UNIVERSIDAD DE GRANADA
FACULTAD DE CIENCIAS

DEPARTAMENTO DE ESTADÍSTICA E INVESTIGACIÓN OPERATIVA

APORTACIONES A LA FORMULACIÓN DE MODELOS, ESTUDIO DE LA ASOCIACIÓN Y ESTIMACIÓN EN ANÁLISIS DE DATOS DE SUPERVIVENCIA MULTIVARIANTE

TESIS DOCTORAL

Esteban Navarrete Álvarez

GRANADA 1997

APORTACIONES A LA FORMULACIÓN DE MODELOS, ESTUDIO DE LA ASOCIACIÓN Y ESTIMACIÓN EN ANÁLISIS DE DATOS DE SUPERVIVENCIA MULTIVARIANTE

Memoria presentada para optar al grado de Doctor en Ciencias (Matemáticas) por Esteban Navarrete Álvarez.

V° B°
Directores de Tesis:

Prof^a Dr^a D^a Julia García Leal

Prof. Dr. D. Jorge Ollero Hinojosa

DEPARTAMENTO DE ESTADÍSTICA E INVESTIGACIÓN OPERATIVA

FACULTAD DE CIENCIAS
UNIVERSIDAD DE GRANADA

Deseo expresar mi agradecimiento a todos mis compañeros del Departamento por su ayuda en la realización de esta memoria. A su Director, D. Ramón Gutiérrez Jáimez, por el constante interés y ánimo que me ha transmitido.

Asimismo quiero agradecer a D.R. Cox las sugerencias recibidas y a Dª Ana Mª Lara Porras y a D. Jose M. Quesada Rubio la colaboración prestada.

Por último, y muy especialmente, reconozco y agradezco la inestimable ayuda que los directores de esta memoria, Profª Julia García Leal y Prof. Jorge Ollero Hinojosa, me han proporcionado durante la realización de la misma.

A Isabel.

ÍNDICE

INTRODUCCIÓN GENERAL

La necesidad de generalizar los estudios del análisis de datos de supervivencia al caso multivariante ha surgido de forma natural. Como en otras ramas de la Estadística las situaciones reales imponen su criterio. El investigador ha obtenido datos de tiempos de fallo correspondientes a varias variables con posible relación entre ellas. Así, por ejemplo, datos de tipo familiar donde se mide un tiempo de fallo a cada miembro de una familia o grupo afín. No es ésta la única situación multivariante que se encuentra, como se verá al inicio del *Capítulo 1 (Fundamentos Generales)* donde se consideran diversos tipos de datos.

Modelos multivariantes habían surgido en los años cincuenta y sesenta en forma aislada y en un contexto que traspasaba la noción de supervivencia, como puede ser el modelo de choque de Marshall, A.W. y Olkin, I. (1967). Pero es en los años setenta y, sobre todo, a partir de los ochenta cuando se ha generalizado el estudio propiamente dicho de la supervivencia multivariante. Autores como Cox, D.R. (1972), que presenta su modelo general bivariante enunciando los conceptos básicos en el contexto de los modelos de Regresión, Johnson, N.L. y Kotz, S. (1975), que inciden en el llamado vector razón de azar multivariante, y Clayton, D.G. (1978), que estudia datos de supervivencia de tipo epidemiológico y presenta su modelo (*Modelo de Clayton*), son algunos de los más importantes en esta primera etapa. Le siguieron autores como Prentice, R.L., Williams, B.J. & Peterson, A.L. (1981) y sobre todo Oakes, D. (1989) con el ya conocido estudio de los *modelos frailty*, Hougaard, P. (1984,1986,1989,1992) que de una forma bastante prolífica hace un estudio variado incidiendo en su modelo (*modelo de Hougaard*). Otros autores que es necesario resaltar en los estudios modernos del análisis de datos de supervivencia multivariante son Genest, C. y Mackay, J. (1986) (con un estudio de las llamadas *cópulas arquimedianas*), Marshall, A.W. y Olkin, I. (1988) (con su enfoque de las distribuciones generadas por sus marginales), Basu, A.P. (1988) (que estudia familias de funciones exponenciales bivariantes), Klein, J.P., Lee, S., Lin, D.Y., etc.

Nuestra intención ha sido hacer una revisión de los estudios de supervivencia multivariante (paramétrica y semiparamétrica) y crear un esquema general a partir del cual pueda hacerse un estudio detallado de los distintos puntos de investigación. Hemos revisado conceptos básicos como el vector función de azar y la razón de fallo presentando algunas modificaciones en ellos, hemos formulado nuevos modelos y hemos estudiado elementos concretos en estudios de asociación o dependencia. Asimismo hemos planteado el esquema general de la inferencia en algunos modelos antes citados.

Esta memoria tiene la estructura general siguiente. En cada uno de los capítulos que a continuación se describen comenzamos con los fundamentos correspondientes a dicho capítulo y concluimos con las aportaciones, intentando respetar los bloques temáticos. Hemos de decir que, aunque la estructura general es la dicha, en los fundamentos van incluidos algunos desarrollos y comentarios propios, en aras de no perjudicar dicha estructura. Asimismo incluimos algunos enfoques o esquemas propios para presentar conocimientos ya asumidos.

En el *Capítulo 1 (Fundamentos generales)* se comienza con una clasificación del tipo de datos usual en supervivencia, de acuerdo a lo comunmente usado a lo largo de la reciente historia. Se sigue con una simple clasificación de los que serán nuestros objetivos fundamentales.

Se continúa presentando los conceptos y definiciones básicos haciendo referencia a la nomenclatura utilizada en Cox (1972). Nuestra aportación en este punto se referirá a una nueva expresión de la función de azar y de la densidad en función de ésta.

Se termina dando unas nociones de lo que consideramos necesario conocer para el desarrollo de esta memoria (distribuciones, copulas, mixturas, etc) y se presenta una revisión bibliográfica de los diversos temas tratados en supervivencia multivariante.

En el *Capítulo 2* se comienza presentando una clasificación de las distintas vías que existen para construir modelos. Se inicia con el modelo de Hougaard y con el ya clásico de Marshall y Olkin (1967), modelos obtenidos directamente a partir de variables aleatorias independientes. Se continúa con otro enfoque, el denominado de las distribuciones condicionadas donde se presenta el modelo de Clayton (1978). Nosotros presentamos una comparación entre las dos formas clásicas de llegar a este modelo (el método de las distribuciones condicionadas anteriormente citado y el método frailty mencionado a continuación) observando el comportamiento de las distribuciones marginales. Se hace un estudio especial del método de la independencia condicionada o método frailty. Nos detenemos en los modelos más importantes de este método (frailties gamma, estable e inversa gaussiana). A continuación se trata una generalización del método anterior, los modelos generados vía distribuciones marginales (Marshall y Olkin (1988) y Genest y Mackay (1986)).

Se hace un estudio amplio de las diversas formas que existen para introducir las covariables en modelos clásicos (modelo de Clayton, modelo de Hougaard).

Nosotros introducimos un modelo frailty con distribución uniforme en presencia de covariables explicativas que suponemos bajo bajo hipótesis de azar proporcional y estudiamos el problema de la identificabilidad. Hacemos mención a un modelo frailty con distribución estable y marginales Pareto. Presentamos los modelos de vida acelerada y destacamos las dificultades que ocasionan. También presentamos una clasificación para introducir covariables: formas multiplicativas puras tanto de las covariables externas como del frailty, formas multiplicativas con covariables y frailty aditivos entre sí, formas no multiplicativas (azar no proporcional), etc; este último caso lo tratamos con un poco más de profundidad en el último capítulo, Temas abiertos. Presentamos los denominados por nosotros modelos mixtos: modelos de azar proporcional con covariables bajo hipótesis de vida acelerada. En este apartado consideraremos en especial un modelo frailty con distribución estable positiva y estudiamos su identificabilidad. Seguidamente consideramos el caso en el que se introducen covariables explicativas en el denominado parámetro de asociación y estudiamos en especial un modelo mixto con frailty gamma. Por último, en este capítulo, presentamos un modelo general de frailty inverso de gran utilidad en casos donde interesa conocer un modelo donde se considera un frailty que es función de otro del cual conocemos una formulación.

En el *Capítulo 3* se hace un estudio pormenorizado del problema de la asociación o dependencia. Se clasifican los distintos métodos seguidos. Coeficientes como el de correlación o el de variación iniciarán el estudio. Se sigue con el más importante, el coeficiente de concordancia de Kendall, y con el denominado coeficiente de asociación de Oakes. Se calculan para los modelos más usuales (frailty estable, gamma e inversa gaussiana). Se mencionan otras medidas o métodos para el estudio de la asociación como son la dependencia cuadrante positiva, el concepto de variables asociadas y la positividad total. La conclusión final sobre estas medidas de asociación en los modelos frailty será fundamental en esta memoria.

Nosotros presentamos en este capítulo un estudio de las distintas medidas de dependencia en el modelo frailty uniforme haciendo notar las dificultades que presentan. Especial estudio se hará sobre el coeficiente de concordancia de Kendall. Asimismo estudiamos medidas de dependencia en dos de los modelos construidos en el capítulo anterior: un modelo mixto con frailty estable positiva y un modelo mixto con frailty gamma y covariables en el parámetro de asociación. Por último, en este capítulo, estudiamos la dependencia cuadrante en los modelos mixtos.

En el *Capítulo 4* estudiaremos los problemas de la estimación. Se inicia distinguiendo la forma de abordarla según los tipos de modelos (paramétricos, semiparamétricos y no paramétricos) continuando con la construcción de la función de verosimilitud en el modelo general paramétrico o semiparamétrico definido a partir de la función de supervivencia. Se presentan los modelos semiparamétricos definidos a partir de las funciones de azar marginales y la metodología general que se ha de seguir con ellos, así como con los no paramétricos.

Nosotros presentamos un estudio de estimación en tres de los modelos considerados: un modelo frailty uniforme con covariables, sujeto a censuras y marginales Weibull; un modelo mixto con frailty estable positiva, marginales Weibull y sujeto a censuras y un modelo mixto con frailty gamma, sujeto a censuras, marginales Weibull y covariables también en el parámetro de asociación.

En el *Capítulo 5*, *Temas abiertos*, presentaremos algunas lineas de investigación. Comenzamos estudiando un modelo con frailty aditivo. A continuación presentamos alternativas en la introducción de las covariables según se vió en el capítulo 2. Se continúa con el problema de los datos agrupados y problemas de simulación. Terminaremos presentando algunos otros puntos de interés objeto de nuestra línea de investigación.

CAPÍTULO 1
FUNDAMENTOS GENERALES

1.1 INTRODUCCIÓN

En este capítulo presentamos las bases necesarias para el desarrollo de esta memoria. Comenzamos describiendo los diversos tipos de datos que se presentan en supervivencia multivariante. A continuación, establecemos los objetivos y exponemos conceptos básicos tales como función de supervivencia multivariante, vector función de azar y razón de fallo, así como la interrelación entre ellos.

Seguidamente exponemos algunos conceptos e instrumentos que posteriormente utilizaremos, tales como distribuciones de probabilidad usuales en supervivencia (*Gamma, Inversa Gaussiana, Pareto y Estable Positiva*), y los instrumentos *mixtura de distribuciones y cópulas arquimedianas*.

Por último, incluimos una revisión bibliográfica sobre estudios de supervivencia multivariante.

1.1.1 TIPOS DE DATOS

El análisis de datos de supervivencia multivariante surge ante la necesidad de considerar dos o más tiempos de supervivencia simultáneamente o bien al considerar varias causas distintas de fallo para uno o más tiempos. Ante estas situaciones, hay que tener en cuenta la naturaleza de los datos a tratar, es decir, el tipo de datos que se manejan, lo cual influirá en su tratamiento posterior. Otro aspecto importante a tener en cuenta es la posible presencia de datos *censurados* debido a la falta de información sobre algunos individuos. Es decir, a veces surgen datos incompletos ya que los tiempos de supervivencia no son conocidos para todos los individuos. La forma más habitual es la censura a la derecha en la que el tiempo observado es menor que el tiempo de supervivencia. También existen censuras a la izquierda, donde el tiempo observado es mayor que el tiempo de supervivencia. Existen otros tipos de censuras, aunque menos frecuentes. Los distintos tipos de censuras afectarán también al tratamiento estadístico de los datos.

Teniendo en cuenta la diversidad de datos que se observan en la literatura sobre supervivencia, consideramos conveniente establecer una clasificación sobre los distintos tipos de datos existentes. Dicha clasificación se muestra a continuación.

a) *Casos de varios individuos*

a_1) *Individuos emparentados.*- En primer lugar, se considera el caso en que se mide un mismo tiempo de fallo a distintos individuos relacionados. Es decir, en varios individuos se mide el mismo evento. Por ejemplo, cuando en varios miembros de una misma familia se mide el tiempo de aparición de una cierta enfermedad. En esta situación suelen estar presentes factores genéticos que inducen cierta dependencia en los tiempos de fallo de los distintos individuos considerados (*datos familiares*).

a_2) *Componentes idénticas, repetidas o en paralelo.*- En segundo lugar, se considera el caso en que se produzcan fallos en diferentes componentes de un sistema. Por ejemplo, si consideramos a una persona como el sistema, los ojos serán las componentes del sistema y el interés puede estar en el instante o momento en el que aparece cierta dolencia en alguno de ellos. Igualmente se pueden considerar como componentes a los riñones. Este tipo de datos es tambien frecuente en fiabilidad industrial utilizándose, por ejemplo, modelos de choque donde se supone que el sistema sufre daños por sucesivos impactos que pueden deteriorar uno o varios componentes. Los fallos podrán ocurrir simultáneamente.

b) *Casos de un solo individuo*

b_1) *Distintos eventos (eventos encadenados).-* Se considera ahora el caso en que en un mismo individuo se miden los instantes de fallo de distintos eventos. Por ejemplo, se mide en un mismo individuo el instante de aparición de cierto nivel de hipertensión, infarto de miocardio y la muerte por esta causa.

b_2) *Eventos repetidos.-* Se considera el caso en que en un mismo individuo puedan darse distintas ocurrencias del mismo evento. Por ejemplo, los instantes de ocurrencia de sucesivos infartos, duración de períodos en paro, tiempos entre accidentes de tráfico, etc. En este caso, obviamente, no es posible que dos fallos ocurran en el mismo tiempo.

b_3) *Causas del mismo evento.-* En este caso se considera, aunque bastantes autores no lo consideran exactamente un problema de supervivencia multivariante, la situación en la que en un individuo se observa el mismo fallo, centrándose la atención en la causa que lo ocasiona. Sólo una vez ocurre el fallo en cada individuo pero puede producirse por diversas causas. Así, por ejemplo, la avería de un automóvil se produce por diversas causas.

b_4) *Distintas causas y eventos diferentes.-* Esta situación es similar a la anterior, donde se consideran distintas causas posibles en cada uno de diferentes eventos.

Estas situaciones anteriores son las más usuales aunque pueden darse casos concretos de datos que no encajen en tal clasificación. Los objetivos que se tengan dependerán del tipo de datos a tratar. Antes de condiderar los objetivos generales en el apartado siguiente veamos algunas situaciones especiales en este sentido.

* Si se dispone de datos del tipo a_2), diferentes componentes de un sistema, y se

observa que es factible hacer un diseño para comparar dos tratamientos por medio de cierta aleatorización, como es el caso de la ceguera en ambos ojos donde se puede aplicar un tratamiento a un ojo y otro distinto al otro, el interés puede estar en hacer un test que indique si un tratamiento reduce el tiempo de ceguera.

* En el caso de la existencia de un factor genético entre individuos emparentados, factor de dependencia común, datos del tipo a_1), el interés se centra principalmente en estudiar la interrelación entre ellos, abordada por la asociación, dependencia o correlación.

* En el caso de haber observado el tiempo de fallo de varios individuos, o de varias o múltiples ocurrencias en el mismo individuo, como los sucesivos infartos, períodos de paro, etc, o bien, varios eventos distintos en el mismo individuo, caso de aparición de hipertensión e incidencia de infarto, por ejemplo, el interés puede estar en hacer predicciones. Así, la incidencia de una enfermedad en un individuo puede dar información sobre la incidencia en su hermano, su hijo, etc. En tales problemas de predicción (regresión) juega un papel fundamental la presencia de posibles *covariables*.

Las covariables o variables explicativas son cualquier tipo de variables distintas a los tiempos de fallo y que pudieran influir de alguna manera en el valor de dichos tiempos. Representan propiedades intrínsecas del individuo, factores exógenos o tratamientos experimentales aplicados. Es decir, características tales como el sexo, la clase social, profesión, etc. Se pueden clasificar las covariables de diferentes formas. Así, por ejemplo, están las covariables constantes y covariables dependientes del tiempo. Dentro de este último tipo están las covariables externas (aquellas no relacionadas directamente con el mecanismo de fallo, por ejemplo, la edad de un individuo) e internas (necesitan la supervivencia del individuo para su existencia y llevan información de la variable tiempo de supervivencia). A su vez, las covariables externas pueden ser covariables fijadas (su valor está medido con antelación y fijado para la duración del estudio), covariables definidas (determinadas de antemano para cada individuo) y ancillarys (salida de un proceso estocástico ajena al individuo bajo estudio, por ejemplo, la contaminación atmosférica). Para más detalle puede verse Lara (1995). Existen otras clasificaciones de las covariables: determinísticas o aleatorias, etc. De especial importancia es una covariable aleatoria que se utilizará en esta memoria. Se trata de una covariable aleatoria que no es observada, denominada *frailty*. Es una característica común que poseen los individuos de una pareja o grupo que explica la posible dependencia entre sus tiempos de supervivencia. Suele representar un factor de riesgo entre los individuos. Un ejemplo puede ser la componente genética entre individuos emparentados.

1.1.2 OBJETIVOS

Partiendo de hipótesis o planteamientos distintos se puede llegar al mismo resultado o formulación de un *modelo*. Puesto que hemos mencionado este último término, se plantean ahora cuáles serán los objetivos del análisis de datos de supervivencia multivariante. Se resumen en la enumeración siguiente:

a) Formulación de un modelo
b) Estudio de la asociación
c) Estimación
d) Predicción

En primer lugar, interesará determinar una familia de distribuciones multivariantes que se ajuste lo mejor posible a los datos. Para ello se deberá buscar una formulación de dicha familia a partir de la cual se efectuará todo tipo de actuaciones tales como el estudio de la dependencia entre las variables, etc. Generalmente dar una formulación de un modelo consiste en encontrar la relación que ligue la función de supervivencia multivariante con todo tipo de parámetros que intervengan en el modelo, en especial los que describan la dependencia. Se puede clasificar la formulación de un modelo como:

- paramétrica
- semiparamétrica
- no paramétrica

Tanto en una formulación paramétrica como en una semiparamétrica pueden suponerse dos casos: que se desee obtener una sola expresión donde estén ligados cualquier tipo de parámetros (incluido un parámetro que explique la asociación entre los tiempos de fallo) junto a las funciones de supervivencia, de distribución, de densidad o de azar marginales, o bien que no se desee disponer de dicha única expresión, sino dos expresiones, una para cada función de azar marginal con posibles parámetros de regresión y un factor de riesgo común que ligue ambos tiempos de fallo.

En el primer caso, si están determinadas las familias de distribuciones de los tiempos de fallo por separado, esto es, las marginales, el modelo se denominará paramétrico. La metodología usada para hacer inferencias sobre los parámetros se basa en la función de verosimilitud. Éste será el caso tratado en esta memoria. Si no están determinadas dichas familias de distribuciones y sólo se dispone de la expresión que liga los tiempos de supervivencia el modelo se denominará semiparamétrico. En este caso el estudio irá dirigido hacia el parámetro que explica la asociación y en su caso sobre posibles covariables explicativas. La metodología usada está basada en la denominada verosimilitud parcial.

En el segundo caso, si las funciones de azar base marginales están determinadas (se conocen las familias de distribuciones a las que pertenecen) el modelo se denominará paramétrico. La metodología usada se basará nuevamente en la función de verosimilitud. Si las funciones de azar base marginales no están determinadas el modelo se denominará semiparamétrico. El interés estará en las covariables explicativas, la metodología se basa nuevamente en la verosimilitud parcial.

En los modelos no paramétricos la metodología es distinta a la de los casos paramétrico y semiparamétrico. Está basada en estimaciones directas de la función de supervivencia por medio del conteo de los datos. No se tratará en esta memoria.

Como segundo objetivo se plantea el problema relativo a la *asociación o dependencia* entre los tiempos de fallo o supervivencia. Existen diversas maneras de abordar este punto, fundamentalmente por el estudio de diferentes coeficientes de asociación tales como el coeficiente de Kendall, el denominado coeficiente de Oakes, etc. En el capítulo 4 de esta memoria se detallarán dichos coeficientes.

La *estimación* de los parámetros (de las distribuciones marginales si están presentes, de asociación o de regresión) se inscribe dentro del problema de la *inferencia* y *predicción*. Este último objetivo ha sido menos tratado en la literatura que los precedentes. En los modelos estudiados en esta memoria, modelos paramétricos, construimos la función de verosimilitud para datos censurados a la derecha y en presencia de covariables explicativas. El estudio de la verosimilitud parcial en los modelos semiparamétricos no se tratará en esta memoria.

1.2 CONCEPTOS BÁSICOS

1.2.1 DEFINICIONES

Vamos a dar una formulación de diversos conceptos básicos de supervivencia multivariante útiles en cualquier enfoque que posteriormente hagamos. Nos ceñiremos al caso bivariante, el cual se generaliza sin dificultad a cualquier número de variables implicadas.

Sea una variable aleatoria bidimensional continua

$$(T_1, T_2) \quad , \quad T_i \geq 0 \quad , \quad i:1,2$$

donde T_i denota el tiempo de fallo de la componente i.

Sea la función de distribución conjunta de la variable (T_1, T_2)

$$F(t_1, t_2) = P\,[T_1 \leq t_1 \;,\; T_2 \leq t_2]$$

y sea la función de supervivencia conjunta de la variable (T_1, T_2)

$$S(t_1, t_2) = P\,[T_1 \geq t_1 \;,\; T_2 \geq t_2] \quad . \qquad [1.1]$$

Entre ambas funciones se verifica que

$$S(t_1, t_2) = 1 - F_1(t_1) - F_2(t_2) + F(t_1, t_2)$$
$$F(t_1, t_2) = 1 - S_1(t_1) - S_2(t_2) + S(t_1, t_2)$$

donde $F_i(t_i)$ y $S_i(t_i)$ son las funciones de distribución marginales y las funciones de supervivencia marginales, respectivamente, $i:1,2$. Además, se verifica que

$$\begin{aligned} S(t_1, 0) &= P[T_1 \geq t_1, T_2 \geq 0] = S_1(t_1) \\ S(0, t_2) &= P[T_1 \geq 0, T_2 \geq t_2) = S_2(t_2) \quad . \end{aligned} \qquad [1.2]$$

Propiedades elementales de la teoría de la probabilidad muestran que

$$S(t_1, t_2) \leq \min[S_1(t_1), S_2(t_2)] \quad \text{y}$$
$$S(t_1, t_2) \geq S_1(t_1) + S_2(t_2) - 1 \quad .$$

Sea $f(t_1, t_2)$ la función de densidad conjunta de la variable (T_1, T_2). Se verifica

$$f(t_1, t_2) = \frac{\partial^2 F(t_1, t_2)}{\partial t_1 \partial t_2} = \frac{\partial^2 S(t_1, t_2)}{\partial t_1 \partial t_2}$$

o lo que es lo mismo

$$F(t_1, t_2) = \int_0^{t_1}\int_0^{t_2} f(t_1, t_2)\, dt_1 dt_2$$

$$S(t_1,t_2)=\int_{t_1}^{\infty}\int_{t_2}^{\infty}f(t_1,t_2)\,dt_1dt_2$$

1.2.2 FUNCIÓN DE AZAR Y RAZÓN DE FALLO

Vamos a introducir la denominada *función de azar*. En Cox (1972) o en Cox & Oakes (1984) está definida de la siguiente manera:

$$h_{i0}(t)=\lim_{\Delta t\to 0+}\frac{P(t\leq T_i<t+\Delta t/\,t\leq T_1,\,t\leq T_2)}{\Delta t}\quad,\qquad (i=1,2)$$

$$h_{21}(t/u)=\lim_{\Delta t\to 0+}\frac{P(t\leq T_2<t+\Delta t/\,t\leq T_2,\,T_1=u)}{\Delta t}\quad,\qquad (u<t)$$

y analogamente $h_{12}(t/u)$. El significado intuitivo de la función de azar se verá en la próxima definición, donde consideramos dicha función de azar con una notación diferente. Según la definición anterior, la función de densidad conjunta es

$$f(t_1,t_2)=\exp\left[-\int_0^{t_1-0}(h_{10}(u)+h_{20}(u))\,du-\int_{t_1+0}^{t_2-0}h_{21}(u/t_1)\,du\right].$$
$$.h_{10}(t_1).h_{21}(t_2/t_1),\quad \text{si}\quad t_2\geq t_1$$

y analogamente para $t_2\leq t_1$.

Además, se verifica que

$$h_{10}(t)=-\frac{\left.\frac{\partial S(t,u)}{\partial t}\right|_{u=t}}{S(t,t)}\quad,\quad h_{12}(t/u)=-\frac{\frac{\partial^2 S(t,u)}{\partial t\partial u}}{\frac{\partial S(t,u)}{\partial u}}\quad .$$

Para un mejor desarrollo de la memoria, nosotros hemos adoptado la siguiente notación que se adapta mejor a la empleada por los autores más importantes y retomamos las siguientes definiciones para dicha notación.

Definición: Se denomina *riesgo instantaneo de fallo* de la componente $i:1,2$, dado que antes no ha fallado ninguna, respectivamente, a la función

$$h_1(t_1,t_2)=\lim_{\Delta\to 0^+}\frac{P[t_1\le T_1<t_1+\Delta/T_1\ge t_1,T_2\ge t_2]}{\Delta}$$

y

$$h_2(t_1,t_2)=\lim_{\Delta\to 0^+}\frac{P[t_1\le T_2<t_2+\Delta/T_1\ge t_1,T_2\ge t_2]}{\Delta} \quad . \qquad [1.3]$$

Estas funciones son las componentes de un vector denominado *función de azar (o vector razón de azar)*.

Es fácil comprobar que

$$h_1(t_1,0)=h_1(t_1)$$
$$h_2(0,t_2)=h_2(t_2)$$

es decir, coinciden con las funciones de azar marginales de las variables T_1 *y* T_2, respectivamente.

De acuerdo a la definición anterior [1.3], vamos a emplear también la notación siguiente:

$$h_1(t_1,t_2)=h_1(t_1/T_2\ge t_2)$$
$$h_2(t_1,t_2)=h_2(t_2/T_1\ge t_1) \quad .$$

Se puede expresar el riesgo instantaneo de fallo en términos de la función de supervivencia conjunta, como se muestra a continuación:

$$h_1(t_1,t_2)=h_1(t_1/T_2\geq t_2)=-\frac{\dfrac{\partial S(t_1,t_2)}{\partial t_1}}{S(t_1,t_2)}=\frac{-\partial \lg S(t_1,t_2)}{\partial t_1}$$

y de igual manera

$$h_2(t_1,t_2)=h_2(t_2/T_1\geq t_1)=-\frac{\dfrac{\partial S(t_1,t_2)}{\partial t_2}}{S(t_1,t_2)}=\frac{-\partial \lg S(t_1,t_2)}{\partial t_2} \quad . \qquad [1.4]$$

Definición: Se denomina *riesgo instantáneo de fallo* de la componente i: 1,2 , dado que ésta no ha fallado y la otra lo ha hecho en un tiempo dado, respectivamente, a la función

$$h_1'(t_1,t_2)=h_1(t_1/T_2=t_2)=\lim_{\Delta\to 0^+}\frac{P[t_1\leq T_1<t_1+\Delta/T_1\geq t_1,T_2=t_2]}{\Delta},\quad t_1\geq t_2$$

y

$$h_2'(t_1,t_2)=h_2(t_2/T_1=t_1)=\lim_{\Delta\to 0^+}\frac{P[t_2\leq T_2<t_2+\Delta/T_2\geq t_2,T_1=t_1]}{\Delta},\quad t_2\geq t_1 \; .$$

Por otra parte, también se puede expresar este riesgo instantaneo de fallo en términos de la función de supervivencia conjunta como mostramos a continuación:

$$h_1'(t_1,t_2)=\frac{\dfrac{\partial^2 S(t_1,t_2)}{\partial t_1\partial t_2}}{-\dfrac{\partial S(t_1,t_2)}{\partial t_2}}=-\frac{\partial}{\partial t_1}\lg\left(-\frac{\partial S(t_1,t_2)}{\partial t_2}\right)\quad ,\quad t_1\geq t_2 \quad .$$

Análogamente, se verifica

$$h_2'(t_1,t_2)=\frac{\dfrac{\partial^2 S(t_1,t_2)}{\partial t_1\partial t_2}}{-\dfrac{\partial S(t_1,t_2)}{\partial t_1}}=-\frac{\partial}{\partial t_2}\lg\left(-\frac{\partial S(t_1,t_2)}{\partial t_1}\right)\quad ,\quad t_2\geq t_1 \quad . \qquad [1.5]$$

De acuerdo a estas definiciones anteriores se verifica que T_1 y T_2 son independientes si y solo si

$$h_1(t_1,t_2)=h_1'(t_1,t_2)$$

$$h_2(t_1,t_2)=h_2'(t_1,t_2)$$

Definición: Se denomina *razón de fallo* al riesgo instantaneo de fallo de ambas componentes dado que antes no ha fallado ninguna, es decir, a la función

$$r(t_1,t_2)=\lim_{\substack{\Delta'\to 0^+\\ \Delta\to 0^+}}\left[\frac{P(t_1\leq T_1<t_1+\Delta,\ t_2\leq T_2<t_2+\Delta' / T_1\geq t_1,\ T_2\geq t_2)}{\Delta.\Delta'}\right] .$$

La razón de fallo puede expresarse de la siguiente manera

$$r(t_1,t_2)=\frac{\dfrac{\partial^2 S(t_1,t_2)}{\partial t_1\partial t_2}}{S(t_1,t_2)}=\frac{f(t_1,t_2)}{S(t_1,t_2)} . \qquad [1.6]$$

Es decir, la razón de fallo es el cociente entre la función de densidad y la función de supervivencia. En supervivencia univariante dicha razón de fallo y la función de azar coinciden, siendo, como se puede observar, significativamente distintas en supervivencia multivariante. La relación entre la razón de fallo y la función de azar es la siguiente:

$$r(t_1,t_2)=h_1(t_1,t_2)h_2(t_1,t_2)-\frac{\partial^2(-lgS(t_1,t_2))}{\partial t_1\partial t_2} . \qquad [1.7]$$

En el desarrollo de esta memoria usamos fundamentalmente los conceptos de función de azar y función de supervivencia así como la razón de fallo. No es usual utilizar la función de densidad. A pesar de ello, es conveniente disponer de una expresión de la función de densidad en función de las componentes del vector función de azar. Para ello, consideremos la expresión

$$f(t_1,t_2)=\frac{\partial^2 S(t_1,t_2)}{\partial t_1\partial t_2}=-h_1'(t_1,t_2).\frac{\partial S(t_1,t_2)}{\partial t_2}=$$

$$=h_1'(t_1,t_2)\,.\,h_2(t_1,t_2)\,.\,S(t_1,t_2) \quad , \quad t_1 \geq t_2$$

Se trata ahora de encontrar una expresión de la función de supervivencia conjunta en función de las componentes del vector función de azar. Por [1.7] se verificaba que

$$r(t_1,t_2)=h_1(t_1,t_2)\,.\,h_2(t_1,t_2)-\frac{\partial^2(-lgS(t_1,t_2))}{\partial t_1\partial_2} \quad \Longrightarrow$$

$$\frac{\partial^2(-lgS(t_1,t_2))}{\partial t_1\partial t_2} = h_1(t_1,t_2)h_2(t_1,t_2)-r(t_1,t_2)$$

Integrando esta última igualdad se verifica

$$lg\frac{S_1(t_1)\,.\,S_2(t_2)}{S(t_1,t_2)}=$$

$$=\int_0^{t_1}\int_0^{t_2}h_1(t_1,t_2)h_2(t_1,t_2)\,dt_1dt_2-\int_0^{t_1}\int_0^{t_2}r(t_1,t_2)\,dt_1dt_2 \quad ,$$

y tomando exponenciales en los dos miembros resulta

$$\frac{S_1(t_1)S_2(t_2)}{S(t_1,t_2)}=$$

$$=\exp\left[\int_0^{t_1}\int_0^{t_2}h_1(t_1,t_2)h_2(t_1,t_2)\,dt_1dt_2-\int_0^{t_1}\int_0^{t_2}r(t_1,t_2)\,dt_1dt_2\right] \quad \Longrightarrow$$

$$S(t_1,t_2)=S_1(t_1)S_2(t_2)\exp\Big[\int_0^{t_1}\int_0^{t_2}r(t_1,t_2)\,dt_1dt_2-$$

$$-\int_0^{t_1}\int_0^{t_2}h_1(t_1,t_2)h_2(t_1,t_2)\,dt_1dt_2\Big]=$$

(expresando las funciones de supervivencia marginales en función de la función de azar)

$$=\exp\Big[-\int_0^{t_1}h_1(t_1)\,dt_1-\int_0^{t_2}h_2(t_2)\,dt_2+$$

$$+\int_0^{t_1}\int_0^{t_2}r(t_1,t_2)\,dt_1dt_2-\int_0^{t_1}\int_0^{t_2}h_1(t_1,t_2)h_2(t_1,t_2)\,dt_1dt_2\Big]\quad .$$

Si se sustituye esta expresión de la función de supervivencia conjunta en la expresión de la función de densidad inicial resulta

$$f(t_1,t_2)=h_1'(t_1,t_2)h_2(t_1,t_2)\exp\Big[-\int_0^{t_1}h_1(t_1)\,dt_1-\int_0^{t_2}h_2(t_2)\,dt_2-$$

$$-\int_0^{t_1}\int_0^{t_2}\frac{\partial^2(-\lg S(t_1,t_2))}{\partial t_1\partial t_2}\,dt_1dt_2\Big]=$$

(expresando la integral doble en términos de la función de azar)

$$=h_1'(t_1,t_2)h_2(t_1,t_2)\exp\Big[-\int_0^{t_1}h_1(t_1)\,dt_1-\int_0^{t_2}h_2(t_2)\,dt_2-$$

$$-\int_0^{t_1}[h_1(t_1,t_2)-h_1(t_1)]\,dt_1\Big]=$$

$$=h_1'(t_1,t_2)h_2(t_1,t_2)\exp\Big[-\int_0^{t_2}h_2(t_2)\,dt_2-\int_0^{t_1}h_1(t_1,t_2)\,dt_1\Big]\quad ,\quad t_1\geq t_2\quad .$$

Análogamente, se obtiene

$$f(t_1,t_2)=h_2'(t_1,t_2)h_1(t_1,t_2)\exp\left[-\int_0^{t_1}h_1(t_1)dt_1-\right.$$
$$\left.-\int_0^{t_2}h_2(t_1,t_2)dt_2\right]\quad,\quad t_2\geq t_1 \qquad [1.8]$$

Johnson & Kotz (1975) estudiaron propiedades del vector función de azar (véase la revisión bibliográfica). Así, por ejemplo, demuestran, en su propia notación que nosotros hemos adaptado a la nuestra, que si T_1 y T_2 son independientes se verifica

$$h_1(t_1,t_2)=h_1(t_1)$$
$$h_2(t_1,t_2)=h_2(t_2)$$

ya que

$$h_1(t_1,t_2)=\frac{-\partial lgS(t_1,t_2)}{\partial t_1}=\frac{-\partial lg[S_1(t_1)S_2(t_2)]}{\partial t_1}=$$

$$=\frac{-\partial[lgS_1(t_1)+lgS_2(t_2)]}{\partial t_1}=\frac{-\partial lgS_1(t_1)}{\partial t_1}=h_1(t_1)\quad .$$

Mencionemos en este momento otros autores que estudian propiedades de este vector, como se verá también en la revisión bibliografica. Basu (1971,1988) considera distintas distribuciones exponenciales bivariantes (Gumbel, Freund, Marshall y Olkin, Block-Basu, Friday y Patil, Dowton, etc.) y sus aplicaciones en fiabilidad. Block (1973), Dykstra y otros (1973), Lehmann (1966) y Marshall (1973) estudian el vector función de azar. Puri y otros (1974) demuestran que la razón de fallo es constante si y sólo si la distribución es una mixtura de distribuciones exponenciales. Sun & Basu (1993) hacen un estudio, en el mismo sentido que Basu (1988), de la razón de fallo y la función de azar de una familia de distribuciones bivariantes exponenciales.

