %. Los valores encontrados en el estudio de Monte Carlo, tanto para el modelo AG (-25%), como para el modelo OK (-29%), concuerdan con este resultado.

Modelos Marginales

A diferencia de los modelos condicionales, en los modelos marginales no se modela cada una de las componentes del vector intensidad $\lambda(\mathbf{t})$, sino los "hazard" correspondientes a las distribuciones marginales de los tiempos de falla (totales o intereventos). Por esta razón, las funciones $h^*(\theta)$ y $h(\theta)$ de la ecuación (5.6.2) representan en este caso los "hazard" marginal de la distribución con y sin memoria marginal respectivamente.

De las tres distribuciones consideradas, la distribución debida a Sarkar es la única que goza de la MLMP; entonces, para evaluar el sesgo en los modelos marginales se usará en el denominador el "hazard" marginal correspondiente a ella.

Pero, dada la diferencia que existe entre los modelos WLW y PC respecto a cómo se considera los tiempos de falla de la segunda recurrencia (X_1+X_2 para WLW y X_2 para PC), para obtener una cota para el sesgo se debe analizar cada uno por separado.

En el caso del modelo WLW se debe tener en cuenta que cuando se estima el parámetro de la segunda recurrencia se introducen dos tipos de sesgos:

- (a). El producido por el hecho que, cuando se estima el parámetro, se considera que los individuos que no experimentaron la primera falla están bajo el riesgo de la segunda recurrencia.
- (b). El que se genera si la estructura subyacente de los datos carece de la propiedad de falta de memoria marginal.

Por esta razón, para evaluar el sesgo en la estimación del parámetro b cuando se considera la segunda recurrencia, se usa la expresión (5.6.4), ya que, el estimador del parámetro en el denominador se ve afectado por el sesgo de tipo (a), independiente de la memoria de la distribución, y por lo tanto para la segunda falla, la razón de los "hazard" cuando $Z=1$ se expresa como

$$\frac{e^{b_2^*}}{e^{\tilde{b}_2}} = \frac{h_2^*(\theta)}{h_2(\theta)}, \quad (5.6.1.8)$$

donde \tilde{b}_2 es el estimador del parámetro en el segundo evento, cuando la distribución subyacente es Sarkar. Aplicando logaritmo, y evaluando la esperanza en ambos miembros, se tiene que

$$E(b_2^*) - E(\tilde{b}_2) = E \left\{ \ln \left[\frac{h_2^*(\theta)}{h_2(\theta)} \right] \right\}.$$

Además, en la Sección 4.1 del Capítulo IV se mostró que, en el caso de independencia, con el modelo WLW, $E(\tilde{b}_2) \leq b - \frac{1}{2} \ln 2$, y que este tipo de sesgo

disminuye a medida que aumenta la correlación; de esta manera, reemplazando la expresión de $E(\hat{b}_2)$ en la igualdad anterior se obtiene:

$$E(b_2^*) - b \geq E \left\{ \ln \left[\frac{h_2^*(\theta)}{h_2(\theta)} \right] \right\} - \frac{1}{2} \ln 2. \quad (5.6.1.9)$$

Como en el estimador de la primera recurrencia sólo se introduce el sesgo de tipo (b), se obtiene que

$$E(b_1^*) - b = E \left\{ \ln \left[\frac{h_1^*(\theta)}{h_1(\theta)} \right] \right\}.$$

Luego, el sesgo total del estimador del parámetro del modelo WLW, es el promedio de las últimas expresiones de acuerdo a (5.6.1). Considerando $\theta_1 = \theta_2 = \theta_c$ para todas las distribuciones y, además, $s_1 = s_2 = s$ para la distribución Ryu, se obtiene una expresión para el sesgo *en el caso de independencia* entre los tiempos de falla, dada por

$$E(b_{WLW}^*) - b \geq E \left\{ \ln \left[\frac{h^*(\theta)}{h(\theta)} \right] \right\} - \frac{1}{4} \ln 2. \quad (5.6.1.10)$$

Si los X_i están *fuertemente correlacionados*, el segundo término se aproxima a cero, en consecuencia, la expresión del sesgo en este caso sería

$$E(b_{WLW}^*) - b = E \left\{ \ln \left[\frac{h^*(\theta)}{h(\theta)} \right] \right\}. \quad (5.6.1.11)$$

En el caso del modelo PC, la expresión para el sesgo viene dada por (5.6.1.11) puesto que el tiempo considerado para la segunda recurrencia es el tiempo interevento (X_2), y entonces, no se introducen sesgos del tipo (a).

Bajo estas consideraciones, aún si la distribución subyacente de los datos es Sarkar y con una correlación pequeña, los estimadores del parámetro de regresión cuando se usa el modelo WLW son sesgados y su $SR \geq 17\%$

⇒ *Sesgo para la Distribución Block-Basu*

El "hazard" marginal de la distribución debida a Block y Basu se obtiene a partir de la sobrevida marginal dada en (5.3.1.2) para el caso $\theta_1 = \theta_2 = \theta_c$ y de su correspondiente función de densidad, que toma la forma

$$f(t) = \frac{2\theta_c + \theta_{12}}{2} e^{-(\theta_c + \theta_{12})t} [\theta_c + \theta_{12} - \theta_{12} e^{-\theta_c t}],$$

de donde $h^*(\theta)$ viene dada por

$$h^*(t) = \frac{f(t)}{S(t)} = \frac{(2\theta_c^* + \theta_{12}^*)(\theta_c^* + \theta_{12}^* - \theta_{12}^* e^{-\theta_c^* t})}{(2\theta_c^* + \theta_{12}^* - \theta_{12}^* e^{-\theta_c^* t})}$$

Una cota para $h^*(t)$ independiente de t se puede obtener considerando que t varía entre 0 e ∞ , así si $t \rightarrow \infty$ entonces $h^*(t) \rightarrow \theta_c^* + \theta_{12}^*$, mientras que si $t \rightarrow 0$, $h^*(t) \rightarrow \theta_c^* + \frac{1}{2} \theta_{12}^*$; en consecuencia,

$$\theta_c^* + \frac{1}{2} \theta_{12}^* \leq h^*(t) \leq \theta_c^* + \theta_{12}^*. \quad (5.6.1.12)$$

El "hazard" marginal correspondiente para la distribución Sarkar, para cualquiera de los eventos es tal que

$$h(t) = \theta_c + \theta_{12} \quad (5.6.1.13)$$

Pero, bajo el supuesto de independencia el parámetro $\theta_{12} = 0$ en todas las distribuciones, y entonces, el "hazard" marginal $h(t) = \theta_c$, de aquí que, usando (5.6.1.10), el sesgo del estimador *bajo independencia* se pueda escribir como

$$E(b_{WLW}^*) - b \geq \ln \left[\frac{\theta_c^*}{\theta_c} \right] - \frac{1}{4} \ln 2 . \quad (5.6.1.14)$$

Sin embargo, si los tiempos de falla *están fuertemente correlacionados* entonces, reemplazando (5.6.1.12) y (5.6.1.13) en la igualdad (5.6.1.11) se tiene

$$\ln \left[\frac{\theta_c^* + \frac{1}{2} \theta_{12}^*}{\theta_c + \theta_{12}} \right] \leq E(b_{WLW}^*) - b \leq \ln \left[\frac{\theta_c^* + \theta_{12}^*}{\theta_c + \theta_{12}} \right] .$$

Para el estudio por simulaciones, si bien las X_i no son independientes, la correlación es pequeña, de manera que, usando (5.6.1.14) se obtiene que

$$E(b_{WLW}^*) - b \geq -1.15,$$

de aquí que la cota para el SR, en el caso $b = -2$, será

$$SR_{WLW} \leq 58 \% .$$

Es decir, se sobrestima el efecto de tratamiento como máximo en un 58 %, valor que concuerda con el estudio por Monte Carlo.

Para el modelo PC, considerando que no se introducen errores de tipo (a), la expresión para el sesgo está dada por

$$\ln \left[\frac{\theta_c^* + \frac{1}{2} \theta_{12}^*}{\theta_c + \theta_{12}} \right] \leq E(b_{PC}^*) - b \leq \ln \left[\frac{\theta_c^* + \theta_{12}^*}{\theta_c + \theta_{12}} \right] . \quad (5.6.1.15)$$

Y, particularizando para los parámetros dados en la Tabla 4.1.1, se obtiene que

$$-0.61 \leq E(b_{PC}^*) - b \leq -0.18,$$

de aquí que la cota para el SR, en el caso $b = -2$, será

$$9 \% \leq SR_{PC} \leq 30 \%,$$

es decir que, en este caso se sobrestima el efecto de tratamiento como máximo en un 30 %, lo que concuerda con el valor estimado por simulación (23 %).

⇒ *Sesgo para la Distribución Ryu*

Al analizar el sesgo en los modelos condicionales se pudo observar que la esperanza del "hazard" marginal de la distribución Ryu estaba dada por

$$E[h^*(t)] < \theta_c^* + \theta_{12}^* + \frac{1}{2} s^*.$$

Reemplazando esta desigualdad en (5.6.1.11), se tiene que la expresión para el sesgo estaría dada por

$$E(b_{WLW}^*) - b = E \left[\ln \frac{h^*(t)}{h(t)} \right] < \ln \left[\frac{\theta_c^* + \theta_{12}^* + \frac{1}{2} s^*}{\theta_c + \theta_{12}} \right]. \quad (5.6.1.16)$$

Evaluando el sesgo en el caso particular de los parámetros usados para las simulaciones, usando la desigualdad (5.6.1.16) se obtiene que

$$E(b_{WLW}^*) - b < 0.35$$

De aquí que la cota para el SR, en el caso $b = -2$, será

$$SR_{WLW} > -17 \%.$$

Es decir, se subestima el efecto de tratamiento a lo sumo en un 17 %. Lo que concuerda con el valor estimado por simulación (-9 %).

Sin embargo, si X_1 y X_2 son independientes, entonces el "hazard" marginal de la distribución Ryu es igual a θ_c^* . Por lo tanto, bajo supuesto de independencia, el sesgo que se genera viene dado por

$$E(b_{WLW}^*) - b \geq \ln \left[\frac{\theta_c^*}{\theta_c} \right] - \frac{1}{4} \ln 2. \quad (5.6.1.17)$$

El SR calculado por esta última expresión para el modelo WLW, usando los mismos valores de los parámetros que en el estudio de Monte Carlo, toma el valor, $SR_{WLW} \leq -2.5\%$.

La expresión para el sesgo en el modelo PC es igual a la dada en (5.6.1.16) y a partir de ella se obtiene, al igual que para el modelo WLW,

$$E(b_{PC}^*) - b < 0.35,$$

de aquí que la cota para el SR, en el caso $b = -2$, será

$$SR_{PC} > -17\%.$$

La diferencia entre esta cota inferior y el valor estimado por simulación se puede explicar si se considera la variabilidad de estos estimadores (0.225).

6.2. Sesgos en la Situación de Eventos Diferentes

Por las razones expuestas en la Sección 4.2 del Capítulo IV, para evaluar el sesgo de los modelos en la situación de eventos diferentes, se toma como referencia al modelo WLW, lo que equivale a considerar en el denominador de la expresión (5.6.2) el "hazard" marginal de la distribución sin memoria marginal; es decir, el correspondiente a la distribución Sarkar. Por lo tanto, cuando la distribución subyacente de los datos es Sarkar y se está en la situación de eventos diferentes, no se generan sesgos por memoria del proceso. Luego, los sesgos estimados en el estudio de Monte Carlo cuando se usa esta distribución se generan por la correlación entre los tiempos de falla, que como se vio en la Sección 5 de este capítulo, es igual a 0.19; y, por ordenamiento de los datos.

Al igual que el caso de eventos recurrentes, se considera por separado los modelos marginales de los condicionales, para analizar el sesgo.

Modelos condicionales

Como ya se analizó en el capítulo anterior, el ordenamiento de los datos requerido para aplicar este tipo de modelos, usando el software desarrollado por Therneau, conduce a generar un sesgo adicional al introducido por la falta de cumplimiento de los supuestos de los modelos. De esta manera, para evaluar los sesgos que se producen en los estimadores del parámetro b en la situación de eventos diferentes, se debe contemplar no sólo el que se genera por no cumplir con el supuesto de falta de memoria marginal, sino también el que se introduce por el ordenamiento de los datos para el cálculo computacional.

➤ *Evaluación del Sesgo por Ordenamiento*

En la Sección 4.2 del Capítulo IV se llegó a la conclusión que la aplicación de los modelos condicionales a eventos diferentes usando el software de Therneau, conduce a considerar los datos como provenientes de muestras de la distribución conjunta de las variables $W = \min(X_1, X_2)$ y $Z = |X_1 - X_2|$, donde X_1 y X_2 tienen distribución Sarkar. Por lo tanto, para evaluar el sesgo por ordenamiento y partiendo de las expresiones (4.3.2.1) se necesita encontrar los "hazard" correspondientes a las variables W y Z .

Por las propiedades de la distribución Sarkar, las variables W y Z son independientes y sus respectivas funciones de sobrevida son

$$S_W(t_1) = e^{-(2\theta_c + \theta_{12})t_1} = e^{-v t_1}$$

$$S_Z(t_2) = e^{-(\theta_c + \theta_{12})t_2}$$
(5.6.2.2)

Como ambas distribuciones son exponenciales, entonces sus funciones "hazard" son iguales al parámetro de dichas distribuciones, es decir,

$$h_W(t_1) = v = 2\theta_c + \theta_{12} \quad (5.6.2.2)$$

$$h_Z(t_2) = \theta_c + \theta_{12}$$

A partir de estas expresiones se obtiene la ecuación del sesgo por ordenamiento (4.3.2.2) dada por

$$E(\hat{b}) - b = \frac{1}{2} \ln \left[\frac{2\theta_c + \theta_{12}}{\theta_c + \theta_{12}} \right] \quad (5.6.2.3)$$

Reemplazando los valores de los parámetros de la distribución Sarkar usados en el estudio de Monte Carlo, se tiene que el sesgo por ordenamiento es en este caso

$$E(\hat{b}) - b = 0.26 \quad (5.6.2.4)$$

➤ Evaluación del Sesgo por Memoria

Cuando la distribución subyacente de los datos tiene memoria marginal y se usan modelos condicionales en la situación de eventos diferentes, se debe usar la siguiente expresión para evaluar el sesgo

$$\frac{e^{b_i}}{e^b} = \frac{h_i^*(\theta)}{h_i(\theta)}, \quad i = 1, 2, \quad (5.6.2.5)$$

donde \hat{b} es el estimador cuando el proceso tiene la propiedad de falta de memoria marginal, y cuyo sesgo es generado por ordenamiento de los datos, h_i^* es el "hazard" condicional de la distribución que se evalúa, y h_i es el "hazard" marginal de la distribución Sarkar. Evaluando la esperanza de esta expresión y

reemplazando $E(b)$ obtenida de (5.6.2.3), se llega a que el sesgo de los modelos condicionales en la situación de eventos diferentes se expresa como

$$E(b^*) - b = \frac{1}{2} E \left[\ln \left(\frac{h_1^*}{h_1} \right) + \ln \left(\frac{h_2^*}{h_2} \right) \right] + \frac{1}{2} \ln \left[\frac{2\theta_c + \theta_{12}}{\theta_c + \theta_{12}} \right] \quad (5.6.2.6)$$

⇒ *Sesgo para la Distribución Sarkar*

Usando la expresión (5.6.2.6) y reemplazando los "hazard" condicionales dados en (5.6.1.3) se obtiene que el sesgo del estimador del parámetro b cuando la distribución subyacente de los datos es Sarkar, se expresa como

$$E(b^*) - b \geq \frac{1}{2} \left[\ln \left(\frac{\theta_c + \frac{1}{2}\theta_{12}}{\theta_c + \theta_{12}} \right) \right] + \frac{1}{2} \ln \left[\frac{2\theta_c + \theta_{12}}{\theta_c + \theta_{12}} \right] \quad (5.6.2.7)$$

Reemplazando los valores de los parámetros usados en el estudio de Monte Carlo se obtiene que

$$E(b^*) - b \geq 0.16,$$

de donde

$$SR \leq -8\%.$$

Lo cual concuerda con el valor estimado por simulación (-15%).

⇒ *Sesgo para la Distribución Block-Basu*

Si la distribución subyacente de los datos es Block-Basu, usando la expresión (5.6.2.6) y los valores de los respectivos "hazard" condicionales, se tiene que, una cota para el sesgo que se introduce en los modelos condicionales en la situación de eventos diferentes y cuando la distribución de los datos es Block-Basu viene dada por

$$E(\hat{b}^*) - b \geq \frac{1}{2} \ln \left[\frac{\theta_c^* + \frac{1}{2}\theta_{12}^*}{\theta_c + \theta_{12}} \right] + \frac{1}{2} \ln \left[\frac{2\theta_c + \theta_{12}}{\theta_c + \theta_{12}} \right] \quad (5.6.2.8)$$

El sesgo para el caso particular de los parámetros de la Tabla 5.5.1 será entonces

$$E(\hat{b}^*) - b \geq -0.05,$$

de donde

$$SR \leq 2.6 \%,$$

lo que concuerda con lo estimado por simulación.

⇒ *Sesgo para la Distribución Ryu*

Para evaluar el sesgo de los modelos condicionales en la situación de eventos diferentes, si (X_1, X_2) tiene distribución Ryu, a partir de la expresión (5.6.2.6) se debe reemplazar las h_i^* ($i=1, 2$) por las expresiones de las funciones "hazard" condicionales correspondientes a la distribución Ryu; luego, la expresión para el sesgo en este caso toma la forma siguiente,

$$E(\hat{b}^*) - b = \frac{1}{2} E \left[\ln \left(\frac{p_1^{(1)*}}{\theta_c + \theta_{12}} \right) + \ln \left(\frac{p_2^{(1)*}}{\theta_c + \theta_{12}} \right) \right] + \frac{1}{2} \ln \left[\frac{2\theta_c + \theta_{12}}{\theta_c + \theta_{12}} \right]. \quad (5.6.2.9)$$

Ahora, puesto que la función $\ln(x)$ es una función cóncava, la expresión (5.6.2.9) se puede escribir como

$$E(\hat{b}^*) - b \leq E \left[\ln \left(\frac{p_1^{(1)*} + p_2^{(1)*}}{2(\theta_c + \theta_{12})} \right) \right] + \frac{1}{2} \ln \left[\frac{2\theta_c + \theta_{12}}{\theta_c + \theta_{12}} \right].$$

Por la desigualdad de Jensen para las funciones cóncavas, se tiene que

$$E(b^*) - b \leq \ln \left[E \left(\frac{p_1^{(1)*} + p_2^{(1)*}}{2(\theta_c + \theta_{12})} \right) \right] + \frac{1}{2} \ln \left[\frac{2\theta_c + \theta_{12}}{\theta_c + \theta_{12}} \right]$$

El numerador del primer término de esta última expresión es el "hazard" marginal de la distribución Ryu, luego usando la expresión (5.6.1.7) para la esperanza de "hazard" marginal, se tiene que

$$E(b^*) - b \leq \ln \left[\frac{\theta_c^* + \theta_{12}^* + \frac{1}{2}s}{(\theta_c + \theta_{12})} \right] + \frac{1}{2} \ln \left[\frac{2\theta_c + \theta_{12}}{\theta_c + \theta_{12}} \right] \quad (5.6.2.10)$$

Analíticamente, la cota del sesgo en el caso particular de los parámetros usados en el estudio por simulación será en este caso

$$E(b^*) - b \leq 0.60,$$

de donde,

$$SR \geq -30\%.$$

El valor estimado por simulación (-36 %) no es significativamente menor que la cota inferior encontrada analíticamente, si se tiene en cuenta la variabilidad de este estimador.

Modelos marginales

Tomando como referencia los "hazard" marginales de la distribución Sarkar, se evalúa el sesgo por memoria de los modelos marginales para la distribución Block-Basu y Ryu.

Cabe recordar, que en el modelo PC se introduce un sesgo adicional por ordenamiento de los datos, el cual es igual a los correspondientes en los modelos condicionales. Entonces, aún si la distribución subyacente de los datos fuera

Sarkar, el modelo PC genera estimadores sesgados, y el sesgo se puede obtener a través de la expresión (5.6.2.3). En el caso particular de los parámetros usados en el estudio por simulación, este sesgo toma el valor de 0.26, de donde el $SR = -13\%$. La diferencia entre el sesgo calculado y el estimado se puede explicar por la variabilidad del estimador.

⇒ *Sesgo para la Distribución Block-Basu*

Usando el hecho que para la distribución Block-Basu se verifica que su "hazard" marginal cumple con

$$\theta_c^* + \frac{1}{2} \theta_{12}^* \leq h^*(t) \leq \theta_c^* + \theta_{12}^*,$$

se tiene que

$$\ln \left[\frac{\theta_c^* + \frac{1}{2} \theta_{12}^*}{\theta_c^* + \theta_{12}^*} \right] \leq E(b_{WLW}^*) - b \leq \ln \left[\frac{\theta_c^* + \theta_{12}^*}{\theta_c^* + \theta_{12}^*} \right]. \quad (5.6.2.11)$$

Reemplazando los valores de los parámetros dados en la Tabla 5.5.1, se tiene que

$$-0.60 \leq E(b_{WLW}^*) - b \leq -0.18$$

de donde se sigue que,

$$9\% \leq SR_{WLW} \leq 30\%,$$

que incluye al valor estimado (18%) en el estudio de Monte Carlo.

Considerando el sesgo generado por ordenamiento de los datos cuando se usa el modelo PC, la expresión del sesgo para este modelo viene dada por

$$\ln \left[\frac{\theta_c^* + \frac{1}{2}\theta_{12}^*}{\theta_c + \theta_{12}} \right] \leq E(b_{PC}^*) - E(\hat{b}) \leq \ln \left[\frac{\theta_c^* + \theta_{12}^*}{\theta_c + \theta_{12}} \right]$$

donde \hat{b} es el estimador del parámetro, en el cual se genera sesgo por ordenamiento. Reemplazando $E(\hat{b})$ que se obtiene a partir de (5.6.2.3) se tiene que

$$\ln \left[\frac{\theta_c^* + \frac{1}{2}\theta_{12}^*}{\theta_c + \theta_{12}} \right] + \frac{1}{2} \ln \left[\frac{2\theta_c + \theta_{12}}{\theta_c + \theta_{12}} \right] \leq E(b_{PC}^*) - b$$

y

(5.6.2.12)

$$E(b_{PC}^*) - b \leq \ln \left[\frac{\theta_c^* + \theta_{12}^*}{\theta_c + \theta_{12}} \right] + \frac{1}{2} \ln \left[\frac{2\theta_c + \theta_{12}}{\theta_c + \theta_{12}} \right].$$

Usando la desigualdad (5.6.2.12) en el caso particular de los parámetros de la Tabla 5.5.1, se llega a que

$$-0.36 \leq E(b_{PC}^*) - b \leq 0.07,$$

de donde se sigue que,

$$-4\% \leq SR_{PC} \leq 18\%.$$

El SR estimado por simulación (27%), resultó significativamente mayor que la cota superior. Sin embargo, esto se explica si se considera que el sesgo estimado por ordenamiento de los datos, es menor que el calculado por la expresión (5.6.2.3), de manera que si en ambos miembros de la desigualdad (5.6.2.12) se reemplaza el segundo término por el sesgo estimado (0.16) se obtiene que

$$-0.46 \leq E(b_{PC}^*) - b \leq -0.17,$$

lo cual conduce a un aumento de las cotas del SR, y de aquí una mayor concordancia entre el valor estimado por simulación y el calculado analíticamente.

⇒ *Sesgo para la Distribución Ryu*

De acuerdo a los resultados en la sección anterior sobre la forma de las marginales de la distribución Ryu, el sesgo debido a la memoria marginal de esta distribución para el modelo WLW se puede expresar como

$$E(\hat{b}_{WLW}^*) - b = E\left[\ln \frac{h^*(t)}{h(t)}\right] < \ln \left[\frac{\theta_c^* + \theta_{12}^* + \frac{1}{2}s}{\theta_c + \theta_{12}} \right]. \quad (5.6.2.13)$$

Luego, reemplazando los valores de los parámetros de la distribución Ryu considerados en el estudio por simulación se obtiene

$$E(\hat{b}_{WLW}^*) - b < 0.35,$$

de donde se sigue que,

$$SR_{WLW} > -17\%.$$

La diferencia entre este valor y el estimado por simulación (-25%) se puede explicar por la variabilidad del estimador (0.105).