1.3 CONCEPTOS E INSTRUMENTOS NECESARIOS PARA EL DESARROLLO DE ESTA MEMORIA

A lo largo de esta memoria utilizaremos algunos conceptos e instrumentos de la Estadística Matemática (distribuciones importantes en análisis de supervivencia, mixtura de distribuciones, cópulas arquimedianas) que resumimos a continuación.

1.3.1 DISTRIBUCIONES

Las distribuciones gamma, inversa gaussiana y estable positiva son fundamentales para la construcción de un tipo de modelos, los modelos frailty, que veremos en el capítulo 2. Ésas son las más empleadas como distribución de un frailty. La distribución de Pareto se utiliza, en algunos casos, como distribución de la supervivencia marginal en un modelo frailty con distribución estable positiva, como también veremos en ese capítulo.

a) DISTRIBUCIÓN GAMMA

Una variable aleatoria Z sigue una distribución gamma si su función de densidad es

$$f(z)=\frac{\theta^{\delta} z^{\delta-1}.\exp(-\theta z)}{\Gamma(\delta)}\ , \qquad z\geq 0 \qquad [1.9]$$

donde los parámetros verifican

$$\theta>0\ ,\ \delta>0\ .$$

Se verifica que

$$E(z)=\frac{\delta}{\theta}$$
$$Var(z)=\frac{\delta}{\theta^2}$$

y, por lo tanto,

$$\theta=1 \quad \Rightarrow \quad E(z)=\delta\ , \qquad Var(z)=\delta$$

La distribución gamma verifica la siguiente propiedad (falta de memoria):

$$P(Z\geq b+a/\, z\geq a)=P(Z\geq b)\ .$$

Casos particulares son los sigientes:

$$1)\ \delta=n/2,\ \theta=1/2 \Rightarrow Z \rightarrow \chi^2(n)$$

$$2)\ \delta=n \in Z,\ \theta=n\mu \Rightarrow Z \rightarrow \mathrm{Erlang}$$

$$3)\ \delta=1 \Rightarrow Z \rightarrow \mathrm{exponencial} \quad .$$

b) DISTRIBUCIÓN INVERSA GAUSSIANA

Una variable aleatoria Z sigue una distribución inversa gaussiana (Hougaard (1984)) si su función de densidad es

$$f(z)=(\psi/\pi)^{1/2}.\exp(4\psi\theta)^{1/2}.z^{-3/2}.\exp(-\theta z-\psi/z), \quad z>0\ ,\ \theta\geq 0\ ,\ \psi>0 \quad . \qquad [1.10]$$

Si $\theta=0$, entonces, la variable 1/z sigue una distribución gamma. Además, se verifica que

$$E(z)=\left(\frac{\psi}{\theta}\right)^{1/2}\ ,\quad \theta>0$$
$$E(z)=\infty\ ,\quad \theta=0$$

y si $\psi=\theta$ entonces se verifica que

$$E(z)=1$$
$$Var(z)=1/2.\psi^{1/2}\theta^{-3/2}$$
$$CV=2^{-1/2}(\psi\theta)^{-1/4} \quad .$$

c) DISTRIBUCIÓN DE PARETO

Una variable aleatoria X sigue una distribución de Pareto si su función de densidad es de la forma

$$f(x)=\frac{\alpha}{x_0}\left(\frac{x_0}{x}\right)^{\alpha+1}\ ,\quad x\geq x_0>0\ ,\quad \alpha>0 \quad . \qquad [1.11]$$

Se verifica que

$$E(X)=\frac{\alpha x_0}{\alpha-1} \quad y \quad Var(X)=\frac{\alpha x_0^2}{(\alpha-1)^2(\alpha-2)}\ ,\quad \alpha>2 \ .$$

d) DISTIBUCIÓN ESTABLE

Existen situaciones reales en las que se manifiestan muchas pequeñas variables actuando aditivamente. Por este motivo, junto a la facilidad de su manejo en un caso concreto, consideremos un tipo de variable aleatoria que se define a continuación.

Definición: Una distribución es estrictamente estable (Feller (1971) y Hougaard (1986)) si la suma de variables independientes e idénticamente distribuidas sigue la misma distribución.

Definición: Una clase de distribuciones se llama estable si las sumas de funciones lineales de variables aleatorias de dicha clase, independientes e idénticamente distribuidas, pertenecen a la misma clase de distribuciones.

A continuación se presenta una caracterización de dichas distribuciones. Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes e idénticamente distribuidas. Dicha distribución es estable si $\forall n \ \exists \ c_n = n^{1/\alpha}$, $\alpha \in (0,2]$, tal que

$$\text{Distr.}(X_1 + \ldots + X_n) = \text{Distr.}(c_n X_1) .$$

Casos particulares son $\alpha=2$ y $\alpha=1$ (distribución normal y degenerada, respectivamente). Si $\alpha \in (0,1]$ se le denomina distribución *estable positiva* y se formula a través de su transformada de Laplace

$$\phi(s) = E[\exp(-sX)] = \exp(-s^{\alpha}) , \quad s \geq 0 \qquad [1.12]$$

Tal expresión facilitará enormemente los cálculos en la construcción de un tipo de modelos de supervivencia multivariante, como veremos en el capítulo 2.

1.3.2 MIXTURAS DE DISTRIBUCIONES

Sea $S(x,\alpha)$ la función de supervivencia (univariante) de la variable aleatoria X, siendo α un parámetro que se considera una variable aleatoria con función de distribución $F(\alpha)$. Se llama mixtura de funciones de supervivencia a

$$\mathcal{S}(x) = \int_{-\infty}^{\infty} S(x,\alpha)\, dF(\alpha)$$

Sean X y Z dos variables aleatorias univariantes y se supone que la función de azar de X está basada en hipótesis de azar proporcional respecto a Z, es decir $h(x/z)=z.h(x)$ o, lo que que es lo mismo, $P[\ X>x\ /\ Z=z\ ] = [S_0(x)]^z$, siendo $S_0(x)$ una función de supervivencia base; entonces, si $F(z)$ es la función de distribución de la variable no negativa Z, la función de supervivencia de la variable X es

$$P(X>x) = \int_0^{\infty} [S_0(x)]^z dF(z) \quad .$$

Sean $X_1,\ldots,X_n$ n variables aleatorias independientes para un valor fijado $Z=z$, donde Z es una variable aleatoria. Si dado $Z=z$ cada X_i, $i:1,\ldots,n$, se comporta según un modelo de azar proporcional, es decir

$$P[X_i>x_i/Z=z] = [S_0(x_i)]^z \;, \quad i:1,\ldots,n$$

entonces la mixtura vendrá determinada por la expresión

$$P[X_1>x_1,\ldots,X_n>x_n) = \int\ldots\int [S_0(x_1)\ldots S_0(x_n)]^z dG(z) \quad . \qquad [1.13]$$

Un estudio detallado de mixturas puede verse en Mei-Ling Ting Lee (1993).

1.3.3 CÓPULAS ARQUIMEDIANAS

Definición: Se llama cópula a una función de distribución bivariante concentrada sobre el cuadrado unidad y con marginales uniformes (Genest & Mackay (1986 a, b)). Toda función de distribución $F(t_1,t_2)$ con marginales continuas se puede representar por una cópula de forma única (Sklar (1959)) así

$$H(x,y)=F(F_1^{-1}(x),F_2^{-1}(y))$$

$$F_i^{-1}(t)=\text{ínf}\,\{u\in\mathbb{R}\,:\,F_i(u)\geq t,\;i{:}1,2\}\quad .$$

$H(x,y)$ es la representación uniforme de F y X e Y son uniformes en [0,1] . Si las variables X e Y son independientes, entonces H(x,y)=x · y .

Sea una función estrictamente creciente y continua

$$\lambda{:}[0,1]\rightarrow[0,1]\ \text{ tal que }\begin{cases}\lambda(0)=0\\\lambda(1)=1\end{cases}$$

y supongamos que se verifica

$$\lambda[F(x,y)]=\lambda[F_1(x)].\lambda[F_2(y)]\quad .$$

Por otro lado, sea $\phi(t)=-lg\lambda(t),\quad \forall t\in(0,1]$. Entonces se verifica

$$\phi[H(x,y)]=\phi(x)+\phi(y)\quad .$$

Pues bien, toda función de distribución H tal que

$$H(x,y)=\phi^{-1}[\phi(x)+\phi(y)]$$

es una cópula arquimediana, según la terminología de Ling (1965).

A continuación se muestra una condición necesaria y suficiente para que una función ϕ genere una cópula arquimediana.

Teorema: Sea una función

$$\phi{:}[0,1]\rightarrow[0,\infty]\ \text{ tal que }\ \phi(1)=0$$

continua, estrictamente decreciente y convexa. Condición necesaria y suficiente para que esta función genere una cópula arquimediana H(x,y) (definida en el cuadrado unidad) es que

$$H(x,y)=\phi^{-1}[\phi(x)+\phi(y)]$$

y donde $H(x,y)=0$ si $\phi(x)+\phi(y)\geq 0$. La función ϕ^{-1} tiene las mismas propiedades de ϕ pero con $\phi^{-1}(0)=1$.

Supongamos que, además, ϕ posea dos derivadas continuas en (0,1). Una cópula H es arquimediana si tiene dos derivadas parciales y si existe

$$f:(0,1)\rightarrow(0,\infty) \text{ integrable y tal que } f(y)\frac{\partial H}{\partial x}=f(x)\frac{\partial h}{\partial y} \quad .$$

La función ϕ que engendra H es

$$\phi(t)=\int_t^1 f(s)\,ds \quad .$$

Las cuestiones anteriores son válidas también si en lugar de funciones de distribución se consideran funciones de supervivencia. En este caso se tiene

$$S(t_1,t_2)=\phi^{-1}[\phi(S_1(t_1))+\phi(S_2(t_2))] \quad .$$

1.4 REVISIÓN BIBLIOGRÁFICA

A continuación hacemos un breve estudio de los antecedentes más importantes en supervivencia multivariante. De acuerdo a los objetivos anteriormente descritos consideramos como estudios más destacados los que a continuación citamos, por orden cronológico dentro de cada tema. En esta revisión distinguimos los siguientes temas: conceptos básicos, construcción de modelos, estudio de la dependencia y estimación.

a) CONCEPTOS BÁSICOS

En el apartado de conceptos básicos debemos mencionar inicialmente a Cox (1972) quien, en su ya clásico artículo, trata los modelos de supervivencia univariantes considerando la presencia de covariables explicativas. Considera el análisis de la regresión en el modelo de azar proporcional, posteriormente llamado de Cox, e introduce conceptos de verosimilitud condicional. El modelo de azar proporcional se define como aquel que verifica la condición

$$h(t;z)=\exp(z\beta)h_0(t)$$

donde β es un vector de parámetros desconocidos, $h_0(t)$ es una función de azar base (función de azar para z=0) y z es un vector de covariables explicativas que toma valores en cada individuo. La función de azar va tomando un valor proporcional al azar base sin más que multiplicar por un factor (usualmente en forma exponencial) que varía de individuo a individuo según el valor de las covariables. En este artículo utiliza también métodos no paramétricos (método del *producto límite*) para obtener una estimación de la función de supervivencia.

Establece los conceptos fundamentales de supervivencia bivariante definiendo el vector función de azar tal como lo presentamos en el apartado 1.2.2. Da una expresión de la función de densidad de los tiempos de supervivencia en función de las componentes del vector función de azar. Utiliza datos relativos a enfermos de leucemia.

En Puri & Rubin (1974) se establece una caracterización de la familia de distribuciones con razón de fallo multivariante constante demostrando que si dicha razón de fallo es constante entonces se trata de mixturas de distribuciones exponenciales. En Johnson & Kotz (1975) se da una definición del vector razón de azar multivariante expresándolo en la siguiente forma:

$$h_x(x)=\left[-(\partial/\partial x_1),\ldots,-(\partial/\partial x_m)\right].\log F_x(x)=-\nabla \log F_x(x)$$

$$\text{donde} \quad F_x(x)=P(X_i<x_i,\, i:1,\ldots,m) \quad .$$

Inician estudios sobre dependencia en algunas distribuciones importantes tales como la distribución normal bivariante, Pareto multivariante, exponenciales bivariantes (Gumbel, Marshall y Olkin), logística bivariante, etc. Otros autores que estudian dicho vector pueden verse al final del apartado 1.2 .

b) CONSTRUCCIÓN DE MODELOS

En Marshall & Olkin (1967) se estudia una distribucion exponencial multivariante fuera del contexto de la supervivencia pero desde el punto de vista de su aplicabilidad, basándose en los llamados modelos de choque, modelos en los que las componentes de un sistema sufren sucesivos deterioros por distintas causas, y calculan la función generatriz de momentos de dicha distribución. Bajo estas hipótesis, la función de supervivencia resultante es

$$P[X > s\ ,\ Y > t] = \exp[-\lambda_1 s - \lambda_2 t - \lambda_{12}\max(s,t)],\quad s,t>0\ .$$

Mediante un cambio de variables obtienen la siguiente distribución, denominada Weibull multivariante

$$S(x,y) = \exp[-\lambda^1 x^{\beta} - \lambda^2 y^{\gamma} - \lambda^{12}\max(x^{\beta},y^{\gamma})]\ .$$

Pueden considerarse los primeros modelos propiamente dichos de supervivencia multivariante.

Siguiendo con la linea anterior, en Downton (1970) se aborda el estudio de las distribuciones exponenciales bivariantes, dentro del contexto de la Fiabilidad, considerando un modelo basado en la existencia de sucesivos daños que sufre un sistema. Estudia la propiedad de que la suma de variables correladas tiene una distribución que puede expresarse como convolución de dos variables exponenciales independientes. Hace un estudio de la regresión y correlación y obtiene los momentos y los compara con los de la distribución de Marshall y Olkin. En Block (1977) se estudian igualmente algunas propiedades de familias de distribuciones exponenciales bivariantes de tiempos de fallo que generalizan los modelos de Marshall y Olkin, Downton y otros, siguiendo la linea de los modelos de choque anteriormente considerados. Dichas propiedades se refieren a la forma de la función característica. Deduce que sólo la distribución de Downton es infinitamente divisible. En Barlow & Proschan (1977) se estima el vector función de azar con la intención de determinar una formulación de la distribución conjunta, todo ello en un modelo que sigue una distribución exponencial. Utilizan datos relativos a los tiempos de fallo de las transmisiones de tractores oruga. En Friday & Patil (1977) se estudian los modelos exponenciales bivariantes formulando un nuevo modelo, la distribución exponencial bivariante extendida, que incluye algunas distribuciones especiales de tiempos de vida como la distribución exponencial bivariante de Marshall y Olkin, las distribuciones exponenciales bivariantes de Block y Basu, la distribución exponencial bivariante de Freund, etc.

En el importante trabajo de Clayton (1978) se describe su conocido modelo de asociación bivariante. Con unas pocas hipótesis y resolviendo algunas ecuaciones diferenciales consigue formular un modelo. A este modelo se le prestará especial atención en el capítulo 2. Este modelo es muy estudiado posteriormente por diversos autores, obteniéndose, además, por otras vías, por el método de las mixturas de distribuciones o método frailty. Utiliza datos censurados de naturaleza epidemiológica en enfermedades crónicas no infecciosas en padre-hijo.

En Oakes (1982) se estudia el modelo de Clayton y se critica la función de verosimilitud propuesta por Clayton en el caso de distribuciones marginales desconocidas. La crítica se refiere a errores que, según su criterio, tiene la verosimilitud parcial obtenida por Clayton en dicho modelo semiparamétrico. En (1989), en un ya clásico artículo, estudia los modelos frailty introduciendo una función (*cross-ratio function*) que caracteriza los distintos modelos de tiempos de fallo. Dicha función adopta la siguiente forma

$$\theta^* = \frac{S(t)\,(D_1 D_2 S(t))}{(D_1 S(t))\,(D_2 S(t))} .$$

Para diversas expresiones de esta función obtiene formulaciones de modelos distintos.

En Hougaard (1984) se inicia un estudio sobre supervivencia bivariante. En este artículo se tratan modelos univariantes, introduciendo los conceptos de heterogeneidad y de frailties (factores de riesgo, covariables aleatorias no observadas), siempre en modelos con azar multiplicativo, esto es, modelos de Cox. Considera la distribución inversa gaussiana para dicho frailty y utiliza datos de tipo biológico. En el artículo (1986a) se establece un modelo, denominado de Hougaard, un modelo frailty con distribución estable positiva. Considera el caso en que las distribuciones marginales son Weibull y lo aplica a datos de aparición de tumores en ratas. En el artículo (1986b) estudia una familia de distribuciones que engloba a los modelos con frailty con distribución gamma, con distribución estable y con distribución inversa Gaussiana. Dicha familia está caracterizada por su transformada de Laplace, de donde deduce propiedades relativas a sus momentos, convoluciones, divisibilidad infinita, etc. Las aplica al estudio de datos de enfermos de infarto de miocardio. En el artículo de (1989) revisa propiedades anteriormente consideradas en el modelo de Hougaard de la referencia (1986a). Tales propiedades se refieren a los momentos de la distribución. Especial interés muestra por un caso particular, la distribución de Weibull multivariante, la cual también puede obtenerse como un modelo de vida acelerada, modelo que citaremos en la siguiente referencia.

En Clayton & Cuzick (1985), en un también ya clásico artículo, se retoma el modelo de Cox y se generalizan al caso multivariante los modelos de azar proporcional mediante la inclusión de un efecto aleatorio que representa la heterogeneidad. Se extiende, asimismo, el modelo considerando la presencia de covariables. El modelo de azar proporcional bivariante que generaliza el modelo de Cox verifica

$$h_i(t/z)=h_i^0(t)\exp(a_i z), \quad i:1,2$$
$$h_i(t)=E[h_i(t/z)]$$

donde $(h_1(t),h_2(t))$ es el vector función de azar, z es una covariable aleatoria no observada, $h_i(t/z)$ es la función de azar condicionada a un valor de z y $h_i^0(t)$, la función de azar base. Los métodos usados son comparados y validados con simulaciones por el método de Montecarlo. Tratan datos censurados a la derecha relativos a enfermos (padre-hijo) de cancer y del corazón.

En Crowder (1985) se estudia un modelo similar al de Clayton (frailty con distribución gamma) suponiendo que las marginales son Weibull y datos de naturaleza química sobre la concentración de plomo en ratas. Hace un estudio de los momentos, distribuciones condicionadas, la razón de fallo y la distribución del mínimo.

En Genest & Mackay (1986a,b) se estudia una clase de distribuciones bivariantes cuyas marginales son uniformes en el intervalo unidad, denominadas cópulas arquimedianas, de gran importancia para el estudio de los modelos frailty. Dan asimismo una interpretación geométrica del coeficiente de concordancia de Kendall. En Lindley & Singpurwalla (1986) se estudian nuevos modelos frailty tales como un modelo frailty con distribución gamma generalizada y un modelo frailty con distribución beta. Obtienen la correspondiente formulación del modelo. En Basu (1988) se hace un estudio de distribuciones exponenciales multivariantes en el contexto de las aplicaciones en Fiabilidad, pero sin utilizar la noción de supervivencia. En Lee & Klein (1988) se sugiere una distribución uniforme para modelos frailty. Inciden sobre todo en el caso de marginales Weibull. Llegan al importante resultado de obtener una expresión del coeficiente de correlación en un modelo frailty general con distribuciones marginales Weibull. En Aalen (1988) se propone una clase de modelos frailty que engloba algunas distribuciones usadas por Hougaard (gamma, estable y Poisson compuesta). El planteamiento se basa en la llamada *función intensidad*, que no es sino una nueva formulación del modelo. Dicha función viene definida por

$$\mu(t)=\frac{\lambda(t)}{1+\delta\alpha^{-1}\Lambda(t)^{\alpha}}, \quad \alpha,\delta>0 \quad .$$

Estudia datos de enfermos de cancer. En Marshall & Olkin (1988) se hace un exhaustivo estudio de las familias de distribuciones multivariantes generadas por sus marginales, es decir,

$$H(x_1,x_2/\theta)=\psi(H_1(x_1),H_2(x_2)/\theta) \quad .$$

Utilizan para ello la teoría de la mixtura de distribuciones. Presentan multitud de ejemplos generados de esta forma, muchos de ellos modelos ya conocidos (modelo de Clayton, modelo de Hougaard, etc.). En Crowder (1989) se generalizan los modelos frailty con distribución estable y distribuciones marginales Weibull estudiando sus distribuciones marginales y condicionadas. Dicha generalización consiste en lo siguiente

$$S(t)=\exp[k^{\nu}-(k+s)^{\nu}] \quad .$$

Utiliza datos sobre la absorción de plomo en ratas. En Whitmore & Mei-Ling Ting Lee (1991) se estudia un modelo frailty con distribución inversa gaussiana y distribuciones marginales exponenciales. Presentan una fórmula general de los momentos. Utilizan datos de accidentes en individuos. En Guo & Rodríguez (1992) se hace un estudio con datos agrupados, datos considerados a intervalos de tiempo fijo (semanales, mensuales), correspondientes a tiempos de vida de niños en Guatemala, y modelos con frailty con distribución gamma o frailty con distribución desconocida.

En Costigan & Klein (1993) se retoman algunos modelos frailty con distribuciones ya consideradas por otros autores (gamma, estable, estable generalizada, inversa gausiana, etc.) y los presentan en un contexto general. En Mei-Ling Ting Lee (1993) se hace un estudio de los diversos tipos de mixturas de distribuciones de tiempos de vida enfocado a su posterior utilización en supervivencia. En Sun & Basu (1993) se estudian algunas familias de distribuciones exponenciales bivariantes, incidiendo sobre todo en la razón de fallo de dichas distribuciones. En Guo & Lin (1993) se estudia el siguiente modelo de regresión para datos agrupados de supervivencia multivariante

$$\lambda_m(t,z_m)=\lambda_{m0}(t)\exp[\beta' z_m(t)], \quad m:1,\ldots,M \ .$$

En Sinha, Tanner & Hall (1994) se hace un estudio con datos agrupados, utilizando el concepto de la verosimilitud parcial, concepto introducido por Cox (1972) para estimar los parámetros de regresión en los modelos semiparamétricos. En Anderson & Louis (1995) se presentan los modelos frailty pero considerados como modelos de vida acelerada y estudian propiedades básicas para estos modelos, ya estudiadas para los modelos de azar proporcional, tales como medidas de asociación (Oakes), identificabilidad, etc. Los modelos de vida acelerada bivariantes verifican la siguiente igualdad en términos de la función de supervivencia

$$S(t_1,t_2)=\int_0^\infty \exp[-(H_1(zt_1)+H_2(zt_2))]\,dF(z)$$

donde z es un frailty. En el modelo de vida acelerada zt_i actúa "dentro" de la función de azar como una sola variable. Por último, en Koehler & Symanowski (1995) se presentan nuevos métodos para construir distribuciones multivariantes con distribuciones marginales especificadas y muestran gráficos de las correspondientes densidades. Posteriormente lo aplican a casos de supervivencia. Este artículo sigue la metodología del estudio de Marshall y Olkin de 1988. Las distribuciones son del tipo

$$F(x_1,\ldots,x_p)=\prod_{i=1}^{p} F_i(x_i)\prod\prod_{i<j} c_{ij}^{-\alpha_{ij}}$$

donde

$$c_{ij}=F_i(x_i)^{\frac{1}{\alpha_{i+}}}+F_j(x_j)^{\frac{1}{\alpha_{j+}}}-F_i(x_i)^{\frac{1}{\alpha_{i+}}}F_j(x_j)^{\frac{1}{\alpha_{j+}}}$$

y

$$\alpha_{ij}=\alpha_{ji}\geq 0,\ \forall i,j \quad y \quad \alpha_{i+}=\alpha_{i1}+\ldots+\alpha_{ip}>0,\ \forall i=1,\ldots,p.$$

c) ESTUDIO DE LA DEPENDENCIA

En el apartado sobre la asociación o dependencia mencionemos, en primer lugar, a Lehmann (1966), que sienta las bases de algunos conceptos de dependencia en modelos generales, uno de los objetivos fundamentales de la supervivencia multivariante. Trata entre otros conceptos la *dependencia cuadrante (quadrant dependent)* definida por la desigualdad

$$F(x,y)\geq F(x)F(y), \quad \forall x,y \quad ,$$

el coeficiente de Kendall, de Spearman y otras medidas de asociación. En Johnson & Kotz (1975) se estudia el vector razón de azar multivariante y se consideran medidas de asociación iniciadas por Lehmann (véase apartado 1.2.2). En Schweizer & Wolf (1981) se tratan importantes medidas de asociación (coeficiente de correlación de Pearson, Spearman y Kendall) y la noción de cópulas orientada hacia la dependencia. Demuestran que una cópula de dos variables aleatorias es invariante bajo transformaciones estrictamente crecientes. En Genest & Mackay (1986a,b), en un estudio de las cópulas arquimedianas, se considera un coeficiente de asociación como es el coeficiente de Kendall. Dan una interpretación geométrica de dicho coeficiente. En Lee & Klein (1988) se estudian también propiedades de dependencia en el modelo frailty con marginales Weibull dando una expresión del coeficiente de correlación. En Hougaard, Harvald & Holm (1992) se comparan, en un estudio sobre los tiempos de vida de los gemelos daneses nacidos en un periodo determinado, ciertas medidas de dependencia en diversos modelos frailty tales como el coeficiente de variación, el coeficiente de correlación y el coeficiente de Kendall. En Lindeboom & Van den Berg (1994) se generaliza la idea de frailty al caso de dos variables no observadas y hacen un estudio de la correlación. Es decir, se considera una covariable aleatoria no observada para cada variable tiempo de supervivencia. Es, en cierto sentido, una idea curiosa aunque se pierde la noción intuitiva de frailty, ya que existe un frailty para cada variable. En Manatunga & Oakes (1996) se estudia el coeficiente de concordancia de Kendall para modelos frailty bivariantes. Obtienen una expresión de la varianza del estimador del citado coeficiente. Particularizan al caso de un frailty con distribución estable positiva. La expresión que obtienen para dicha varianza es

$$V(U)=\frac{2(1-\theta^2)}{n(n-1)}+\frac{4(n-2)}{n(n-1)}\left[-\frac{1}{3}+8V_1+8V_2-\theta^2\right]$$

donde

$$\theta=E(U)=4\left[\frac{W_1^2}{(W_1+W_2)^2}\right]-1$$

$$V_1=E\left[\frac{W_1^2W_2^2}{(W_1+W_2+W_3)^2(W_1+W_3)^2}\right]$$

$$V_2=E\left[\frac{W_1W_2^2W_3}{(W_1+W_2+W_3)^2(W_1+W_2)(W_2+W_3)}\right]$$

siendo W_1,W_2 y W_3 variables aleatorias independientes e idénticamente distribuidas correspondientes a la distribución del frailty.

d) ESTIMACIÓN

En el apartado de la estimación, Cox (1975) trata con rigor el concepto de la verosimilitud parcial, generalizando las ideas de la verosimilitud condicional y marginal, básicas en un estudio semiparamétrico de los modelos. La verosimilitud para el vector de los coeficientes de regresión es

$$\prod_j \frac{\exp(\beta^T z_{(j)})}{\sum_{k\in R_j}\exp(\beta^T z_k)} \quad .$$

En Barlow & Proschan (1977) se hacen estimaciones en el modelo mencionado en esta misma referencia del apartado MODELOS de esta sección. En Clayton (1978) se hace una estimación del parámetro de asociación dentro del contexto de los modelos semiparamétricos. Construye una verosimilitud muy criticada por diversos autores. En Prentice, Williams & Peterson (1981) se estudian los modelos semiparamétricos bajo el concepto de la verosimilitud parcial y utilizan datos relativos a enfermos de anemia y leucemia. Se Oakes (1982) se hacen inferencias sobre el parámetro de asociación en el modelo de Clayton en el caso de marginales exponenciales y se critica el método de estimación de Clayton en el caso semiparamétrico formulando una alternativa no paramétrica basada en el coeficiente de Kendall. En el artículo de (1986) hace inferencia semiparamétrica en el modelo de Clayton. En Clayton & Cuzick (1985) se hacen inferencias en modelos semiparamétricos, utilizando los conceptos de verosimilitud parcial. En Basu (1988) se hacen estimaciones en los modelos mencionados en esta referencia del apartado MODELOS. En Wei, Lin & Weissfeld (1989) se hace inferencia semiparamétrica en modelos de Regresión de Cox, utilizando conceptos de verosimilitud parcial y lo aplican a datos de cancer. En Costigan & Klein (1993) se muestran algunos métodos de estimación semiparamétrica en su artículo de revisión de algunos modelos anteriormente estudiados por Clayton, Hougaard, etc. En Wassell & Moeschberger (1993) se describe un ejemplo de datos, realmente ilustrativo, para un modelo con frailty gamma. Dichos datos se refieren a tiempos de fallo de enfermos de hipertensión y del corazón. En Maguluri (1993) se hace un estudio completo de inferencia semiparamétrica, a partir del modelo de Clayton, utilizando la verosimilitud parcial.

CAPÍTULO 2
MÉTODOS DE CONSTRUCCIÓN DE MODELOS

2.1 INTRODUCCIÓN

En el capítulo 1 de esta memoria hemos visto que uno de los objetivos fundamentales en análisis de datos de supervivencia multivariante es el planteamiento, la formulación y estudio de modelos. Interesa construir un modelo y a partir de, por ejemplo, la función de supervivencia hacer el estudio de la dependencia entre las variables tiempos de fallo, la estimación de los parámetros, etc. Se trata, por consiguiente, de determinar esta función. La expresión que adopte la función de supervivencia dependerá del tipo de datos disponibles, de las hipótesis impuestas y de otras consideraciones. Esta función deberá ajustarse lo mejor posible a los datos.

En este capítulo presentamos una breve clasificación de los distintos métodos de construcción de modelos, de acuerdo a lo mostrado en la literatura. Nos detendremos especialmente en el método denominado *frailty* y en el método de modelos generados por distribuciones marginales.

En este capítulo presentamos como aportaciones los siguientes puntos: el estudio de un modelo frailty con distribución uniforme, el planteamiento de un modelo paramétrico con frailty estable y marginales Pareto, el planteamiento de modelos basados en hipótesis de vida acelerada (poco tratados en la literatura), la introducción de modelos *mixtos* (frailty vía azar proporcional y covariables bajo hipótesis de vida acelerada) y el estudio del problema de un *frailty inverso*.

En todos los casos del párrafo anterior nosotros consideramos la función de supervivencia del modelo en la que estarán presentes los parámetros (incluido alguno que describa la asociación entre los tiempos de fallo) junto a las funciones de supervivencia marginales (véase apartado 1.1.2). Consideramos modelos paramétricos, es decir, supondremos conocida la familia a la que pertenecen las distribuciones marginales aunque sus parámetros sean desconocidos.

Nosotros no consideramos el otro enfoque usualmente utilizado, es decir, partir de las funciones de azar marginales en función de las funciones de azar base y de cierta función de covariables explicativas. El objetivo en este caso sería conseguir estimaciones sobre los parámetros asociados a las covariables, denominados parámetros de regresión. Según se conozca o no la familia a la que pertenecen las distribuciones de las funciones de azar base se denominarían modelos paramétricos o semiparamétricos, respectivamente. Este último caso se mencionará en el capítulo 4. Aproximaciones no paramétricas de la formulación del modelo se consiguen con una metodología sensiblemente diferente a los casos anteriores y no las consideraremos en esta memoria.

Hemos de observar que con planteamientos totalmente distintos, desde distintos modelos, se llega en muchos casos a la misma función de supervivencia, es decir, con distintas hipótesis iniciales puede llegarse al mismo resultado. Presentamos a continuación una clasificación de los métodos de construcción de modelos, de acuerdo a lo que ha sido más utilizado por los distintos autores. Sólo nos detendremos en los casos que hemos considerado de mayor interés.

a) Modelos generados por medio de transformaciones de variables aleatorias independientes. En Marshall & Olkin (1967) se estudia una distribución generada por este método. En Klein, Keiding y Kamby (1989) y en Hougaard (1986) se construyen modelos generados también por este método.

b) Modelos basados en distribuciones condicionadas. En Freund (1961), en Leurgans y otros (1982) y en Clayton (1978) se construyen modelos correspondientes a este caso.

c) Modelos generados por medio de mixturas de distribuciones o modelos de efecto aleatorio (*frailties*). Esta clase de modelos también se denomina modelos de *independencia condicionada*. En Clayton (1978), Clayton & Cuzick (1985), Lindley & Singpurwalla (1986), Aalen (1988), Lee & Klein (1988), Hougaard (1984,1986,1987,1989), Hougaard y otros (1992), Oakes (1982, 1986, 1989) y Klein y otros (1991) se estudian estos métodos. De entre todos ellos destacan los modelos de Clayton y de Hougaard.

d) Modelos generados a partir de las distribuciones marginales. En Marshall & Olkin (1988) y en Genest & Mackay (1986) se hace un estudio exhaustivo de estos casos.

e) Modelos basados en hipótesis de vida acelerada. En los métodos anteriores se obtienen modelos generalmente de azar proporcional. Si se parte de hipótesis de vida acelerada se obtienen ecuaciones sensiblemente diferentes. Estos modelos han sido muy poco tratados aún, destacando recientemente el trabajo de Anderson & Louis (1995).

f) Modelos basados en motivaciones de carácter físico. En Gumbel (1960) y en Dowton (1970) se estudian estos modelos.

g) Modelos basados en aplicaciones de los procesos de Markov, Poisson y Polya. En Platz (1984) se estudian modelos de este tipo.

2.2 MODELOS BASADOS EN TRANSFORMACIONES DE VARIABLES ALEATORIAS INDEPENDIENTES

En esta familia de modelos destacan por su importancia los modelos *de Hougaard y de Marshall y Olkin* que describimos a continuación.