Considerando el sesgo generado por ordenamiento de los datos, cuando se usa el modelo PC, la expresión del sesgo para este modelo viene dada por

$$E(\hat{b}_{PC}^*) - b < \ln \left[\frac{\theta_c^* + \theta_{12}^* + \frac{1}{2}s}{\theta_c + \theta_{12}} \right] + \frac{1}{2} \ln \left[\frac{2\theta_c + \theta_{12}}{\theta_c + \theta_{12}} \right]. \quad (5.6.2.13)$$

Si se reemplaza en esta expresión los valores de los parámetros de la Tabla 5.5.1 se obtiene

$$E(\hat{b}_{PC}^*) - b < 0.60,$$

y, entonces

$$SR_{PC} > -30\%.$$

El valor encontrado por simulación (-29 %) concuerda con lo encontrado por simulación.

En la siguiente tabla, a manera de resumen, se muestra los valores estimados del SR en el estudio por simulación y las cotas analíticas, para cada situación (eventos ordenados y no ordenados), cada modelo y cada tipo de distribución subyacente de los datos.

Tabla 5.6.1. Cuadro comparativo entre el valor estimado por simulación y la cota obtenida analíticamente para el sesgo relativo (SR) de los estimadores del parámetro para cada una de las situaciones consideradas.

Distribución	Eventos Ordenados o Recurrentes							
	AG		OK		PC		WLW	
	Monte Carlo	Cota	Monte Carlo	Cota	Monte Carlo	Cota	Monte Carlo	Cota
Sarkar	- 6 %	≤ - 6%	- 9 %	≤ - 6%	2 %		19 %	≥ 17 %
Block y Basu	- 2 %		3 %		23 %	[9%, 30%]	40 %	≤ 58%
Ryu	- 25 %	> - 48%	- 29 %	> - 48%	- 23 %	> - 17%	- 9 %	> - 17%

Distribución	Eventos No Ordenados o Diferentes							
	AG		OK		PC		WLW	
	Monte Carlo	Cota	Monte Carlo	Cota	Monte Carlo	Cota	Monte Carlo	Cota
Sarkar	- 15 %	≤ - 8 %	- 15 %	≤ - 8%	- 8 %	≈ - 13 %	0 %	
Block y Basu	1 %	≤ 2.6 %	2 %	≤ 2.6 %	27 %	[-4%, 18%]	18 %	[9%, 30%]
Ryu	- 36 %	> - 30 %	- 36 %	> - 30 %	- 29 %	> - 30 %	- 25 %	> - 17%

Las celdas sombreadas indican la concordancia entre el tipo de modelo (marginal o condicional) y el cumplimiento de la propiedad de falta de memoria de la distribución (marginal o conjunta).

7. Conclusiones

Del análisis de los resultados surge que:

A. En la situación de *eventos recurrentes*:

- ✓ Si la distribución subyacente de los datos *goza de MLMP* pero no cumple con la propiedad de falta de memoria conjunta (Sarkar), los modelos condicionales y el modelo PC tienen en general mejor comportamiento que el modelo WLW; pero, esto se debe fundamentalmente al sesgo que en este último modelo se introduce por ordenamiento. Sin embargo, como se discute en el Capítulo IV, si hay una fuerte correlación entre los tiempos de falla este tipo de sesgo llega a ser insignificante.
- ✓ Si la distribución subyacente de los datos *cumple con JLMP*, pero carece de MLMP (Block-Basu) el mejor comportamiento lo presentan los modelos condicionales.
- ✓ Si la distribución subyacente de los *datos posee memoria conjunta y marginal* (Ryu), el sesgo producido por ordenamiento en el modelo WLW compensa en gran medida la memoria del proceso, generando el menor sesgo en los estimadores. Sin embargo, si existe una fuerte correlación entre los tiempos de falla, debido a que el sesgo por ordenamiento se hace despreciable, los modelos condicionales podrían producir el menor sesgo en esta situación.
- ✓ La memoria marginal y conjunta de los procesos no sólo afecta sensiblemente el sesgo de los estimadores sino también su variabilidad.

- ✓ Para los eventos recurrentes si las distribuciones marginales son exponenciales, es decir la distribución tiene MLMP, la recomendación es usar el modelo PC. Si por el contrario, sus marginales se apartan sensiblemente del modelo exponencial, pero la distribución goza de JLMP, el menor SR se consigue usando los modelos condicionales. Pero, si la distribución tiene tanto memoria marginal como conjunta se debería usar el modelo WLW.

B. En la situación de *eventos diferentes*

- ✓ El ECM de todos los modelos es sensiblemente mayor si la distribución posee memoria marginal y conjunta, debido tanto al sesgo como a la variabilidad de los estimadores.
- ✓ Si la distribución subyacente de los datos *goza de MLMP* pero no cumple con la propiedad de falta de memoria conjunta (Sarkar), los modelos marginales son los que producen el menor sesgo relativo si en el modelo PC se elimina el sesgo por ordenamiento, lo que equivale a tomar como segundo tiempo de falla el tiempo total y no el tiempo interevento.
- ✓ Si la distribución subyacente de los datos *cumple con JLMP*, pero carece de MLMP (Block-Basu), los modelos condicionales son los que generan el menor SR.
- ✓ Si la distribución subyacente de los *datos posee memoria conjunta y marginal* (Ryu), el modelo WLW y el modelo PC generan el menor SR, si en este último se corrige el sesgo por ordenamiento.
- ✓ Resumiendo, en el caso de eventos diferentes es conveniente usar el modelo WLW, excepto en la situación en que las marginales presenten un gran alejamiento del modelo exponencial, en cuyo caso cualquiera de los modelos condicionales (AG o OK) resulta adecuado.

CAPITULO VI

TAMAÑO DE MUESTRA Y CENSURAS

1. Introducción

En los capítulos precedentes se analizó en detalle el efecto de la correlación entre los tiempos de falla y la memoria de los procesos subyacentes en el sesgo de los estimadores de los parámetros de regresión, considerando siempre que se dispone de tamaños de muestras grandes ($n = 500$) y porcentajes de censuras no superiores al 25%. Sin embargo, en las muchas aplicaciones, debido a la naturaleza del problema bajo estudio, no es posible alcanzar tamaños de muestras grandes y porcentajes de censuras bajas. Por ejemplo, si la frecuencia de individuos en la población, que cumplen con los criterios de inclusión en el estudio es pequeña, resulta casi imposible disponer de una cantidad importante de sujetos para el estudio. Por otra parte, si la incidencia del evento es baja, se necesitaría períodos de tiempos muy largos para obtener un pequeño porcentaje de censuras. Por lo tanto, resulta de interés estudiar el efecto de la disminución del tamaño de muestra y el aumento del porcentaje de censuras en el sesgo de los estimadores del parámetro regresión, con el uso de los diferentes modelos considerados.

En la Sección 2, se estudia la variación del *Sesgo Relativo*, *Error Estándar* y *Error Cuadrático Medio* de los estimadores del parámetro de los modelos de "hazard" proporcional, cuando disminuye el tamaño de muestra mediante un estudio de Monte Carlo. En la Sección 3, se analiza, mediante simulaciones, el efecto del aumento del porcentaje de censuras tanto en el Sesgo como en el Error Cuadrático Medio de los estimadores de los parámetros en los modelos de regresión. Por último, en la Sección 4, se discute, comparativamente los resultados presentados en las secciones previas.

2. Efecto del tamaño de muestra

Al igual que en los capítulos previos, para llevar a cabo el estudio, se simulan datos como proveniente de la aplicación de dos tratamientos hipotéticos, reflejados en la covariable Z que toma valores cero-uno. Se considera el problema de dos fallas, es decir $k=2$. Y, se simulan 100 muestras de diferentes tamaños $n = 500, 200, 100$ y 50 , para cada una de las siguientes situaciones:

- I) *eventos recurrentes* en los cuales los tiempos de falla intereventos tienen conjuntamente la distribución Sarkar, es decir, $(X_1, X_2) \sim S(0.04, 0.04, 0.01)$.
- II) *eventos diferentes* donde los tiempos de falla totales se distribuyen conjuntamente como una distribución Sarkar, es decir $(T_1, T_2) \sim S(0.04, 0.04, 0.01)$.

En todos los casos se considera el parámetro del modelo de regresión $b_k = -2$ ($k=1, 2$); y un modelo de aleatorias uniformes con un porcentaje de censuras promedio del 10%. Además, al igual que en los previos estudios de Monte Carlo, en la situación de eventos ordenados, si un individuo tiene la primera recurrencia censurada, entonces tiene también censurada la segunda recurrencia.

A cada conjunto de datos simulados se aplica los cuatro modelos de regresión usando el software desarrollado por Therneau (1995) para S-Plus.

También, para cada tamaño de muestra considerado y cada tipo de modelo de regresión, se estima el Error Cuadrático Medio (ECM), el Error Estándar (SE) y, por último, el Sesgo Relativo (SR) del estimador \hat{b} del parámetro del modelo usando la expresión (4.2.1) del Capítulo IV. Se calcula, además, el cambio relativo del ECM respecto a $n = 500$ mediante

$$\text{ECM relativo} = \text{ECM}(n_i)/\text{ECM}(500),$$

donde $\text{ECM}(500)$ es el valor del ECM cuando se considera que se tiene un tamaño de muestra $n = 500$ y $\text{ECM}(n_i)$ el valor del ECM correspondiente a un tamaño de

muestra $n = n_i$. Por otra parte, La variación del SR con el tamaño de muestra se obtiene a través de la siguiente expresión

$$\text{Cambio SR} = \text{SR}(n_i) - \text{SR}(500).$$

Los valores del ECM estimado a partir de las simulaciones, para los diferentes tamaños de muestra, se muestran en la Tabla 6.2.1, y se representan gráficamente en las Figuras 6.2.1 y 6.2.2; los correspondientes valores del SR se presentan en la Tablas 6.2.3, y en las figuras 6.2.5 y 6.2.6. Los valores del SE, se muestran en la Tabla 6.2.5, y Figuras 6.2.9 y 6.2.10. El ECM relativo y el cambio del SR se presentan en las Tablas 6.2.2 y 6.2.4 respectivamente, con sus correspondientes figuras.

Tabla 6.2.1. Error cuadrático Medio (ECM) de los estimadores del parámetro $b = -2$ para las diferentes situaciones según el tamaño de muestra.

Tamaño de Muestra	Error Cuadrático Medio							
	Eventos Ordenados			Eventos no Ordenados				
	AG	OK	PC	WLW	AG	OK	PC	WLW
500	0,014	0,026	0,003	0,170	0,139	0,132	0,053	0,008
200	0,024	0,033	0,023	0,277	0,099	0,097	0,034	0,021
100	0,032	0,040	0,048	0,305	0,111	0,110	0,048	0,038
50	0,079	0,093	0,083	0,457	0,115	0,101	0,059	0,069

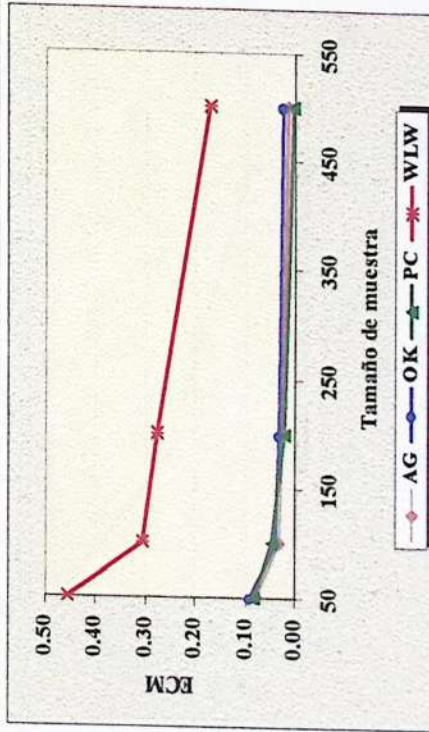
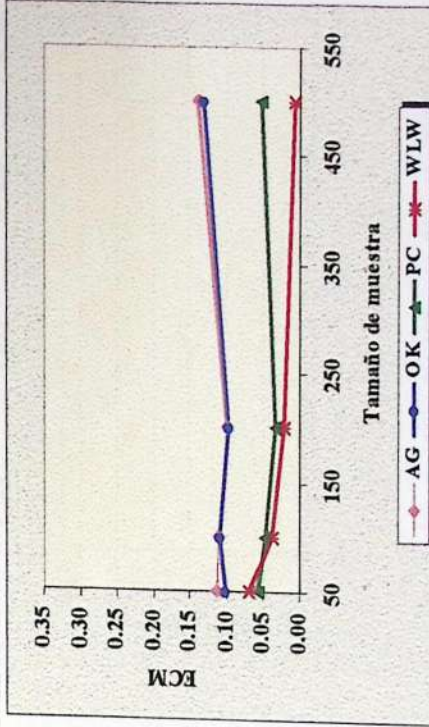
Figura 6.2.1.: ECM del estimador de b para cada modelo según el tamaño de muestra, en eventos recurrentes.Figura 6.2.2.: ECM del estimador de b para cada modelo según el tamaño de muestra, en eventos diferentes.