2.2.1 MODELO DE HOUGAARD

El modelo de *Hougaard* es uno de los más utilizados en supervivencia. Su importancia estriba en que se obtuvo posteriormente como un modelo frailty (este método se verá en el apartado 2.4) con distribución estable positiva, una de las distribuciones más utilizadas en supervivencia. El procedimiento original de obtención por Hougaard (1986) fué el siguiente: Sean X_1 y X_2 dos variables aleatorias independientes con distribución de Weibull, es decir, con funciones de supervivencia

$$S_1(x_1)=\exp(-\varepsilon_1 x_1^{\gamma}) \quad , \quad S_2(x_2)=\exp(-\varepsilon_2 x_2^{\gamma})$$

respectivamente. Sea Y una variable aleatoria con distribución estable positiva, es decir, definida por medio de la transformada de Laplace

$$E[\exp(-sY)]=\exp(-s^{\alpha}) \;,\; \alpha \in (0,1]$$

y verificándose además que X_1, X_2 e Y sean independientes. Se definen las variables

$$T_1=X_1 \cdot Y^{-1/\gamma}$$

$$T_2=X_2 \cdot Y^{-1/\gamma} \quad .$$

En estas condiciones, la función de supervivencia condicionada será

$$S(t_1, t_2/Y)=P(T_1\geq t_1, T_2\geq t_2)=P(Y^{-1/\gamma}X_1\geq t_1, Y^{-1/\gamma}X_2\geq t_2)=$$

$$=P(X_1\geq Y^{1/\gamma}t_1, X_2\geq Y^{1/\gamma}t_2)=P(X_1\geq Y^{1/\gamma}t_1)\,P(X_2\geq Y^{1/\gamma}t_2)=$$

$$=\exp[-\varepsilon_1(Y^{1/\gamma}t_1)^\gamma]\exp[-\varepsilon_2(Y^{1/\gamma}t_2)^\gamma]=\exp[-Y(\varepsilon_1 t_1^\gamma+\varepsilon_2 t_2^\gamma)]$$

de donde se obtiene que la función de supervivencia conjunta es

$$S(t_1,t_2)=\int\exp[-Y(\varepsilon_1 t_1^\gamma+\varepsilon_2 t_2^\gamma)]\,dF(Y)=$$

$$=E[\exp(-Y(\varepsilon_1 t_1^\gamma+\varepsilon_2 t_2^\gamma))]=\exp[-(\varepsilon_1 t_1^\gamma+\varepsilon_2 t_2^\gamma)^\alpha]\ . \qquad [2.1]$$

La distribución resultante se denomina Weibull bivariante.

2.2.2 MODELO DE MARSHALL & OLKIN

El modelo descrito por Marshall & Olkin (1967) se formuló fuera del contexto de la supervivencia, dentro del contexto de los modelos de choque, y fué retomado posteriormente para esta disciplina. La función de supervivencia que se obtiene de este modelo resulta ser una de las distribuciones exponenciales bivariantes más importantes.

El procedimiento para su obtención se describe a continuación. Se considera que las variables aleatorias X_1, X_2 y X_3 son independientes y con distribución exponencial de parámetros

$$\lambda_1\ ,\ \lambda_2\ \text{ y }\ \lambda_3$$

respectivamente (dichos parámetros son, pués, las funciones de azar de dichas variables aleatorias). Sean T_1 = mín (X_1,X_3) y T_2 = mín (X_2,X_3). En estas condiciones, las variables T_1, T_2 y T = mín (T_1,T_2) siguen también distribuciones exponenciales y la función de supervivencia conjunta de (T_1,T_2) tiene la expresión

$$S(t_1,t_2)=P(T_1\geq t_1, T_2\geq t_2)=\exp[-(\lambda_1 t_1+\lambda_2 t_2+\lambda_3 \text{máx}(t_1,t_2))]\ . \qquad [2.2]$$

Obsérvese que el caso de independencia se presenta cuando el parámetro $\lambda_3=0$.

Este modelo, también llamado modelo de choque, tiene interesantes propiedades como la "falta de memoria" expresada en la forma

$$P(T_1 \geq s+t_1, T_2 \geq s+t_2 / T_1 \geq s, T_2 \geq s) = P(T_1 \geq t_1, T_2 \geq t_2) \quad .$$

2.3 MODELOS BASADOS EN DISTRIBUCIONES CONDICIONADAS

Dentro de este tipo de modelos vamos a describir los debidos a Freund y especialmente el debido a Clayton.

2.3.1 MODELO DE FREUND

La función de densidad del modelo de Freund adopta, al igual que la del modelo de Marhsall y Olkin, una de las expresiones que definen una distribución exponencial bivariante. El procedimiento de obtención de este modelo es el siguiente. Sean dos componentes A y B cuyos respectivos tiempos de fallo T_1 y T_2 se distribuyen exponencialmente:

$$f(t_1) = \alpha . \exp(-\alpha t_1) \; , \; f(t_2) = \beta . \exp(-\beta t_2) \quad \text{con} \quad \alpha, \beta, t_1, t_2 > 0 \quad .$$

Se considera que el fallo de una componente cambia el parámetro de la distribución de la otra componente. Si A falla, el parámetro β pasa a ser β', y recíprocamente. Es decir, la función de azar de la variable T_1 es α si la otra componente no ha fallado y α' si lo ha hecho. Nótese que el planteamiento de este modelo está presente, por ejemplo, en el funcionamiento de los riñones de un individuo, en mecanismos técnicos en la industria, etc. La función de densidad conjunta queda

$$f(t_1, t_2) = \begin{cases} \alpha\beta' \exp[-\beta' t_2 - (\alpha+\beta-\beta') t_1] \; , & \text{si} \quad 0<t_1<t_2 \\ \alpha'\beta \exp[-\alpha' t_1 - (\alpha+\beta-\alpha') t_2] \; , & \text{si} \quad 0<t_2<t_1 \end{cases} \qquad [2.3]$$

2.3.2 MODELO DE CLAYTON

El modelo denominado *de Clayton* merece una atención especial. La función de supervivencia fué obtenida posteriormente por otro procedimiento, por el método frailty, resultando un modelo frailty con distribución gamma, otra de las distribuciones más usuales en supervivencia. En cambio, Clayton (1978) la obtuvo por un procedimiento directo, como veremos a continuación.

Las hipótesis que se suponen inicialmente son las siguientes:

- La asociación entre T_1 y T_2 sólo vendrá explicada por cierto parámetro ϕ.

- Las funciones de supervivencia marginales no dependen de este parámetro.

- Debe cumplirse que

$$\text{máx}[0, S_1(t_1)+S_2(t_2)-1] \le S(t_1, t_2) \le \text{mín}[S_1(t_1), S_2(t_2)]$$

donde los casos extremos se alcanzan en el caso de haber una posible mínima o máxima asociación, respectivamente.

- Debe considerarse que el siguiente cociente es constante,

$$\frac{h_1'(t_1, t_2)}{h_1(t_1, t_2)} = (1+\phi)$$

(donde estas funciones corresponden a las definidas en 1.2.2), o lo que es lo mismo

$$h_1(t_1/T_2=t_2) = (1+\phi)\, h_1(t_1/T_2 \ge t_2)$$

donde ϕ es un valor constante. Se puede expresar, según vimos, mediante la notación condicional

$$\frac{h_1(t_1/T_2=t_2)}{h_1(t_1/T_2\geq t_2)}=1+\phi \quad .$$

Clayton considera que $\phi \geq 0$. Por lo tanto, este cociente anterior ha de ser una constante mayor o igual que la unidad (en general no tiene porqué serlo, es una función de t_1 y t_2). Obsérvese que si el parámetro toma el valor cero se trataría del caso de independencia de variables.

En Lee & Klein (1988) se demuestra, para modelos frailty que veremos en el apartado siguiente, que las funciones

$$h_1(t_1/T_2=t_2) \text{ y } h_1(t_1/T_2\geq t_2)$$

son decrecientes en t_2 para cualquier valor t_1 y además se cumple

$$h_1(t_1/T_2=t_2) > h_1(t_1/T_2\geq t_2) \;, \forall\; t_1, t_2 \quad .$$

Para valores cada vez mayores del parámetro existirá mayor asociación entre las variables correspondiendo el extremo al valor mín($S_1(t_1)$,$S_2(t_2)$).

Asociaciones negativas o inversas implican que

$$\phi \leq 0$$

pero estos casos no se tienen en cuenta en este modelo.

Veamos en forma resumida el procedimiento para la obtención de la función de supervivencia del modelo. Sustituyendo [1.4] y [1.5] en el cociente anterior se tiene

$$-\frac{\partial}{\partial t_1}\lg\left[-\frac{\partial S(t_1,t_2)}{\partial t_2}\right]=(1+\phi)\left[-\frac{\partial}{\partial t_1}\lg S(t_1,t_2)\right] \quad .$$

Integrando dicha ecuación diferencial se obtiene la función de supervivencia bivariante

$$S(t_1,t_2)=\left[\frac{1}{[S_1(t_1)]^{\phi}}-\frac{1}{[S_2(t_2)]^{\phi}}-1\right]^{-\frac{1}{\phi}}\ ,\quad \phi>0\quad . \qquad [2.4]$$

Observemos que para llegar a la expresión anterior también se podía haber partido de la expresión

$$\frac{\partial^2 S(t_1,t_2)}{\partial t_1 \partial t_2}S(t_1,t_2)=(1+\phi)\frac{\partial S(t_1,t_2)}{\partial t_2}\frac{\partial S(t_1,t_2)}{\partial t_1}$$

que dividida por $S^2(t_1,t_2)$ quedaría así

$$r(t_1,t_2)=(1+\phi)h_1(t_1,t_2)h_2(t_1,t_2)\quad ,$$

ecuación diferencial que integrada conduce a

$$S(t_1,t_2)=[1+\phi(\Lambda_1(t_1)+\Lambda_2(t_2))]^{-\frac{1}{\phi}} \qquad [2.5]$$

con $\phi>0$ y donde $\Lambda_i(t_i)$ es la función de azar acumulativa marginal de T_i . Se observa que si $t_2=0$ entonces la función de supervivencia marginal es

$$S_1(t_1)=[1+\phi\Lambda_1(t_1)]^{-\frac{1}{\phi}} \implies \Lambda_1=\frac{[S_1(t_1)]^{-\phi}-1}{\phi}\quad . \qquad [2.6]$$

El modelo de Clayton se puede generalizar simplemente considerando que $1+\phi$ en lugar de ser constante sea función de t_1 y t_2 . Para formas sencillas de esta función, por ejemplo lineales, la resolución de la correspomdiente ecuación diferencial es abordable. Sin embargo, el camino para obtener la función de supervivencia será otro ya que relacionaremos este método con el método frailty, que veremos a continuación.

2.4 MODELOS FRAILTY

2.4.1 PLANTEAMIENTO GENERAL

Uno de los procedimientos más importantes para la construcción de modelos de supervivencia multivariante se basa en la idea debida a Vaupel, Manton & Stallard (1979) para explicar la heterogeneidad de la población en modelos univariantes. Rápidamente esta idea fué extendida a supervivencia bivariante. Nosotros la consideramos de especial importancia ya que será el método que usaremos para nuestras principales aportaciones.

La idea en la que se basa este método es la siguiente. La asociación posible entre los tiempos de supervivencia, T_1 y T_2, se explica a través de una característica común que poseen los individuos de una pareja. Esta característica común induce cierta dependencia entre los tiempos. La heterogeneidad se modeliza por un factor de riesgo, una covariable aleatoria no observable, denominada frailty. Para cada individuo de una pareja concreta el frailty toma el mismo valor y distinto entre individuos de parejas distintas. Si se denota por Z la variable aleatoria frailty, se considera que T_1 y T_2 son condicionalmente independientes, es decir, son independientes para cada valor fijo de Z. En cierto sentido, como se verá a continuación, se podría considerar el frailty como una covariable aleatoria no observada y se suele considerar que actúa multiplicativamente en la función de azar, es decir, bajo la hipótesis de azar proporcional.

La idea intuitiva de este efecto aleatorio es la siguiente. La asociación entre los tiempos de supervivencia proviene de un factor desconocido denominado frailty y no porque la incidencia de T_1 influya directamente en el valor de T_2; es decir, por ejemplo, que un padre se muera de infarto no influye de una forma directa en que el hijo lo vaya a hacer pero sí existe un factor que poseen ambos, por ejemplo, un factor genético, que liga al padre con el hijo.

Vamos a considerar el caso bivariante (dos tiempos de fallo) e hipótesis de azar proporcional, es decir, cada función de azar marginal (univariante) se comporta, en cuanto al frailty, como un modelo de azar proporcional. En lo que sigue consideraremos que E(Z) = 1 lo cual no resta generalidad. *Si no fuera así*, en lugar de $h_i(t_i)$ habría de tomarse un azar base o inicial $h_{i0}(t_i)$. Es decir

$$h_i(t_i/Z)=Zh_i(t_i) \quad , \quad i:1,2$$

o, lo que es lo mismo, la función de supervivencia es

$$S_i(t_i/Z)=\exp[-z\int_0^{t_i}h_i(t_i)\,dt_i]=\exp[-Z\Lambda_i(t_i)]=[S_i(t_i)]^Z$$

donde

$$\Lambda_i(t_i)$$

es la función de azar acumulativa marginal correspondiente a la variable T_i .

En Hougaard (1986a) se demuestra que si el frailty actúa multiplicativamente en las componentes del vector función de azar, así lo hace también en las componentes del vector función de azar marginal aunque con distinta constante de proporcionalidad.

La función de supervivencia condicionada es

$$S(t_1,t_2/Z)=P(T_1\geq t_1,T_2\geq t_2/Z)=$$
$$=P(T_1\geq t_1/Z)P(T_2\geq t_2/Z)=[S_1(t_1)]^Z[S_2(t_2)]^Z=$$

$$=[S_1(t_1)S_2(t_2)]^Z=\exp[-Z\Lambda_1(t_1)]\exp[-Z\Lambda_2(t_2)]=$$

$$=\exp[-Z(\Lambda_1(t_1)+\Lambda_2(t_2))]$$

y realizando la mixtura respecto a la distribución del frailty Z, se verificará que la función de supervivencia conjunta es

$$S(t_1,t_2)=\int S(t_1,t_2/Z)\,dF(z)=E[\exp(-Z(\Lambda_1(t_1)+\Lambda_2(t_2)))] \qquad [2.7]$$

El tratamiento del modelo dependerá, por la expresión anterior, de la distribución del frailty Z y, sobre todo, de la resolubilidad de la integral anterior. Ésta no es fácil de resolver en general, pero veremos más adelante un método para solucionar este problema.

Veamos antes un caso sencillo, concreto, estudiado por Hougaard (1986). Sea Z una variable aleatoria con distribución estable positiva. La definición de esta distribución puede verse en Feller (1971), véase el Capítulo 1, que por comodidad volvemos a definir. Una distribución es estable si existe una constante c, que depende del número de variables consideradas, tal que la distribución de la suma de variables coincide con la distribución de la constante por una de ellas. La constante c adopta la forma

$$c=n^{\frac{1}{\alpha}}\ ,\quad \alpha\in(0,2] \quad .$$

Casos importantes de esta distribución se obtienen para valores particulares del parámetro.
Si $\alpha=2$ se obtiene la distribución normal.
Si $\alpha=1$ se obtiene la distribución degenerada.
Si $\alpha\in(0,1]$ se obtiene la distribución estable positiva definida mediante su transformada de Laplace

$$E[\exp(-sZ)]=\exp(-s^{\alpha}) \quad .$$

Si la distribución del frailty es estable positiva la función de supervivencia conjunta queda

$$S(t_1,t_2)=E[\exp(-Z(\Lambda_1(t_1)+\Lambda_2(t_2)))]=\exp[-(\Lambda_1(t_1)+\Lambda_2(t_2))^{\alpha}] \qquad [2.8]$$

expresión que define el modelo denominado de Hougaard.

Hemos de notar que este modelo se pudo obtener también desde el enfoque de los modelos basados en la transformación de variables aleatorias independientes utilizando distribuciones concretas de Weibull (véase [2.1]). Asimismo, para ciertas expresiones de

$$h_1(t_1/T_2>t_2) \quad y \quad h_1(t_1/t_2=t_2) \quad ,$$

es posible obtener este modelo; es decir, puede obtenerse desde el enfoque de distribuciones condicionadas.

Volvamos al caso general, un frailty cualquiera. A continuación se verá un procedimiento para la obtención de la función de supervivencia en este tipo de modelos. Se ha visto que, para este modelo, la función de supervivencia se expresa como

$$S(t_1,t_2)=\int[S_1(t_1)S_2(t_2)]^z dF(z) \quad . \qquad [2.9]$$

Sea u una variable definida de la siguiente forma

$$u=-lg[S_1(t_1)S_2(t_2)]=-lgS_1(t_1)-lgS_2(t_2)=\Lambda_1(t_1)+\Lambda_2(t_2)$$

y sea la transformada de Laplace de Z

$$\phi(x)=E[\exp(-xZ)]=\int\exp(-xZ)\,dF(z) \quad ,$$

con lo que resulta que

$$\phi(u)=\int\exp(-uZ)\,dF(z)=S(t_1,t_2) \quad . \qquad [2.10]$$

Así pues se trata de obtener la transformada de Laplace de la distribución del frailty.

Antes de abordar este problema se presenta ahora el importante problema de la *identificabilidad* del modelo, en el sentido del siguiente conocido teorema que verifican las transformadas de Laplace.

Teorema: Sea Z una variable aleatoria no negativa y sean f(z) y g(z) dos funciones continuas correspondientes a dos funciones de densidad. Si se verifica que $\phi_{f(z)}(u)=\phi_{g(z)}(u)$ entonces f(z)=g(z) para cualquier z.

Si no se conoce la distribución de la variable Z, F(z), y, por lo tanto, tampoco su transformada de Laplace, se procede como sigue. Se considera una clase más general que la clase de funciones de supervivencia generadas por frailties. Dicha clase está constituida por las distribuciones arquimedianas (véase el Capítulo 1). En esta clase la función de supervivencia verifica

$$S(t_1,t_2)=\phi[\phi^{-1}(S_1(t_1))+\phi^{-1}(S_2(t_2))] \qquad [2.11]$$

donde ϕ es una función decreciente no negativa con derivada segunda no negativa cumpliendo que

$$\phi(0)=1 \ .$$

Si $\phi(u)$ es una transformada de Laplace, la expresión [2.11] es equivalente a [2.9]. Es decir, cuando esta función sea una transformada de Laplace se tratará de un modelo frailty. Así pues, se parte de la expresión [2.11] para obtener un procedimiento que permita obtener la función de supervivencia. Para ello Oakes (1989) define la siguiente función

$$\theta^*(t_1,t_2)=\frac{[S(t_1,t_2)]\left[\dfrac{\partial^2 S(t_1,t_2)}{\partial t_1 \partial t_2}\right]}{\left[\dfrac{-\partial S(t_1,t_2)}{\partial t_1}\right]\left[\dfrac{-\partial S(t_1,t_2)}{\partial t_2}\right]}=\frac{h_1(t_1/T_2=t_2)}{h_1(t_1/T_2\geq t_2)} \qquad [2.12]$$

y demuestra que si $S(t_1,t_2)$ pertenece a la clase de distribuciones arquimedianas se verifica que la función [2.12] depende de los tiempos de supervivencia a través de la función de supervivencia, es decir

$$\theta^*(t_1,t_2)=\theta[S(t_1,t_2)]$$

y se cumple también el resultado recíproco. En otras palabras, θ^* depende de t_1 y t_2 a través de la función $S(t_1,t_2)$. Esta función

$$\theta[S(t_1,t_2)]$$

determina la función

$$\phi(u)$$

por medio de su funcion inversa ϕ^{-1} de la siguiente manera

$$\phi^{-1}(v)=\int_v^1 \exp\left[\int_x^1 \frac{\theta(y)}{y}\,dy\right]dx \quad . \qquad [2.13]$$

Una consecuencia fundamental de esto anterior es lo siguiente: dada una expresión de la función

$$\theta[S(t_1,t_2)] \quad ,$$

pueden calcularse las funciones

$$\phi^{-1} \text{ y } \phi$$

que determinan la función de supervivencia del modelo.

En resumen, existen dos caminos para encontrar la función de supervivencia del modelo:

a) Determinar la transformada de Laplace de la distribución del frailty. Luego la cuestión es determinar la distribución del frailty.

b) Partiendo de una expresión para $\theta(t_1,t_2)$, llegar a la función de supervivencia del modelo por el procedimiento que acabamos de describir.

Es necesario observar que, por la propia definición de la función θ, ésta informa acerca de la asociación entre las variables, lo cual puede orientar sobre la forma que deberá adoptar dicha función.

Los procedimientos anteriores son, en cierto modo, inversos lo cual puede servir para considerar una distribución del frailty, conocida la expresión de

$$\theta(t_1,t_2)$$

y viceversa. Este hecho va a explicar el porqué de la elección, en cierta manera arbitraria, de la distribución de Z. Éste procedimiento será el que nosotros emplearemos para proponer nuevos modelos.

2.4.2 EJEMPLOS

Exponemos seguidamente tres modelos clásicos en la literatura sobre supervivencia: el modelo frailty con distribución gamma, el modelo frailty con distribución estable positiva y el modelo frailty con distribución inversa gaussiana.

a) Frailty con distribución gamma

Si el coeficiente es constante, es decir

$$\theta(S(t_1,t_2))=c \quad (cte)$$

el resultado de

$$\phi(u) \quad \text{es} \quad \left[\frac{1}{1+u}\right]^{\frac{1}{c-1}} \quad \text{con} \quad c>1$$

lo cual coincide con la transformada de Laplace de la distribución gamma

$$f(z)=\frac{z^{\frac{1}{\alpha}-1}\exp(-z/\alpha)}{\Gamma\left(\frac{1}{\alpha}\right)\alpha^{\frac{1}{\alpha}}} \quad \text{con} \quad \alpha=c-1 \quad .$$

Es decir, con un frailty gamma resulta

$$\theta(t_1,t_2)=c$$

$$S(t_1,t_2)=\left[\frac{1}{[S_1(t_1)]^{c-1}}+\frac{1}{[S_2(t_2)]^{c-1}}-1\right]^{-\frac{1}{c-1}}$$

expresión que coincide con la del modelo de Clayton, obtenido anteriormente. Algunos casos particulares son

i) Si $c=1$ $\quad S(t_1,t_2)=S_1(t_1)S_2(t_2)$ $\quad$ independencia

ii) Si $c\to\infty$ $\quad S(t_1,t_2)\to \text{mín}[S_1(t_1),S_2(t_2)]$

iii) Si $c<1$ $\quad S(t_1,t_2)=\text{máx}[(S_1(t_1))^{1-c}+(S_2(t_2))^{1-c}-1,0]^{\frac{1}{1-c}}$

iv) Si $c\to 0$ $\quad S(t_1,t_2)\to \text{máx}[S_1(t_1)+S_2(t_2)-1,0]$

A continuación se presenta el procedimiento a la inversa. Si

$$f(z)=\frac{\theta^{\delta}z^{\delta-1}\exp(-\theta z)}{\Gamma(\delta)}$$

entonces

$$\phi(u)=\int_0^{\infty}\exp(-zu)f(z)\,dz=(1+\theta^{-1}u)^{-\delta} \quad .$$

Es importante tener en cuenta que

$$E(z)=\int_0^{\infty}zf(z)\,dz=\frac{\delta}{\theta}$$

$$Var(z)=\frac{\delta}{\theta^2}$$

y

$$\delta=\theta \Rightarrow E(z)=1 \ , \quad Var(z)=\frac{1}{\delta} \quad .$$

Hay que observar que en el modelo de Clayton se supone que

$$\theta=\delta \quad ,$$

lo cual no impide que pueda generalizarse al caso de no ser iguales.

Además, se observa que

$$(1+\theta^{-1}(\Lambda_1+\Lambda_2))^{-\delta} = \left(1+\frac{\Lambda_1}{\theta}+\frac{\Lambda_2}{\theta}\right)^{-\delta} =$$

$$= \left[\frac{1}{(S_1(t_1))^{\frac{1}{\delta}}}+\frac{1}{(S_2(t_2))^{\frac{1}{\delta}}}-1\right]^{-\delta} = [1+(S_1(t_1))^{-\frac{1}{\delta}}-1+(S_2(t_2))^{-\frac{1}{\delta}}-1]^{-\delta}$$

es decir,

$$(S_1(t_1))^{-\frac{1}{\delta}}-1=\frac{\Lambda_1}{\theta} \quad y \quad S_1(t_1)=(1+\frac{\Lambda_1}{\theta})^{-\delta} \quad .$$

Un caso particular se presenta cuando los parámetros verifican

$$\theta=\lambda$$

$$\delta=1$$

es decir, Z sigue una distribución exponencial

$$f(z)=\lambda\exp(-\lambda z) \quad .$$

En este caso

$$\phi(u)=\frac{\lambda}{u+\lambda}$$

b) Frailty con distribución estable positiva

Se obtiene el modelo de Hougaard si

$$\theta[S(t_1,t_2)]=1+\frac{1-\alpha}{-\alpha lgS(t_1,t_2)} \quad , \quad \alpha\in(0,1)$$

con lo que resulta que

$$\phi(u)=S(t_1,t_2)=\exp(-u^{\alpha}) \quad . \qquad [2.14]$$

Este caso es más intuitivo verlo en sentido inverso, es decir, partiendo de dicha transformada de Laplace. Es fácil observar que

Si $\alpha=1$, existe independencia

Si $\alpha\to 0$, entonces $S(t_1,t_2) \to \text{mín}[S_1(t_1),S_2(t_2)]$.

c) Frailty con disribución inversa gaussiana

Si Z sigue una distribución inversa gaussiana, es decir, con función de densidad

$$f(z)=(\eta\pi)^{-1/2}\exp(2/\eta)\,z^{-3/2}\exp(-\eta^{-1}z-\eta^{-1}z^{-1})$$

resulta que la función de supervivencia es

$$S(t_1,t_2)=\exp\left[\frac{2}{\eta}-2\left(\frac{1}{\eta^2}+\frac{1}{\eta}(\Lambda_1(t_1)+\Lambda_2(t_2))\right)^{1/2}\right] \qquad [2.15]$$

correspondiéndole una función

$$\theta[S(t_1,t_2)]=1+\frac{\eta}{2-\eta\, lgS(t_1,t_2)} \quad .$$

Estos ejemplos de modelos frailty se pueden englobar en una familia de distribuciones triparamétrica notada por

$$P(\gamma,\sigma,\rho)$$

para la cual se conoce la expresión de la función de densidad general y su transformada de Laplace. Para esta familia se cumple que

- Si $\gamma=0$ resulta el frailty gamma.
- Si $\gamma\in(0,1)$ tenemos el modelo de Hougaard.

- Si $\gamma=1/2$ resulta un frailty con distribución inversa gaussiana.

(Véase Hougaard (1986) y Costigan & Klein (1993)).

Aalen (1988) generalizó aún más esta fórmulación incluyendo valores de γ mayores que la unidad.

Otros casos aún poco estudiados de modelos frailty son aquellos que suponen para Z una variable gamma con media mayor que uno, una distribución beta y otras distribuciones.

2.5 MODELOS GENERADOS VÍA DISTRIBUCIONES MARGINALES

2.5.1 PLANTEAMIENTO GENERAL

En numerosos modelos de supervivencia, la formulación del modelo, bien a través de la función de supervivencia o a través de la función de distribución, viene explícitamente determinada por las distribuciones marginales de los tiempos de fallo. Los ejemplos de modelos frailty vistos anteriormente, e incluso algunos ejemplos vistos en los apartados correspondientes a los métodos basados en la transformación de variables aleatorias independientes y en hipótesis de tipo condicional, cumplen esta propiedad. Esto ha hecho que se investiguen métodos generales que cumplan tal propiedad. Así, en Genest & Mackay (1986) se enfoca el problema utilizando tanto las distribuciones arquimedianas como las distribuciones cuyas marginales son uniformes. En Marshall & Olkin (1988), en cambio, se hace un enfoque basado en la mixtura de distribuciones.

En cierta forma estos métodos extienden o generalizan los modelos frailty en el sentido de que éstos están incluidos en esta nueva clase de modelos. La generalización se basa en considerar frailties para cada una de las componentes del modelo multivariante, perdiéndose en parte la idea intuitiva de la noción de frailty. Esta nueva idea influirá bastante en conceptos que se desarrollarán más adelante como son la dependencia o asociación e incluso métodos de simulación.

A continuación se expone el método propuesto por Marshall & Olkin (1988). Sea F(x) una función de distribución univariante cuya función de supervivencia S(x) la notaremos también por $\overline{F}(x)$. Es sencillo comprobar que la potencia

$$[\overline{F}(x)]^{\theta} \quad , \quad \theta>0$$

es también una función de supervivencia. Si se considera que $\theta>0$ es una variable aleatoria con

$$G(\theta)$$

su función de distribución, entonces

$$\overline{H}(x)=\int \overline{F}^{\theta}(x)\,dG(\theta) \qquad [2.16]$$

es una mixtura de distribuciones, siendo

$$\overline{H}(x)$$

una función de supervivencia y H(x) la correspondiente función de distribución. Y recíprocamente, dada una función de distribución G tal que

$$\overline{G}(0)=1$$

y dada una función de distribución H, existe una función de distribución F tal que se satisface [2.16]. Además resulta que

$$\overline{F}(x)=\exp[-\phi^{-1}\overline{H}(x)]$$

donde

$$\phi$$

es la transformada de Laplace de G (similar a [2.9] y [2.10]). Esta propiedad del caso univariante se puede generalizar de la siguiente forma. Sea G una función de distribución bivariante y sea

$$\overline{G}$$

la correspondiente función de supervivencia. Por consiguiente se cumple que

$$\overline{G}(0,0)=1 \quad .$$

Sean G_i, i:1,2 , las distribuciones marginales de G. Sea F una función de distribución bivariante con funciones de distribución marginales F_i , i:1,2. Se define la mixtura de distribuciones como

$$\overline{H}(t_1,t_2)=\int\int\overline{F}_1^{\theta_1}(t_1)\overline{F}_2^{\theta_2}(t_2)\,dG(\theta_1,\theta_2) \qquad [2.17]$$

De esta manera se genera una nueva función de supervivencia bivariante con funciones de supervivencia marginales

$$\overline{H}(t_i)=\int\overline{F}_i^{\theta_i}(t_i)\,dG_i(\theta_i) \qquad i:1,2$$

cumpliéndose que

$$\overline{F}_i(t_i)=\exp[-\phi_i^{-1}\overline{H}_i(t_i)]$$

donde ϕ_i es la transformada de Laplace de G_i , i:1,2 .

La expresión [2.17] puede considerarse la función de supervivencia de un modelo de azar proporcional bivariante puesto que

$$h_i(t_i/\theta_i)=\theta_i h_i(t_i) \;\Rightarrow\; \overline{F}_i(t_i/\theta_i)=\overline{F}_i^{\theta_i}(t_i) \quad .$$

Si en la expresión [2.17] se considera

$$\theta_1=\theta_2 \quad ,$$

tal expresión coincide con [2.9] a partir de la cual se generan los modelos frailty.

Veamos a continuación otra propiedad importante que verifican las funciones de supervivencia bivariantes. Sean H_1 y H_2 dos funciones de distribución univariantes, G una función de distribución bivariante tal que

$$\overline{G}(0,0)=1$$

con marginales G_i , ϕ la transformada de Laplace de G , ϕ_i la transformada de Laplace de G_i y K una función de distribución bivariante con marginales uniformes en [0,1].

Entonces si

$$F_i(t_i)=\exp[-\phi_i^{-1}H_i(t_i)] \quad .$$

se verifica

$$\overline{H}(t_1,t_2)=\iint K(\overline{F}_1^{\theta_1}(t_1),\overline{F}_2^{\theta_2}(t_2))\,dG(\theta_1,\theta_2) \quad . \qquad [2.18]$$

Esta expresión generaliza [2.17] al introducir el factor K. La función

$$\overline{H}(t_1,t_2)$$

es una función de supervivencia, con marginales

$$\overline{H}_1 \text{ y } \overline{H}_2 \quad .$$

Bastará seleccionar G y K para generar funciones de supervivencia bivariantes con marginales prefijadas. Además, se verifica que

$$\overline{H}(t_1,t_2)=\phi(\phi_1^{-1}\overline{H}_1(t_1),\phi_2^{-1}\overline{H}_2(t_2)) \qquad [2.19]$$

sin más que tomar K(a.b) = a.b .

En caso de que

$$G_1=G_2 \quad y \quad G(\theta_1,\theta_2)=\min(G_1(\theta_1),G_2(\theta_2))$$

se verifica

$$\overline{H}(t_1,t_2)=\iint K(\overline{F}_1^{\theta}(t_1),\overline{F}_2^{\theta}(t_2))\,dG_1(\theta)$$

y si además K(a,b)=a.b , entonces

$$\overline{H}(t_1,t_2)=\phi(\phi^{-1}\overline{H}_1(t_1)+\phi^{-1}\overline{H}_2(t_2)) \quad , \qquad [2.20]$$

donde ϕ es la transformada de Laplace de G_1. Esta expresión, obtenida con las dos restricciones anteriores, coincide con la expresión de los modelos frailty [2.11]. En este caso, bastará conocer ϕ (la transformada de Laplace de G_1) para generar la función de supervivencia del modelo. Sin embargo, en el caso general, se deberán conocer

$$\phi \; , \; \phi_i \; \text{y} \; K \quad .$$

2.5.2 EJEMPLOS

A continuación se verán algunos ejemplos según la forma que adopten estas funciones anteriores.

a) Se considera el caso más usual para la función K, un solo frailty y una función exponencial para la transformada de Laplace. Es decir,

$$K(a,b)=a.b$$
$$\phi(s)=\exp(-s^{\alpha}) \quad , \quad \alpha\in(0,1]$$
$$G_1=G_2 \qquad G=\min(G_1(t_1),G_2(t_2)) \quad .$$

Entonces, la función de supervivencia adopta la expresión

$$S(t_1,t_2)=\exp[-(\Lambda_1(t_1)+\Lambda_2(t_2))^{\alpha}]$$

que coincide con la del modelo propuesto por Hougaard.

b) Se considera el mismo caso anterior pero considerando otra transformada de Laplace. Es decir,

$$K(a,b)=a.b$$
$$G_1=G_2 \qquad G=\min(G_1(t_1),G_2(t_2))$$
$$\phi(s)=(1+s)^{-\alpha} \quad , \quad \alpha>0$$

es decir, G sigue una distribución gamma. Entonces

$$S(t_1,t_2)=\left[\frac{1}{(S_1(t_1))^{1/\alpha}}+\frac{1}{(S_2(t_2))^{1/\alpha}}-1\right]^{-\alpha} \quad ,$$

modelo que coincide con el propuesto por Clayton.

c) En este caso, se mantiene la forma de la función K pero se consideran dos frailties". Es decir,

$$K(a,b)=a.b$$
$$\psi_i(s)=\exp(-\alpha_i s^{1/m}) \qquad \alpha_i\geq 0 \quad i:0,1,2$$
$$\alpha_{10}=\alpha_0+\alpha_1>0$$
$$\alpha_{20}=\alpha_0+\alpha_2>0$$
$$\psi_i(s)=\psi_i(s)\psi_0(s) \quad ,$$

y, entonces, resulta que

$$-\lg H(t_1,t_2)=-\frac{\alpha_1}{\alpha_{10}}\lg H_1(t_1)-\frac{\alpha_2}{\alpha_{20}}\lg H_2(t_2)+$$
$$+\alpha_0\left[\left(\frac{-\lg H_1(t_1)}{\alpha_{10}}\right)^m+\left(\frac{-\lg H_2(t_2)}{\alpha_{20}}\right)^m\right]^{1/m} \quad .$$

d) Se consideran otros "frailties". Si se verifica

$$K(a,b)=a.b$$
$$\psi(s_1,s_2)=\frac{(\lambda+s_1+s_2)(\lambda_1+\lambda_{12})(\lambda_2+\lambda_{12})+(\lambda_{12}s_1s_2)}{(\lambda+s_1+s_2)(\lambda_1+\lambda_{12}+s_1)(\lambda_2+\lambda_{12}+s_2)}$$
$$\lambda_1,\lambda_2,\lambda_{12}\geq 0$$
$$\lambda_1+\lambda_{12}>0$$
$$\lambda_2+\lambda_{12}>0$$
$$\lambda=\lambda_1+\lambda_2+\lambda_{12} \quad ,$$

entonces resulta

$$\phi(s_1,s_2)=[\psi(s_1,s_2)]^m \quad con \quad m\in\mathbb{N}^*$$

entonces la función de distribución del modelo es

$$H(t_1, t_2) = H_1(t_1) + H_2(t_2) - 1 + \beta\left(\frac{1}{\overline{H}_1(t_1)}, \frac{1}{\overline{H}_2(t_2)}\right)$$

siendo β la función beta.

Como se observa, con sólo variar las funciones K ó G se generan distintos modelos de supervivencia multivariante donde las marginales están prefijadas. Algunos de estos modelos ya fueron generados por otros métodos. En Lindeboom y Van den Gerg (1994) se expone un planteamiento similar en cuanto a la generalización del frailty .

que coincide con la distribución exponencial bivariante de Marshall & Olkin (1967).

e) En este caso se considera otra función para K. Si

$$K(a,b)=ab[1+\beta(1-a)(1-b)]$$
$$0\leq a,b\leq 1 \qquad -1\leq\beta\leq 1$$
$$\phi_i(s)=\exp(-s^{1/m}) \qquad m\geq 1 \quad ,$$

entonces resulta

$$H(t_1,t_2)=(1+\beta)H_1(t_1)H_2(t_2)-\beta H_1(t_1)H_2^{\gamma}(t_2)-\beta H_1^{\gamma}(t_1)H_2(t_2)+$$
$$+\beta H_1^{\gamma}(t_1)H_2^{\gamma}(t_2) \qquad \text{con} \qquad \gamma=2^{1/m} \qquad 0\leq\gamma\leq 2 \quad .$$

f) Si se consideran las siguientes expresiones

$$K(a,b)=\text{mín}(a,b) \qquad 0\leq a,b\leq 1$$
$$G(\theta_1,\theta_2)=G_1(\theta_1)G_2(\theta_2)$$
$$G_i(\theta_i) \quad \text{exponencial} \quad i:1,2 \quad ,$$

entonces resulta

$$H(t_1,t_2)=H_1(t_1)H_2(t_2)\left[1+\frac{\overline{H}_1(t_1)\overline{H}_2(t_2)}{1-\overline{H}_1(t_1)\overline{H}_2(t_2)}\right] \quad .$$

g) Por último, se considera otra función K y otra distribución bivariante para los "frailties". Sea

$$K(a,b)=\text{máx}(a+b-1,0) \qquad 0\leq a,b\leq 1$$
$$G(\theta_1,\theta_2)=G_1(\theta_1)G_2(\theta_2)$$
$$G_i(\theta_i) \quad \text{exponenciales} \quad ,$$

2.6 INTRODUCCIÓN DE COVARIABLES

Hasta ahora se han visto algunos procedimientos para la construcción de modelos en los cuales la función de supervivencia bivariante depende de un parámetro que describe la asociación entre las variables. En la práctica existen ciertas variables explicativas o covariables que caracterizan a los distintos individuos, es decir, son características propias del individuo, factores exógenos, tratamientos experimentales aplicados, etc. Ejemplos de estas características son la edad, el sexo, clase social, etc. Si se tienen dos tiempos de supervivencia que miden, por ejemplo, la aparición de una enfermedad cardiovascular y la aparición de hipertensión, respectivamente, características asociadas a estos individuos que pueden afectar a los tiempos de fallo podrían ser la edad, hábitos de fumar, sexo, una medicación específica, índice de colesterol, etc. Un ejemplo ciertamente interesante puede verse en Wassell (1993)). Una clasificación detallada de los distintos tipos de covariables puede verse en Lara (1995).

Se denota por

$$x_1=(x_{11},\ldots,x_{1p})^T$$
$$x_2=(x_{21},\ldots,x_{2q})^T$$

a dos vectores de covariables correspondientes a dos individuos, por ejemplo, padre-hijo (o a un solo individuo al que se le miden dos tiempos de fallo de distinta índole). El efecto de las covariables suele introducirse en el modelo mediante una función que usualmente es

$$\psi_i(x_i)=\exp(\beta_i x_i) \quad ,i:1,2$$

donde $\beta_1=(\beta_{11},\ldots,\beta_{1p})$ y $\beta_2=(\beta_{21},\ldots,\beta_{2q})$ son los parámetros asociados a las covariables, también llamados coeficientes de regresión.

Se considera un modelo de azar proporcional (modelo de regresión de Cox (1972)), es decir, las covariables y el frailty actúan multiplicativamente en la función de azar marginal, es decir

$$h_i(t_i, \psi_i(x_i), z) = z.\psi_i(x_i).h_i(t_i) \qquad i:1,2 \quad .$$

En este caso, las funciones de supervivencia marginales condicionadas a un valor fijo de Z son

$$S_i(t_i, \psi_i(x_i)/z) = \exp\left[-z\int_0^{t_i}\psi_i(x_i)h_i(t_i)dt_i\right] =$$
$$= \exp[-z\psi_i(x_i).\Lambda_i(t_i)] \quad .$$

La función de supervivencia conjunta es

$$S(t_1, t_2) = \int \exp[-z(\exp(\beta_1 x_1)\Lambda_1(t_1) + \exp(\beta_2 x_2)\Lambda_2(t_2))]dF(z) \qquad [2.21]$$

donde $\Lambda_i(t_i)$ corresponde a una función de azar base acumulativa.

A continuación se presentan dos ejemplos muy importantes reflejados en la literatura: un modelo con distribución del frailty estable positiva y otro con distribución gamma. Ambos se consideran en el capítulo sobre Estimación, especialmente el segungo caso, donde, además, consideraremos covariables que se introducen en el parámetro de asociación.

a) Frailty con distribución estable positiva

Supongamos que el frailty Z sigue una distribución estable positiva. En este caso, la función de supervivencia conjunta es

$$S(t_1, t_2) = \exp[-(\exp(\beta_1 x_1)\Lambda_1(t_1) + \exp(\beta_2 x_2)\Lambda_2(t_2))^\alpha] \ , \quad \alpha\in(0,1] \ [2.22]$$

Es importante observar que, si $t_i=0$, entonces

$$S_i(t_i) = \exp[-\exp(\alpha\beta_i x_i).(\Lambda_i(t_i))^\alpha] \quad , \qquad [2.23]$$

donde en esta función de supervivencia marginal está presente el parámetro de asociación. Por lo tanto, la función de azar marginal acumulativa es

$$\Lambda_i(t_i)=\exp(\alpha\beta_i x_i)(\Lambda_i(t_i))^{\alpha} \quad ,$$

entendiendo que la nueva función de azar marginal acumulativa depende, entre otros, de la función de azar base marginal acumulativa. Por lo tanto, derivando, la función de azar marginal queda

$$h_i(t_i)=\alpha\exp(\alpha\beta_i x_i)(\Lambda_i(t_i))^{\alpha-1}h_i(t_i) \quad .$$

Si se particulariza al caso donde los tiempos de fallo puedan modelizarse con distribuciones Weibull, las funciones de azar marginales y de azar acumulativo marginales serán respectivamente

$$h_i(t_i)=\epsilon\gamma t_i^{\gamma-1}$$

$$\Lambda_i(t_i)=\epsilon t_i^{\gamma} \quad .$$

Considerando el efecto del frailty Z con distribución estable positiva quedaría

$$h_i(t_i)=\epsilon^{\alpha}\gamma\alpha t_i^{\gamma\alpha-1}$$
$$\Lambda_i(t_i)=(\epsilon t_i^{\gamma})^{\alpha}$$

y la función de supervivencia conjunta sería

$$S(t_1,t_2)=\exp[-(\exp(\beta_1 x_1)t_1^{\gamma}+\exp(\beta_2 x_2)t_2^{\gamma})^{\alpha}] \qquad [2.24]$$

quedando la función de supervivencia marginal como

$$S_i(t_i)=\exp[-\exp(\alpha\beta_i x_i)t_i^{\gamma\alpha}] \qquad [2.25]$$

donde en el factor $\exp(\beta_i x_i)$ se integra al parámetro de escala de la distribución Weibull. Es decir, en vez de

$$(\exp(\beta_i x_i))^{\alpha}\epsilon^{\alpha}=(\epsilon\exp(\beta_i x_i))^{\alpha}$$

se considera

$$\epsilon \exp(\beta_i x_i) \equiv \exp(\beta_i x_i) \quad .$$

b) Frailty con distribución gamma

Si el frailty Z puede modelizarse mediante una distribución gamma, la función de supervivencia conjunta condicionada a un valor fijo de Z es

$$S(t_1, t_2/z) = \exp[-z(\exp(\beta_1 x_1)\Lambda_1(t_1) + \exp(\beta_2 x_2)\Lambda_2(t_2))]$$

y la función de supervivencia conjunta queda así

$$S(t_1, t_2) = \left[1 + \frac{\exp(\beta_1 x_1)\Lambda_1(t_1) + \exp(\beta_2 x_2)\Lambda_2(t_2)}{\theta}\right]^{-\theta} \qquad [2.26]$$

siendo la función de supervivencia marginal

$$S_i(t_i) = \left[1 + \frac{\exp(\beta_i x_i)\Lambda_i(t_i)}{\theta}\right]^{-\theta} \quad . \qquad [2.27]$$

En el caso particular en el que se consideran distribuciones Weibull para las marginales se procede igual que en el apartado a).

En este modelo, que coincide con el de Clayton, la función de supervivencia puede expresarse así

$$S(t_1, t_2) = \left[\frac{1}{(S_1(t_1))^{\frac{1}{\theta}}} + \frac{1}{(S_2(t_2))^{\frac{1}{\theta}}} - 1\right]^{-\theta}$$

y si se consideran distribuciones weibull para las marginales, es decir

$$S_i(t_i)=\exp(-\epsilon_i t_i^{\gamma_i}) \quad ,$$

las funciones de supervivencia marginales quedan

$$S_i(t_i)=\exp[-\exp(\beta_i x_i)\, t_i^{\gamma_i}]$$

donde el parámetro de escala queda integrado en las covariables.

Wassell (1993) propone la introducción de covariables también en el parámetro de asociación. Nosotros trataremos este caso en el capítulo sobre Estimación.

2.7 APORTACIONES EN CONSTRUCCIÓN DE MODELOS

A continuación nosotros tratamos algunos modelos que completaremos en el capítulo sobre Dependencia y sobre Estimación. Son los siguientes:

a) Consideramos, en primer lugar, un modelo con un frailty con distribución uniforme. En este modelo consideramos covariables bajo hipótesis de azar proporcional y marginales Weibull y estudiamos problemas de identificabilidad.

b) Presentamos un modelo con un frailty con distribución estable positiva en el caso concreto en el que las marginales se modelizan con distribuciones de Pareto.

c) Tratamos el problema de los modelos bajo hipótesis de vida acelerada haciendo hincapié en la dificultad de su tratamiento.

d) Exponemos algunas formas alternativas en la introducción de las covariables explicativas.

e) Especial tratamiento tendrán los modelos denominados *mixtos* por nosotros. Entre ellos, consideramos un modelo frailty con distribución estable positiva y marginales Weibull.

f) Presentamos la posibilidad de introducir las covariables también en el parámetro de asociación estudiando sobre todo un modelo mixto con frailty gamma.

g) Por último, consideramos el problema donde existe una función del frailty que actúa multiplicativamente en la función de azar marginal.

2.7.1 MODELO FRAILTY UNIFORME

Vamos a estudiar el siguiente modelo sugerido por Lee & Klein (1988) y aún no tratado en profundidad. Vamos a plantear un modelo frailty con distribución uniforme. Consideramos que el factor no observado que crea la dependencia, el frailty, en algunas situaciones o tipo de datos se puede modelizar con esta distribución; por ejemplo, la infuencia genética entre individuos puede considerarse uniforme en un intervalo de valores [a,b]. Estos valores a y b han de ser positivos. Su función de densidad es

$$f(z)=\begin{cases}\dfrac{1}{b-a} & \text{si} \quad z\in[a,b]\\ 0 & \text{si} \quad z\notin[a,b]\end{cases}$$

con $a\geq 0$. En estas condiciones, la transformada de Laplace de Z es

$$\phi(u)=\int_a^b \exp(-zu)\,\frac{1}{b-a}\,dz=\frac{\exp(-bu)-\exp(-au)}{(a-b)\,u} \qquad [2.28]$$

con $u=\Lambda_1(t_1)+\Lambda_2(t_2)$. Según esto, la función de supervivencia conjunta quedará así

$$S(t_1,t_2)=\frac{\exp[-b(\Lambda_1(t_1)+\Lambda_2(t_2))]-\exp[-a(\Lambda_1(t_1)+\Lambda_2(t_2))]}{(a-b)(\Lambda_1(t_1)+\Lambda_2(t_2))}. \qquad [2.29]$$

Nosotros hemos considerado de especial interés el caso de un modelo frailty uniforme, marginales con distribución de Weibull y en presencia de covariables explicativas introducidas vía azar proporcional. Este modelo volverá a ser tratado en los capítulos 3 (Dependencia) y 4 (Estimación). La expresión de la función de supervivencia es la siguiente:

$$S(t_1,t_2)=\frac{e^{-b(t_1^{\gamma}e^{\beta_1 x_1}+t_2^{\gamma}e^{\beta_2 x_2})}-e^{-a(t_1^{\gamma}e^{\beta_1 x_1}+t_2^{\gamma}e^{\beta_2 x_2})}}{(a-b)(t_1^{\gamma}e^{\beta_1 x_1}+t_2^{\gamma}e^{\beta_2 x_2})}.$$

Vamos a tratar el problema de la identificabilidad del modelo. La identificabilidad ha sido tratada, en general en los modelos frailty, entre otros, por Marshall & Olkin (1988) y por Oakes (1989). El modelo frailty de azar proporcional es identificable por el siguiente

Teorema: Si f(z) y g(z) son dos funciones continuas en el intervalo $(0,\infty)$ entonces

$$\phi_{f(z)}(u)=\phi_{g(z)}(u) \Rightarrow f(z)=g(z) \quad .$$

Este resultado de las transformadas de Laplace asegura que dos modelos frailty de azar proporcional "idénticos" sólo pueden proceder del mismo frailty. En otras palabras, el modelo es identificable en cuanto al frailty. Todo modelo de la forma [2.28] proviene de un frailty con distribución uniforme en [a,b].

Veamos si en cuanto a las distribuciones marginales también lo es. Sabemos que éstas se introducen a través del argumento u. Debemos asegurarnos de que

$$\phi_{f(z)}(u)=\phi_{f(z)}(u') \Rightarrow u=u' \quad .$$

En el caso uniforme no se puede asegurar en principio. Pero si las marginales son Weibull, Heckman & Singer (1982) establecen que los modelos frailty de azar proporcional son identificables. Es decir si ϵ y γ son los parámetros de la distribución Weibull se verifica que

$$S(t_1,t_2,\epsilon,\gamma)=S(t_1,t_2,\epsilon',\gamma') \Rightarrow \epsilon=\epsilon' \ , \ \gamma=\gamma' \quad \forall t_1,t_2.$$

Ahora bién, en el modelo con marginales Weibull y covariables se verifica que

$$u=\exp(\beta_1 x_1)\,t_1^{\gamma}+\exp(\beta_2 x_2)\,t_2^{\gamma}.$$

Debemos asegurarnos que $\exp(\beta_i x_i)=\exp(\beta'_i x_i) \implies \beta_i=\beta'_i,\ \forall x_i$. Ello es evidente sin más que darle distintos valores a x_i .

En el apartado sobre modelos mixtos consideraremos este modelo con las covariables no en forma multiplicativa, azar proporcional, sino según hipótesis de vida acelerada.

Por sugerencia de D.R. Cox la distribución uniforme podemos normalizarla sin más que tomar la transformación monótona creciente

$$\lambda : [a,b] \to [0,b-a]$$

$$\lambda(a)=0$$
$$\lambda(b)=b-a$$
$$\lambda(x)=x-a \quad .$$

Así, resulta una distribución uniforme en el intevalo [0,b-a]. El modelo, en este caso, quedaría con un sólo parámetro c=b-a, notablemente más simple, que nos facilita el trabajo, aunque al final debamos deshacer dicha transformación.

A continuación, mostramos las gráficas de algunas funciones de supervivencia conjunta y de una función de supervivencia marginal. En primer lugar, veamos la gráfica de la función de supervivencia conjunta en el caso de marginales Weibull, con parámetros $\epsilon=0.5$, $\gamma=2$, con a=0.5 y b=1.5

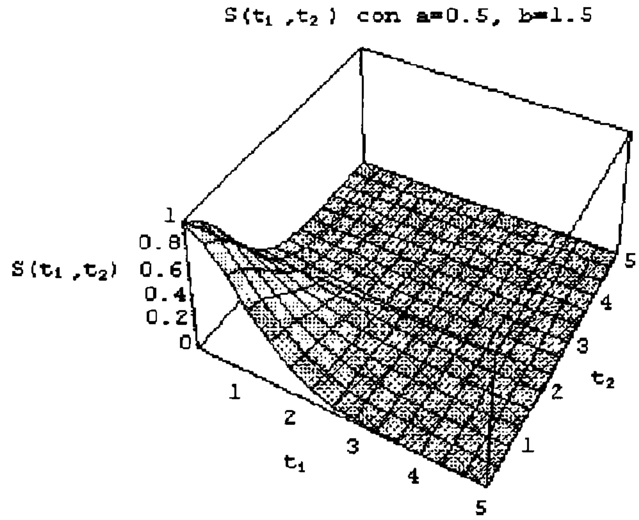

Fig. 1

La función de supervivencia conjunta, en el mismo caso, pero con a=1.25 y b=1.75 es

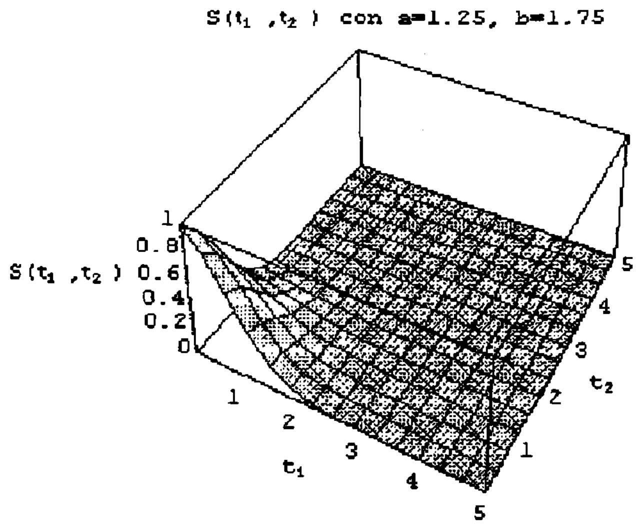

Fig. 2

La función de supervivencia conjunta, en el mismo caso, con a=0.25 y b=0.75 es

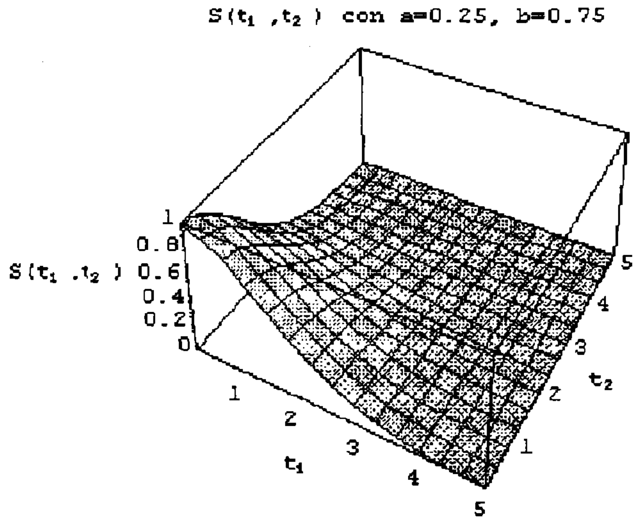

Fig. 3

Veamos ahora la función de supervivencia conjunta con a=0.5, b=1.5 pero con $\epsilon=0.5$ *y* $\gamma=0.5$

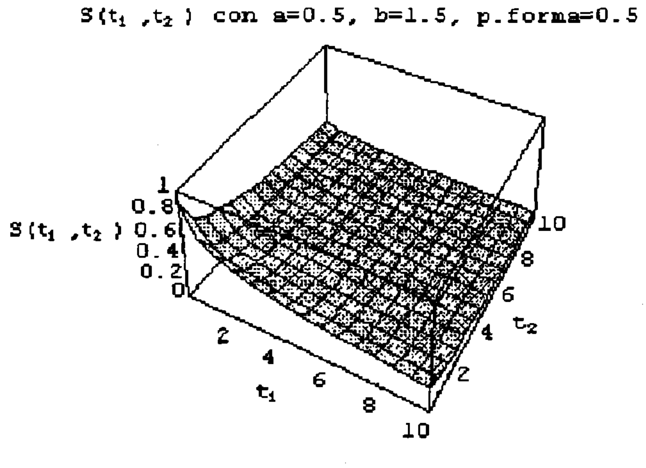

Fig. 4

Por último veamos la gráfica de la función de supervivencia marginal correspondiente a los parámetros a=0.5, b=1.5, $\epsilon=0.5$ *y* $\gamma=2$

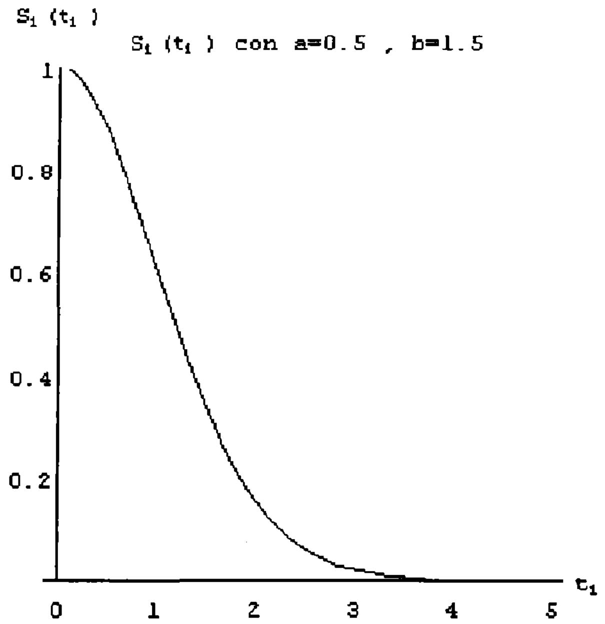

Fig. 5

2.7.2 MODELO FRAILTY ESTABLE CON MARGINALES PARETO

Proponemos el siguiente modelo: un frailty con distribución estable positiva, covariables según un modelo de azar proporcional y marginales con distribuciones de Pareto. En el capítulo 1, puede verse una justificación de esta distribución. La función de supervivencia conjunta es

$$S(t_1,t_2)=\exp\left[-(\exp(\beta_1 x_1).lg(\frac{t_1}{t_{10}})^{\delta}+\exp(\beta_2 x_2).lg(\frac{t_2}{t_{20}})^{\delta})^{\alpha}\right] \qquad [2.30]$$

quedando como supervivencia marginal

$$S_i(t_i)=\exp\left[-(\exp(\beta_i x_i).lg(\frac{t_i}{t_{io}})^{\delta})^{\alpha}\right] \qquad [2.31]$$

$$(t_j=t_{jo}\ ,\ i,j:1,2\ ,\ i\neq j)\ .$$

2.7.3 MODELOS DE VIDA ACELERADA

En casi todos los métodos de construcción de modelos vistos hasta ahora la hipótesis de partida se consideraba de azar proporcional, es decir, eran modelos de azar proporcional en cuanto se verificaba que

$$h_i(t_i,z)=zh_i(t_i) \quad i:1,2 \qquad [2.32]$$

o bien

$$h_i(t_i,z_i)=z_i h_i(t_i) \quad i:1,2$$

donde z es el frailty y z_i es el frailty en el caso más general en el que se considera un frailty para cada tiempo de fallo (véase el método de los modelos generados vía distribuciones marginales). Así ocurría en los modelos frailties tratados y en los modelos definidos a partir de las marginales. Incluso algunos modelos construidos vía variables aleatorias independientes o vía variables aleatorias condicionadas pueden considerarse de azar proporcional.

Nosotros ahora proponemos hipótesis de vida acelerada, muy tratada en supervivencia univariante pero no en supervivencia multivariante. Según esto, la función de azar marginal cumple

$$h_i(t_i,z)=zh_i(zt_i) \quad i:1,2 \qquad [2.33]$$

o bien, la función de supervivencia marginal cumple

$$S_i(t_i,z)=[S_i(zt_i)] \quad i:1,2$$

o bien, la función de azar marginal acumulativa cumple

$$\Lambda_i(t_i,z)=\Lambda_i(zt_i) \qquad i:1,2 \quad .$$

La función de supervivencia conjunta condicionada a un valor fijo de Z es

$$S(t_1,t_2/z)=S_1(t_1/z)S_2(t_2/z)=$$

$$=\exp[-z\int_0^{t_1}h_1(zt_1)dt_1].\exp[-z\int_0^{t}h_2(zt_2)dt_2]=$$

$$=\exp[-z(\int_0^{t_1}h_1(zt_1)dt_1+\int_0^{t_2}h_2(zt_2)dt_2)]=\exp[-(\Lambda_1(zt_1)+\Lambda_2(zt_2))] \quad .$$

La función de supervivencia conjunta será

$$S(t_1,t_2)=\int\exp[-z(\int_0^{t_1}h_1(zt_1)dt_1+\int_0^{t_2}h_2(zt_2)dt_2)]dF(z)=$$

$$=\int\exp[-(\Lambda_1(zt_1)+\Lambda_2(zt_2))]dF(z)=$$

$$=\int S_1(zt_1).S_2(zt_2)dF(z)=\int[S_1'(zt_1)S_2'(zt_2)]^z dF(z) \qquad [2.34]$$

donde hemos notado

$$S_i'(zt_i)=\exp[-\int_0^{t_i}h_i(zt_i)dt_i] \quad .$$

En cualquiera de estas expresiones anteriores se observa que no se puede aplicar el razonamiento de los de frailties puesto que no son transformadas de Laplace de z al no "separarse" z de una expresión que no dependa de z, como ocurría en el caso de azar proporcional. En todo caso, estas expresiones anteriores ofrecen un planteamiento teórico para obtener un modelo de vida acelerada, el problema está en la resolución de la integral. Nosotros estamos investigando un procedimiento para no tener que llegar a esta integral. El procedimiento está basado en utilizar el método Blup (Best linear umbiased prediction) para estimar los parámetros asociados al modelo.

Hougaard (1986,1989) presenta su modelo general

$$S(t_1,t_2)=\exp[-(\Lambda_1+\Lambda_2)^{\alpha}]$$

como un modelo de vida acelerada considerando distribuciones Weibull para las marginales. En general se verifica que un modelo frailty con marginales Weibull puede considerarse de azar proporcional o de vida acelerada.

2.7.4 ALTERNATIVAS EN LA INTRODUCCIÓN DE COVARIABLES

En los modelos frailty se suelen considerar hipótesis de azar proporcional, esto es, dicho frailty actuando multiplicativamente en la función de azar marginal. Las covariables admiten diversas posibilidades en cuanto a la forma de su introducción en el modelo. A continuación consideramos distintos planteamientos para introducir dichas covariables. Nosotros mantenemos en lo que sigue a continuación la hipótesis de azar proporcional, es decir, el frailty y las covariables explicativas actuarán multiplicativamente, de alguna forma, en la función de azar marginal. En el capítulo 3 (Estudio de la dependencia) y sobre todo en el capítulo 4 (Estimación) consideramos covariables que se introducen también en el parámetro de asociación. En el capítulo 5 (Temas abiertos) consideramos además diversas posibilidades en cuanto a azares no proporcionales.

a) Consideremos en primer lugar una generalización del caso mas usual. Hemos visto que frailty y covariables actúan multiplicativamente entre ellos y con la función de azar marginal. Es decir

$$h_i(t_i, \psi_i(x_i), z) = z\exp(\beta_i x_i) h_i(t_i) .$$

Si consideramos el frailty en forma exponencial podemos integrarlo con las covariables explicativas, es decir

$$y = lg(z) \Rightarrow z = \exp(y) \Rightarrow$$

$$h_i(t_i, \psi_i(x_i) z) = \exp(y)\exp(\beta_i x_i) h_i(t_i) = \exp(\beta_i x_i + y) h_i(t_i)$$

y así englobamos las covariables y el frailty en el modelo de azar proporcional. Notemos por

$$\eta = \beta x + y .$$

Es decir, estamos considerando un factor multiplicativo

$$g(\eta) = \exp(\eta) .$$

Así pues, el modelo se generaliza sin más que considerar funciones no necesariamente exponenciales para $g(\eta)$ en las que se cumpla que sean no negativas y tomen el valor uno en las condiciones estandar, esto es, para X=0 y Z=1.

b) Consideremos el siguiente caso de azar proporcional: el frailty Z y las covariables actuando de forma aditiva entre sí. De esta manera los factores exógenos X y el efecto aleatorio no observable Z "se acumulan" antes de actuar sobre una función de azar base marginal. Este caso admite diversas posibilidades, por ejemplo, que el factor multiplicativo sea

$$z + \beta_i x_i .$$

En este caso la función de azar base marginal se alcanzará para $x_i=0$ y Z=1. Además admite la posibilidad de un frailty negativo. Nosotros en cambio consideraremos el caso siguiente:

$$h_i(t_i, z, \psi_i(x_i)) = (z + \exp(\beta_i x_i)) h_{i0}(t_i) .$$

La función de azar base marginal se alcanzará para valores $x_i=0$ y $Z=0$. En este caso la función de supervivencia condicionada marginal es

$$S_i(t_i, z) = [S_{i0}(t_i)]^{z+\exp(\beta_i x_i)}$$

con valores en el intervalo [0,1]. La función de supervivencia conjunta condicionada es

$$S(t_1, t_2/z, \psi_1(x_1), \psi_2(x_2)) =$$

$$= \exp[-(z+\exp(\beta_1 x_1))\Lambda_{10}(t_1) - (z+\exp(\beta_2 x_2))\Lambda_{20}(t_2)] =$$

$$= \exp[-\exp(\beta_1 x_1)\Lambda_{10}(t_1) - \exp(\beta_2 x_2)\Lambda_{20}(t_2) - z(\Lambda_{10}(t_1) + \Lambda_{20}(t_2))] \quad .$$

La función de supervivencia conjunta será

$$S(t_1, t_2) = \int \exp[-\exp(\beta_1 x_1)\Lambda_{10}(t_1) - \exp(\beta_2 x_2)\Lambda_{20}(t_2)] .$$

$$. \exp(-z(\Lambda_{10}(t_1) + \Lambda_{20}(t_2))) dF(z) =$$

$$= \exp[-\exp(\beta_1 x_1)\Lambda_{10}(t_1) - \exp(\beta_2 x_2)\Lambda_{20}(t_2)] .$$

$$. \int \exp(-z(\Lambda_{10}(t_1) + \Lambda_{20}(t_2))) dF(z) \qquad [2.35]$$

siendo esta integral una transformada de Laplace de z. Si z sigue una distribución estable positiva entonces

$$S(t_1, t_2) = \exp[-\exp(\beta_1 x_1)\Lambda_{10}(t_1) - \exp(\beta_2 x_2)\Lambda_{20}(t_2)] .$$

$$. \exp[-(\Lambda_{10}(t_1) + \Lambda_{20}(t_2))^{\alpha}]$$

modelo que es significativamente distinto al obtenido en el caso multiplicativo. Las funciones de azar acumulativo marginal pueden modelizarse mediante distribuciones Weibull.

2.7.5 MODELOS MIXTOS

Nosotros vamos a construir un modelo, al que denominaremos modelo *mixto*, en el que las covariables actúan según un modelo de vida acelerada en tanto el frailty lo hace multiplicativamente en la función de azar marginal, esto es, es de azar proporcional. Es decir, partiremos de la hipótesis siguiente:

$$h_i(t_i, z, \psi_i(x_i)) = z.\exp(\beta_i x_i) h_i(t_i.\exp(\beta_i x_i)) \quad . \qquad [2.36]$$

La función de supervivencia marginal condicionada será

$$S_i(t_i, \psi_i(x_i), z) = \exp[-z.\Lambda_i(t_i \exp(\beta_i x_i))] \quad .$$

La función de supervivencia conjunta condicionada será

$$S(t_1, t_2 / z, \psi_1(x_1), \psi_2(x_2)) =$$
$$= \exp[-z exp(\beta_1 x_1) \int_0^{t_1} h_1(t_1 \exp(\beta_1 x_1)) dt_1 -$$
$$- z exp(\beta_2 x_2) \int_0^{t_2} h_2(t_2 \exp(\beta_2 x_2)) dt_2] =$$

$$= \exp[-z(\int_0^{t_1} h_1(t_1 \exp(\beta_1 x_1)) d(\exp(\beta_1 x_1) t_1) +$$
$$+ \int_0^{t_2} h_2(t_2 \exp(\beta_2 x_2)) d(\exp(\beta_2 x_2) t_2))] =$$

$$= \exp[-z(\Lambda_1(t_1 \exp(\beta_1 x_1)) + \Lambda_2(t_2 \exp(\beta_2 x_2)))] \quad .$$

Realizando la mixtura de distribuciones, la función de supervivencia quedará

$$S(t_1,t_2)=\int \exp[-z(\Lambda_1(t_1\exp(\beta_1x_1))+\Lambda_2(t_2\exp(\beta_2x_2)))]\,dF(z) \quad . \quad [2.37]$$

Supongamos,a continuación, que modelizamos el frailty Z con las distribuciones estable positiva, gamma y uniforme.

a) Si z sigue la distribución estable positiva, la función de supervivencia queda

$$S(t_1,t_2)=\exp[-(\Lambda_1(t_1\exp(\beta_1x_1))+\Lambda_2(t_2\exp(\beta_2x_2)))^{\alpha}]. \quad [2.38]$$

Si las marginales son Weibull, es decir

$$\Lambda_1(t_1)=\epsilon_1 t_1^{\gamma_1}$$
$$\Lambda_2(t_2)=\epsilon_2 t_2^{\gamma_2}$$

entonces la expresión de la función de supervivencia conjunta será

$$S(t_1,t_2)=\exp[-(\epsilon_1 t_1^{\gamma_1}\exp(\gamma_1\beta_1x_1)+\epsilon_2 t_2^{\gamma_2}\exp(\gamma_2\beta_2x_2))^{\alpha}] \quad . \quad [2.39]$$

Se tratará de forma especial en el capítulo de la Dependencia y de la Estimación.