Tabla 6.2.2. Error cuadrático Medio relativo (ECM relativo) de los estimadores del parámetro $b = -2$, con respecto a $n = 500$, para las diferentes situaciones y según el tamaño de muestra.

Tamaño de muestra	Error Cuadrático Medio Relativo							
	Eventos Ordenados				Eventos no Ordenados			
	AG	OK	PC	WLW	AG	OK	PC	WLW
200	1.701	1.290	6.765	1.625	0.713	0.736	0.650	2.552
100	2.270	1.574	14.351	1.792	0.801	0.830	0.908	4.630
50	5.617	3.627	24.731	2.684	1.165	1.041	1.709	3.289

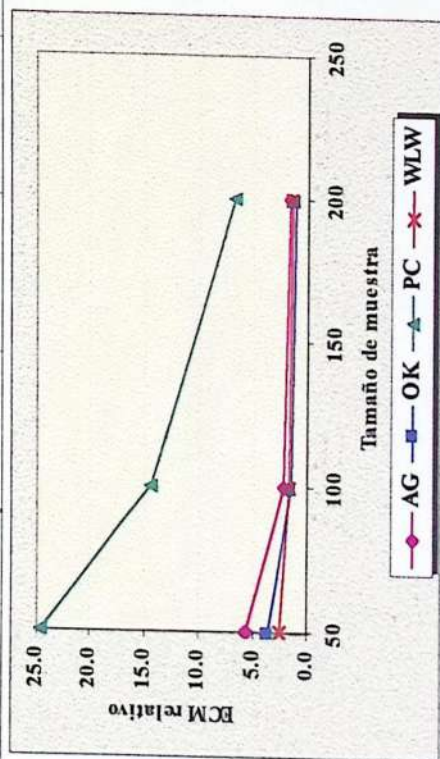


Figura 6.2.3.: ECM relativo del estimador de b para cada modelo según el tamaño de muestra, en eventos recurrentes.

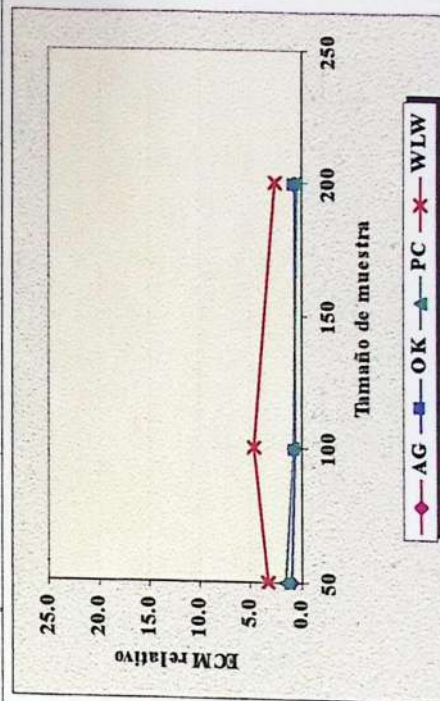


Figura 6.2.4.: ECM relativo del estimador de b para cada modelo según el tamaño de muestra, en eventos diferentes.

Tabla 6.2.3. Sesgo Relativo (SR) de los estimadores del parámetro $b = -2$ para las diferentes situaciones y según el tamaño de muestra.

Tamaño de muestra	Sesgo Relativo (SR)							
	Eventos Ordenados			Eventos no Ordenados				
	AG	OK	PC	WLW	AG	OK	PC	WLW
500	-4%	-8%	1%	20%	-19%	-18%	-11%	-4%
200	-2%	-4%	3%	24%	-14%	-14%	-6%	2%
100	1%	-5%	4%	25%	-14%	-14%	-7%	2%
50	0%	-3%	7%	28%	-14%	-12%	-3%	5%

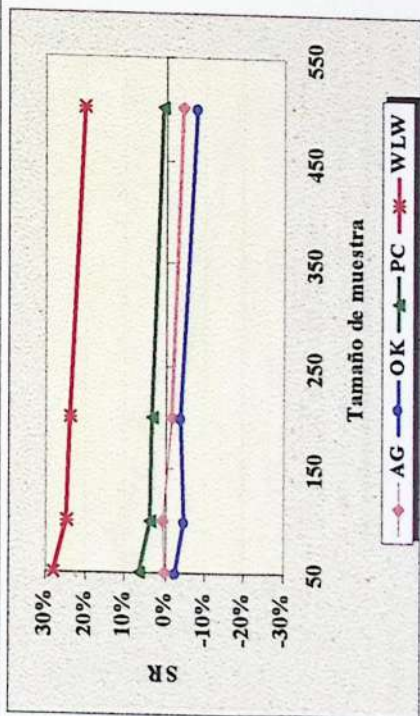


Figura 6.2.5.: SR del estimador de b para cada modelo según el tamaño de muestra, en eventos recurrentes.

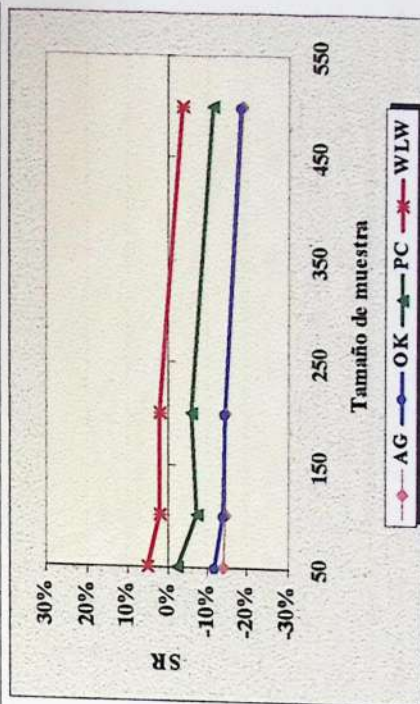


Figura 6.2.6.: SR del estimador de b para cada modelo según el tamaño de muestra, en eventos diferentes.

Tabla 6.2.4. Cambio del Sesgo Relativo (SR) de los estimadores del parámetro $b = -2$ para las diferentes situaciones y según el tamaño de muestra

Tamaño de muestra	Cambio del Sesgo Relativo (SR)							
	Eventos Ordenados			Eventos no Ordenados				
	AG	OK	PC	WLW	AG	OK	PC	WLW
500	0%	0%	0%	0%	0%	0%	0%	0%
200	3%	4%	3%	4%	4%	4%	5%	6%
100	5%	3%	3%	5%	4%	4%	5%	6%
50	5%	5%	6%	8%	5%	7%	9%	9%

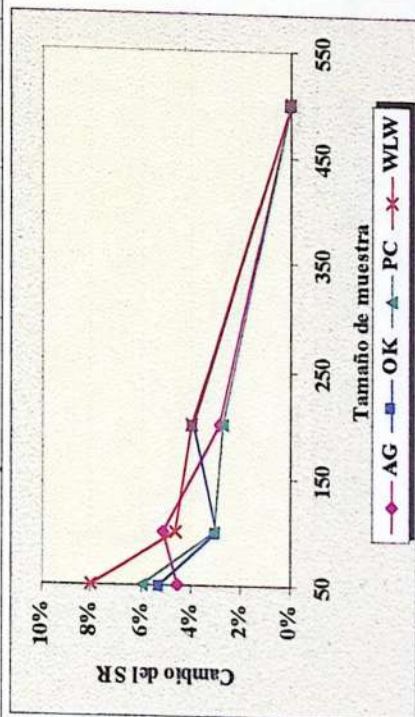
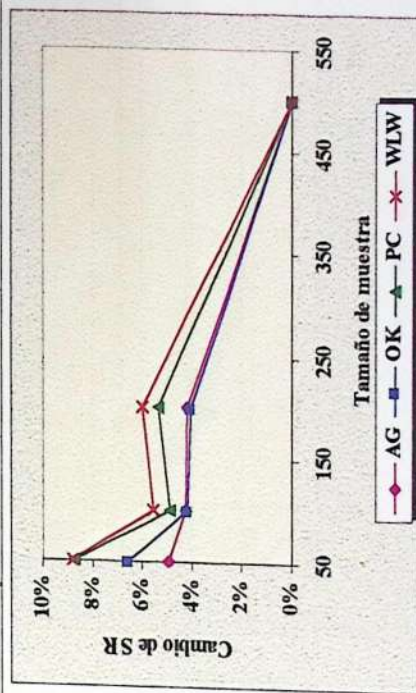
Figura 6.2.7.: Cambio SR del estimador de b para cada modelo según el tamaño de muestra, en eventos recurrentes.Figura 6.2.8.: Cambio del SR del estimador de b para cada modelo según el tamaño de muestra, en eventos diferentes.

Tabla 6.2.5. Error Estándar (SE) de los estimadores del parámetro $b = -2$ para las diferentes situaciones y según el tamaño de muestra.

Tamaño de muestra	Eventos Ordenados				Eventos no Ordenados			
	AG	OK	PC	WLW	AG	OK	PC	WLW
500	0.08	0.03	0.06	0.07	0.04	0.04	0.04	0.05
200	0.15	0.16	0.14	0.20	0.13	0.14	0.14	0.14
100	0.18	0.18	0.21	0.24	0.17	0.18	0.17	0.19
50	0.28	0.30	0.26	0.37	0.20	0.22	0.24	0.24

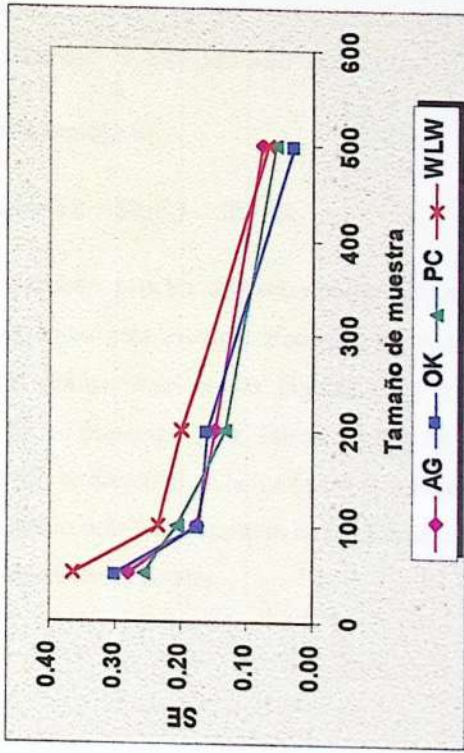


Figura 6.2.9.: SE del estimador de b para cada modelo según el tamaño de muestra, en eventos recurrentes.

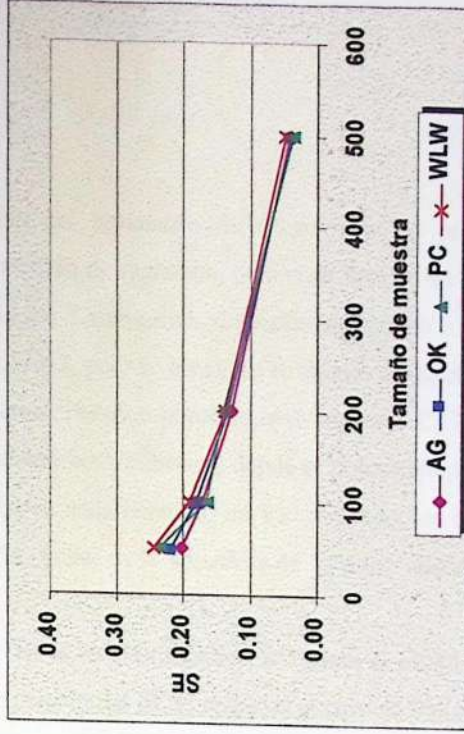


Figura 6.2.10.: SE del estimador de b para cada modelo según el tamaño de muestra, en eventos diferentes.

3. Efecto de las censuras

El efecto del aumento del porcentaje de censuras en la precisión del estimador del parámetro del modelo de regresión, para cada una de las propuestas, se analiza mediante simulaciones. Las muestras simuladas responden a situaciones como las descritas en la Sección 2, pero en este caso se mantiene fijo el tamaño de muestra, $n = 500$, y se consideran diferentes porcentajes de censuras C_i . Para ello, se usa un esquema de censuras aleatorias Uniformes, donde el porcentaje promedio de datos censurados considerados en cada conjunto de 100 muestras, es de 0%, 1%, 10%, 21%, 25%, 48% y 60%, tanto en la situación de eventos diferentes como eventos recurrentes.