Si, además, los dos individuos tienen tiempos de fallo con idéntica distribución podemos suponer que

$$\epsilon_1=\epsilon_2=\epsilon$$
$$\gamma_1=\gamma_2=\gamma$$

y entonces la función de supervivencia conjunta es

$$S(t_1,t_2)=\exp[-\epsilon^{\alpha}(t_1^{\gamma}\exp(\gamma\beta_1 x_1)+t_2^{\gamma}\exp(\gamma\beta_2 x_2))^{\alpha}] \quad . \qquad [2.40]$$

Las funciones de supervivencia marginales serán

$$S_i(t_i)=\exp[-\epsilon^{\alpha}t_i^{\gamma\alpha}\exp(\alpha\gamma\beta_i x_i)]$$

e integrando el parámetro de escala en la forma paramétrica de las covariables

$$S_i(t_i)=\exp[-t_i^{\alpha\gamma}\exp(\alpha\gamma\beta_i x_i)] \quad .$$

Veamos la identificabilidad del modelo cuya expresión de la función de supervivencia es [2.40]. La novedad ahora es que

$$u=\epsilon\exp(\gamma\beta_1 x_1)t_1^{\gamma}+\epsilon\exp(\gamma\beta_2 x_2)t_2^{\gamma} \quad .$$

En este caso se verifica

$$\epsilon\exp(\gamma\beta_i x_i)=\epsilon\exp(\gamma\beta'_i x_i) \quad \Rightarrow \quad \beta_i=\beta'_i, \forall x_i$$

demostrable sin más que darle valores a x_i .

b) Si el frailty Z se modeliza con una distribución gamma, la función de supervivencia quedará

$$S(t_1,t_2)=\left[1+\frac{\Delta_1(t_1\exp(\beta x))+\Delta_2(t_2\exp(\gamma y))}{\theta}\right]^{-\theta} \qquad [2.41]$$

donde

$$S_1(t_1)=\left[1+\frac{\epsilon_1 t_1^{\delta_1}\exp(\delta_1\beta x)}{\theta}\right]^{-\theta} \quad .$$

En este mismo caso, pero considerando el modelo inicialmente así

$$S(t_1,t_2)=\left[\frac{1}{S_1(t_1)^{\delta}}+\frac{1}{S_2(t_2)^{\delta}}-1\right]^{-\frac{1}{\delta}}$$

se tendría

$$S_1(t_1)=\exp[-\Delta_1(t_1\exp(\beta x)]\quad ,$$

expresión coherente con la definición original de supervivencia marginal.

c) Si consideramos el caso de un frailty con distribución uniforme, la función de supervivencia conjunta será

$$S(t_1,t_2)=\frac{\exp[-b(\Lambda_1(t_1\exp(\beta_1x_1))+\Lambda_2(t_2\exp(\beta_2x_2)))]}{(a-b)(\Lambda_1(t_1\exp(\beta_1x_1))+\Lambda_2(t_2\exp(\beta_2x_2)))}-\frac{\exp[-a(\Lambda_1(t_1\exp(\beta_1x_1))+\Lambda_2(t_2\exp(\beta_2x_2)))]}{(a-b)(\Lambda_1(t_1\exp(\beta_1x_1))+\Lambda_2(t_2\exp(\beta_2x_2)))}\,. \quad [2.42]$$

Consideremos que las marginales siguen distribuciones weibull, es decir

$$\Lambda_i(t_i)=\epsilon_i t_i^{\gamma_i}\ ,\quad i:1,2$$

y en este caso la función de supervivencia quedará así

$$S(t_1,t_2)=\frac{\exp[-b(\epsilon_1t_1^{\gamma_1}\exp(\gamma_1\beta_1x_1)+\epsilon_2t_2^{\gamma_2}\exp(\gamma_2\beta_2x_2))]}{(a-b)(\epsilon_1t_1^{\gamma_1}\exp(\gamma_1\beta_1x_1)+\epsilon_2t_2^{\gamma_2}\exp(\gamma_2\beta_2x_2))}-\frac{\exp[-a(\epsilon_1t_1^{\gamma_1}\exp(\gamma_1\beta_1x_1)+\epsilon_2t_2^{\gamma_2}\exp(\gamma_2\beta_2x_2))]}{(a-b)(\epsilon_1t_1^{\gamma_1}\exp(\gamma_1\beta_1x_1)+\epsilon_2t_2^{\gamma_2}\exp(\gamma_2\beta_2x_2))}$$

y englobando el parámetro de escala de la distribución Weibull en la forma paramétrica de las covariables quedaría

$$S(t_1,t_2)=\frac{\exp[-b(t_1^{\gamma_1}\exp(\gamma_1\beta_1x_1)+t_2^{\gamma_2}\exp(\gamma_2\beta_2x_2))]}{(a-b)(t_1^{\gamma_1}\exp(\gamma_1\beta_1x_1)+t_2^{\gamma_2}\exp(\gamma_2\beta_2x_2))}-$$

$$-\frac{\exp[-a(t_1^{\gamma_1}\exp(\gamma_1\beta_1x_1)+t_2^{\gamma_2}\exp(\gamma_2\beta_2x_2))]}{(a-b)(t_1^{\gamma_1}\exp(\gamma_1\beta_1x_1)+t_2^{\gamma_2}\exp(\gamma_2\beta_2x_2))}$$

y en el caso particular de coincidir los parámetros de forma de dichas distribuciones weibull, es decir

$$\gamma_1=\gamma_2=\gamma$$

quedaría

$$S(t_1,t_2)=\frac{\exp[-b(t_1^{\gamma}\exp(\gamma\beta_1x_1)+t_2^{\gamma}\exp(\gamma\beta_2x_2))]}{(a-b)(t_1^{\gamma}\exp(\gamma\beta_1x_1)+t_2^{\gamma}\exp(\gamma\beta_2x_2))}-$$

$$-\frac{\exp[-a(t_1^{\gamma}\exp(\gamma\beta_1x_1)+t_2^{\gamma}\exp(\gamma\beta_2x_2))]}{(a-b)(t_1^{\gamma}\exp(\gamma\beta_1x_1)+t_2^{\gamma}\exp(\gamma\beta_2x_2))}$$

quedando como supervivencia marginal

$$S_i(t_i)=\frac{\exp(-bt_i^{\gamma}\exp(\gamma\beta_ix_i))-\exp(-at_i^{\gamma}\exp(\gamma\beta_ix_i))}{(a-b)(t_i^{\gamma}\exp(\gamma\beta_ix_i))}.$$

Observación: En la expresión de $\phi(u)$ no influye la presencia de covariables, es decir, la "forma" de la transformada de Laplace es la misma, tanto en el caso de introducir las covariables según un modelo de azar proporcional como de vida acelerada.

2.7.6 COVARIABLES EN EL PARÁMETRO DE ASOCIACIÓN

Hemos visto que las covariables explicativas se suelen introducir a través de las distribuciones marginales vía azar proporcional o de vida acelerada. Así lo hemos considerado hasta ahora, concretamente en el caso de marginales Weibull. Nosotros presentamos a continuación otra forma de hacerlo. Se trata de introducirlas a través del parámetro de asociación. De esta manera, como veremos en el capítulo sobre Dependencia, involucramos a dichas covariables en la posible asociación entre los tiempos de fallo, es decir, estas covariables explicarán el grado de dependencia entre los tiempos de fallo.

Nosotros hemos considerado los dos casos que exponemos a continuación donde además de las covariables ya introducidas anteriormente introducimos ahora otras, vía parámetro de asociación. Estos dos casos corresponden a un frailty estable positiva y a un frailty gamma.

a) La expresión de la función de supervivencia en un modelo frailty estable positiva, marginales Weibull y mixto era la siguiente:

$$S(t_1,t_2)=\exp[-\epsilon^{\alpha}(t_1^{\gamma}\exp(\gamma\beta_1 x_1)+t_2^{\gamma}\exp(\gamma\beta_2 x_2))^{\alpha}]\,.$$

El parámetro de asociación α proviene de la estable positiva. Su rango de valores es el intervalo $[0,1]$. Consideraremos que la expresión que introduce las covariables es la siguiente

$$\alpha=\frac{1}{1+\exp(c+\beta_3 x_3)}\quad .$$

En este caso la función de supervivencia será

$$S(t_1,t_2)=\exp\left[-(\epsilon t_1^{\gamma}\exp(\gamma\beta_1 x_1)+\epsilon t_2^{\gamma}\exp(\gamma\beta_2 x_2))^{\frac{1}{1+\exp(c+\beta_3 x_3)}}\right]\quad .$$

b) En un modelo frailty gamma, mixto y marginales Weibull la expresión que introduce las covariables en el parámetro de asociación suponemos que es

$$\theta-1=\exp(c+\gamma y)$$

$$lg(\theta-1)=c+\gamma y=\alpha\quad .$$

En este caso la expresión de la función de supervivencia es

$$S(t_1,t_2)=$$

$$[\exp(\epsilon\exp(\eta_1\beta_1x_1+\alpha)\,t_1^{\eta_1})+\exp(\epsilon\exp(\eta_2\beta_2x_2+\alpha)\,t_2^{\eta_2})-1]^{-\frac{1}{\exp\alpha}} \quad .$$

Este modelo lo trataremos en los capítulos sobre Dependencia y sobre Estimación.

La identificabilidad del modelo está asegurada teniendo en cuenta lo visto anteriormente. Lo novedoso ahora sería ver si

$$(1+\alpha u)^{-\frac{1}{\alpha}}=(1+\alpha' u)^{-\frac{1}{\alpha'}} \Longrightarrow \alpha=\alpha' \quad .$$

Esto está asegurado desde el inicio ya que, por el teorema de identificabilidad de las transformadas de Laplace, el frailty gamma ha de ser el mismo y con los mismos parámetros.

2.7.7 UN FRAILTY INVERSO

Es interesante considerar la siguiente situación sugerida por D.R. Cox en un reciente seminario con nuestra Línea de investigación. Hay situaciones en las cuales el frailty, que induce la dependencia entre las variables, actúa de forma inversa. Es decir, si Z es el frailty, 1/Z actuando multiplicativamente en la función de azar marginal. Nosotros planteamos esta situación en el caso más general: si conocemos un frailty Z y su correspondiente transformada de Laplace, ¿podremos aprovechar esta información para construir un modelo si en realidad es una función de este frailty lo que actúa multiplicativamente en la función de azar marginal?. Veámoslo en el caso del frailty inverso, esto es, 1/Z. Hemos de modificar las hipótesis de partida en el sentido que vemos a continuación.

Partamos de la hipótesis siguiente: sea la función de azar marginal

$$h_i(t_i, z) = z^{-1} \cdot h_i(t_i); \quad i:1,2 \quad .$$

La función de supervivencia marginal condicionada es

$$S_i(t_i/z) = \exp[-\frac{1}{z}\int_0^{t_i} h_i(t_i)\,dt_i] = \exp[-\frac{1}{z}\Lambda_i(t_i)] = [S_i(t_i)]^{\frac{1}{z}}$$

con lo que la función de supervivencia conjunta condicionada es

$$S(t_1, t_2/z) = P(T_1 \geq t_1, T_2 \geq t_2/z) =$$

$$= P(T_1 \geq t_1/z) \cdot P(T_2 \geq t_2/z) = [S_1(t_1)]^{\frac{1}{z}}[S_2(t_2)]^{\frac{1}{z}} =$$

$$= [S_1(t_1)S_2(t_2)]^{\frac{1}{z}} = \exp[-\frac{1}{z}(\Lambda_1(t_1) + \Lambda_2(t_2))] \quad .$$

Realizando la correspondiente mixtura de distribuciones respecto a Z la función de supervivencia conjunta quedará

$$S(t_1, t_2) = \int S(t_1, t_2/z)\,dF(z) =$$

$$= E[\exp(-\frac{1}{z}(\Lambda_1(t_1) + \Lambda_2(t_2))] = \int [S_1(t_1)S_2(t_2)]^{\frac{1}{z}}\,dF(z) \quad .$$

Obsérvese que esta última expresión no es una transformada de Laplace de Z. Si hacemos el cambio de variable Y = 1/Z y considerando un valor y = 1/z resulta que

$$F_Y(y) = P(Y \leq y) = P(\frac{1}{Z} \leq y) = P(\frac{1}{y} \leq Z) = 1 - P(Z \leq \frac{1}{y}) = 1 - F_Z(\frac{1}{y}) = 1 - F_Z(z) \,.$$

Tomando diferenciales queda

$$dF_Z(z) = d(1 - F_Y(y)) = -dF_Y(y)$$

con lo que la función de supervivencia conjunta quedará así

$$S(t_1,t_2)=-[-\int[S_1(t_1).S_2(t_2)]^y dF_Y(y)]=\phi(u)$$

donde

$$u=\Lambda_1(t_1)+\Lambda_2(t_2)$$

y ϕ es la transformada de Laplace de la variable Y=1/Z .

Se pueden presentar dos inconvenientes: que no conozcamos la distribución de la variable Y, o bien, que no conozcamos su transformada de Laplace. Veamos si podemos aprovechar la transformada de Laplace de Z. Es decir, intentemos conocer esta transformada de Laplace de Y a partir de la transformada de Laplace de Z.

$$\phi(u)=\int[S_1(t_1)S_2(t_2)]^y dF_Y(y)=\int[S_1(t_1)S_2(t_2)]^y f_Y(y)\,dy$$

$$dF_Z(z)=-dF_Y(y)\Rightarrow f_Z(z)\,dz=-f_Y(y)\,dy$$

$$f_Z(z)\,dz=f_Z(1/y)\,d(1/y)=(-1/y^2)\,f_Z(1/y)\,dy$$

luego

$$f_Y(y)=(1/y^2)\,f_Z(1/y)\ \Rightarrow$$

$$f_Y(y)=z^2 f_Z(z)\quad .$$

Usaremos a continuación los siguientes resultados:

Teorema.- Si $\phi_z(u)$ es la transformada de Laplace de $f_z(z)$ entonces $[\phi_z(u)]''$ es la transformada de Laplace de la función $g(z)= z^2 f_z(z)$. (Véase Apostol (1972)).

Corolario.- Conocida la transformada de Laplace de $f_z(z)$ se conoce la de la función

$$g(z)=g(1/y)=f_Y(y) \quad .$$

Observación: Se verifica que

$$\int [S_1(t_1)S_2(t_2)]^{1/y} g(1/y)\,d(1/y)=\int [S_1(t_1)S_2(t_2)]^{y} f_Y(y)\,dy.$$

Por consiguiente, la función de supervivencia conjunta que describe el modelo

$$S(t_1,t_2)=\phi_y(u) \quad ,$$

o transformada de Laplace de Y, es la derivada segunda de la transformada de Laplace de Z.

Como aplicación, si Z sigue una distribución estable positiva, es decir

$$\phi_z(u)=\exp(-u^{\alpha})$$

entonces la expresión de la transformada de Laplace que describe el modelo frailty inverso sería

$$\phi_z''(u)=\exp(-u^{\alpha})[\alpha(\alpha-1)(-u)^{\alpha-2}+\alpha^2(-u)^{2(\alpha-1)}]$$

que con

$$u=\Lambda_1(t_1)+\Lambda_2(t_2)$$

resulta la función de supervivencia conjunta buscada.

Este resultado es importante, nos permite obtener una expresión del modelo frailty inverso a partir de una expresión de un modelo frailty inicial. Lo más importante es que nos induce a generalizar este resultado en el sentido de poder considerar cualquier función de un frailty Z. Es decir, conocido el modelo para un frailty conocido Z, esto es, conocida su transformada de Laplace, podemos buscar un modelo para una función genérica del frailty original.

En el capítulo 3 continuaremos tratando el modelo frailty inverso en problemas de dependencia.

A continuación presentamos un modelo frailty con algunas variaciones importantes en las hipótesis iniciales. Consideraremos dos frailties que actúan multiplicativamente en las funciones de azar marginales, respectivamente. Un frailty será Z y el otro será su inverso 1/Z. De esta manera el riesgo de fallo aumenta en la primera componente al aumentar el valor del frailty y al mismo tiempo disminuye en la segunda componente. El tratamiento de este modelo presenta serios problemas, como veremos a continuación. Partimos de

$$h_1(t_1,z)=z.h_1(t_1)$$
$$h_2(t_2,z)=(1/z).h_2(t_2)$$

y entonces, la función de supervivencia conjunta condicionada es

$$S(t_1,t_2/z)=\exp[-z\Lambda_1(t_1)-(1/z)\Lambda_2(t_2)]$$

y la función de supervivencia conjunta será

$$S(t_1, t_2) = \int \exp[-z\Lambda_1(t_1) - (1/z)\Lambda_2(t_2)]\, dF(z)$$

que no es una transformada de Laplace. Deberemos resolver la integral. En el capítulo 3 veremos algunas cuestiones de dependencia en este modelo.

CAPÍTULO 3
ESTUDIO DE LA DEPENDENCIA

3.1 INTRODUCCIÓN

Como ya hicimos notar al principio de esta memoria un objetivo fundamental en el análisis de datos de supervivencia multivariante es el estudio de la dependencia o asociación entre los tiempos de fallo o supervivencia. Hemos visto en la construcción de modelos que juega un importante papel el tipo de datos considerados así como la estructura específica que refleja la posible dependencia entre los tiempos de fallo. Es decir, ya se enfocaba el problema hacia este objetivo fundamental. El estudio de la dependencia, comunmente denominada asociación, está muy abierto. No hay reglas generales para todo tipo de modelos por la dificultad del problema. Será necesario distinguir entre los distintos criterios de construcción de modelos, o bién, concretar las distribuciones marginales, para hacer un estudio mínimamente satisfactorio de la dependencia.

En este capítulo se comienza presentando en esquema las distintas vías abordadas para estudiar la dependencia. A continuación se presenta un estudio de cada una de estas vías y, por último, pasaremos a nuestras aportaciones: consideramos el modelo frailty uniforme y estudiamos algunas de estas medidas de asociación como son el coeficiente de correlación lineal, el coeficiente de variación, el coeficiente de Kendall y el coeficiente de Oakes. Estudiamos la asociación en un modelo mixto y mencionaremos el hecho de que con la forma usual de introducir las covariables no se recoge la posible dependencia que estas producen y sí en cambio si se introducen en el parámetro de asociación. También tratamos el estudio de la dependencia cuadrante en los modelos mixtos.

Las distintas vías usadas por los autores se pueden clasificar de la siguiente manera:

a) El coeficiente de correlación lineal de Pearson es estudiado por Hougaard y otros (1992) y Lindeboom & Van den Berg (1994) que plantean las limitaciones que presenta.

b) El cuadrado del coeficiente de variación de Pearson del frailty es estudiado por los mismos autores del apartado anterior como una medida de asociación en los modelos frailties.

c) El uso del coeficiente de concordancia de Kendall está tomando gran aceptación en la actualidad. Es una medida de asociación global en cuanto no depende del momento considerado. Hougaard (1986, 1992), Genest & Mackay (1986), Oakes (1982, 1989), Lemman (1966), Schweizer & Woeff (1981), Costigan & Klein (1993) son algunos de los autores que lo estudian.

d) La siguiente función fué introducida por Oakes (1989):

$$\frac{h_1(t_1/T_2=t_2)}{h_1(t_1/T_2>t_2)}=\theta(t_1,t_2).$$

Es una medida de dependencia local. Es de suma importancia dada su relación con los modelos frailty.

e) Una medida de correlación local la introdujo Bjerve & Doksum (1991). Otro autor que la usa es Mei-Ling Ting Lee (1993). Es una medida de correlación local que definiremos más adelante.

f) La siguiente función

$$\frac{F(t_1,t_2)}{F_1(t_1)\cdot F_2(t_2)}$$

es usada por Crowder (1989), Johnson (1975) y Barlow & Proschan (1977) como una medida de dependencia.

g) Los conceptos de positividad total de orden r (TP_r) para estudiar la dependencia es estudiada por Marshall & Olkin (1988), Lee & Klein (1988), Whitmore & Mei-Ling (1991) y otros.

Al igual que en los métodos de construcción de modelos la metodología que emplearemos en el desarrollo de esta memoria será la de ir describiendo cada caso anterior, deteniendonos en las medidas de asociación que consideramos de más importancia y presentándolas en los casos para los cuales no se han considerado o no se ha hecho lo suficientemente. Los casos c) y d) serán los que más desarrollaremos puesto que serán los más utilizados en nuestras aportaciones.

3.2 EL COEFICIENTE DE CORRELACIÓN DE PEARSON EN SUPERVIVENCIA

Una primera forma de abordar el problema de la asociación es considerar el coeficiente de correlación lineal de Pearson. Dado que se tiene una variable (T_1,T_2) que denota dos tiempos de fallo, es lógico pensar en el coeficiente de correlación de Pearson entre T_1 y T_2

$$\rho_{T_1T_2}=\frac{cov(T_1,T_2)}{\sigma_{T_1}\sigma_{T_2}}$$

expresión que coincide con

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[F(t_1,t_2)-F_1(t_1)F_2(t_2)]dt_1dt_2 \qquad [3.1]$$

según Lehmann (1966) y Schweizer y Wolff (1981).

Este coeficiente es más idoneo para datos normales bivariantes que para datos de supervivencia.

Un cálculo de este coeficiente supone conocer la distribución conjunta además de la dificultad de resolver dicha integral. Sólo se han llegado a obtener resultados importantes considerando fuertes restricciones en las hipótesis.

Se considera que en el modelo la función $\psi_i(x_i,\beta_i)$ modeliza las covariables y z_1 y z_2 son dos covariables no observadas que actúan de la siguiente manera:

$$h_i(t_i/z_i,\psi_i(x_i,\beta_i))=z_i\cdot\psi_i(x_i,\beta_i)\cdot h_i(t_i) \quad , \quad i:1,2$$

es decir, es un modelo de azar proporcional.

Se considera que a cada tiempo t_i solo le afecta la variable z_i , esto es, la función de densidad marginal condicionada es

$$f(t_i/z_1,z_2,\psi_1(x_1,\beta_1),\psi_2(x_2,\beta_2))=f(t_i/z_i,\psi_i(x_i,\beta_i)) \quad , \quad i:1,2.$$

Se considera la hipótesis de independencia condicionada, esto es, la función de densidad conjunta condicionada es

$$f(t_1,t_2/z_1,z_2,\psi_1(x_1,\beta_1),\psi_2(x_2,\beta_2))=$$
$$=f(t_1/z_1,z_2,\psi_1(x_1,\beta_1),\psi_2(x_2,\beta_2))\cdot$$
$$\cdot f(t_2/z_1,z_2,\psi_1(x_1,\beta_1),\psi_2(x_2,\beta_2))$$

es decir, hay independencia condicionada a z_1, z_2 y $\psi_i(x_i, \beta_i)$.

Observamos que estas hipótesis coinciden con las hipótesis del método de construcción de modelos vía marginales desarrollado entre otros por Marshall & Olkin (1988). Las variables z_1 y z_2 hacen en cierto modo el papel de frailties. La función de supervivencia conjunta sería

$$S(t_1, t_2) = \int_0^\infty \int_0^\infty S(t_1, t_2 / z_1, z_2) f(z_1, z_2) dz_1 dz_2 .$$

Se supone que las $h_i(t_i)$ son constantes, por ejemplo, iguales a 1. Lindeboom & Van den Berg (1994) demuestran que

$$\rho_{lgT_1, lgT_2} = \frac{cov(lgz_1, lgz_2)}{\left[(var(lgz_1) + \frac{\Pi^2}{6}) (var(lgz_2) + \frac{\Pi^2}{6}) \right]^{1/2}} \qquad [3.2]$$

que da una expresión de la correlación entre las variables lgT_1 y lgT_2 .

Por otra parte se ha deducido que

$$\rho_{T_1, T_2} = \frac{cov(1/z_1, 1/z_2)}{[(var(1/z_1) + E(1/z_1)^2) (var(1/z_2) + E(1/z_2)^2)]^{1/2}} \qquad [3.3]$$

y que

$$cov(lgT_1, lgT_2) = cov(lgz_1, lgz_2) .$$

Según esto, nosotros hemos deducido un resultado importante en uno de los modelos considerados en el capítulo 2. Se trata del caso en el que $Z_1 = Z$ y $Z_2 = 1/Z$. Es decir, hay un frailty inverso en la segunda componente. En este caso $cov(lgZ, lg1/Z) = -var(lgZ)$. Por consiguiente la asociación es negativa.

En el caso de que $z_1=z_2=z$ se verifica

$$cov(lgT_1, lgT_2)=var(lgz)$$

con lo que se observa que la posible correlación es positiva. Este resultado es muy importante y se corrovorará posteriormente al considerar otras medidas de asociación. Además la dependencia entre T_1 y T_2 está directamente determinada por la varianza de la distribución del frailty.

En general, se cumple que, sin ningún tipo de restricciones,

$$\rho_{T_1, T_2}=0$$

si y solo si z_1 y z_2 son degeneradas, por lo que si estas no son degeneradas habrá cierta correlación entre T_1 y T_2 .

Si $h_i(t_i)$ no es constante no es posible dar unas expresiones de la correlación análogas a las anteriores puesto que en la deducción de tales fórmulas era necesario suponer la hipótesis de ser constante. Haciendo una transformación de los ejes podríamos calcular

$$\rho_{\Lambda_1(t_1), \Lambda_2(t_2)}$$

o bien

$$\rho_{lg\Lambda_1(t_1), lg\Lambda_2(t_2)}$$

expresiones válidas por la monotonía de dichas transformaciones. Así se elimina la necesidad de conocer las distribuciones marginales. Es decir, si se supone que se parte de datos que corresponden a los azares acumulativos marginales, el coeficiente de correlación de estos es indicativo del correspondiente de los datos de los tiempos de fallo. Hougaard (1986,1992) lo calcula para el caso de un frailty con distribución estable positiva

$$\rho_{\Lambda_1(t_1),\Lambda_2(t_2)} = \frac{2\Gamma(1+\alpha)^2}{\Gamma(1+2\alpha)} - 1$$

$$\rho_{lg\Lambda_1(t_1),lg\Lambda_2(t_2)} = 1 - \alpha^2$$

expresiones que, como se ve, dependen del parámetro llamado de asociación, el cual es el parámetro que define la distribución del frailty, en este caso el frailty estable.

3.3 EL COEFICIENTE DE VARIACIÓN

Algunas consideraciones anteriores nos sugieren en cierto modo el coeficiente de variación como una medida de la asociación. Los modelos frailty se obtienen por medio de una mixtura de distribuciones. Se suele usar la varianza de la distribución del frailty como una medida de asociación. Éste es el que crea la dependencia entre las variables. Es lógico pensar en su varianza. Se suele utilizar mejor aún el coeficiente de variación del frailty, o mejor, su cuadrado

$$CV^2 = \frac{varZ}{EZ^2}. \qquad [3.4]$$

Presenta problemas calcularlo para el frailty estable, uno de los modelos más clásicos, y para el frailty inversa gaussiana. Hay ocasiones en las que, como alternativa, es mejor calcular Var(lgZ).

3.4 EL COEFICIENTE DE KENDALL EN SUPERVIVENCIA

3.4.1 INTRODUCCIÓN

Veamos a continuación uno de los métodos más importantes para el estudio de la asociación. Se trata del uso del coeficiente de concordancia de Kendall. Es más idóneo para datos de supervivencia que el coeficiente de correlación de Pearson.

Sea

$$(T_1^{(1)}, T_2^{(1)}), \ldots\ldots, (T_1^{(n)}, T_2^{(n)})$$

una muestra aleatoria simple de la variable (T_1, T_2).

Definición: El coeficiente de correlación por rangos de Kendall, $\hat{\tau}$, está basado en la frecuencia del suceso

$$[(T_1^{(i)} - T_1^{(j)})(T_2^{(i)} - T_2^{(j)}) > 0] \quad .$$

Se calcula disponiendo la muestra en forma de dos sucesiones

$$\begin{array}{c} T_1^{(1)} \ldots\ldots T_1^{(n)} \\ T_2^{(1)} \ldots\ldots T_2^{(n)} \end{array} \quad ,$$

siendo la primera sucesión los valores ordenados de las primeras componentes y la segunda la sucesión de las correspondientes parejas de los valores de la primera sucesión.

A cada diferencia

$$T_2^{(j)} - T_2^{(i)}$$

se le asigna el valor $+1$ si

$$T_2^{(j)} - T_2^{(i)} > 0 \quad (j > i)$$

y el valor -1 si

$$T_2^{(j)}-T_2^{(i)}<0 \quad (j>i)\,.$$

Sea P la suma de los valores +1 (pares concordantes) y Q la suma de los valores -1 (pares discordantes). Entonces se define

$$\hat{\tau}=\frac{P-Q}{n(n-1)/2} \quad .$$

Varía entre -1 y +1 al ser una media de dichos valores.

Si la ordenación es totalmente opuesta entonces vale -1. Si la ordenación es coincidente entonces vale 1. Si el valor es cero no hay relación entre las ordenaciones. Además se observa que

$$P+Q=\frac{n(n-1)}{2}=\binom{n}{2}$$

y si

$$v(a,b,c,d)=\begin{cases}1 & si \quad (a-b)(c-d)>0\\ -1 & si \quad (a-b)(c-d)<0\end{cases}$$

entonces

$$P-Q=\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} v(T_1^{(i)},T_1^{(j)},T_2^{(i)},T_2^{(j)}) \quad .$$

En el caso de que algunos valores de T_1 y T_2 estén ligados, P sería el n° de veces que ocurra

$$(T_1^{(j)}-T_1^{(i)})(T_2^{(j)}-T_2^{(i)})>0 \qquad (j>i)$$

y Q sería el n° de veces que

$$(T_1^{(j)}-T_1^{(i)})(T_2^{(j)}-T_2^{(i)})<0 \quad .$$

Si este producto es cero no se contabiliza. Es decir, si ahora se considera

$$v^*(a,b,c,d)=\begin{cases} 1 & si \quad (a-b)(c-d)>0 \\ 0 & si \quad =0 \\ -1 & si \quad <0 \end{cases}$$

entonces

$$P-Q=\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} v^*(T_1^{(i)},T_1^{(j)},T_2^{(i)},T_2^{(j)})$$

y en este caso

$$\hat{\tau}=\frac{P-Q}{\sqrt{\frac{1}{2}[n(n-1)-\sum_{i=1}^{g}t_i(t_i-1)]}\sqrt{\frac{1}{2}[n(n-1)-\sum_{i=1}^{h}u_i(u_i-1)]}}$$

donde

g = nº de grupos de observaciones ligadas de T_1

t_i = nº de observaciones ligadas en el grupo i

h = nº de grupos de observaciones ligadas de T_2

u_i = nº de observaciones ligadas en el grupo i

(Véase Cuadras (1991), Vélez Ibarrola y García Pérez, Daniel (1990), Harnett & Murphy (1987), Tsokos (1987) y Siegel (1970)).

A continuación se dará una definición del coeficiente de concordancia τ de Kendall (poblacional). Se verifica que $E(\hat{\tau})=\tau$.

Se puede contrastar la hipótesis

$$H_0: \tau=0$$

a partir de la distribución del estadístico

$$\frac{\hat{\tau}}{\sqrt{\frac{2(2n+5)}{9n(n-1)}}} \underset{(n\geq 8)}{\rightarrow} N(0,1)$$

siendo además posible calcular $var(\hat{\tau})$, como puede verse en Manatunga & Oakes (1996).

El coeficiente de correlación por rangos $\hat{\tau}$ de Kendall es un coeficiente de correlación no paramétrico bién definido para calcularse a partir de los datos de una muestra. Servirá para estimar el correspondiente valor poblacional τ. Se ha definido dicho coeficiente para todos los datos de una población.

Definición: Sea la variable aleatoria (T_1,T_2), y sean

$$(T_1^{(1)}, T_2^{(1)}) \quad y \quad (T_1^{(2)}, T_2^{(2)})$$

dos realizaciones independientes (muestras) de dicha variable aleatoria. Se define el coeficiente τ como la diferencia entre la probabilidad de concordancia y la probabilidad de discordancia de estas dos observaciones, es decir

$$\tau=P[(T_1^{(2)}-T_1^{(1)})(T_2^{(2)}-T_2^{(1)})\geq 0]-P[(T_1^{(2)}-T_1^{(1)})(T_2^{(2)}-T_2^{(1)})<0] \qquad [3.5]$$

expresión que es equivalente a la siguiente

$$\tau = 2P[(T_1^{(2)} - T_1^{(1)})(T_2^{(2)} - T_2^{(1)}) \geq 0] - 1 \qquad [3.6]$$

y también a la siguiente

$$\tau = E\ signo[(T_1^{(2)} - T_1^{(1)})(T_2^{(2)} - T_2^{(1)})]. \qquad [3.7]$$

Propiedades: Es fácil comprobar que

$$-1 \leq \tau \leq 1$$

y que este coeficiente es invariante frente a transformaciones monótonas de T_1 y T_2.

Se verifica además que

$$P[(T_1^{(2)} - T_1^{(1)})(T_2^{(2)} - T_2^{(1)}) \geq 0] = 1 \Rightarrow \tau = 1$$
$$P[(T_1^{(2)} - T_1^{(1)})(T_2^{(2)} - T_2^{(1)}) \geq 0] = 0 \Rightarrow \tau = -1$$
$$P[(T_1^{(2)} - T_1^{(1)})(T_2^{(2)} - T_2^{(1)}) \geq 0] = 1/2 \Rightarrow \tau = 0.$$

Si T_1 y T_2 son independientes, entonces

$$\tau = 0 \ .$$

Vimos en el apartado 2.4 que los modelos frailty están incluidos el la clase de las funciones arquimedianas. Este hecho será de gran utilidad para calcular el coeficiente de Kendall en dichos modelos. Como vimos en el apartado 1.3.3, Genest & Mackay (1986) definen una cópula arquimediana como una distribución definida de la siguiente forma

$$F(t_1,t_2)=\begin{cases}\phi^{-1}(\phi(t_1)+\phi(t_2)) & \text{si} \quad \phi(t_1)+\phi(t_2)\leq\phi(0)\\ 0 & \text{en otro caso}\end{cases} \qquad [3.8]$$

donde ϕ es una función definida de la siguiente manera:

$$\phi:[0,1]\rightarrow[0,\infty]$$

con las dos primeras derivadas continuas en (0,1] y cumpliéndose

$$\begin{aligned}&\phi(1)=0\\&\phi'(t)<0\\&\phi''(t)>0 \quad \forall t\in(0,1).\end{aligned}$$

En esta clase de distribuciones el coeficiente de Kendall adopta la forma

$$\tau=4E[F(t_1,t_2)]-1=4\iint_{[0,1]^2}F(t_1,t_2)\,dF(t_1,t_2)-1 \qquad [3.9]$$

expresión de suma importancia que liga el coeficiente de Kendall con la función de distribución conjunta.

Esta clase de distribuciones (cópulas arquimedianas) posee la propiedad de que dichas distribuciones tienen las marginales uniformes en el intervalo (0,1) y se verifica que

$$S(t_1,t_2)=F(t_1,t_2)-t_1-t_2+1$$

$$máx(0,t_1+t_2-1)\leq F(t_1,t_2)\leq mín(t_1,t_2).$$

Por otra parte se deduce que

$$\tau=4\int_0^1\frac{\phi(t)}{\phi'(t)}dt+1 \qquad [3.10]$$

ya que

$$E[F(t_1,t_2)]=\iint_{\phi(t_1)+\phi(t_2)\leq\phi(0)} F(t_1,t_2)\,dF(t_1,t_2)=$$

$$=\iint_{\phi(t_1)+\phi(t_2)\leq\phi(0)} F(t_1,t_2)\,f(t_1,t_2)\,dt_1dt_2.$$

Tal expresión será de suma utilidad puesto que da una fórmula práctica para calcular el coeficiente. Además, usando dicha fórmula, se observa que el area bajo la curva $\phi(t)/\phi'(t)$ en (0,1) da un valor del coeficiente, con lo cual se tiene una interpretación geométrica del coeficiente.

Estas expresiones del coeficiente de Kendall pueden utilizarse en los modelos frailty. Además, en estos modelos, Oakes (1989) expresa el coeficiente de Kendall de la siguiente manera

$$\tau=4\int_0^\infty u.\phi^{-1}(u)\phi^{-1''}(u)\,du - 1. \qquad [3.11]$$

De esta manera se tiene una expresión para calcular el coeficiente de Kendall en función de ϕ^{-1}.

Veamos el siguiente resultado obtenido a partir de los resultados de 2.7.7. Sea Z un frailty cualquiera y sea Y=1/Z. Si

$$\phi_z(u)$$

es la transformada de Laplace de Z entonces el coeficiente de asociación de Kendall de un modelo frailty Y se puede expresar en función de esta transformada de Laplace de la sigiente manera

$$\tau_Y=4\int_0^\infty u\phi_z''(u)\,.\,\phi_z^{IV)}(u)\,du-1$$

de acuerdo a lo estudiado en el apartado 2.7.7 .

3.4.2 EJEMPLOS Y COMENTARIOS

A continuación se presenta el valor del coeficiente de Kendall en algunos de los modelos clásicos: Modelo de Clayton, frailty inversa gaussiana y frailty estable positiva. En segundo lugar se presenta dicho coeficiente según distintas formas de la función ϕ. Por último, se darán algunos comentarios de gran importancia sobre este coeficiente.

Los modelos clásicos más importantes son los siguientes:

a) Para el modelo de Clayton, véase 2.3.2 y 2.4.2, se verifica

$$\theta(S(t_1,t_2))=\frac{h_1(t_1/T_2=t_2)}{h_1(t_1/T_2\geq t_2)}=c$$

donde

$$\tau=\frac{c-1}{c+1}$$

fácilmente comprobable ya que

$$\phi^{-1}(u)=(1+\theta^{-1}u)^{-\delta} \quad \Rightarrow$$

$$\phi(t)=\theta(t^{-\frac{1}{\delta}}-1) \quad \Rightarrow$$

$$\tau=1-2\delta+\frac{4\delta^2}{1+2\delta}=\frac{c-1}{c+1} \quad , \quad (\delta=\frac{1}{c-1}).$$

En el caso especial de un frailty con distribución exponencial se verifica

$$\phi^{-1}(u)=\frac{\lambda}{u+\lambda}$$

$$\phi(u)=\frac{\lambda-\lambda u}{u} \quad \Rightarrow \quad \tau=\frac{1}{3}.$$

b) En un frailty con distribución inversa gaussiana se verifica

$$\tau=\frac{1}{2}-\eta+\eta^2\exp(2\eta)E_1(2\eta) \qquad \text{con}$$

$$E_1(k)=\int_k^\infty\frac{(\exp(-u)}{u}du.$$

Hougaard y otros (1992) generalizan este coeficiente en el sentido de que consideran una clase de distribuciones que incluye a la distribución inversa gaussiana y calculan dicho coeficiente en esta clase.

c) En un frailty con distribución estable positiva

$$\tau=1-\alpha$$

ya que

$$\phi^{-1}(t)=\exp(-t^\alpha) \quad , \quad \phi(t)=(-lgt)^{\frac{1}{\alpha}}.$$

Veamos a continuación la siguiente definición. Una distribución $F(t_1,t_2)$ definida por [3.8] se dice que es singular si

$$\phi(t_1)+\phi(t_2)=\phi(0).$$

Genest & Mackay (1986) demuestran que en las distribuciones sigulares se verifica que

$$\frac{\phi(0)}{\phi'(0)}\neq 0.$$

Veamos cómo queda el coeficiente de Kendall según la forma que adopten las funciones ϕ:

1. Si se tiene la expresión

$$\phi(t)=(t^{-\alpha}-1)/\alpha \qquad \alpha>0$$

entonces la función de distribución conjunta es

$$F(t_1, t_2) = (1/t_1^{\alpha} + 1/t_2^{\alpha} - 1)^{-1/\alpha}$$

que coincide con la expresión de la función de distribución del modelo de Clayton, en el cual

$$\phi(0)/\phi'(0) = 0$$

es decir, es no singular y

$$\tau = \frac{\alpha}{\alpha+2}.$$

Si se expresa el parámetro de la siguiente forma

$$\alpha = c - 1$$

el coeficiente de Kendall coincide con la fórmula vista en el apartado a) anterior. Este coeficiente de Kendall toma el valor particular

$$\tau = 1 \text{ si } \alpha \to \infty \Rightarrow F(t_1, t_2) = \text{mín}(t_1, t_2).$$

2. Si se tiene la expresión

$$\phi(t) = (t^{-\alpha} - 1)/\alpha \quad \text{con} \quad -1 < \alpha \leq 0$$

el modelo que resulta también es el de Clayton. Se cumple que

$$\tau = -1 \text{ si } \alpha = -1 \text{ y } \tau = 0 \text{ si } \alpha = 0.$$

Además, para $\alpha=0$ se tiene que

$$\phi(t)=\text{lím}_{\alpha\to 0}(t^{\alpha}-1)/\alpha=-\lg t$$

3. Si se tiene la expresión

$$\phi(t)=(1-t)^{\alpha}\ ,\ \alpha\geq 1$$

entonces la función de distribución queda

$$F(t_1,t_2)=\text{máx}[0,1-((1-t_1)^{\alpha}+(1-t_2)^{\alpha})^{1/\alpha}]$$

donde para $\alpha=1$ la distribución que resulta es singular. El coeficiente de Kendall es

$$\tau=1-2/\alpha$$

verificándose que

$$\alpha=2\ \Rightarrow\ \tau=0$$

quedando en este caso la función de distribución de la siguiente forma

$$F(t_1,t_2)=\text{máx}[0,1-\sqrt{(1-t_1)^2+(1-t_2)^2}].$$

4. Si se tiene la expresión

$$\phi(t)=|\lg t|^{\alpha+1}\ ,\ 0<t\leq 1\ ,\ \alpha\geq 0$$

se llega al modelo de Gumbel

$$|lgF(t_1,t_2)|^{\alpha+1}=|lgF_1(t_1)|^{\alpha+1}+|lgF_2(t_2)|^{\alpha+1}$$

donde ϕ^{-1} es la transformada de Laplace de la distribución estable positiva.

$$\text{Si}\quad \alpha'=\frac{1}{\alpha+1} \quad\Rightarrow\quad \tau=1-\alpha'=\frac{\alpha}{\alpha+1}.$$

5. Si se tiene la expresión

$$\phi(t)=-lg[\frac{1-\alpha^t}{1-\alpha}]$$

resulta la función de distribución

$$F(t_1,t_2)=lg_\alpha[1+\frac{(\alpha^{t_1}-1)(\alpha^{t_2}-1)}{\alpha-1}].$$

Una propiedad muy interesante es que si ϕ_α y ϕ_β son dos funciones del tipo [3.8] tal que $\phi_\alpha(t)/\phi_\beta(t)$ es monótona creciente en t para $\alpha<\beta$ resulta que el coeficiente τ de Kendall es estrictamente creciente en α,β. Esto es de mucha utilidad en el caso del modelo de Gumbel donde se verifica tal hipótesis.

Como se ve, distintas formas de la función $\phi(t)$ generan un modelo con su correspondiente coeficiente de Kendall. O mejor, cada modelo con marginales uniformes en [0,1] proviene de su correspondiente función $\phi(t)$ al que le corresponde un coeficiente de Kendall.

Costigan & Klein (1993) dan una versión condicional del coeficiente de Kendall adaptándolo al caso en el que el sistema ha sobrevivido al tiempo t. Igualmente se define el coeficiente de Kendall dado que una de las componentes ha fallado en el tiempo t. Para el modelo de frailty gamma estos coeficientes son constantes. En el frailty estable esto no ocurre, son funciones decrecientes en el tiempo. Igual le ocurre al frailty inversa gaussiana.

Oakes (1982) utiliza la definición del coeficiente de Kendall muestral como un estimador del parámetro de asociación para el modelo de Clayton. Haremos referencia a esto en el capítulo de Estimación.

Todas las fórmulas vistas consideradas del coeficiente de Kendall son de sumo interés puesto que nos dan una expresión de la asociación en el modelo donde tal expresión depende del parámetro de asociación. Pero el problema se puede ver desde una óptica diferente. Como hemos mencionado, habrá que hacer inferencias sobre este parámetro de asociación que aparece en el modelo. Además de las técnicas usuales de inferencia que utilizaremos en el capítulo de Estimación se puede ver de la siguiente manera. El coeficiente de Kendall es una función que depende de este parámetro. Podríamos estimar dicho coeficiente por su correspondiente valor muestral y a partir de ahí obtener una estimación del parámetro de asociación.

El coeficiente de Kendall es la medida de asociación más usada en análisis de supervivencia multivariante. No obstante también se han considerado otras medidas similares tales como el coeficiente de correlación de Spearman, el coeficiente de correlación medial de Blomquist, etc. (Véase Genest & Mackay (1986), Schweizer & Wolff (1981), Harnett & Murphy (1987), Daniel (1990) y Tsokos (1987)).

3.5 EL COEFICIENTE DE ASOCIACIÓN DE OAKES

3.5.1 INTRODUCCIÓN

Veamos a continuación otro coeficiente para el estudio de la asociación entre los dos tiempos de supervivencia. Se trata del coeficiente de asociación definido por Oakes (1989) de la siguiente forma:

$$\frac{h_1(t_1/T_2=t_2)}{h_1(t_1/T_2>t_2)}=\theta(t_1,t_2). \qquad [3.12]$$

Es una medida de asociación local. Lee & Klein (1988) demuestran que, para modelos frailty, las funciones $h_1(t_1/T_2=t_2)$ y $h_1(t_1/T_2>t_2)$ son decrecientes en t_2 , para cualquier t_1 y se verifica que

$$h_1(t_1/T_2=t_2)\geq h_1(t_1/T_2>t_2) \quad \forall t_1, t_2.$$

Por lo tanto, se verifica que

$$\theta(t_1,t_2)\geq 1.$$

Oakes (1989) demostró que efectivamente depende sólo de t_1 y t_2. El interés fundamental de esta función $\theta(t_1,t_2)$ es que verdaderamente mide la dependencia entre las variables como bién puede observarse directamente de la propia definición.

En el modelo de Clayton resulta que $\theta(t_1,t_2)=constante$, la asociación no depende de los tiempos. Clayton construyó su modelo a partir esta hipótesis anterior. Del mismo modo se podrían construir otros modelos considerando diversas funciones para el coeficiente $\theta(t_1,t_2)$, pero este camino suele resultar dificultoso por la complejidad de las ecuaciones diferenciales.

Es fácil observar que este coeficiente también puede expresarse como

$$\theta(t_1,t_2)=\frac{S(t_1,t_2)\dfrac{\partial^2 S(t_1,t_2)}{\partial t_1\partial t_2}}{\dfrac{\partial S(t_1,t_2)}{\partial t_1}\dfrac{\partial S(t_1,t_2)}{\partial t_2}}. \qquad [3.13]$$

Veamos a continuación un resultado importante. Se trata de la forma que adopta esta función anterior cuando $S(t_1,t_2)$ es arquimediana.

Sea $S(t_1,t_2)$ una función arquimediana , es decir, que pueda expresarse de la forma siguiente

$$S(t_1,t_2)=\phi(\phi^{-1}S_1(t_1)+\phi^{-1}S_2(t_2))$$

siendo ϕ una función decreciente, no negativa, y tal que

$$\phi(0)=1,\quad \phi''(0)>0.$$

Entonces el coeficiente $\theta(t_1,t_2)$ depende de t_1 y t_2 sólo a través de $S(t_1,t_2)$, es decir

$$\theta(t_1,t_2)=\theta^*(S(t_1,t_2)).$$

Por otra parte sabemos que si $S(t_1,t_2)$ se ha obtenido de un modelo frailty entonces esta función es arquimediana. Es decir, los modelos frailty son una subclase de las funciones arquimedianas. El recíproco se verifica cuando la función ϕ sea una transformada de Laplace.

En resumen, si $S(t_1,t_2)$ proviene de un modelo frailty, $S(t_1,t_2)$ es arquimediana. Ahora bién, si $S(t_1,t_2)$ es arquimediana se verifica que

$$\theta(t_1,t_2)=\frac{-S(t_1,t_2)(\phi^{-1})''(S(t_1,t_2))}{(\phi^{-1})'(S(t_1,t_2))} \qquad [3.14]$$

y según vimos en el capítulo de construcción de modelos, dada una función

$$\theta(t_1,t_2)=\theta(S(t_1,t_2))$$

se puede llegar a construir el modelo, y viceversa, dada una función ϕ (que define el modelo), podemos calcular

$$\theta(t_1,t_2)=\theta(S(t_1,t_2)).$$

Luego a partir de la transformada de Laplace de una distribución cualquiera (el frailty) se puede obtener el coeficiente

$$\theta(t_1, t_2)$$

que proporciona la medida de asociación.

3.5.2 EJEMPLOS

Veamos a continuación la forma que adopta este coeficiente en algunos modelos importantes ya estudiados.

a) Modelo de Clayton, $\theta(t_1, t_2)=\text{cte}$

b) Frailty inversa gaussiana, $\theta(t_1, t_2)=1+\dfrac{1}{\eta-lgS(t_1, t_2)}$

c) Frailty estable, $0<\alpha<1$, $\theta(t_1, t_2)=1+\dfrac{1-\alpha}{-\alpha lgS(t_1, t_2)}$.

En este último caso, frailty estable, las expresiones de la función de supervivencia conjunta y el coeficiente de asociación de Oakes quedan

$$S(t_1, t_2)=\exp[-(\Lambda_1+\Lambda_2)^{\alpha}]$$

$$\theta(t_1, t_2)=1+\frac{1-\alpha}{-\alpha lg\exp(-(\Lambda_1+\Lambda_2)^{\alpha})}=1+\frac{1-\alpha}{\alpha(\Lambda_1+\Lambda_2)^{\alpha}}$$

y si las marginales son weibull entonces

$$\theta(t_1, t_2)=1+\frac{1-\alpha}{\alpha(\xi_1 t_1^{\eta_1}+\xi_2 t_2^{\eta_2})^{\alpha}}.$$

Oakes (1989) demuestra que, para modelos frailies,

$$\theta(t_1, t_2)$$

puede utilizarse para definir una versión condicional del coeficiente de Kendall. Se trata de la expresión

$$\frac{\theta(S(t_1, t_2))-1}{\theta(S(t_1, t_2))+1}$$

en la que se necesita conocer la expresión de $S(t_1,t_2)$.

3.6 PROCEDIMIENTOS ALTERNATIVOS PARA EL ESTUDIO DE LA ASOCIACIÓN

A continuación se describen otras medidas o formas alternativas en el estudio de la asociación entre dos variables que denotan tiempos de fallo. Se trata de un coeficiente de correlación introducido por Bjerve & Doksum en 1991, el concepto de la dependencia cuadrante, el concepto de variables asociadas y, por último, la positividad total. Se resumen asimismo las implicaciones fundamentales de estas medidas en los modelos frailty.

3.6.1 COEFICIENTE DE CORRELACIÓN DE BJERVE & DOKSUM

Bjerve & Doksum (1991) introdujeron la siguiente medida de correlación local. La función de correlación local entre T_2 y T_1 con $T_1=t_1$ se define como

$$\rho(t_1)=\frac{\sigma_1\beta(t_1)}{[(\sigma_1\beta(t_1))^2+\sigma^2(t_1)]^{1/2}} \qquad [3.15]$$

donde

$$\beta(t_1)=\frac{d\,E(T_2/T_1=t_1)}{dt_1}$$

$$\sigma_1^2=var\ T_1$$

$$\sigma^2(t_1)=var\ (T_2/T_1=t_1)$$

observándose que en esta medida intervienen las distribuciones marginales de los tiempos de fallo.

Mei-Ling Ting Lee (1993) calcula esta medida en un caso particular, cuando se tiene un frailty gamma y la función de supervivencia marginal condicionada es

$$S_i(t_i/z)=\exp(-zt_i)$$

es decir, considera un frailty gamma y realiza una mixtura de exponenciales.

3.6.2 DEPENDENCIA CUADRANTE

Veamos a continuación otro enfoque distinto para estudiar una posible asociación o dependencia entre las variables. Lehmann (1966) define la *dependencia cuadrante* de la siguiente manera: $F(t_1,t_2)$ tiene una dependencia cuadrante positiva si se cumple que

$$\frac{F(t_1,t_2)}{F_1(t_1)\cdot F_2(t_2)}\geq 1 \quad \forall t_1,t_2 \qquad [3.16]$$

y será negativa en el caso de ser menor o igual que la unidad.

La definición es equivalente si en lugar de considerar funciones de distribución suponemos funciones de supervivencia. Johnson & Kotz (1975) relacionan esta definición con el vector razón de azar. Barlow & Proschan (1977) establecieron una condición suficiente para que F sea dependiente cuadrante negativa. Crowder (1989) estudia esta medida de asociación para el caso de un modelo frailty estable con marginales weibull, es decir un modelo en el que la función de supervivencia conjunta es

$$S(t_1,t_2)=\exp[-(\xi_1 t_1^{\phi_1}+\xi_2 t_2^{\phi_2})^{\upsilon}]$$

es decir, una distribución de weibull bivariante. Posteriormente considera el modelo mas general siguiente, una función de supervivencia conjunta

$$S(t_1,t_2)=\exp[k^{\upsilon}-(k+s)^{\upsilon}] \quad \text{con} \quad k\geq 0$$

$$s=\xi_1 t_1^{\phi_1}+\xi_2 t_2^{\phi_2}.$$

Llega a la conclusión de que si $\upsilon < 1$ existe dependencia positiva, si es mayor que uno dependencia negativa e independencia si es igual a uno.

3.6.3 VARIABLES ASOCIADAS

Veamos el enfoque que presentan Marshall & Olkin (1988) para el estudio de la dependencia. T_1 y T_2 (o sus distribuciones) se dicen asociadas si se verifica que

$$cov[\,u(T_1,T_2),v(T_1,T_2)\,]\geq 0$$

para cualesquiera u,v funciones crecientes. Algunas consecuencias de esta definición son las siguientes:

a) Si T_1 y T_2 son asociadas implica que toda correlación es no negativa.

b) Si T_1 y T_2 son asociadas se verifica que tienen una dependencia cuadrante positiva, es decir, se verifica que

$$\frac{S(t_1,t_2)}{S_1(t_1)\cdot S_2(t_2)}\geq 1.$$

c) Si T_1 y T_2 son asociadas e incorreladas, son independientes.

d) En los modelos generados vía marginales (véase el capítulo 2) se cumple que si G y K son asociadas implica que la función de supervivencia obtenida por medio de la mixtura de distribuciones [2.18] es asociada.

e) Si G es asociada y la función de supervivencia se obtuvo a través de la fórmula [2.19], esta función de supervivencia es asociada.

3.6.4 POSITIVIDAD TOTAL

También Marshall & Olkin (1988) utilizan la siguiente definición anteriormente definida por Karlin (1968). Una función K(x,y) no negativa se dice totalmente positiva de orden 2 (TP_2) si para

$$x_1 \le x_2$$
$$y_1 \le y_2$$

implica que

$$\begin{vmatrix} K(x_1, y_1) & k(x_1, y_2) \\ K(x_2, y_1) & K(x_2, y_2) \end{vmatrix} \ge 0 \ .$$

La positividad total de orden r (TP_r) se define en términos de los determinantes de orden 1,2,.....r . Una función es totalmente positiva de orden infinito, TP_∞ , si es totalmente positiva de cualquier orden finito. Un resultado importante es que un modelo frailty definido a través de la función de supervivencia

$$S(t_1, t_2) = \int [S_1(t_1) S_2(t_2)]^\theta dG(\theta)$$

verifica que ésta es totalmente positiva de orden infinito.

Esta última propiedad generaliza el resultado de Lee & Klein (1988) que formuló que los modelos frailty son TP_2. Además, este hecho implica que T_1 y T_2 son asociadas y cuadrante dependiente positivas. Estos resultados son fundamentales y hacen que en los modelos frailty todas las medidas de dependencia, por ejemplo el coeficiente de Kendall, sean no negativas. En estos modelos todas estas medidas de asociación valen cero (uno en el caso de la dependencia cuadrante) en el caso de que el frailty sea degenerado y en tal caso T_1 y T_2 son independientes.

3.7 ESTUDIO DE LA ASOCIACIÓN EN EL MODELO FRAILTY UNIFORME

A continuación nosotros presentamos aportaciones sobre la determinación de algunas medidas de dependencia en un modelo frailty con distribución uniforme. Estudiaremos el coeficiente de correlación tratado al inicio de este capítulo y veremos algunas limitaciones que presenta. A diferencia del anterior, el coeficiente de variación no presenta estas dificultades, será fácil de calcular. Finalmente calculamos el coeficiente de Kendall y exponemos un método para determinar el coeficiente de Oakes.

3.7.1 CORRELACIÓN

El coeficiente de correlación es una medida de dependencia global. Es por lo general difícil de calcular porque depende de las distribuciones marginales. Solo para algunos frailties concretos se ha podido eliminar esta dependencia de las marginales y calcular $\rho_{\Lambda_1(t_1),\Lambda_2(t_2)}$. Lee & Klein (1988) establecen que en un modelo frailty general pero con marginales Weibull con parámetros de forma iguales se verifica que el coeficiente de correlación de Pearson vale

$$\rho = \frac{\Gamma(1+\frac{1}{\gamma})^2 var(z^{-\frac{1}{\gamma}})}{var(z^{-\frac{1}{\gamma}})\Gamma(1+\frac{2}{\gamma}) + \left[\Gamma(1+\frac{2}{\gamma}) - \Gamma(1+\frac{1}{\gamma})^2\right] E(z^{-\frac{1}{\gamma}})^2}$$

siendo z el frailty.

En nuestro caso, Z uniforme, y con marginales Weibull, efectuando los oportunos cálculos obtenemos el coeficiente de correlación donde hemos calculado previamente

$$E(z^{-\frac{1}{\gamma}})^2 \quad y \quad var(z^{-\frac{1}{\gamma}}).$$

Hemos obtenido que

$$E(z^{-\frac{1}{\gamma}})^2=\frac{1}{b-a}\left[\frac{b^{-\frac{2}{\gamma}+1}-a^{-\frac{2}{\gamma}+1}}{-\frac{2}{\gamma}+1}\right]$$

y

$$var(z^{-\frac{1}{\gamma}})=\frac{1}{b-a}\left[\frac{b^{-2/\gamma+1}-a^{-2/\gamma+1}}{-2/\gamma+1}-\frac{2kb^{-1/\gamma+1}-2ka^{-1/\gamma+1}}{-1/\gamma+1}+b-a\right]$$

siendo $k=\dfrac{b^{-1/\gamma+1}-a^{-1/\gamma+1}}{(b-a)(-1/\gamma+1)}$. De esta manera la expresión final del coeficiente es

$$\rho=\frac{\Gamma(1+\frac{1}{\gamma})^2}{\Gamma(1+\frac{2}{\gamma})+\dfrac{\Gamma(1+\frac{2}{\gamma})-\Gamma(1+\frac{1}{\gamma})^2}{1+\dfrac{(b-a)(-\frac{2}{\gamma}+1)}{b^{-\frac{2}{\gamma}+1}-a^{-\frac{2}{\gamma}+1}}-\dfrac{2(-\frac{2}{\gamma}+1)(b^{-\frac{1}{\gamma}+1}-a^{-\frac{1}{\gamma}+1})^2}{(b-a)(-\frac{1}{\gamma}+1)^2(b^{-\frac{2}{\gamma}+1}-a^{-\frac{2}{\gamma}+1})}}}$$

que no depende de las covariables puesto que éstas van implícitas en el parámetro de escala de la distribución de Weibull.

Por otra parte, a partir de las expresiones [3.2] y [3.3] deberemos calcular el valor de

$$\frac{var(lgz)}{var(lgz)+\pi^2/6} \quad y \quad \frac{var(1/z)}{var(1/z)+E(1/z)^2} \qquad [3.17]$$

para un frailty Z que nosotros condideramos con distribución uniforme. Para ello habremos de sustituir en las anteriores expresiones el resultado de calcular var(lgz), var(1/z) y $E(1/z)^2$ cuyos valores tras laboriosos aunque sencillos cálculos son respectivamente

$$var(lgz) = \frac{1}{b-a}\Big[k^2 b - 2kb(lgb-1) + b(lgb-1)\,lgb + b - b(lgb-1) -$$

$$-k^2 a + 2ka(lga-1) - a(lga-1)\,lga - a + a(lga-1)\Big],$$

$$var(1/z) = \frac{1}{b-a}\Big[-\frac{1}{b} + \frac{(lgb-lga)^2}{(b-a)^2}b - \frac{2(lgb-lga)}{b-a}lgb + \frac{1}{a} -$$

$$-\frac{(lgb-lga)^2}{(b-a)^2}a + \frac{2(lgb-lga)}{b-a}lga\Big]$$

y

$$E(1/z)^2 = 1/(ab)$$

siendo $k=b[lg(b)-1]-a[lg(b)-1]$; $a>0$.

3.7.2 COEFICIENTE DE VARIACIÓN

En un modelo frailty con distribución uniforme el cálculo del coeficiente de variación del frailty es sencillo, tomando su cuadrado la expresión siguiente

$$CV^2 = \frac{(b-a)^3}{4(b^3-a^3)}.$$

3.7.3 EL COEFICIENTE DE KENDALL

Hemos visto en la primera parte de este capítulo la importancia del cálculo de este coeficiente para explicar la dependencia entre las variables. El uso de la expresión [3.10] supone conocer la inversa de la transformada de Laplace del frailty uniforme. Nosotros utilizamos, dada esta dificultad, la expresión [3.11]. En el frailty uniforme el cálculo de esta expresión es laborioso. Si ϕ^{-1} es la transformada de Laplace de un frailty uniforme, entonces

$$\phi^{-1}(u)=\frac{\exp(-bu)-\exp(-au)}{(a-b)u}$$

siendo su derivada segunda

$$(\phi^{-1})''(u)=\frac{1}{u^3(a-b)}\Big[u^2[b^2\exp(-bu)-a^2\exp(-au)]-$$

$$-2[(-b.\exp(-bu)+a.\exp(-au))u-(\exp(-bu)-\exp(-au))]\Big]$$

quedando la expresión [3.11] de la siguiente forma

$$\tau=\int_0^\infty\Big[\frac{u^2[\exp(-bu)-\exp(-au)][b^2\exp(-bu)-a^2\exp(-au)]}{u^3(a-b)}-$$

$$-\frac{2u[\exp(-bu)-\exp(-au)][-b\exp(-bu)+a\exp(-au)]}{u^3(a-b)^2}+$$

$$+\frac{2[\exp(-bu)-\exp(-au)]^2}{u^3(a-b)^2}\Big]du-1=$$

$$=\int_0^\infty\frac{1}{(a-b)^2}\Big[\frac{b^2\exp(-2bu)}{u}-\frac{a^2\exp(-(a+b)u)}{u}-\frac{b^2\exp(-(a+b)u)}{u}+$$

$$+\frac{a^2\exp(-2au)}{u}+\frac{2b\exp(-2bu)}{u^2}-\frac{2a\exp(-(a+b)u)}{u^2}-$$

$$-\frac{2b\exp(-(a+b)u)}{u^2}+\frac{2a\exp(-2au)}{u^2}+\frac{2\exp(-2bu)}{u^3}+$$

$$+\frac{2\exp(-2au)}{u^3}-\frac{4\exp(-(a+b)u)}{u^3}\Big]du-1$$

donde todas las integrales anteriores pueden aproximarse a partir de la integral

$$\int_\epsilon^\infty\frac{\exp(-2ku)}{u}du$$

donde K=a,b ó (a+b)/2 y ϵ una constante próxima a cero. Para un valor de epsilon suficientemente pequeño obtenemos un valor aproximado del coeficiente.

Es de observar que el cálculo de $\int_0^{\infty}(e^{-t}/t)\,dt$ es equivalente al cálculo de $\int_{-\infty}^{0}(e^{x}/x)\,dx$. Para el cálculo de esta integral se tiene que $\int_{-\infty}^{x}(e^{t}/t)\,dt = C + \ln|x| + \frac{x}{1.1!} + \frac{x^2}{2.2!} + \ldots$ siendo C la constante de Euler.

3.7.4 EL COEFICIENTE DE OAKES

En el caso del frailty uniforme el cálculo de este coeficiente de asociación presenta el problema de la inversión de la función transformada de Laplace [2.28]. Es decir no podemos usar la fórmula [3.14] y habremos de usar [3.13]. Se trata pues de derivar la función de supervivencia que define el modelo frailty uniforme y sustituir en dicha expresión [3.13].

Si $c = t_1^{\gamma} e^{\beta_1 x_1} + t_2^{\gamma} e^{\beta_2 x_2}$ entonces

$$\theta(t_1, t_2) = \frac{[e^{-bc} - e^{-ac}][e^{-bc}(c^2 b^2 + 2cb + 2) - e^{-ac}(c^2 a^2 + 2ca + 2)]}{[(ac+1)e^{-ac} - (bc+1)e^{-bc}]^2}$$

coeficiente que depende de a, b, t_1, t_2 y las covariables.

En este apartado mencionemos un hecho importante en el que insistiremos en el próximo capítulo. Se trata de la posible dependencia que inducen las covariables explicativas. En un modelo frailty general la asociación o dependencia la induce el frailty. La posible dependencia que puedan inducir las covariables no está recogida en algunos de los coeficientes de asociación vistos. Es necesario modificar un poco el modelo para recoger este hecho. Para ello nosotros hemos introducido también las covariables en el parámetro de asociación presentando una forma de hacerlo en dos casos vistos en el capítulo anterior.

3.8 EL COEFICIENTE DE OAKES EN UN MODELO MIXTO CON FRAILTY ESTABLE POSITIVA

Hemos visto que en un modelo frailty estable positiva el coeficiente de asociación global τ de Kendall es $1-\alpha$. No depende de t_1 y t_2 , de las marginales o de las covariables que se introducen a través de éstas. El coeficiente de variación es infinito. El coeficiente de correlación presenta el problema de que en la estable positiva no es posible dar una expresión de su media o varianza. Sí se ha conseguido obtener el siguiente resultado cuando las marginales son Weibull:

$$\rho_{lgT_1,\, lgT_2} = 1-\alpha^2 .$$

También se ha obtenido, sin hipótesis sobre las marginales, que

$$\rho_{\Lambda_1(t_1),\Lambda_2(t_2)} = \frac{2\Gamma(1+\alpha)^2}{\Gamma(1+2\alpha)} - 1 .$$

Para involucrar a la covariables obtendremos el coeficiente local de Oakes:

$$\theta(t_1, t_2) = 1 + \frac{1-\alpha}{-\alpha\, lgS(t_1, t_2)} = 1 + \frac{1-\alpha}{\alpha[\epsilon t_1^{\gamma}\exp(\gamma\beta_1 x_1) + \epsilon t_2^{\gamma}\exp(\gamma\beta_2 x_2)]^{\alpha}}$$

que depende del parámetro de la distribución estable positiva, de los parámetros de la distribución Weibull, de los coeficientes de regresión y de los tiempos de fallo.

Sea $S(t_1,t_2)$ una función absolutamente continua correspondiente a la función de supervivencia de un modelo con frailty estable positiva, distribuciones marginales Weibull y covariables externas según hipótesis de vida acelerada. El coeficiente de asociación de Oakes $\theta(t_1,t_2)$ decrece desde infinito hasta la unidad en tanto la función de supervivencia lo hace desde 1 a 0. La justificación es la siguiente: Todo modelo frailty implica una función de supervivencia arquimediana. Para estas funciones se verifica que la dependencia del coeficiente de Oakes de los tiempos de fallo lo es a través de la función de supervivencia. Es decir $\theta(t_1,t_2)=\theta^*(S(t_1,t_2))$. Oakes (1989) calcula este coeficiente para un frailty estable positiva

$$\theta(t_1,t_2)=1+\frac{1-\alpha}{-\alpha\log S(t_1,t_2)}$$

con lo cual si tenemos marginales Weibull (con parámetros coincidentes lo cual no resta generalidad) y covariables bajo hipótesis de vida acelerada queda

$$\theta(t_1,t_2)=1+\frac{1-\alpha}{\alpha(\epsilon t_1^{\gamma}\exp(\gamma\beta_1 x_1)+\epsilon t_2^{\gamma}\exp(\gamma\beta_2 x_2))^{\alpha}}.$$

Si $\alpha=1$, frailty degenerado, el coeficiente es 1 y estamos en el caso de independencia entre T_1 y T_2. Si $\alpha\to 0$ estaríamos en el caso de la máxima dependencia y la función de supervivencia coincidiría con el valor mínimo de las supervivencia marginales. Por otro lado, para valores fijos de α, si la función de supervivencia vale 1 (tiempos de fallo 0) el coeficiente vale infinito. Al crecer los tiempos de fallo (función de supervivencia tendiendo a cero) el coeficiente tiende a 1.

3.9 LA ASOCIACIÓN EN UN MODELO FRAILTY GAMMA , MIXTO, MARGINALES WEIBULL Y COVARIABLES TAMBIÉN EN EL PARÁMETRO DE ASOCIACIÓN

Hemos visto que en un modelo frailty gamma el coeficiente de asociación global de Kendall es $\tau=\frac{\theta-1}{\theta+1}$. Al introducir las covariables en el parámetro de asociación lo hacemos de la sigiente forma:

$$\theta-1=\exp(c+\eta y).$$

Así, éstas explican la asociación. El coeficiente de variación en este caso es $CV^2=\theta-1$. Para calcular el coeficiente de correlación con las marginales Weibull resulta (véase Lee & Klein (1988))

$$\rho=\frac{cov(T_1,T_2)}{\sqrt{varT_1}\sqrt{varT_2}}$$

con

$$varT_i=$$

$$[\epsilon(\theta-1)]^{-2/\gamma}\left[\frac{\Gamma(1+2/\gamma)\Gamma(\frac{1}{\theta-1}-\frac{2}{\gamma})}{\Gamma(\frac{1}{\theta-1})}-\left(\frac{\Gamma(1+1/\gamma)\Gamma(\frac{1}{\theta-1}-\frac{1}{\gamma})}{\Gamma(\frac{1}{\theta-1})}\right)^2\right]$$

y

$$cov(T_1,T_2)=[\epsilon(\theta-1)]^{-1/\gamma}[\epsilon(\theta-1)]^{-1/\gamma}[\Gamma(1+\frac{1}{\gamma})^2].$$

$$\cdot\left[\frac{\Gamma(\frac{1}{\theta-1}-\frac{1}{\gamma}-\frac{1}{\gamma})}{\Gamma(\frac{1}{\theta-1})}-\frac{\Gamma(\frac{1}{\theta-1}-\frac{1}{\gamma})\Gamma(\frac{1}{\theta-1}-\frac{1}{\gamma})}{\Gamma(\frac{1}{\theta-1})}\right]$$

con $\frac{1}{\theta-1} > \frac{2}{\gamma}$. Este coeficiente depende de los parámetros de la distribución Weibull y del parámetro de la distribución gamma.

Para involucrar a las covariables en la asociación obtenemos el coeficiente de Oakes. Éste resulta ser θ =cte. Luego en este caso no depende del tiempo, es global.

3.10 DEPENDENCIA CUADRANTE POSITIVA EN LOS MODELOS MIXTOS

Nuestro objetivo es estudiar la dependencia cuadrante en los modelos mixtos. Hemos visto que los modelos frailty (azar proporcional) son dependientes cuadrante positivos (Marshall & Olkin (1988) y Lee & Klein (1988)). Es decir, verifican que

$$S(t_1,t_2) \geq S_1(t_1).S_2(t_2) \;, \; \forall\; t_1,t_2.$$

Recientemente Anderson & Louis (1995) han demostrado que los modelos frailty de vida acelerada también lo son. Veamos el siguiente resultado que afirma lo mismo para los modelos mixtos.