También, al igual que en la sección anterior, se estima el ECM, SR, y SE. Además, se calcula la tasa de cambio del ECM respecto al caso de 0% de censuras mediante, la expresión

$$\text{ECM relativo} = \text{ECM}(C_i) / \text{ECM}(0\%),$$

y el cambio del SR a través de la expresión

$$\text{Cambio SR} = \text{SR}(C_i) - \text{SR}(0\%).$$

Los valores del ECM estimado, para los diferentes porcentajes de censuras, tanto para eventos recurrentes como para eventos diferentes, se muestran en la Tabla 6.3.1, y se representan gráficamente en las Figuras 6.3.1 y 6.3.2; los correspondientes valores del SR se presentan en las Tablas 6.3.3, y en las figuras 6.3.5 y 6.3.6. Los valores del SE, se muestran en la Tabla 6.3.5, y Figuras 6.3.9 y 6.3.10. El ECM relativo y el cambio del SR se presentan en las Tablas 6.3.2 y 6.3.4 respectivamente, con sus correspondientes figuras.

Tabla 6.3.1. Error cuadrático Medio (ECM) de los estimadores del parámetro $b = -2$ para las diferentes situaciones según el porcentaje de censuras.

Porcentaje de Censuras	Error Cuadrático Medio									
	Eventos Ordenados					Eventos no Ordenados				
	AG	OK	PC	WLW	WLV	AG	OK	PC	WLW	WLV
0%	0.020	0.025	0.016	0.352	0.352	0.070	0.069	0.013	0.004	0.004
1%	0.011	0.007	0.003	0.327	0.327	0.076	0.075	0.015	0.003	0.003
10%	0.009	0.009	0.005	0.331	0.331	0.083	0.083	0.018	0.002	0.002
21%	0.023	0.020	0.011	0.405	0.405	0.102	0.099	0.028	0.002	0.002
25%	0.037	0.031	0.016	0.482	0.482	0.099	0.094	0.031	0.003	0.003
48%	0.061	0.082	0.041	0.721	0.721	0.141	0.132	0.074	0.013	0.013
60%	0.104	0.176	0.132	1.097	1.097	0.139	0.124	0.077	0.015	0.015

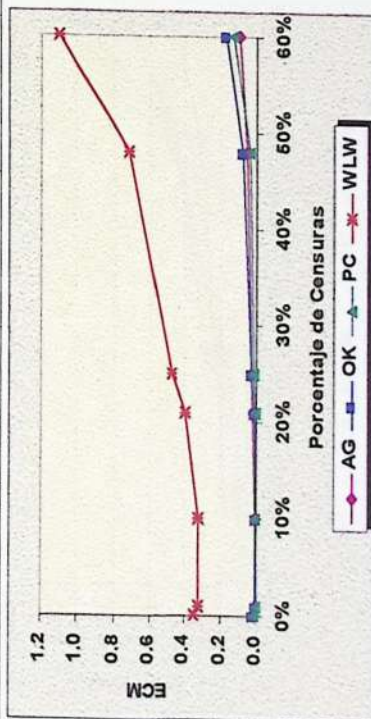


Figura 6.3.1.: ECM del estimador de b para cada modelo según el porcentaje de censuras, en eventos recurrentes.

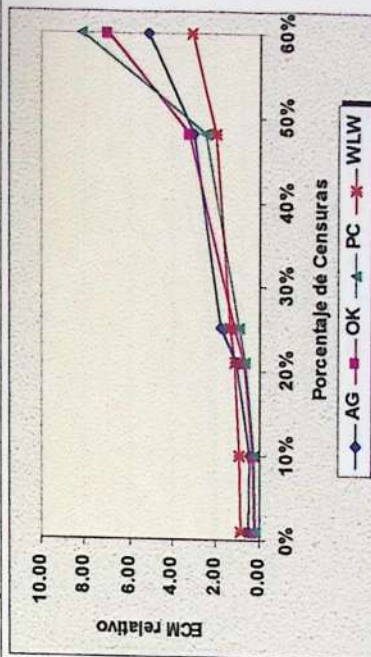


Figura 6.3.2.: ECM del estimador de b para cada modelo según el porcentaje de censuras, en eventos diferentes.

Tabla 6.3.2. Error cuadrático Medio relativo (ECM) de los estimadores del parámetro $b = -2$ para las diferentes situaciones según el porcentaje de censuras.

Porcentaje de Censuras	Error Cuadrático Medio relativo							
	Eventos Ordenados			Eventos no Ordenados				
	AG	OK	PC	WLW	AG	OK	PC	WLW
1%	0.563	0.275	0.203	0.927	1.088	1.088	1.184	0.645
10%	0.466	0.354	0.290	0.939	1.187	1.206	1.403	0.573
21%	1.143	0.779	0.703	1.151	1.458	1.438	2.148	0.507
25%	1.811	1.209	0.993	1.367	1.425	1.371	2.352	0.743
48%	3.025	3.247	2.495	2.046	2.026	1.914	5.684	3.056
60%	5.140	6.966	8.096	3.113	1.995	1.810	5.883	3.420

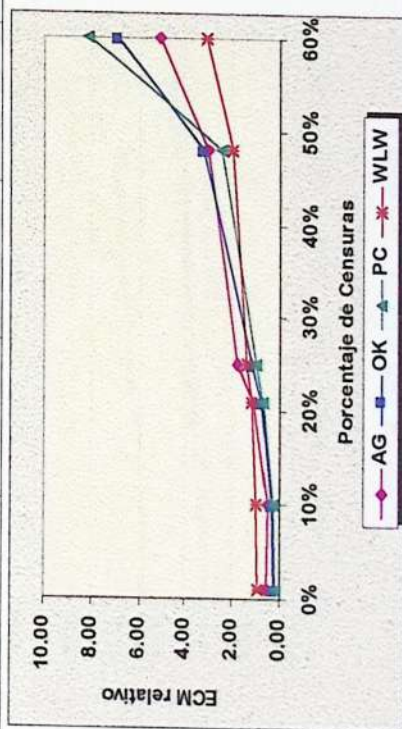


Figura 6.3.3: ECM relativo del estimador de b para cada modelo según el porcentaje de censuras, en eventos recurrentes.

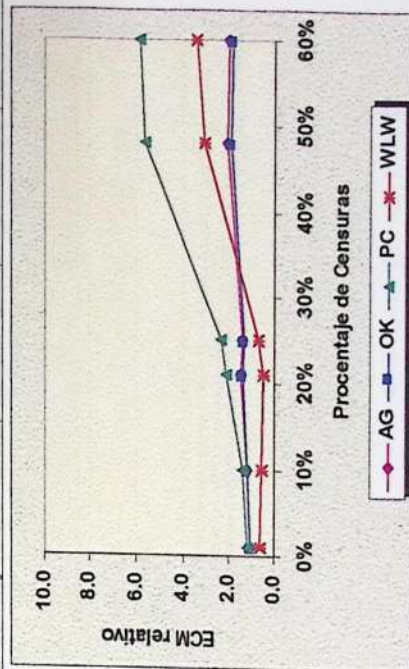


Figura 6.3.4: ECM relativo del estimador de b para cada modelo según el porcentaje de censuras, en eventos diferentes.

Tabla 6.3.3. Sesgo Relativo (SR) de los estimadores del parámetro $b = -2$ para las diferentes situaciones y según el porcentaje de censuras.

Porcentaje de Censuras	Sesgo Relativo (SR)							
	Eventos Ordenados			Eventos no Ordenados				
	AG	OK	PC	WLW	AG	OK	PC	WLW
0%	2%	3%	3%	28%	-13%	-13%	-6%	3%
1%	3%	2%	3%	28%	-14%	-14%	-6%	2%
10%	3%	2%	3%	28%	-14%	-14%	-7%	2%
21%	5%	5%	5%	31%	-16%	-16%	-8%	0%
25%	7%	7%	6%	34%	-16%	-15%	-8%	0%
48%	9%	11%	9%	41%	-19%	-18%	-13%	-4%
60%	14%	20%	17%	51%	-18%	-17%	-14%	-5%

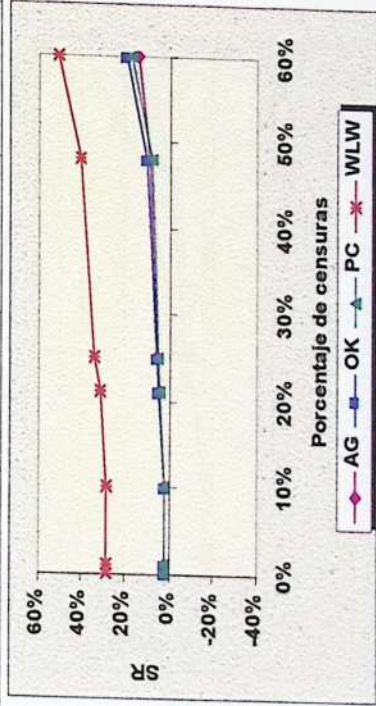


Figura 6.3.5.: SR del estimador de b para cada modelo según porcentaje de censuras, en eventos recurrentes.

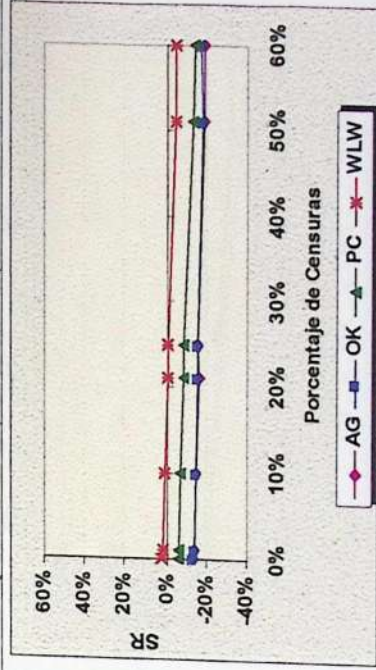


Figura 6.3.6.: SR del estimador de b para cada modelo según el porcentaje de censuras, en eventos diferentes.

Tabla 6.3.4. Cambio del Sesgo Relativo (SR) de los estimadores del parámetro $b=-2$ para las diferentes situaciones y según el porcentaje de censuras.

Porcentaje de Censuras	Cambio del Sesgo Relativo (SR)									
	Eventos Ordenados					Eventos no Ordenados				
	AG	OK	PC	WLW	WLV	AG	OK	PC	WLV	WLV
0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
1%	0%	0%	0%	0%	0%	-1%	-1%	-1%	-1%	-1%
10%	0%	0%	0%	0%	0%	-1%	-1%	-1%	-1%	-1%
21%	3%	3%	2%	3%	3%	-3%	-3%	-3%	-3%	-3%
25%	4%	4%	3%	6%	6%	-2%	-2%	-3%	-4%	-4%
48%	7%	8%	6%	12%	12%	-5%	-5%	-8%	-7%	-7%
60%	12%	17%	15%	23%	23%	-5%	-4%	-8%	-8%	-8%

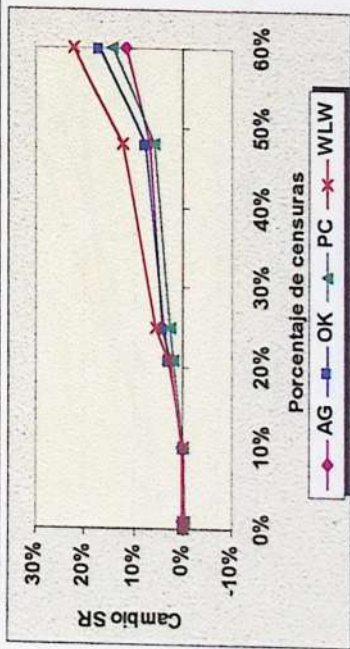


Figura 6.3.7.: Cambio SR del estimador de b para cada modelo según el porcentaje de censuras, en eventos recurrentes

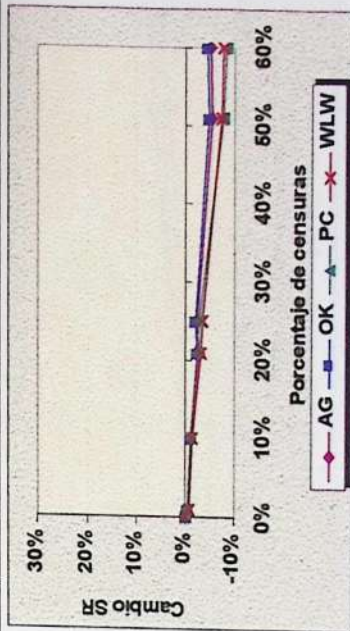


Figura 6.3.8.: Cambio SR del estimador de b para cada modelo según el porcentaje de censuras, en eventos diferentes

Tabla 6.3.5. Error Estándar (SE) de los estimadores del parámetro $b=-2$ para las diferentes situaciones y según el porcentaje de censuras.