Teorema: Las variables que definen los modelos mixtos son dependientes cuadrante positivas.

Demostración: Tenemos que ver que se verifica la siguiente desigualdad, consecuencia inmediata de la definición de Lehmann (1966):

$$S(t_1,t_2) \geq S_1(t_1).S_2(t_2) \;, \; \forall\; t_1,t_2$$

la cual es equivalente a

$$E_z\, S_1(t_1/z,x_1) \;.\; S_2(t_2/z,x_2) \geq E_z S_1(t_1/z,x_1) \;.\; E_z S_2(t_2/z,x_2).$$

En los modelos mixtos las funciones de supervivencia marginales son

$$S_i(t_i/z,x_i)=\exp(-z\Lambda_i(t_1\exp(\beta_i x_i)) \quad , \quad i:1,2$$

y sabemos que los modelos frailty de azar proporcional son dependientes cuadrante positivos, resultado que se formula sobre la variable (T_1, T_2). Lehmann (1966) demuestra que si esta variable es dependiente cuadrante positiva entonces la variable $(r(T_1),s(T_2))$ también lo es si y solo si r y s son funciones no crecientes. Por lo tanto

$$(S_1(t_1/z,x_1),S_2(t_2/z,x_2))$$

es una variable dependiente cuadrante positiva. Lehmann (1966) establece también que, en este caso, se verifica

$$E\,S_1(t_1/z,x_1).S_2(t_2/z,x_2) \geq E\,S_1(t_1/z,x_1).E\,S_2(t_2/z,x_2)$$

que es lo que queríamos demostrar.

Nosotros hemos considerado un modelo con una variante respecto a los modelos mixtos. Se trata de la inclusión de covariables en el parámetro de asociación en un modelo frailty gamma. En un modelo frailty gamma, mixto con marginales Weibull y covariables también en el parámetro de asociación, la función

$$S_i(t_i/z,x_i)=[\exp(\epsilon\exp(\eta_i\beta_i x_i+\alpha)\,t_i^{\eta_i})-1]^{-\frac{1}{\exp\alpha}}$$

con $\alpha=c+\gamma y$ es no creciente y por lo tanto verifica el teorema anterior.

CAPÍTULO 4
ESTIMACIÓN

4.1 INTRODUCCIÓN

En el análisis de datos de supervivencia multivariante el problema de la estimación ha sido relativamente poco estudiado, se ha tratado más el problema de construcción de modelos y el de asociación o dependencia. El objetivo de este capítulo es el estudio de la estimación en algunos modelos que mencionaremos a continuación. Se comienza, en esta misma introducción, recordando una clasificación de modelos que se consideró en el capítulo 1. Esta clasificación servirá para delimitar en cada caso sobre qué se efectuarán las estimaciones. Además, se obtendrá la expresión general que adoptará la función de verosimilitud en algunos de estos modelos, los modelos definidos a partir de su función de supervivencia conjunta, paramétricos y teniendo en cuenta la posibilidad de censuras.

En el apartado 4.2 se trata la metodología general que se sigue en los modelos semiparamétricos definidos a partir de las funciones de azar marginales y se hará una breve mención a los modelos no paramétricos. En ambos casos se mencionarán las referencias más importantes. Seguidamente, en el apartado 4.3, nosotros haremos algunas aportaciones en un modelo paramétrico definido a partir de la función de supervivencia conjunta, ya considerado en esta memoria. Concretamente, trataremos el problema de la estimación en un modelo frailty uniforme con covariables y marginales Weibull. En este modelo consideraremos que dichas covariables explicativas se introducen de la manera usual, esto es, según un modelo de azar proporcional. En el apartado 4.4 trataremos un modelo mixto con frailty estable positiva y marginales Weibull. Por último, en el apartado 4.5 consideraremos un modelo mixto con frailty gamma y con covariables, también, en el parámetro de asociación.

Como ya indicamos al inicio de esta memoria (véase Capítulo 1) se puede hacer la siguiente clasificación de los modelos: semiparamétricos, paramétricos o no paramétricos. A su vez, los dos primeros casos se pueden considerar de dos formas: definidos a partir de la función de supervivencia conjunta o definidos a partir de las funciones de azar marginales. Así, por ejemplo, en el primer caso está el modelo de Clayton, obtenido en 1978 y deducido posteriormente por caminos diferentes, cuya función de supervivencia se puede expresar así

$$S(t_1,t_2)=\left[\frac{1}{(S_1(t_1))^{\theta-1}}+\frac{1}{(S_2(t_2))^{\theta-1}}-1\right]^{-\frac{1}{\theta-1}} \quad , \quad \theta>1$$

es decir, se trata de una expresión semiparamétrica en la que se observa que $S(t_1,t_2)$ es función de los tiempos de supervivencia a través de las marginales y aparece un parámetro de asociación. No son conocidas las familias a las que pertenecen las distribuciones de dichos tiempos (marginales). Estas funciones de supervivencia marginales actúan como funciones "nacimiento" que generan el modelo. Interesará estimar sobre todo dicho parámetro de asociación, hacer inferencia sobre él. La técnica más usual de inferencia en este modelo semiparamétrico será el uso de la verosimilitud parcial (marginal o condicional). En el caso de actuar covariables se hacen inferencias, además, sobre los parámetros asociados a estas covariables, denominados parámetros de regresión.

Un enfoque semiparámetrico a partir de las funciones de azar marginales se plantea de la siguiente forma. Supongamos hipótesis de azar proporcional, esto es, donde las covariables actúan multiplicativamente en las funciones de azar marginales

$$h_i(t_i,\psi_i(x_i))=\psi_i(x_i)\,.\,h_i(t_i) \quad , \quad i:1,2$$

siendo $h_i(t_i)$ funciones de azar base marginales de familias de distribuciones desconocidas. En este caso, si las covariables se introducen mediante la expresión

$$\psi_i(x_i)=\exp(\beta_i' x_i)$$

la inferencia se hará sobre los parámetros de regresión, el vector β.

Volvamos al planteamiento anterior, los modelos semiparamétricos definidos a partir de la función de supervivencia. Si ya construido el modelo se conoce la familia de distribuciones a la que pertenecen las funciones de supervivencia marginales, se estará ante un modelo paramétrico. Éste será el caso que trataremos fundamentalmente. En el modelo de Clayton anterior la inferencia se hará sobre el parámetro de asociación y sobre los parámetros de dichas distribuciones marginales. Además, si hay covariables, sobre los parámetros de regresión. La técnica más usual es la estimación máximo verosímil.

Modelos no paramétricos serán aquellos en los cuales no se formulan hipótesis para construir un modelo sino que se estima la función de supervivencia directamente.

Es conveniente hacer la siguiente observación relacionada con algunos de los modelos considerados en la clasificación inicial de este apartado. Hemos visto que una de las técnicas más importantes de construcción de modelos era el uso de frailties. A partir de un frailty se genera el modelo semiparamétrico o paramétrico en el sentido visto anteriormente. El frailty en definitiva es una variable aleatoria que actúa de forma común en los tiempos de supervivencia de los dos individuos. Por lo general este frailty es conocido, es decir, se conoce la familia de distribuciones a la que pertenece. Pero podría darse el caso de no ser conocida la distribución del frailty. Sería un frailty no paramétrico. No es común que esto ocurra. Puede verse un caso en Guo & Rodriguez (1992). Presenta la ventaja de la flexibilidad pero no tiene las importantes ventajas de las propiedades que posee un frailty conocido.

Supongamos para comenzar que se conoce la expresión de la función de supervivencia bivariante $S(t_1,t_2)$. Se supone que, además, se está ante una situación paramétrica según el sentido visto anteriormente. Se necesita hacer inferencias sobre un conjunto, que posiblemente sea grande, de parámetros, ya sean parámetros de regresión, parámetros asociados a las distribuciones marginales o parámetros de asociación. En supervivencia multivariante los tamaños muestrales han de ser relativamente más grandes que en supervivencia univariante. Es decir, se necesita un conjunto de datos multivariantes suficientemente alto. En la práctica se observa que la toma de datos presenta ciertos problemas. Uno de ellos, muy frecuente, es que no se conocen algunos tiempos de supervivencia. Los datos son "incompletos". Se dice que se tienen datos censurados. En supervivencia bivariante esto ocurre cuando alguna componente o las dos del vector de datos no se conoce. La censura mas usual es la censura a la derecha en la cual el tiempo observado es menor que el tiempo de supervivencia real. Es decir, al medir aún no ha fallado tal componente. Sería censura a la izquierda si el tiempo observado es mayor que el tiempo de supevivencia real. Al ir a medir ya había fallado. Existen otros tipos de censuras, por intervalos, de tipo I, tipo II, aleatoria, etc. Consideremos censuras a la derecha en lo que sigue. Se dispone de un dato expresado por el vector (t_1,t_2). Estableceremos cuatro situaciones distintas definidas por los siguientes grupos:

Grupo 1 si t_1 y t_2 no son censuras
Grupo 2 si t_1 no es censura pero t_2 sí lo es
Grupo 3 si t_1 es censura pero t_2 no lo es
Grupo 4 si t_1 y t_2 son censuras.

Usaremos el indicador de censura siguiente:

$$I(j)=\begin{cases}1 & \text{si} \quad (t_1,t_2) \text{ está en el grupo } j\\ 0 & \text{si} \quad (t_1,t_2) \text{ no está en el grupo } j\end{cases}$$
$$j{:}1,2,3,4.$$

La estimación de los parámetros suele hacerse por el método de la máxima verosimilitud. Cada tipo de datos aporta a la verosimilitud una cantidad distinta. Así, un dato no censurado (grupo 1) aporta a la función de verosimilitud la expresión

$$g1=\frac{\partial^2 S(t_1,t_2)}{\partial t_1 \partial t_2}$$

es decir, el valor correspondiente de la función de densidad. Un dato del grupo 2 aporta la expresión

$$g2 = \frac{-\partial S(t_1, t_2)}{\partial t_1}.$$

Un dato del grupo 3 aporta la expresión

$$g3 = \frac{-\partial S(t_1, t_2)}{\partial t_2}$$

y un dato del grupo 4 aporta el valor correspondiente de la función de supervivencia, es decir, la expresión

$$g4 = S(t_1, t_2).$$

En resumen, si se tiene la muestra

$$(t_{1i}, t_{2i}) \quad , \quad i:1, \ldots., n$$

la función de verosimilitud es

$$L = \prod_{i=1}^{n} \left[\left(\frac{\partial^2 S(t_{1i}, t_{2i})}{\partial t_{1i} \partial t_{2i}} \right)^{I(1)} . \left(\frac{-\partial S(t_{1i}, t_{2i})}{\partial t_{1i}} \right)^{I(2)} . \right.$$
$$\left. . \left(\frac{-\partial S(t_{1i}, t_{2i})}{\partial t_{2i}} \right)^{I(3)} . \left(S(t_{1i}, t_{2i}) \right)^{I(4)} \right] \qquad [4.1]$$

entendiendo tales derivadas respecto t_1 ó t_2 valuadas en t_{1i} ó t_{2i}, respectivamente.

4.2 MODELOS SEMIPARAMÉTRICOS Y NO PARAMÉTRICOS

Se considera ahora el caso semiparamétrico. Como hemos visto, se puede plantear de dos formas. Una de ellas es considerar el modelo definido a partir de la función de supervivencia, con las distribuciones marginales de familias de distribuciones desconocidas. Se hace inferencia sobre el parámetro de asociación y sobre los parámetros asociados a las covariables. Un estudio semiparamétrico de este tipo puede verse en Maguluri (1993) y en Oakes (1986). Otra forma es considerar directamente las hipótesis del modelo de regresión de Cox, o modelo de azar proporcional, donde

$$h_i(t_1, \psi_i(x_i)) = \psi_i(x_i) \cdot h_i(t_i) \quad , \quad i:1,2$$

y donde

$$\psi_i(x_i) \quad y \quad h_i(t_i)$$

son una expresión parámetrica de las covariables y una distribución de azar base desconocida, respectivamente. En este caso la inferencia se hace sobre los parámetros de las covariables, esto es, los parámetros de regresión. Se estima el azar base por métodos directos no paramétricos. La metodología usual es utilizar la técnica de la verosimilitud parcial (marginal o condicional) formulada por Cox (1972, 1975).

Así, si

$$\psi_i(x_i) = \exp(\beta_i x_i)$$

cada tiempo de fallo contribuye a la verosimilitud con la cantidad

$$\frac{\exp(\beta_i x_i)}{\sum_{l \in \Re(t_{(i)})} \exp(\beta_l x_l)} \qquad [4.6]$$

siendo $\Re(t_{(i)})$ el conjunto de riesgo, conjunto de individuos que no han fallado o son censurados. La verosimilitud es

$$L(\beta) = \sum_{i=1}^{k} \beta_i x_i - \sum_{i=1}^{k} lg[\sum_{l \in \Re(t_{(i)})} \exp(\beta_l x_l)] . \qquad [4.7]$$

Un estudio de este tipo puede verse en Wei, Lin & Weissfeld (1989), Prentice, Williams & Peterson (1981), Oakes (1986), Kung-yee Liang, Self & Yue Cune Chang (1992) y en Sinha, Tanner & Hall (1994).

La inferencia no paramétrica analiza la función de supervivencia directamente, haciendo estimaciones sobre ella,

$$\hat{S}(t_1, t_2)$$

y calculando

$$Var(\hat{S}(t_1, t_2)).$$

Así, puede verse Miller Jr (1981), Lawless (1982), Cox & Oakes (1984), Kalbfleisch & Prentice (1980) y Cox (1972).

Una forma, por ejemplo, es considerar como función de supervivencia la expresión

$$\hat{S}(x,y) = \frac{\text{número de observaciones } (t_1, t_2) \text{ con } t_1>x, t_2>y}{n}.$$

4.3 ESTIMACIÓN EN UN MODELO FRAILTY UNIFORME CON COVARIABLES

En este apartado pretendemos estimar los parámetros desconocidos del siguiente modelo de azar proporcional ya introducido en el capítulo 2: un modelo frailty con distribución uniforme, en presencia de covariables y marginales Weibull con parámetros de forma iguales. En este caso la función de supervivencia será

$$S(t_1, t_2) = \frac{e^{-b(t_1^{\gamma}e^{\beta_1 x_1}+t_2^{\gamma}e^{\beta_2 x_2})} - e^{-a(t_1^{\gamma}e^{\beta_1 x_1}+t_2^{\gamma}e^{\beta_2 x_2})}}{(a-b)(t_1^{\gamma}e^{\beta_1 x_1}+t_2^{\gamma}e^{\beta_2 x_2})}$$

donde $0 \leq a < b$. Los parámetros a estimar son los parámetros de asociación a y b , el parámetro de forma γ y los parámetros asociados a las covariables

$$\beta_1 = (\beta_{11}, \ldots, \beta_{1k_1}) \text{ y } \beta_2 = (\beta_{21}, \ldots, \beta_{2k_2}).$$

En total tenemos $1+1+1+k_1+k_2$ parámetros.

La expresión desarrollada de la forma paramétrica de las covariables es

$$\exp(\beta_j x_j) = \exp(\beta_{j1} x_{j1} + \ldots + \beta_{j,k_j} x_{j,k_j}) \quad , \quad j:1,2$$

donde x_{jk} (k:1,...,k_j , j:1,2) en el término i de la muestra (i:1,...,n) toma el valor x_{jki}. En lo que sigue, y para facilitar la notación, no ponemos, y lo suponemos, el subindice i en cualquier tiempo de supervivencia t_j y en cualquier covariable x_{jk}.

Considerando la presencia de datos censurados, en el sentido visto en 4.1, la función de verosimilitud y la función de log verosimilitud quedarán respectivamente

$$L(a,b,\gamma,\beta_1,\beta_2) = \prod_{i=1}^{n} \left[\left(\frac{\partial^2 S(t_{1i},t_{2i})}{\partial t_{1i} \partial t_{2i}} \right)^{I(1)} \cdot \left(\frac{-\partial S(t_{1i},t_{2i})}{\partial t_{1i}} \right)^{I(2)} \cdot \right.$$
$$\left. \cdot \left(\frac{-\partial S(t_{1i},t_{2i})}{\partial t_{2i}} \right)^{I(3)} \cdot \left(S(t_{1i},t_{2i}) \right)^{I(4)} \right]$$

y

$$lgL(a,b,\gamma,\beta_1,\beta_2) = \sum_{i=1}^{n} lg[(g1)^{I(1)} \cdot (g2)^{I(2)} \cdot (g3)^{I(3)} \cdot (g4)^{I(4)}] =$$

$$= \sum_{i=1}^{n} [I(1)\,lg(g1) + I(2)\,lg(g2) + I(3)\,lg(g3) + I(4)\,lg(g4)] =$$

$$= I(1) \sum_{i=1}^{n} lg(g1) + I(2) \sum_{i=1}^{n} lg(g2) + I(3) \sum_{i=1}^{n} lg(g3) + I(4) \sum_{i=1}^{n} lg(g4)$$

entendiendo por g1,g2,g3 y g4 aquellas derivadas parciales definidas en el apartado 4.1 .

Llamaremos

$$C = t_1^{\gamma} e^{\beta_1 x_1} + t_2^{\gamma} e^{\beta_2 x_2}$$

$$D = \frac{\partial C}{\partial \gamma} = t_1^{\gamma}\, lg t_1\, e^{\beta_1 x_1} + t_2^{\gamma}\, lg t_2\, e^{\beta_2 x_2}$$

$$E = \frac{\partial C}{\beta_{1k}} = t_1^{\gamma} e^{\beta_1 x_1} x_{1k}.$$

La derivada respecto t_1 es

$$\frac{\partial S(t_1,t_2)}{\partial t_1}=\frac{(a-b)\gamma t_1^{\gamma-1}e^{\beta_1 x_1}[(aC+1)e^{-aC}-(bC+1)e^{-bC}]}{(a-b)C^2}.$$

La derivada respecto t_2 es

$$\frac{\partial S(t_1,t_2)}{\partial t_2}=\frac{(a-b)\gamma t_2^{\gamma-1}e^{\beta_2 x_2}[(aC+1)e^{-aC}-(bC+1)e^{-bC}]}{(a-b)C^2}$$

y la derivada segunda será

$$\frac{\partial^2 S(t_1,t_2)}{\partial t_1\partial t_2}=$$

$$=[\gamma^2 t_1^{\gamma-1}t_2^{\gamma-1}e^{\beta_1 x_1}e^{\beta_2 x_2}(a-b)^3[e^{-aC}(aC^2-C^2a(aC+1)-2C(aC+1))+$$

$$+e^{-bC}(C^2b(bC+1)+2C(bC+1)-C^2b)]]/[(a-b)^4C^4].$$

La función de log verosimilitud será

$$LogL(a,b,\gamma,\beta_1,\beta_2)=$$

$$=I(1)[2nlg\gamma+(\gamma-1)\sum lgt_1+(\gamma-1)\sum lgt_2+\sum\beta_1 x_1+\sum\beta_2 x_2-nlg(b-a)-$$

$$-4\sum lgC+\sum[lg(e^{-aC}(2Csu2a+2C+C^3a^2)+e^{-bC}(-2C^2b-2C-C^3b^2))]]+$$

$$+I(2)[-nlg(b-a)+nlg\gamma+(\gamma-1)\sum lgt_1+\sum\beta_1 x_1-2\sum lgC+$$

$$+\sum lg((aC+1)e^{-aC}-(bC+1)e^{-bC})]+$$

$$+I(3)[-nlg(b-a)+nlg\gamma+(\gamma-1)\sum lgt_2+\sum\beta_2 x_2-2\sum lgC+$$

$$+\sum lg((aC+1)e^{-aC}-(bC+1)e^{-bC})]+$$

$$+I(4)[\sum lg(e^{-aC}-e^{-bC})-nlg(b-a)-\sum lgC].$$

Comencemos a derivar respecto los parámetros. Respecto a será

$$\frac{\partial lgL(a,b,\gamma,\beta_1,\beta_2)}{\partial a}=$$

$$=I(1)[\frac{n}{b-a}+\sum\frac{e^{-aC}(-C)(2C^2a+2C+C^3a^2)+e^{-aC}(2C^2+2C^3a)}{e^{-aC}(2C^2a+2C+C^3a^2)+e^{-bC}(2C^2+2C^3a)}]+$$

$$+I(2)[\frac{n}{b-a}+\sum\frac{Ce^{-aC}+(aC+1)e^{-aC}(-a)}{(aC+1)e^{-aC}-(bC+1)e^{-bC}}]+$$

$$+I(3)[\frac{n}{b-a}+\sum\frac{Ce^{-aC}+(aC+1)e^{-aC}(-a)}{(aC+1)e^{-aC}-(bC+1)e^{-bC}}]+$$

$$+I(4)[\sum\frac{-Ce^{-aC}}{e^{-aC}-e^{-bC}}+\frac{n}{b-a}].$$

La derivada respecto el parámetro b es

$$\frac{\partial lgL(a,b,\gamma,\beta_1,\beta_2)}{\partial b}=$$

$$=I(1)[-\frac{n}{b-a}+\sum\frac{e^{-bC}(-C)(-2C^2b-2C-C^3b^2)+e^{-bC}(-2C^2-2C^3b)}{e^{-aC}(2C^2a+2C+C^3a^2)+e^{-bC}(-2C^2b-2C-C^3b^2)}]+$$

$$+I(2)[-\frac{n}{b-a}+\sum\frac{-Ce^{-bC}-(bC+1)e^{-bC}(-C)}{(aC+1)e^{-aC}-(bC+1)e^{-bC}}]+$$

$$+I(3)\left[-\frac{n}{b-a}+\sum\frac{-Ce^{-bC}-(bC+1)e^{-bC}(-C)}{(aC+1)e^{-aC}-(bC+1)e^{-bC}}\right]+$$

$$+I(4)\left[\sum\frac{Ce^{-bC}}{e^{-aC}-e^{-bC}}-\frac{n}{b-a}\right].$$

La derivada respecto el parámetro de forma de la distibución Weibull es

$$\frac{\partial lgL(a,b,\gamma,\beta_1,\beta_2)}{\partial\gamma}=I(1)\Big[\frac{2n}{\gamma}+\sum lgt_1+\sum lgt_2-4\sum\frac{D}{C}+$$

$$+\sum[e^{-aC}(-aD)(2C^2a+2C+C^3a^2)+e^{-aC}(4aCD+2D+3C^2a^2D)+$$

$$+e^{-bC}(-bD)(-2C^2b-2C-C^3b^2)+e^{-bC}(-4CbD-2D-3C^2b^2D)]/$$

$$/[e^{-aC}(2C^2a+2C+C^3a^2)+e^{-bC}(-2C^2b-2C-C^3b^2)]\Big]+$$

$$+I(2)\left[\frac{n}{\gamma}+\sum lgt_1-2\sum\frac{D}{C}+\right.$$

$$\left.+\sum\frac{aDe^{-aC}+(aC+1)e^{-aC}(-aD)-(bD)e^{-bC}-(bC+1)e^{-bC}(-bD)}{(aC+1)e^{-aC}-(bC+1)e^{-bC}}\right]+$$

$$+I(3)\left[\frac{n}{\gamma}+\sum lgt_2-2\sum\frac{D}{C}+\right.$$

$$\left.+\sum\frac{aDe^{-aC}+(aC+1)e^{-aC}(-aD)-(bD)e^{-bC}-(bC+1)e^{-bC}(-bD)}{(aC+1)e^{-aC}-(bC+1)e^{-bC}}\right]+$$

$$+I(4)\left[\frac{-aDe^{-aC}+bDe^{-bC}}{e^{-aC}-e^{-bC}}-\sum\frac{D}{C}\right].$$

La derivada respecto la componente del vector de parámetros de regresión es

$$\frac{\partial lgL(a,b,\gamma,\beta_1,\beta_2)}{\partial\beta_{1k}}=I(1)\Big[\sum x_{1k}-4\sum\frac{E}{C}+$$

$$+\sum[[e^{-aC}(-aE)(2C^2a+2C+C^3a^2)+e^{-aC}(4CaD+2D+3C^2Da^2)+$$

$$+e^{-bC}(-bD)(-2C^2b-2C-C^3b^2)+e^{-bC}(-4CbD-2D-3C^2Db^2)]/$$

$$/[e^{-aC}(2C^2a+2C+C^3a^2)+e^{-bC}(-2C^2b-2C-C^3b^2)]]\Big]+$$

$$+I(2)[\sum x_{1k}-2\sum\frac{E}{C}+$$

$$+\sum\frac{aEe^{-aC}+(aC+1)e^{-aC}(-aE)-(bE)e^{-bC}-(bC+1)e^{-bC}(-Eb)}{(aC+1)e^{-aC}-(bC+1)e^{-bC}}]+$$

$$+I(3)[-2\sum\frac{E}{C}+$$

$$+\sum\frac{aEe^{-aC}+(aC+1)e^{-aC}(-aE)-(bE)e^{-bC}-(bC+1)e^{-bC}(-Eb)}{(aC+1)e^{-aC}-(bC+1)e^{-bC}}]+$$

$$+I(4)[\sum\frac{e^{-aC}(-aE)-e^{-bC}(-bE)}{e^{-aC}-e^{-bC}}-\sum\frac{E}{C}]$$

con $k:1,\ldots,k_1$, e igualmente $\frac{\partial lgL(a,b,\gamma,\beta_1,\beta_2)}{\partial\beta_{2k}}$, $k:1,\ldots,k_2$.

La solución del sistema de ecuaciones

$$\left.\begin{array}{l}\frac{\partial lgL(a,b,\gamma,\beta_1,\beta_2)}{\partial a}=0\\ \frac{\partial lgL(a,b,\gamma,\beta_1,\beta_2)}{\partial b}=0\\ \frac{\partial lgL(a,b,\gamma,\beta_1,\beta_2)}{\partial\gamma}=0\\ \frac{\partial lgL(a,b,\gamma,\beta_1,\beta_2)}{\partial\beta_j}=0\\ j:1,2\end{array}\right\}$$

proporciona los estimadores de máxima verosimilitud

$$\hat{a}, \hat{b}, \hat{\gamma}, \hat{\beta}_j.$$

La obtención de los estimadores ha de hacerse por métodos numéricos usando, por ejemplo, el método de Newton Raphson.

Para que el valor de estos estimadores corresponda efectivamente a un máximo tendría que verificarse que la matríz hessiana fuera definida negativa, o lo que es igual, el valor negativo de la matriz hessiana definida positiva. Esta matriz es

$$\left(-E\frac{\partial^2 lgL(\theta)}{\partial\theta_i\partial\theta_j}\right)=\left(E\left(\frac{\partial lgL(\theta)}{\partial\theta_i}\frac{\partial lgL(\theta)}{\partial\theta_j}\right)\right)=I(\theta) \qquad \text{donde} \qquad I(\theta)$$

es la matriz de información de Fisher y

$$\theta=(a,b,\gamma,\beta_{11},\ldots,\beta_{21},\ldots).$$

Un resultado importante y abierto en la actualidad es que estos estimadores de máxima verosimilitud verifiquen que

$$\hat{\theta}\rightarrow_{(n\rightarrow\infty)}N(\theta,(nI(\theta))^{-1})$$

es decir, sean asintoticamente centrados, normales y eficientes, es decir

$$Cov(\hat{\theta})=(nI(\theta))^{-1}$$

expresión que da un valor aproximado de la varianza de los estimadores.

Las distribuciones exactas de estos estimadores y sus correspondientes medias y varianzas son por lo general bastante difíciles de conseguir.

Para algunos casos particulares se pueden conseguir expresiones más sencillas de la matríz de información.

4.4 ESTIMACIÓN EN UN MODELO MIXTO CON FRAILTY ESTABLE POSITIVA

Veamos uno de los modelos paramétricos mas comunes. Se trata de un modelo (de azar proporcional) frailty estable positiva, con marginales Weibull, sujeto a censuras y en presencia de covariables que se introducen bajo hipótesis de vida acelerada.

El modelo con un frailty con distribución estable positiva, marginales Weibull y sin covariables es el siguiente

$$S(t_1,t_2)=\exp[-(\epsilon t_1^{\gamma}+\epsilon t_2^{\gamma})^{\alpha}] \ , \quad \alpha\in(0,1].$$

α, el parámetro de asociación, proviene de la distribución del frailty. Si es igual a uno corresponde al caso de variables independientes. Los demás parámetros corresponden a la distribución de Weibull.

En presencia de covariables el modelo queda así

$$S(t_1,t_2)=\exp[-(\epsilon\exp(\gamma\beta_1 x_1)\, t_1^{\gamma}+\epsilon\exp(\gamma\beta_2 x_2)\, t_2^{\gamma})^{\alpha}]$$

siendo β_j un vector de parámetros asociado al vector de covariables x_j (j:1,2) .

Como se ve hemos considerado que las dos distribuciones Weibull correspondientes a las distribuciones marginales de los tiempos de fallo tienen los mismos parámetros. Esto no resta generalidad al modelo.

La expresión desarrollada de la función paramétrica de las covariables es

$$\exp(\beta_j x_j)=\exp(\sum_{k=1}^{k_j}\beta_{jk}x_{jk})=\exp(\beta_{j1}x_{j1}+\ldots+\beta_{jk_j}x_{jk_j}) \ , \quad j:1,2$$

donde la covariable x_{jk} (k:1,...,k_1 ó k:1,...,k_2) en el término i (i:1,...,n) tomará el valor

$$x_{jki}.$$

Los parámetros a estimar son

$$\epsilon\ ,\ \gamma\ ,\ \alpha\ ,\ \beta_1=(\beta_{11},\dots,\beta_{1k_1})\ ,\ \beta_2=(\beta_{21},\dots,\beta_{2k_2})$$

es decir, hay $1+1+1+k_1+k_2$ parámetros a estimar.

Consideremos una muestra

$$(t_{1i},t_{2i}),\quad i:1,\dots,n.$$

Dicha muestra estará formada por un vector de datos bivariantes donde existe posibilidad de censuras, es decir, una componente, las dos o ninguna podrán corresponder a datos censurados.

En estas condiciones la función de verosimilitud es

$$L(\alpha,\gamma,\epsilon,\beta)=\prod_{i=1}^{n}\left[\left(\frac{\partial^2 S(t_{1i},t_{2i})}{\partial t_{1i}\partial t_{2i}}\right)^{I(1)}\cdot\left(\frac{-\partial S(t_{1i},t_{2i})}{\partial t_{1i}}\right)^{I(2)}\cdot\right.$$
$$\left.\cdot\left(\frac{-\partial S(t_{1i},t_{2i})}{\partial t_{2i}}\right)^{I(3)}\cdot\left(S(t_{1i},t_{2i})\right)^{I(4)}\right]$$

y su logaritmo será

$$lgL(\alpha,\gamma,\epsilon,\beta)=\sum_{i=1}^{n}lg[(g1)^{I(1)}\cdot(g2)^{I(2)}\cdot(g3)^{I(3)}\cdot(g4)^{I(4)}]=$$

$$=\sum_{i=1}^{n}[I(1)lg(g1)+I(2)lg(g2)+I(3)lg(g3)+I(4)lg(g4)]=$$

$$=I(1)\sum_{i=1}^{n}lg(g1)+I(2)\sum_{i=1}^{n}lg(g2)+I(3)\sum_{i=1}^{n}lg(g3)+I(4)\sum_{i=1}^{n}lg(g4)$$

entendiendo por g1, g2, g3 y g4 aquellas derivadas parciales definidas anteriormente e i afectando a los tiempos de fallo y a las covariables.

Si derivamos resulta que

$$\frac{\partial S(t_1,t_2)}{\partial t_1}=-\alpha\epsilon\gamma t_1^{\gamma-1}\exp(\gamma\beta_1x_1)[\epsilon\exp(\gamma\beta_1x_1)t_1^{\gamma}+\epsilon\exp(\gamma\beta_2x_2)t_2^{\gamma}]^{\alpha-1}.$$

$$.\exp[-(\epsilon\exp(\gamma\beta_1x_1)t_1^{\gamma}+\epsilon\exp(\gamma\beta_2x_2)t_2^{\gamma})^{\alpha}],$$

$$\frac{\partial S(t_1,t_2)}{\partial t_2}=-\alpha\epsilon\gamma t_2^{\gamma-1}\exp(\gamma\beta_2x_2)[\epsilon\exp(\gamma\beta_1x_1)t_1^{\gamma}+\epsilon\exp(\gamma\beta_2x_2)t_2^{\gamma}]^{\alpha-1}.$$

$$.\exp[-(\epsilon\exp(\gamma\beta_1x_1)t_1^{\gamma}+\epsilon\exp(\gamma\beta_2x_2)t_2^{\gamma})^{\alpha}]$$

y

$$\frac{\partial^2 S(t_1,t_2)}{\partial t_1\partial t_2}=\alpha\epsilon^2\gamma^2 t_1^{\gamma-1}t_2^{\gamma-1}\exp(\gamma\beta_1x_1)\exp(\gamma\beta_2x_2).$$

$$.\exp[-(\epsilon\exp(\gamma\beta_1x_1)t_1^{\gamma}+\epsilon\exp(\gamma\beta_2x_2)t_2^{\gamma})^{\alpha}].$$

$$.[\alpha(\epsilon\exp(\gamma\beta_1x_1)t_1^{\gamma}+\epsilon\exp(\gamma\beta_2x_2)t_2^{\gamma})^{2\alpha-2}-$$

$$-(\alpha-1)(\epsilon\exp(\gamma\beta_1x_1)t_1^{\gamma}+\epsilon\exp(\gamma\beta_2x_2)t_2^{\gamma})^{\alpha-2}]$$

y donde es fácil observar que

$$\frac{-\partial S(t_1,t_2)}{\partial t_j}>0\ ,\quad j:1,2$$

$$\frac{\partial^2 S(t_1,t_2)}{\partial t_1\partial t_2}>0$$

ya que

$$\alpha(\epsilon\exp(\gamma\beta_1 x_1)t_1^{\gamma}+\epsilon\exp(\gamma\beta_2 x_2)t_2^{\gamma})^{2\alpha-2}-$$

$$-(\alpha-1)(\epsilon\exp(\gamma\beta_1 x_1)t_1^{\gamma}+\epsilon\exp(\gamma\beta_2 x_2)t_2^{\gamma})^{\alpha-2}\geq 0$$

necesario para poder tomar logaritmos.