Porcentaje de Censuras	Error Estándar (SE)							
	Eventos Ordenados			Eventos no Ordenados				
	AG	OK	PC	WLW	AG	OK	PC	WLW
0%	0.13	0.15	0.11	0.17	0.02	0.02	0.03	0.03
1%	0.09	0.06	0.02	0.09	0.02	0.02	0.03	0.02
10%	0.08	0.08	0.04	0.09	0.04	0.03	0.03	0.04
21%	0.12	0.09	0.05	0.12	0.04	0.04	0.04	0.05
25%	0.14	0.12	0.06	0.15	0.04	0.05	0.05	0.06
48%	0.17	0.19	0.10	0.23	0.06	0.08	0.06	0.08
60%	0.15	0.15	0.10	0.24	0.07	0.06	0.04	0.08

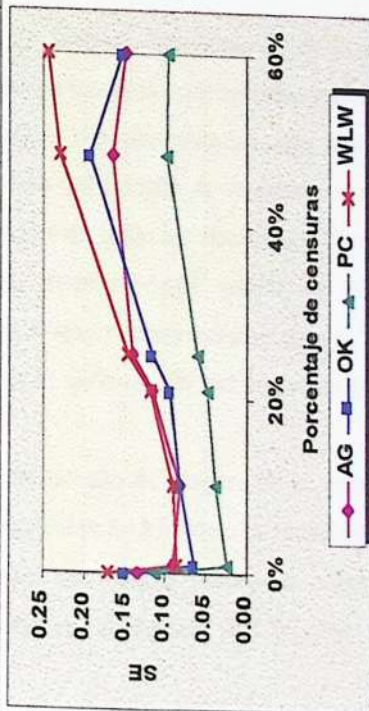


Figura 6.3.9.: SE del estimador de b para cada modelo según el porcentaje de censuras, en eventos recurrentes.

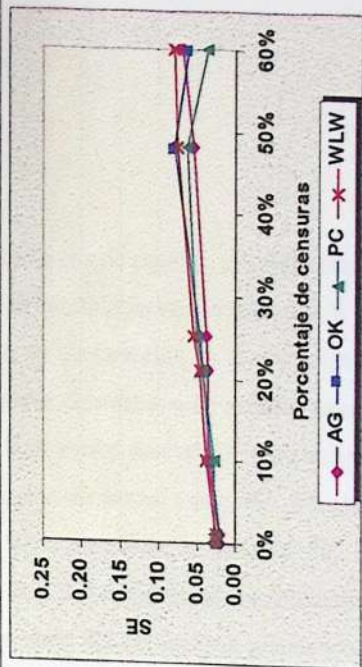


Figura 6.3.10.: SE del estimador de b para cada modelo según el porcentaje de censuras, en eventos diferentes.

4. Conclusiones

⇒ Efecto del tamaño de muestra

A partir de estos resultados, y como era de esperar, se puede observar que la precisión del estimador del parámetro del modelo se ve afectada por la disminución del tamaño de la muestra. Sin embargo, la disminución en la precisión es diferente según el modelo y según se trate de eventos diferentes o de eventos recurrentes.

Basado en el ECM relativo, en la situación de eventos recurrentes, el modelo más afectado por la disminución del tamaño de muestra es el PC. Sin embargo, si se observan los valores de ECM, aún para muestras de tamaño 50, la precisión correspondiente a este modelo es similar al de los modelos condicionales. El menor ECM relativo si $n=50$ lo presenta el modelo WLW, en esta misma situación, pero mantiene los valores más elevados del SR y del SE para cualquier tamaño de muestra.

Mientras que, en la situación de eventos diferentes, llama la atención la disminución del ECM que experimentan los modelos condicionales y el modelo PC, si el tamaño de muestra disminuye hasta $n=100$. Sin embargo, esto se explica si se considera que el efecto de la disminución del tamaño de muestra, es aumentar el valor del estimador en todos los modelos y en todas las situaciones. Y, puesto que, en la situación de eventos diferentes, tales modelos subestiman el valor del parámetro si $n=500$, esto conduce a que el sesgo tienda a disminuir y así aumenta la precisión del estimador. Por otra parte, la varianza del estimador es pequeña aún para tamaños de muestra chicos.

Ahora, con respecto al efecto del tamaño de muestra en la variabilidad del estimador, es similar en todos los modelos si se está frente a la situación de eventos no ordenados. Sin embargo, en el caso de eventos ordenados o recurrentes la variabilidad aumenta más rápidamente en el modelo WLW. Por otra parte,

cualquiera sea la situación, el efecto diferencial del tamaño de muestra en la precisión, es más apreciable en el sesgo que en la variabilidad.

⇒ *Efecto del porcentaje de censuras*

Como se esperaba, el aumento del porcentaje de censuras disminuye la precisión del estimador del parámetro del modelo de regresión, siendo el efecto diferente según se trate de eventos diferentes o eventos recurrentes. Lo cual se explica perfectamente si se tiene en cuenta que, en el caso de eventos ordenados, una censura de la primera recurrencia implica que la segunda recurrencia tampoco se observa, cosa que no ocurre en los eventos no ordenados.

Cabe destacar que, en todos los modelos y en todas las situaciones, si el porcentaje de censuras no supera el 25%, no hay un efecto importante en la precisión, como ya lo afirmaron Prentice y Cai respecto al modelo propuesto por ellos.

En la situación de eventos recurrentes, el cambio relativo del ECM es similar en todos los modelos si el porcentaje de censuras C_i es menor al 50%; pero, para valores mayores comienza a observarse un efecto diferencial entre los modelos. Si bien el menor ECM relativo lo presenta el modelo WLW, cuando el porcentaje de censuras es del orden del 60%, el correspondiente valor del ECM es casi 10 veces mayor que el correspondiente al modelo AG. Por otra parte, los valores menores que uno que presenta el ECM relativo para los modelos cuando C_i es igual al 1% ó 10% se debe principalmente a que si los datos no son censurados, se presentan problemas de convergencia en el algoritmo computacional, y esto produce una mayor variabilidad en el estimador, como se puede observar en la Tabla 6.3.5 de valores del SE. Con respecto al aumento del SR con el porcentaje de censuras, se observa homogeneidad en los modelos condicionales y PC, mientras que en el modelo WLW el SR crece más rápidamente después del 25%. Por último, un aspecto que debe resaltarse, en esta situación, es que el efecto diferencial del aumento del porcentaje de censuras, es más apreciable en el SE que en el SR.

Analizando la situación de eventos diferentes, llama la atención el pequeño efecto que tiene el aumento del porcentaje de censuras en el ECM de los modelos condicionales. Por otra parte, las diferencias observadas del efecto antes mencionado entre los modelos condicionales y marginales, se deben fundamentalmente a que el aumento de las censuras afecta principalmente al SR de estos últimos. Por todo esto, en la situación de eventos diferentes el porcentaje de censuras no sería un criterio a tener en cuenta en la elección de los modelos.

CAPITULO VII

ANALISIS DE LOS RESULTADOS Y CONCLUSIONES

Como ya se precisó en el Capítulo I, para el estudio del efecto de covariables sobre el riesgo de falla, en el caso de tener datos de sobrevida multivariados, varios modelos de regresión han sido propuestos usando una extensión del modelo de "hazard" proporcional debido a Cox para el caso univariado. Sin embargo, en el contexto multivariado surgen algunos aspectos particulares que como ser: (a) la presentación de dos tipos de situaciones marcadamente diferentes, los eventos pueden ser del mismo tipo y en ese caso se presentan de manera ordenada (eventos recurrentes u ordenados), o los eventos pueden ser de diferentes tipos y ellos pueden presentarse en cualquier orden (eventos diferentes o no ordenados); (b) los tiempos de falla dentro de un mismo sujeto pueden estar correlacionados. Atendiendo a estas particularidades, algunas propuestas modelan el "hazard" condicional de los tiempos de falla (modelos condicionales), y de esta manera naturalmente tienen en cuenta el orden de los eventos, mientras que otras modelan el "hazard" marginal de los tiempos de falla (modelos marginales). La discusión sobre la bondad de las diferentes propuestas para estimar de manera eficiente el verdadero efecto de las covariables se centró en la consideración del orden de los eventos y en el manejo de la correlación entre los tiempos de falla dentro de un mismo individuo. Intentando poner luz sobre esta discusión, durante el desarrollo de este trabajo, surgió otro interrogante que no se encontró planteado en la literatura, y es el referido a la suposición que tienen todos los modelos propuestos de que el proceso subyacente que generan los datos no tiene memoria. Usando los conceptos debidos a Ryu (1993) quien define dos tipos de memoria (marginal y conjunta) en una distribución bivariada, se estudió el comportamiento de los diferentes modelos cuando no se cumplía la propiedad de falta de memoria de los procesos subyacentes.

Los estudios de Monte Carlo muestran que el problema más importante que presentan los estimadores de los parámetros de los modelos de regresión tipo Cox,

cuando no se cumplen los supuestos bajo los cuales se derivaron sus propiedades, es la introducción de sesgos, que en algunos casos alcanzan el orden del 50% del verdadero valor del parámetro, ya sea subestimando o sobrestimando dicho valor. Si bien en algunas situaciones el Error Estándar del estimador es casi tres veces mayor que el que se obtiene cuando los datos cumplen con los supuestos del modelo, la variabilidad de los estimadores es tan pequeña que se puede desestimar frente al aumento del sesgo. Esto motivó que en esta tesis el estudio analítico se concentrara en buscar cotas para los sesgos.

Por razones de claridad en el análisis, el estudio del efecto de la correlación entre los tiempos de falla, de la memoria del proceso, del tamaño de muestra y del porcentaje de censuras, se analiza por separado, aunque, como se explicita en el análisis, los sesgos actúan de una manera acumulativa, compensando sus efectos en algunos casos o aumentándolos en otros.

Por otra parte, es importante destacar que en la situación de eventos no ordenados o diferentes además de los sesgos que se producen por falta de cumplimiento de los supuestos, en los modelos AG, OK y PC se generan sesgos por consideraciones respecto a cómo se ordenan los datos para usar el programa computacional propuesto por Therneau (1995), que aquí se denominan *sesgos por ordenamiento*. Sin embargo, estos sesgos por ordenamiento generados por el modelo PC responden más a la sugerencia de Therneau de considerar que el modelo responde a una estructura de proceso Markov, que parece ser más apropiada en el caso de eventos ordenados o recurrentes, y no a un proceso semi Markov como lo puntualizan Prentice y Cai (1992) en su propuesta, que afirman que la elección de uno u otro tipo de modelo dependerá de cada situación particular que se estudie. Otro problema adicional se presenta en el modelo WLW en la situación de eventos ordenados que proviene de considerar que los individuos están expuestos al riesgo de tener cualquiera de las fallas desde el comienzo de la observación. Esto, que tiene sentido en el caso de eventos no ordenados, produce importantes sesgos en el caso de eventos recurrentes, como ya lo sugerían algunos autores y lo enfatizó

Oakes en su publicación de 1997 y en la discusión que él hace del trabajo publicado por Wei y Glidden (1997).

A partir de estas consideraciones y de las expresiones para las cotas encontradas, se puede inferir que en el caso de los modelos marginales (PC y WLW) este tipo de sesgos se puede evitar usando los *tiempos intereventos* de cada falla, en ambos modelos en la situación de eventos ordenados o recurrentes y usando los *tiempos totales* (desde el comienzo de la observación) de las fallas en la situación de eventos no ordenados o diferentes. Por otra parte, el efecto del ordenamiento en los modelos condicionales, es similar al del modelo PC si la correlación es menor de 0.6. Sin embargo, para correlaciones entre los tiempos de falla superiores a 0.6, el valor del sesgo generado por usar el "hazard" condicional en la ecuación de verosimilitud parcial, empieza a ser importante frente al sesgo por ordenamiento, si el parámetro θ_c es pequeño (ecuación 4.4.2.3 y Fig. 4.2.15).

Un análisis global permite afirmar que, en general, los cuatro modelos estudiados son más robustos a la falta de cumplimiento del supuesto de independencia y al aumento del porcentaje de censuras, en la situación de eventos no ordenados o diferentes que en la de eventos ordenados o recurrentes. Sin embargo, de manera inversa, ellos se muestran más robustos en la situación de eventos recurrentes, cuando no se cumple el supuesto de falta de memoria del proceso subyacente. La disminución del tamaño de muestra aumenta el sesgo de todos los modelos de una manera similar en ambas situaciones, aunque el efecto diferencial se observa en el cambio de variabilidad, que es más pronunciado en la situación de eventos recurrentes y se repite cuando aumenta el porcentaje de censuras.

Con respecto al porcentaje de censuras, si éste no es superior al 25% no se observa un efecto diferencial entre los modelos.

Otro punto importante de destacar es que en la literatura no se encontró ninguna preocupación por los problemas que podrían surgir si no se cumple el supuesto de falta de memoria en los procesos subyacentes, que se presenta con frecuencia en la práctica, como lo destaca Ryu (1993). Sin embargo, como se

demuestra en este estudio, la precisión de los estimadores se ve más afectada por la memoria de los procesos que por la falta de correlación entre los tiempos de falla. Este problema resulta aún más complicado si se tiene en cuenta que es más sencillo verificar la presencia de correlación entre los tiempos de falla que la memoria de los procesos, especialmente la memoria conjunta. En efecto, la falta de memoria marginal, equivalente a decir que las marginales son exponenciales, no nos dice nada sobre la de memoria conjunta del proceso, y hasta ahora no se encontró en la literatura ninguna propuesta para someter a un test esta propiedad. De esta manera, sólo el conocimiento "a priori" de la situación particular que se está analizando podría dar alguna información sobre este punto.

Debido a las diferencias de comportamiento de los modelos para las diferentes situaciones, a continuación se analiza por separado cada una de ellas:

A. En la situación de *eventos recurrentes*:

La tasa de cambio del SR, con respecto al sesgo bajo independencia (Fig. 4.2.14), muestra que no hay grandes diferencias entre los modelos con respecto a la velocidad de crecimiento de la subestimación del parámetro, a excepción del modelo OK que aparece como más sensible a los efectos del aumento de la correlación. Esto parece contradecir la propuesta de Oakes que intenta corregir el modelo AG introduciendo un factor que capture el efecto de la correlación dentro de los individuos y así, indirectamente, contemplan la correlación entre los tiempos de falla. Sin embargo, esta aparente contradicción se debe a que los datos tienen una distribución Sarkar, la cual es una distribución con memoria conjunta, y en el estudio sobre el efecto de la memoria del proceso en el sesgo de los estimadores este modelo aparece como el más afectado si el proceso subyacente posee este tipo de memoria.

Si se tiene en cuenta el efecto de la memoria de los procesos en los sesgos de los estimadores, se puede inferir que los modelos condicionales son los más robustos a la presencia de uno sólo de los tipos de memoria (marginal o conjunta), mientras que los modelos marginales son más sensibles a la memoria marginal. Sin

embargo, y a causa del efecto compensador del sesgo por ordenamiento, el modelo WLW es el que presenta el menor SR, si el proceso subyacente de los datos posee ambos tipos de memoria.

Por último, en la situación de eventos recurrentes, y teniendo en cuenta todas las posibles fuentes de sesgos estudiadas, *los modelos AG y PC aparecen como los más recomendables en cualquier caso, a excepción de que se tenga fuerte evidencia de que el proceso posee ambos tipos de memoria y la correlación entre los tiempos de falla sea muy pequeña o nula, y en ese caso se sugiere usar el modelo WLW.*

B. En la situación de *eventos diferentes*:

En la situación de eventos diferentes, la mayor pérdida de precisión de los estimadores de los modelos, llega a través de la memoria del proceso suyaente generador de los datos. Por otra parte, una particularidad que se presenta en esta situación, es la similitud entre los modelos marginales y entre los modelos condicionales, con respecto al cambio en el SR de los modelos cuando aumenta el grado de correlación entre los tiempos de falla, similitud que sería aún más acentuada, si en el modelo PC se considera que los datos responden a un proceso semi-Marcov. Esto se repite, además, cuando se considera el efecto del tamaño de muestra y el aumento del porcentaje de censuras.

En los modelos marginales, que aparecen como muy robustos con el aumento de correlación entre los tiempos de falla, la magnitud del sesgo se hace muy significativa si el proceso posee memoria marginal, sin importar la memoria conjunta, que lo único que produce es un cambio de signo. Por otra parte, si el proceso posee sólo memoria marginal, el sesgo de los modelos condicionales es casi 10 veces menor que el de los modelos marginales.

Por último, en la situación de eventos diferentes, y teniendo en cuenta todas las posibles fuentes de sesgos estudiadas, *los modelos marginales, con la corrección propuesta para el modelo PC, aparecen como los más recomendables si la distribución posee memoria marginal, en caso contrario se debería usar cualquiera de los modelos condicionales.*

Los objetivos de esta tesis quedan hasta aquí alcanzados, pero a partir de los problemas que se ponen en evidencia en el análisis de los resultados y de otros que se insinúan en el desarrollo de esta investigación, surgen nuevos interrogantes que puede dar lugar a futuras investigaciones, como ser:

- ⇒ ¿Cómo se verifica, a partir de los datos, el cumplimiento de la falta de memoria conjunta de procesos multivariados?
- ⇒ Si la distribución subyacente de los datos posee ambos tipos de memoria, ¿qué modificación debería hacerse a los modelo tipo Cox para aumentar su precisión, o sería conveniente usar otro modelo diferente?
- ⇒ ¿Cómo afectaría la violación del supuesto que los hazard condicionales sean proporcionales, fundamentalmente en el caso de eventos ordenados?

ANEXO 1

CONCEPTOS Y RESULTADOS ADICIONALES DE LA TEORÍA DE
 PROCESOS DE CONTEO APLICADOS AL ANÁLISIS DE SOBREVIDA

Definición 1: Un *proceso estocástico* es una familia de variables aleatorias $X = \{X(t): t \in \Gamma\}$ indexada por un conjunto Γ , definidas en un espacio de probabilidad $(\Omega, \mathfrak{F}, P)$.

Aquí se considera $\Gamma = \mathbb{R}^+ = [0, t)$.

Definición 2: $X(\cdot, \omega): \mathbb{R}^+ \rightarrow \mathbb{R}$, $\omega \in \Omega$ se denomina *una trayectoria o una realización de X*.

Definición 3: Sea X un proceso estocástico, entonces, se dice que X es *integrable* si

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

Definición 4: Sea X un proceso estocástico, entonces, se dice que X es *cuadrado integrable* si

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

Definición 5: Sea X un proceso estocástico, entonces, se dice que X es *acotado* si y sólo si existe una constante finita G tal que

$$P \left[\sup_{0 \leq t < \infty} |X(t)| < G \right] = 1$$

Definición 6: Una familia de sub σ -álgebras $\{\mathfrak{F}_t: t \geq 0\}$ de una σ -álgebra de \mathfrak{F} se dice que es *creciente*, si para $s \leq t$, entonces $\mathfrak{F}_s \subset \mathfrak{F}_t$.

Definición 7: Una familia creciente de sub σ -álgebras se llama una *filtración*.

Definición 8: Una filtración $\{\mathfrak{F}_t: t \geq 0\}$ se dice continua a derecha si y sólo si para todo t $\mathfrak{F}_{t+} = \mathfrak{F}_t$.

La mayoría de las filtraciones son historias de procesos estocásticos.

Definición 9: Una *base estocástica* es un espacio de probabilidad $(\Omega, \mathfrak{F}, P)$ equipado con una filtración continua a derecha $\{\mathfrak{F}_t: t \geq 0\}$ y que denotaremos como $(\Omega, \mathfrak{F}, \{\mathfrak{F}_t: t \geq 0\}, P)$.

Definición 10: Una base estocástica se dice *completa* si \mathfrak{F} contiene todos los subconjuntos de P-medida nula.

Definición 11: Un proceso estocástico se dice *adaptado a una filtración* $\{\mathfrak{F}_t: t \geq 0\}$ si para todo $t \geq 0$, $X(t)$ es \mathfrak{F}_t -medible.

Cualquier proceso es adaptado a su historia.

Definición 12: Un proceso estocástico $\{N(t): t \geq 0\}$ es un *proceso de conteo* si:

1. Es adaptado a una filtración $\{\mathfrak{F}_t: t \geq 0\}$
2. $N(0) = 0$
3. $N(t) < \infty$
4. Sus realizaciones cumplen con las siguientes propiedades:
 - i) Son continuas a derecha con probabilidad 1
 - ii) Son constantes a trozo
 - iii) Tienen sólo discontinuidades a salto
 - iv) Los saltos son de tamaño +1

A partir de la definición se puede mostrar que $N(t) - N(s)$ es el número de eventos que ocurren en el intervalo $(s, t]$. Más aún, si N es un proceso de conteo y f una función de t (posiblemente aleatoria), si $0 \leq s < t \leq \infty$ entonces $\int_s^t f(u) dN(u)$ es la *representación integral de Stieljes* de la suma de los valores de f en los saltos de N en el intervalo $(s, t]$.

Para pensar el proceso estocástico en términos de la distribución condicional de su futuro comportamiento dada su evolución actual, es decir, en términos de $X(t+s)$ dado $\{X(u): 0 \leq u \leq t\}$ se hace necesario dar una definición general de esperanza condicional. Esta definición surge, intuitivamente, de la definición estándar si se piensa que condicionar sobre la trayectoria del proceso hasta el tiempo t es equivalente a condicionar sobre la σ -álgebra autogenerada por el mismo proceso, es decir, sobre $\sigma\{X(u): 0 \leq u \leq t\}$.

Definición 13: Sea Y es una variable aleatoria sobre un espacio de probabilidad $(\Omega, \mathfrak{F}, P)$, \wp una sub σ -álgebra de \mathfrak{F} , y X una variable aleatoria que satisface las siguientes condiciones:

1. X es \wp -medible, y
2. $\int_B Y dP = \int_B X dP$ para todo $B \in \wp$.

Entonces la variable X es llamada la *esperanza condicional de Y dado \wp* , y es denotada por $E(Y|\wp)$.

Definición 14: Sea $X = \{X(t): t \geq 0\}$ un proceso estocástico continuo a derecha y con límites a izquierda; y, sea $\{\mathfrak{F}_t: t \geq 0\}$ una filtración, definida sobre un espacio de probabilidad común, entonces, se dice que X es una *martingala con respecto a $\{\mathfrak{F}_t: t \geq 0\}$* si:

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

Definición 15: X es llamada una *submartingala* si la condición 3 de la Definición 14 es reemplazada por

$$E\{X(t+s) | \mathfrak{F}_t\} \geq X(t) \text{ a.s.}$$

Definición 16: X es llamada una *supermartingala* si la condición 3 de la Definición 14 es reemplazada por

$$E\{X(t+s) | \mathfrak{F}_t\} \leq X(t) \text{ a.s.}$$

Definición 17: Sea $(\Omega, \mathfrak{F}, P)$ un espacio de probabilidad con filtración $\{\mathfrak{F}_t; t \geq 0\}$. La σ -álgebra en $[0, \infty) \times \Omega$ generada por todos los conjuntos de la forma

$$[0] \times A, \quad A \in \mathfrak{F}_0 \quad \text{y} \quad (a, b] \times A, \quad 0 \leq a < b < \infty, A \in \mathfrak{F}_a$$

es llamada la σ -álgebra *predictible* para la filtración $\{\mathfrak{F}_t; t \geq 0\}$.

Definición 18: Un proceso X es *predictible con respecto a una filtración*, si X es medible con respecto a la σ -álgebra predictible generada por dicha filtración. Llamaremos entonces a X un *proceso \mathfrak{F}_t -predictible*.

Intuitivamente, un proceso X se dice que es *predictible* si su comportamiento en t es determinado por el comportamiento en $[0, t)$ para cualquier t .

Teorema 1 (descomposición de Doob-Meyer): Sea X una submartingala no negativa continua a derecha con respecto a una base estocástica $(\Omega, \mathfrak{F}, \{\mathfrak{F}_t; t \geq 0\}, P)$. Entonces existe una martingala continua a derecha y un proceso predictible creciente continuo a derecha Λ , tal que $E\Lambda(t) < \infty$ y

$$X(t) = M(t) + \Lambda(t) \text{ a.s.}$$

Para todo $t \geq 0$. Si $\Lambda(0) = 0$ a.s., y si $X = M' + \Lambda'$ es otra descomposición con $\Lambda'(0) = 0$, entonces para todo $t \geq 0$,

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

El teorema de la descomposición establece que para cualquier submartingala X no negativa y continua a derecha, hay un único proceso predecible creciente y continuo a derecha Λ tal que $\Lambda(0) = 0$ y $X - \Lambda$ es una martingala. Ya que cualquier proceso adaptado no negativo y creciente con esperanza finita es una submartingala, entonces, para cualquier proceso de conteo N con esperanza finita, hay un único proceso Λ , tal que $N - \Lambda$ es una martingala.

Corolario 1: Sea $\{N(t): t \geq 0\}$ un proceso de conteo adaptado a una filtración $\{\mathfrak{F}_t: t \geq 0\}$ con $EN(t) < \infty$ para todo t . Entonces, existe un único proceso \mathfrak{F}_t -predecible Λ creciente y continuo a derecha tal que $\Lambda(0) = 0$ a.s., $E\Lambda(t) < \infty$ para todo t , y $\{M(t) = N(t) - \Lambda(t): t \geq 0\}$ es una \mathfrak{F}_t -martingala continua a derecha.

Definición 19: El proceso Λ de la descomposición de Doob-Meyer es llamado el *compensador* para la submartingala X .

El método de martingala para los modelos estadísticos con procesos de conteo, es útil solo en las situaciones en las cuales Λ es conocido o puede ser calculado. Una interpretación más intuitiva de la descomposición de Dobb-Meyer es considerar que M es el proceso puro ruido, entonces la parte sistemática de un proceso de conteo es su compensador: Un proceso predecible que varía suavemente, el cual sustraído del proceso de conteo deja un proceso ruido no predecible de media cero.

Definición 20: Sea M una martingala continua a derecha con respecto a una filtración $\{\mathfrak{F}_t: t \geq 0\}$, si se supone que $EM^2(t) < \infty$ para todo $t \geq 0$; entonces, existe un único proceso predecible continuo a derecha y creciente $\langle M, M \rangle$, llamado la *variación cuadrática predecible* de M o

proceso de variación predecible de M , tal que $(M, M)(0) = 0$ a.s., $E(M, M)(t) < \infty$ para cada t , y $\{M^2(t) - (M, M)(t) : t \geq 0\}$ es una martingala continua a derecha.

Si se considera el proceso M^2 , aunque M sea puro ruido, M^2 tiene una tendencia a crecer con el tiempo. De la definición se puede inferir que la componente sistemática (su compensador) de este proceso es precisamente el proceso de variación predecible. Por otra parte, ya que $d(M, M)(t) = E[dM^2(t)|\mathfrak{F}_{t-}]$, se puede evaluar los incrementos del compensador de M^2 mediante,

$$\begin{aligned} d(M^2)(t) &= M[(t+dt)-]^2 - M(t-)^2 \\ &= [M(t-) + dM(t)]^2 - M(t-)^2 \\ &= [dM(t)]^2 + 2dM(t)M(t-) \end{aligned}$$

Puesto que $E[dM(t)M(t-)|\mathfrak{F}_{t-}] = M(t-)E[(dM(t)|\mathfrak{F}_{t-})] = 0$ por definición de martingala, se tiene que

$$E[d(M^2)(t)|\mathfrak{F}_{t-}] = E[(dM(t))^2|\mathfrak{F}_{t-}] = \text{var}[dM(t)|\mathfrak{F}_{t-}]$$

Esta última igualdad muestra que los incrementos del compensador de M^2 es la varianza condicional de los incrementos de M , o sea que el proceso $(M, M)(t)$ puede ser interpretado como la "suma" de la varianza condicional de $dM(s)$, dada la información hasta el tiempo s . El proceso de conteo N es tal que para cada t , $N(t)$ es una variable aleatoria cero-uno con $E[dN(t)|\mathfrak{F}_{t-}] = d\Lambda(t)$, y la $\text{var}[dN(t) - d\Lambda(t)|\mathfrak{F}_{t-}] = \text{var}[dM(t)|\mathfrak{F}_{t-}] = d\Lambda(t)(1 - d\Lambda(t))$. De esta manera, cuando Λ es continua $\text{var}[dM(t)|\mathfrak{F}_{t-}] \approx d\Lambda(t)$, y entonces, el proceso de variación predecible de M , $(M, M)(t)$, es simplemente $\Lambda(t)$. Este sorprendente resultado se relaciona al hecho que para una variable aleatoria Poisson la media y la varianza coinciden y que un proceso de conteo localmente al tiempo t y condicionado sobre el pasado es un proceso Poisson con tasa $\lambda(t)$.

A partir de las propiedades de las martingalas se puede demostrar el siguiente teorema

Teorema 2: Sean M_1 y M_2 martingalas continuas a derecha con respecto a una base estocástica, tales que $E\{M_i(t)\}^2 < \infty$ para todo $t \geq 0$ e $i=1,2$. Entonces, existe un proceso predecible continuo a derecha $\langle M_1, M_2 \rangle$, llamado *proceso de covariación predecible*, con $\langle M_1, M_2 \rangle(0) = 0$ y con $E\langle M_1, M_2 \rangle(t) < \infty$ tal que:

1. $\langle M_1, M_2 \rangle$ es la diferencia de dos procesos predecibles continuos a derecha y crecientes.
2. $M_1 M_2 - \langle M_1, M_2 \rangle$ es una martingala.

La propiedad 1 implica que $\langle M_1, M_2 \rangle$ tiene trayectorias de variación acotada y además se cumple que $d\langle M_1, M_2 \rangle(t) = \text{cov}\{dM_1, dM_2 | \mathfrak{F}_t\}$ y así es muy útil para calcular la covarianza entre martingalas.

Corolario 3: Sean M_1 y M_2 \mathfrak{F}_t -martingalas continuas a derecha tales que $E\{M_i(t)\}^2 < \infty$ para todo $t \geq 0$ e $i=1,2$. Entonces, el proceso $M_1 M_2$ continuo a derecha es una martingala si y sólo si $\langle M_1, M_2 \rangle \equiv 0$. Y en este caso se dice que M_1 y M_2 son ortogonales.

Dos martingalas ortogonales son no correlacionadas si $M_1(0)$ y $M_2(0)$ son no correlacionadas.

Teorema 3: Sea N un proceso de conteo con $EN(t) < \infty$ para todo t , y sea $\{\mathfrak{F}_t : t \geq 0\}$ una filtración continua a derecha tal que:

1. $M = N - \Lambda$ es una \mathfrak{F}_t -martingala, donde $\Lambda = \{\Lambda(t) : t \geq 0\}$ es un proceso \mathfrak{F}_t -predecible con $\Lambda(0) = 0$.
2. H es un proceso \mathfrak{F}_t -predecible y acotado.

Entonces el proceso L dado por

$$L(t) = \int_0^t H(u) dM(u)$$

es una \mathfrak{F}_t -martingala.

Este teorema adquiere una importancia fundamental en la derivación de los estimadores de los coeficientes de regresión, ya que la verosimilitud parcial toma la forma del proceso $L(t)$.

BIBLIOGRAFÍA

1. Aalen, O. O. 1975. Statistical inference for a family of counting processes. Ph.D. dissertation. University of California. Berkeley.
2. Aalen, O. O. 1988. Heterogeneity in survival analysis. *Statistics in Medicine*. 7: 1121-37.
3. Andersen, P. K., and Gill, R. D. 1982. Cox's regression model for counting processes: A large sample study. *Ann. Statist.* 10: 1100-20.
4. Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. 1995. *Statistical Models Based on Counting Processes*. New York: Springer-Verlag.
5. Block, H. y Basu, A. P. 1974. A continuous bivariate exponential extension. *J. Amer. Statist. Assoc.* 69: 1031-37.
6. Breslow, N. E. 1974. Covariance analysis of censored survival data. *Biometrika*. 30: 89-99.
7. Cox, D. R. 1972. Regression models and life tables (with discussion). *J. Roy. Statist. Soc. B.* 34: 187-220.
8. Cox, D. R. 1975. Partial likelihood. *Biometrika*. 62: 269-76
9. Cox, D. R., and Oakes, D. 1984. *Analysis of Survival Data*. London: Chapman and Hall.
10. Efron, B. 1967. The two-sample problem with censored data. *Proc. Fifth Berkeley Symposium on Mathematical Statistics and Probability* 4: 831-53.
11. Efron, B. 1977. The efficiency of Cox's likelihood function for censored data. *J. Am. Statist. Ass.* 72: 557-65.
12. Fleming, T.R., and Harrington, D.P. 1991. *Counting Processes and Survival Analysis*. New York: Wiley.
13. Gill, R. D. 1992. Marginal partial likelihood, *Sacand. J. Statist.* 19: 133-37.
14. Gill, R.D. and Johansen, S. (1990). A survey of product-integration with a view towards application in survival analysis. *Ann. Statist.* 18: 1501-55.
15. Hougaard, P. 1986. Survival models for heterogeneous populations derived from stable distributions. *Biometrika*. 73: 387-96.

16. Jacod, J. 1975. Multivariate point processes: Predictable projection, Radon-Nikodym derivatives, representation of martingales. *Z. Wahrsch. Verw. Geb.* **31**: 235-53.
17. Karr, A. F. 1987. Maximum likelihood estimation in the multiplicative intensity model via sieves. *Ann. Statist.* **15**: 473-90.
18. Karr, A. F. 1991. *Point Processes and Their statistical Inference* (2da Edición). New York: Dekker.
19. Kalbfleisch, J. D., and Prentice, R. L. 1980. *The Statistical Analysis of Failure Time Data*. New York: Wiley.
20. Marshall, A. W., and Olkin, I. 1967. A generalized bivariate exponential distribution. *Journal of Applied Probability*. **4**:291-302.
21. Lin, D. Y. 1994. Cox regression analysis of multivariate failure time data: the marginal approach. *Statistics in Medicine*. **13**: 2233-47.
22. Oakes, D. 1992. Frailty models for multiple event times. In Klein, J.P. and Goel, P. K. (eds) *Survival Analysis: State of the Art*, Kulwer Academic Publishers: Dordrecht. 371-79.
23. Oakes, D. 1997. Model Based and/or Marginal Analysis for Multiple Event-Time. D.Y.Lin, and Fleming T.R., (eds), *Proceeding of the First Seattle Symposium in Biostatistics: Survival Analysis*. New York: Springer-Verlag, Inc. 85-98.
24. Pedrosa de Lima, A.C. and Sen, P.K. 1998. Bivariate exponential, dual counting processes and their intensities. *Journal of Combinatorics, Information & System Sciences*. **23**: 33-46.
25. Pepe, M. S., and Cai, J. 1993. Some graphical display and marginal regression analyses for recurrent failure times and time-dependent covariates. *J. Amer. Statist. Assoc.* **88**: 811-20.
26. Prentice, R. L., and Cai, J. 1992. Marginal and conditional models for the analysis of multivariate failure time data. In Klein, J. P., and Goel, P. K. (eds), *Survival Analysis: State of the Art*, Kulwer Academic Publishers: Dordrecht. 393-406.
27. Rebolledo, R. T. 1980. Central limit theorems for local martingales. *Z. Wahrsch. Verw. Gebiete*. **51**: 269-86.
28. Ryu, K. 1993. An extension of Marshall and Olkin's bivariate exponential distribution. *J. Amer. Statist. Assoc.* **88**: 1458-65.

29. Sarkar, S. K. 1987. A continuous bivariate exponential distribution. *Journal of American Statistical Association*. **82**: 667-75.
30. Therneau, T. 1995. Extending the Cox model. *Mayo Clinic*, Rochester, Minnesota, USA.
31. Turnbull, B. W. 1992. Multiple Failure Time Analysis: A Discussion of Paper by Oakes; Pons, Kaddour and Turkheim; and Prentice and Cai. In Klein, J.P. and Goel, P. K. (eds), *Survival Analysis: State of the Art*, Kulwer Academic Publishers: Dordrecht. 371-79.
32. Vaupel, J. W., Manton, K. G., and Stallard, E. 1979. The impact of heterogeneity in individual frailty on dynamics of mortality. *Demography*. **16**: 439-54.
33. Wei, L. J., Lin, D. Y., and Weissfeld, L. 1989. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of American Statistical Association*. **84**: 1065-73.
34. Wei, L.J. and Glidden, D. V. 1997. An overview of statistical methods for multiple failure time data in clinical trial and their discussion. *Statistics in Medicine*. **16**: 833-51