Por consiguiente la log-verosimilitud es

$$lgL(\alpha,\gamma,\epsilon,\beta)=$$

$$=I(1)\sum_{i=1}^{n}[lg\alpha+2lg\epsilon+2lg\gamma+(\gamma-1)lg(t_1 t_2)+(\gamma\beta_1 x_1+\gamma\beta_2 x_2)-$$

$$-(\epsilon\exp(\gamma\beta_1 x_1)t_1^{\gamma}+\epsilon\exp(\gamma\beta_2 x_2)t_2^{\gamma})^{\alpha}+$$

$$+lg(\alpha(\epsilon\exp(\gamma\beta_1 x_1)t_1^{\gamma}+\epsilon\exp(\gamma\beta_2 x_2)t_2^{\gamma})^{2\alpha-2}-$$

$$-(\alpha-1)(\epsilon\exp(\gamma\beta_1 x_1)t_1^{\gamma}+\epsilon\exp(\gamma\beta_2 x_2)t_2^{\gamma})^{\alpha-2})]+$$

$$+I(2)\sum_{i=1}^{n}[lg\alpha+lg\epsilon+lg\gamma+(\gamma-1)lgt_1+\gamma\beta_1 x_1+$$

$$+(\alpha-1)lg(\epsilon\exp(\gamma\beta_1 x_1)t_1^{\gamma}+\epsilon\exp(\gamma\beta_2 x_2)t_2^{\gamma})-$$

$$-(\epsilon\exp(\gamma\beta_1 x_1)t_1^{\gamma}+\epsilon\exp(\gamma\beta_2 x_2)t_2^{\gamma})^{\alpha}]+$$

$$+I(3)\sum_{i=1}^{n}[lg\alpha+lg\epsilon+lg\gamma+(\gamma-1)lgt_2+\gamma\beta_2 x_2+$$

$$+(\alpha-1)lg(\epsilon\exp(\gamma\beta_1 x_1)t_1^{\gamma}+\epsilon\exp(\gamma\beta_2 x_2)t_2^{\gamma})-$$

$$-(\epsilon\exp(\gamma\beta_1 x_1)t_1^{\gamma}+\epsilon\exp(\gamma\beta_2 x_2)t_2^{\gamma})^{\alpha}]-$$

$$-I(4)\sum_{i=1}^{n}(\epsilon\exp(\gamma\beta_1 x_1)t_1^{\gamma}+\epsilon\exp(\gamma\beta_2 x_2)t_2^{\gamma})^{\alpha}=$$

$$=I(1)[nlg\alpha+2nlg\epsilon+2nlg\gamma+(\gamma-1)\sum lg(t_1t_2)+\sum(\gamma\beta_1x_1+\gamma\beta_2x_2)-$$

$$-\sum(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)^\alpha+$$

$$+\sum lg(\alpha(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)^{2\alpha-2}-$$

$$-(\alpha-1)(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)^{\alpha-2})]+$$

$$+I(2)[nlg\alpha+nlg\epsilon+nlg\gamma+(\gamma-1)\sum lgt_1+\sum\gamma\beta_1x_1+$$

$$+(\alpha-1)\sum lg(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)-$$

$$-\sum(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)^\alpha]+$$

$$+I(3)[nlg\alpha+nlg\epsilon+nlg\gamma+(\gamma-1)\sum lgt_2+\sum\gamma\beta_2x_2+$$

$$+(\alpha-1)\sum lg(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)-$$

$$-\sum(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)^\alpha]-$$

$$-I(4)\sum(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)^\alpha.$$

Hemos de derivar respecto los parámetros a estimar. Respecto a α será

$$\frac{\partial lgL(\alpha,\gamma,\epsilon,\beta)}{\partial\alpha}=$$

$$=I(1)\Big[\frac{n}{\alpha}-\sum[(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)^\alpha.$$

$$.lg(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)]+$$

$$+\sum\{(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)^{2\alpha-2}+$$

$$+\alpha(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)^{2\alpha-2}lg(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)-$$

$$-(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)^{\alpha-2}-$$

$$-(\alpha-1)(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)^{\alpha-2}.$$

$$.lg(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)\}\ /\ \{\alpha+$$

$$+(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)^{2\alpha-2}-$$

$$-(\alpha-1)(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)^{\alpha-2}\}\Big]+$$

$$+I(2)\left[\frac{n}{\alpha}+\sum lg(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)-\right.$$

$$\left.-\sum(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)^\alpha lg(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)\right]+$$

$$+I(3)\left[\frac{n}{\alpha}+\sum lg(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)-\right.$$

$$\left.-\sum(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)^\alpha lg(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)\right]+$$

$$I(4)\sum(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)^\alpha lg(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)$$

Respecto a gamma será

$$\frac{\partial lgL(\alpha,\gamma,\epsilon,\beta)}{\partial\gamma}=$$

$$I(1)\left[\frac{2n}{\gamma}+\sum lgt_1t_2+\sum(\beta_1x_1+\beta_2x_2)-\sum\alpha(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)^{\alpha-1}\right.$$

$$\cdot(\epsilon\exp(\gamma\beta_1x_1)(\beta_1x_1)t_1^\gamma(1+lgt_1)+\epsilon\exp(\gamma\beta_2x_2)(\beta_2x_2)t_2^\gamma(1+lgt_2))+$$

$$+\sum\{\alpha(2\alpha-2)(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)^{2\alpha-3}\cdot$$

$$\cdot(\epsilon\exp(\gamma\beta_1x_1)(\beta_1x_1)t_1^\gamma(1+lgt_1)+\epsilon\exp(\gamma\beta_2x_2)(\beta_2x_2)t_2^\gamma(1+lgt_2))-$$

$$-(\alpha-1)(\alpha-2)(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)^{\alpha-3}\cdot$$

$$\cdot(\epsilon\exp(\gamma\beta_1x_1)(\beta_1x_1)t_1^\gamma(1+lgt_1)+\epsilon\exp(\gamma\beta_2x_2)(\beta_2x_2)t_2^\gamma(1+lgt_2))\}\ /\ \{\alpha\cdot$$

$$\cdot(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)^{2\alpha-2}-$$

$$\left.-(\alpha-1)(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)^{\alpha-2}\}\right]+$$

$$+I(2)\left[\frac{n}{\gamma}+\sum lgt_1+\sum \beta_1 x_1+\right.$$

$$+(\alpha-1)\sum\frac{(\epsilon\exp(\gamma\beta_1 x_1)(\beta_1 x_1)t_1^{\gamma}(1+lgt_1)+\epsilon\exp(\gamma\beta_2 x_2)(\beta_2 x_2)t_2^{\gamma}(1+lgt_2))}{(\epsilon\exp(\gamma\beta_1 x_1)t_1^{\gamma}+\epsilon\exp(\gamma\beta_2 x_2)t_2^{\gamma})}-$$

$$-\sum\alpha(\epsilon\exp(\gamma\beta_1 x_1)t_1^{\gamma}+\epsilon\exp(\gamma\beta_2 x_2)t_2^{\gamma})^{\alpha-1}\cdot$$

$$\left.\cdot(\epsilon\exp(\gamma\beta_1 x_1)(\beta_1 x_1)t_1^{\gamma}(1+lgt_1)+\epsilon\exp(\gamma\beta_2 x_2)(\beta_2 x_2)t_2^{\gamma}(1+lgt_2))\right]+$$

$$+I(3)\left[\frac{n}{\gamma}+\sum lgt_2+\sum \beta_2 x_2+\right.$$

$$+(\alpha-1)\sum\frac{(\epsilon\exp(\gamma\beta_1 x_1)(\beta_1 x_1)t_1^{\gamma}(1+lgt_1)+\epsilon\exp(\gamma\beta_2 x_2)(\beta_2 x_2)t_2^{\gamma}(1+lgt_2))}{(\epsilon\exp(\gamma\beta_1 x_1)t_1^{\gamma}+\epsilon\exp(\gamma\beta_2 x_2)t_2^{\gamma})}-$$

$$-\sum\alpha(\epsilon\exp(\gamma\beta_1 x_1)t_1^{\gamma}+\epsilon\exp(\gamma\beta_2 x_2)t_2^{\gamma})^{\alpha-1}\cdot$$

$$\left.\cdot(\epsilon\exp(\gamma\beta_1 x_1)(\beta_1 x_1)t_1^{\gamma}(1+lgt_1)+\epsilon\exp(\gamma\beta_2 x_2)(\beta_2 x_2)t_2^{\gamma}(1+lgt_2))\right]+$$

$$-I(4)\sum\alpha(\epsilon\exp(\gamma\beta_1 x_1)t_1^{\gamma}+\epsilon\exp(\gamma\beta_2 x_2)t_2^{\gamma})^{\alpha-1}\cdot$$

$$\cdot(\epsilon\exp(\gamma\beta_1 x_1)(\beta_1 x_1)t_1^{\gamma}(1+lgt_1)+\epsilon\exp(\gamma\beta_2 x_2)(\beta_2 x_2)t_2^{\gamma}(1+lgt_2)).$$

Calculemos la derivada respecto el parámetro de escala de la distribución Weibull; llamaremos

$$a=\epsilon\exp(\gamma\beta_1 x_1)t_1^{\gamma}+\epsilon\exp(\gamma\beta_2 x_2)t_2^{\gamma}$$
$$b=\exp(\gamma\beta_1 x_1)t_1^{\gamma}+\exp(\gamma\beta_2 x_2)t_2^{\gamma}$$

entonces

$$\frac{\partial lgL(\alpha,\gamma,\epsilon,\beta)}{\partial\epsilon}=$$

$$=I(1)\left[\frac{2n}{\epsilon}-\sum\alpha a^{\alpha-1}b+\sum\frac{\alpha(2\alpha-2)a^{2\alpha-3}b-(\alpha-1)(\alpha-2)a^{\alpha-3}b}{\alpha a^{2\alpha-2}-(\alpha-1)a^{\alpha-2}}\right]+$$

$$+I(2)\left[\frac{n}{\epsilon}+(\alpha-1)\sum\frac{b}{a}-\sum\alpha a^{\alpha-1}b\right]+$$

$$+I(3)\left[\frac{n}{\epsilon}+(\alpha-1)\sum\frac{b}{a}-\sum\alpha a^{\alpha-1}b\right]-$$

$$-I(4)\sum\alpha a^{\alpha-1}b.$$

Y, por último, respecto a beta será

$$\frac{\partial lgL(\alpha,\gamma,\epsilon,\beta)}{\partial\beta_{1k}}=$$

$$=I(1)\Big[\sum\gamma x_{1k}-\sum\alpha(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)^{\alpha-1}\epsilon t_1^\gamma\exp(\gamma\beta_1x_1)\gamma x_{1k}+$$

$$+\sum\{\alpha(2\alpha-2)(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)^{2\alpha-3}\epsilon t_1^\gamma\exp(\gamma\beta_1x_1)\gamma x_{1k}-$$

$$-(\alpha-1)(\alpha-2)(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)^{\alpha-3}\epsilon t_1^\gamma\exp(\gamma\beta_1x_1)\gamma x_{1k}\}\;/\;\{\alpha\cdot$$

$$\cdot(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)^{2\alpha-2}-(\alpha-1)\cdot$$

$$\cdot(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)^{\alpha-2}\}\Big]+$$

$$+I(2)\Big[\sum\gamma x_{1k}+(\alpha-1)\sum\frac{\epsilon t_1^\gamma\exp(\gamma\beta_1x_1)\gamma x_{1k}}{(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)}-$$

$$-\sum\alpha(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)^{\alpha-1}\epsilon t_1^\gamma\exp(\gamma\beta_1x_1)\gamma x_{1k}\Big]+$$

$$+I(3)\Big[(\alpha-1)\sum\frac{\epsilon t_1^\gamma\exp(\gamma\beta_1x_1)\gamma x_{1k}}{\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma}-$$

$$-\sum\alpha(\epsilon\exp(\gamma\beta_1x_1)t_1^\gamma+\epsilon\exp(\gamma\beta_2x_2)t_2^\gamma)^{\alpha-1}(\epsilon t_1^\gamma\exp(\gamma\beta_1x_1)\gamma x_{1k})\Big]-$$

$$-I(4)\sum \alpha(\epsilon\exp(\gamma\beta_1 x_1)\, t_1^{\gamma}+\epsilon\exp(\gamma\beta_2 x_2)\, t_2^{\gamma})^{\alpha-1}\epsilon\, t_1^{\gamma}\exp(\gamma\beta_1 x_1)\gamma x_{1k}$$

con

$$k: 1,\ldots,k_1$$

e igualmente

$$\frac{\partial lgL(\alpha,\gamma,\beta)}{\partial\beta_{2k}} \qquad k: 1,\ldots,k_2.$$

La solución del sistema de ecuaciones

$$\left.\begin{array}{l} \frac{\partial lgL(\alpha,\gamma,\epsilon,\beta)}{\partial\alpha}=0 \\ \frac{\partial lgL(\alpha,\gamma,\epsilon,\beta)}{\partial\gamma}=0 \\ \frac{\partial lgL(\alpha,\gamma,\epsilon,\beta)}{\partial\epsilon}=0 \\ \frac{\partial lgL(\alpha,\gamma,\epsilon,\beta)}{\partial\beta_{jk}}=0 \end{array}\right\}$$

proporciona los estimadores de máxima verosimilitud

$$\hat{\alpha},\hat{\gamma},\hat{\epsilon},\hat{\beta}_k.$$

La obtención de los estimadores ha de hacerse por métodos numéricos usando, por ejemplo, el método de Newton Raphson.

Para que el valor de estos estimadores corresponda efectivamente a un máximo tendría que verificarse que la matríz hessiana fuera definida negativa, o lo que es igual, el valor negativo de la matriz hessiana definida positiva. Esta matriz es

$$\left(-E\frac{\partial^2 lgL(\theta)}{\partial\theta_i\partial\theta_j}\right)=\left(E\left(\frac{\partial lgL(\theta)}{\partial\theta_i}\frac{\partial lgL(\theta)}{\partial\theta_j}\right)\right)=I(\theta) \qquad \text{donde} \qquad I(\theta)$$

es la matriz de información de Fisher y

$$\theta=(\alpha,\gamma,\epsilon,\beta_{11},\ldots,\beta_{2k_2}).$$

Un resultado importante y abierto en la actualidad es que estos estimadores de máxima verosimilitud verifiquen que

$$\hat{\theta} \to_{(n\to\infty)} N(\theta, (nI(\theta))^{-1})$$

es decir, sean asintoticamente centrados, normales y eficientes, es decir

$$Cov(\hat{\theta}) = (nI(\theta))^{-1}$$

expresión que da un valor aproximado de la varianza de los estimadores.

Las distribuciones exactas de estos estimadores y sus correspondientes medias y varianzas son por lo general bastante difíciles de conseguir.

Para algunos casos particulares se pueden conseguir expresiones más sencillas de la matríz de información. Así por ejemplo para el caso de independencia, alfa igual a uno.

4.5 ESTIMACIÓN EN UN MODELO MIXTO CON FRAILTY GAMMA Y COVARIABLES TAMBIÉN EN EL PARÁMETRO DE ASOCIACIÓN

Veamos a continuación el siguiente modelo paramétrico. Se trata de un modelo con un frailty (azar proporcional) con distribución gamma, marginales con distribución Weibull, sujeto a censuras y en presencia de covariables que actúan según hipótesis de vida acelerada. Es decir, se trata de un modelo mixto. Además, covariables sobre el parámetro de asociación.

El modelo, sin concretar las marginales y sin covariables, tiene la siguiente función de supervivencia

$$S(t_1, t_2) = \left[S_1(t_1)^{1-\theta} + S_2(t_2)^{1-\theta} - 1\right]^{\frac{1}{1-\theta}} \quad , \quad \theta \geq 1 .$$

Casos especiales son los siguientes:

$$\theta=1 \qquad \text{independencia}$$

$$\theta\to\infty \qquad S(t_1,t_2)=\text{mín}(S_1(t_1),S_2(t_2)).$$

Que haya covariables dentro del parámetro de asociación puede ser debido a que haya covariables que incidan o tengan cierta influencia en la correlación. Es de observar que la media del frailty gamma es uno y la varianza $\exp(c+\gamma x_i)$, según la formulación que exponemos a continuación. Así, las covariables explican la asociación. Dado que en la expresión anterior del modelo aparece el factor

$$\theta-1$$

expresaremos

$$\theta_i-1=\exp(c+\gamma y_i)$$

$$lg(\theta_i-1)=c+\gamma y_i=\alpha.$$

Las funciones de supervivencia marginales son

$$S_1(t_1)=\exp[-(\epsilon\exp(\eta_1\beta_1x_1))t_1^{\eta_1}]$$

$$S_2(t_2)=\exp[-(\epsilon\exp(\eta_2\beta_2x_2))t_2^{\eta_2}]$$

y la función de supervivencia conjunta será

$$S(t_1,t_2)=$$

$$=[\exp((\epsilon\exp(\eta_1\beta_1x_1))t_1^{\eta_1}\exp(c+\gamma y))+$$

$$+\exp((\epsilon\exp(\eta_2\beta_2x_2))t_2^{\eta_2}\exp(c+\gamma y))-1]^{\frac{1}{-\exp(c+\gamma y)}}=$$

$$[\exp(\epsilon\exp(\eta_1\beta_1x_1+\alpha)t_1^{\eta_1})+\exp(\epsilon\exp(\eta_2\beta_2x_2+\alpha)t_2^{\eta_2})-1]^{-\frac{1}{\exp\alpha}}=$$

$$=A^{\frac{1}{-\exp(c+\gamma y)}}.$$

Consideramos una muestra

$$(t_{1i},t_{2i}),\quad i:1,\ldots,n$$

donde además dispondremos de los valores correspondientes de las covariables. Los parámetros a estimar son los siguientes

$$\eta_1,\ \eta_2,\ \epsilon\ ,\ \beta_1=(\beta_{11},\dots,\beta_{1J}),\ \beta_2=(\beta_{21},\dots,\beta_{2K}),\ c,\ \gamma=(\gamma_1,\dots,\gamma_M)$$

es decir 1+1+1+J+K+1+M parámetros, y habremos de calcular

$$\frac{\partial lgL}{\partial\eta_1},\ \frac{\partial lgL}{\partial\eta_2},\ \frac{\partial lgL}{\partial\epsilon}$$

$$\frac{\partial lgL}{\partial\beta_{1j}},(j:1,\dots,J),\ \frac{\partial lgL}{\partial\beta_{2k}},(k:1,\dots,K),$$

$$\frac{\partial lgL}{\partial c},\ \frac{\partial lgL}{\partial\gamma_m},(m:1,\dots,M).$$

Comencemos calculando las siguientes derivadas:

$$\frac{\partial S(t_1,t_2)}{\partial t_1}=$$

$$=-\frac{1}{\exp(\alpha)}A^{-\frac{1}{\exp(\alpha)}-1}\exp(\epsilon t_1^{\eta_1}\exp(\eta_1\beta_1x_1+\alpha))\exp(\eta_1\beta_1x_1+\alpha)\epsilon\eta_1t_1^{\eta_1-1}\ ,$$

$$\frac{\partial S(t_1,t_2)}{\partial t_2}=$$

$$=-\frac{1}{\exp(\alpha)}A^{-\frac{1}{\exp(\alpha)}-1}\exp(\epsilon t_2^{\eta_2}\exp(\eta_2\beta_2x_2+\alpha))\exp(\eta_2\beta_2x_2+\alpha)\epsilon\eta_2t_2^{\eta_2-1}$$

y

$$\frac{\partial^2 S(t_1,t_2)}{\partial t_1 \partial t_2}=-\frac{1}{\exp(\alpha)}\exp(\epsilon t_1^{\eta_1}\exp(\eta_1\beta_1 x_1+\alpha))\exp(\eta_1\beta_1 x_1+\alpha)\,.$$

$$.\epsilon\eta_1 t_1^{\eta_1-1}\exp(\epsilon t_2^{\eta_2}\exp(\eta_2\beta_2 x_2+\alpha))\exp(\eta_2\beta_2 x_2+\alpha)\,.$$

$$.\epsilon\eta_2 t_2^{\eta_2-1}\left(-\frac{1}{\exp(\alpha)}-1\right)A^{-\frac{1}{\exp(\alpha)}-2}.$$

Luego la log-verosimilitud es

$$lgL=\sum_{i=1}^{n}\left[I(1)\,lg\frac{\partial^2 S(t_{1i},t_{2i})}{\partial t_{1i}\partial t_{2i}}+I(2)\,lg\frac{-\partial S(t_{1i},t_{2i})}{\partial t_{1i}}+\right.$$

$$\left.+I(3)\,lg\frac{-\partial S(t_{1i},t_{2i})}{\partial t_{2i}}+I(4)\,lgS(t_{1i},t_{2i})\right]=$$

$$=I(1)\left[\sum lg(1+\exp(\alpha))-2\sum\alpha+\epsilon\sum t_1^{\eta_1}\exp(\eta_1\beta_1 x_1+\alpha)+\right.$$

$$+\sum(\eta_1\beta_1 x_1+\alpha)+n\,lg\epsilon+n\,lg\eta_1+(\eta_1-1)\sum lgt_1+$$

$$+\epsilon\sum t_2^{\eta_2}\exp(\eta_2\beta_2 x_2+\alpha)+\sum(\eta_2\beta_2 x_2+\alpha)+n\,lg\epsilon+n\,lg\eta_2+$$

$$\left.+(\eta_2-1)\sum\ln t_2-\sum\left(\frac{1}{\exp(\alpha)}+2\right)lgA\right]+$$

$$+I(2)\left[-\sum\alpha-\sum\left(\frac{1}{\exp(\alpha)}+1\right)lgA+\epsilon\sum t_1^{\eta_1}\exp(\eta_1\beta_1 x_1+\alpha)+\right.$$

$$\left.+\sum(\eta_1\beta_1 x_1+\alpha)+n\,lg\epsilon+n\,lg\eta_1+(\eta_1-1)\sum lgt_1\right]+$$

$$+I(3)\left[-\sum\alpha-\sum\left(\frac{1}{\exp(\alpha)}+1\right)lgA+\epsilon\sum t_2^{\eta_2}\exp(\eta_2\beta_2 x_2+\alpha)+\right.$$

$$\left.+\sum(\eta_2\beta_2 x_2+\alpha)+n\,lg\epsilon+n\,lg\eta_2+(\eta_2-1)\sum lgt_2\right]-$$

$$-I(4)\sum\frac{1}{\exp(\alpha)}lgA.$$

Calculemos las derivadas respecto los parámetros: llamemos

$$a=\exp[\epsilon\exp(\eta_1\beta_1x_1+\alpha)\,t_1^{\eta_1}]\,.$$

$$\cdot[\epsilon\exp(\eta_1\beta_1x_1+\alpha)\,t_1^{\eta_1}\,lgt_1+\epsilon\,t_1^{\eta_1}\exp(\eta_1\beta_1x_1+\alpha)\beta_1x_1]$$

entonces

$$\frac{\partial lgL}{\partial\eta_1}=I(1)\Big[\epsilon\sum t_1^{\eta_1}lgt_1\exp(\eta_1\beta_1x_1+\alpha)+$$

$$+\epsilon\sum t_1^{\eta_1}\exp(\eta_1\beta_1x_1+\alpha)\beta_1x_1+\sum\beta_1x_1+\frac{n}{\eta_1}+\sum\ln t_1-$$

$$-\sum\left(\frac{1}{\exp(\alpha)}+2\right)\frac{a}{A}\Big]+$$

$$+I(2)\Big[-\sum\left(\frac{1}{\exp(\alpha)}+1\right)\frac{a}{A}+$$

$$+\epsilon\sum t_1^{\eta_1}lgt_1\exp(\eta_1\beta_1x_1+\alpha)+\epsilon\sum t_1^{\eta_1}\exp(\eta_1\beta_1x_1+\alpha)\beta_1x_1+$$

$$+\sum\beta_1x_1+\frac{n}{\eta_1}+\sum lgt_1\Big]+$$

$$+I(3)\left[-\sum\left(\frac{1}{\exp(\alpha)}+1\right)\frac{a}{A}\right]-$$

$$-I(4)\left[\sum\frac{1}{\exp(\alpha)}\frac{a}{A}\right]$$

e igualmente si llamamos

$$b=\exp[\epsilon(\eta_2\beta_2x_2+\alpha)\,t_2^{\eta_2}]\,.$$

$$\cdot[\epsilon\exp(\eta_2\beta_2x_2+\alpha)\,t_2^{\eta_2}\,lgt_2+\epsilon\,t_2^{\eta_2}\exp(\eta_2\beta_2x_2+\alpha)\beta_2x_2]$$

entonces

$$\frac{\partial lgL}{\partial \eta_2}=$$

$$=I(1)[\epsilon\sum t_2^{\eta_2}lgt_2\exp(\eta_2\beta_2x_2+\alpha)+\epsilon\sum t_2\sum \eta_2\exp(\eta_2\beta_2x_2+\alpha)\beta_2x_2+$$

$$+\sum\beta_2x_2+\frac{n}{\eta_2}+\sum lgt_2-\sum(\frac{1}{\exp\alpha}+2)\frac{b}{A}]+$$

$$+I(2)[-\sum(\frac{1}{\exp\alpha}+1)\frac{b}{A}]+$$

$$+I(3)[-\sum(\frac{1}{\exp\alpha}+1)\frac{b}{A}+$$

$$+\epsilon\sum t_2^{\eta_2}lgt_2\exp(\eta_2\beta_2x_2+\alpha)+\epsilon\sum t_2^{\eta_2}\exp(\eta_2\beta_2x_2+\alpha)\beta_2x_2]-$$

$$-I(4)[\sum\frac{1}{\exp\alpha}\frac{b}{A}].$$

Sea ahora

$$d=\exp[\epsilon\exp(\eta_1\beta_1x_1+\alpha)t_1^{\eta_1}]\exp[(\eta_1\beta_1x_1+\alpha)t_1^{\eta_1}]+$$

$$+\exp[\epsilon\exp(\eta_2\beta_2x_2+\alpha)t_2^{\eta_2}]\exp[(\eta_2\beta_2x_2+\alpha)t_2^{\eta_2}]$$

entonces

$$\frac{\partial lgL}{\partial\epsilon}=I(1)[\sum t_1^{\eta_1}\exp(\eta_1\beta_1x_1+\alpha)+\frac{n}{\epsilon}+$$

$$+\sum t_2^{\eta_2}\exp(\eta_2\beta_2x_2+\alpha)+\frac{n}{\epsilon}-\sum(\frac{1}{\exp\alpha}+2)\frac{d}{A}]+$$

$$+I(2)[-\sum(\frac{1}{\exp\alpha}+1)\frac{d}{A}+\sum t_1^{\eta_1}\exp(\eta_1\beta_1x_1+\alpha)+\frac{n}{\epsilon}]+$$

$$+I(3)[-\sum(\frac{1}{\exp\alpha}+1)\frac{d}{A}+\sum t_2^{\eta_2}\exp(\eta_2\beta_2x_2+\alpha)+\frac{n}{\epsilon}]-$$

$$-I(4)[\sum\frac{1}{\exp\alpha}\frac{d}{A}].$$

Sea ahora

$$f=\exp[\epsilon\exp(\eta_1\beta_1x_1+\alpha)t_1^{\eta_1}]\epsilon t_1^{\eta_1}\exp(\eta_1\beta_1x_1+\alpha)+$$

$$+\exp[\epsilon\exp(\eta_2\beta_2x_2+\alpha)t_2^{\eta_2}]\epsilon t_2^{\eta_2}\exp(\eta_2\beta_2x_2+\alpha)$$

entonces

$$\frac{\partial lgL}{\partial c}=I(1)[\sum\frac{\exp\alpha}{1+\exp\alpha}-2n+\epsilon\sum t_1^{\eta_1}\exp(\eta_1\beta_1x_1+\alpha)+$$

$$+n+\epsilon\sum t_2^{\eta_2}\exp(\eta_2\beta_2x_2+\alpha)+n-\sum[-\exp(-\alpha)lgA+(\frac{1}{\exp\alpha}+2)\frac{f}{A}]]+$$

$$+I(2)[-n-\sum(-\exp(-\alpha)lgA+(\frac{1}{\exp\alpha}+1)\frac{f}{A})+$$

$$+\epsilon\sum t_1^{\eta_1}\exp(\eta_1\beta_1x_1+\alpha)+n]+$$

$$+I(3)[-\sum(-\exp(-\alpha)lgA+(\frac{1}{\exp\alpha}+1)\frac{f}{A})+\epsilon\sum t_2^{\eta_2}\exp(\eta_2\beta_2x_2+\alpha)]-$$

$$-I(4)[-\exp(-\alpha)lgA+\frac{1}{\exp\alpha}\frac{f}{A}].$$

Sea ahora

$$h=\exp[\epsilon\exp(\eta_1\beta_1x_1+\alpha)t_1^{\eta_1}]\epsilon t_1^{\eta_1}\exp(\eta_1\beta_1x_1+\alpha)\eta_1x_{11}$$

entonces

$$\frac{\partial\, lgL}{\partial\beta_{11}}=$$

$$=I(1)\,[\,\epsilon\sum t_1^{\eta_1}\exp(\eta_1\beta_1 x_1+\alpha)\eta_1 x_{11}+\sum\eta_1 x_{11}-\sum(\frac{1}{\exp}\alpha)+2)\frac{h}{A}\,]+$$

$$+I(2)\,[\,-\sum(\frac{1}{\exp\alpha}+1)\frac{h}{A}+\epsilon\sum t_1^{\eta_1}\exp(\eta_1\beta_1 x_1+\alpha)\eta_1 x_{11}+\sum\eta_1 x_{11}\,]+$$

$$+I(3)\,[\,-\sum(\frac{1}{\exp\alpha}+1)\frac{h}{A}\,]\quad-\quad I(4)\,[\,\sum\frac{1}{\exp\alpha}\frac{h}{A}\,]$$

y así los demás coeficientes de β_1 y β_2 .

Por último sea

$$k=\exp[\,\epsilon\exp(\eta_1\beta_1 x_1+\alpha)\,t_1^{\eta_1}\,]\,\epsilon\, t_1^{\eta_1}\exp(\eta_1\beta_1 x_1+\alpha)\,y_1+$$

$$+\exp[\,\epsilon\exp(\eta_2\beta_2 x_2+\alpha)\,t_2^{\eta_2}\,]\,\epsilon\, t_2^{\eta_2}\exp(\eta_2\beta_2 x_2+\alpha)\,y_1$$

entonces

$$\frac{\partial\, lgL}{\partial\gamma_1}=$$

$$=I(1)\,[\,\sum\frac{y_1\exp\alpha}{1+\exp\alpha}-2\sum y_1+\epsilon\sum t_1^{\eta_1}\exp(\eta_1\beta_1 x_1+\alpha)y_1+\sum y_1+$$

$$+\epsilon\sum t_2^{\eta_2}\exp(\eta_2\beta_2 x_2+\alpha)y_1+\sum y_1-\sum(-y_1\exp(-\alpha)\,lgA+(\frac{1}{\exp\alpha}+2)\frac{k}{A}\,]+$$

$$+I(2)\,[\,-\sum((-y_1 e^{-\alpha})\,lgA+(\frac{1}{\exp\alpha}+1)\frac{k}{A})+\epsilon\sum t_1^{\eta_1}\exp(\eta_1\beta_1 x_1+\alpha)y_1\,]+$$

$$+I(3)[-\sum(-y_1 e^{-\alpha} lgA+(\frac{1}{\exp\alpha}+1)\frac{k}{A}+\epsilon\sum t_2^{\eta_2}\exp(\eta_2\beta_2 x_2+\alpha)y_1]-$$

$$-I(4)[\sum(-y_1 e^{-\alpha} lgA+\frac{k}{A}\frac{1}{e^{\alpha}})]$$

e igualmente los demás coeficientes del vector γ .

Si

$$\phi=(\eta_1,\eta_2,\epsilon,\beta_{11},\ldots,\beta_{1J},\beta_{21},\ldots,\beta_{2K},c,\gamma_1,\ldots,\gamma_M)=$$
$$=(\phi_1,\ldots,\phi_{4+J+K+M})$$

las soluciones del sistema

$$\frac{\partial lgL}{\partial \phi_i}=0\ ,\qquad i:1,\ldots,4+J+K+M$$

proporcionan los estimadores de máxima verosimilitud. Para que correspondan a un máximo habría de ser definida positiva la matriz

$$I(\phi)=\left(E(\frac{\partial L}{\partial \phi_i})^2\right)=\left(-E\frac{\partial^2 lgL}{\partial \phi_i \partial \phi_j}\right).$$

Un problema abierto es que

$$\hat{\phi}\to N(\phi,(nI(\phi))^{-1}).$$

Las soluciones de las ecuaciones de verosimilitud son bastante complejas. Necesitamos utilizar paquetes matemáticos (Mathematica...) para resolverlas.

En Oakes (1982) se utiliza un método indirecto para estimar el parámetro de asociación en el modelo de Clayton. Está basado en la estimación del coeficiente de Kendall que como vimos es función directa de dicho parámetro.

CAPÍTULO 5
TEMAS ABIERTOS (APÉNDICE)

En este último capítulo presentamos algunas cuestiones abiertas, algunas de ellas continuación natural de problemas estudiados anteriormente. Así, en el apartado 5.1 consideramos un frailty que actúa aditivamente en la función de azar marginal. En el apartado 5.2 presentamos la introducción de covariables explicativas en algunos modelos de azar no proporcional. En los apartados 5.3 y 5.4 consideramos los problemas de los datos agrupados y de la simulación, respectivamente. Por último, en 5.5 enumeramos algunos de los puntos de interés de próximo estudio.

5.1 MODELO FRAILTY ADITIVO

Nosotros hemos considerado, además, otra alternativa para la construcción de modelos. Es la siguiente: suponer que

$$h_i(t_i,z)=z+h_{i0}(t_i)\quad,\quad i:1,2$$

es decir, una forma aditiva en la función de azar marginal. En este caso la función de supervivencia marginal es

$$S_i(t_i,z)=\exp[-\int_0^{t_1}(z+h_{i0}(t_i))dt_i]=\exp[-zt_i-\Lambda_{i0}(t_i)]=$$
$$=S_{i0}(t_i).\exp(-zt_i).$$

Esta función de supervivencia marginal es una verdadera función de supervivencia ya que por ser $z,t_i \geq 0$ se cumple que

$$\frac{S_{i0}(t_i)}{\exp(zt_i)} \in [0,1]$$

y

$$t_i \to 0 \Rightarrow S_i(t_i, z) \to 1$$
$$t_i \to \infty \Rightarrow S_i(t_i, z) \to 0.$$

Obsérvese que, en este modelo, la función de azar base se alcanza para un valor del frailty igual a cero, es decir, las condiciones estandar corresponden a Z=0. En el caso multiplicativo lo era para Z=1. Además, hemos de observar que en este modelo la variable Z no ha de ser obligatoriamente no negativa ya que por hipótesis basta con que se cumpla que $Z > -h_{i0}(t_i)$. Este hecho puede sernos útil en cuanto que nos permite considerar frailties más generales. Sin embargo el tratamiento puede complicarse puesto que, como se ve a continuación, al realizar la mixtura de distribuciones no resulta una transformada de Laplace. Nosotros mantenemos la condición de ser Z no negativa.

Por consiguiente, la función de supervivencia conjunta condicionada a un valor fijo del frailty Z es

$$S(t_1, t_2/z) = S(t_1/z)\, S(t_2/z) =$$

$$= \exp[-\int_0^{t_1}(z + h_{10}(t_1))\,dt_1] . \exp[-\int_0^{t_2}(z + h_{20}(t_2))\,dt_2] =$$

$$= \exp[-\int_0^{t_1}(z + h_{10}(t_1))\,dt_1 - \int_0^{t_2}(z + h_{20}(t_2))\,dt_2] =$$

$$= \exp[-z(\int_0^{t_1}dt_1 + \int_0^{t_2}dt_2) - \int_0^{t_1}h_{10}(t_1)\,dt_1 - \int_0^{t_2}h_{20}(t_2)\,dt_2] =$$

$$= \exp[-z(t_1 + t_2) - \Lambda_{10}(t_1) - \Lambda_{20}(t_2)] =$$
$$= S_{10}(t_1)\, S_{20}(t_2) \exp(-z(t_1 + t_2)).$$

Según esto la función de supervivencia conjunta será

$$S(t_1,t_2)=\int S_{10}(t_1)S_{20}(t_2)\exp(-z(t_1+t_2))\,dF(z)=$$

$$=S_{10}(t_1)S_{20}(t_2)\int \exp(-z(t_1+t_2))\,dF(z)=S_{10}(t_1)S_{20}(t_2)\phi(u)$$

donde

$$u=t_1+t_2$$

y $\phi(u)$ es la transformada de Laplace de z. Esta nueva expresión de la función de supervivencia conjunta es significativamente distinta al caso habitual de un frailty que actúa multiplicativamente en la función de azar marginal. En este caso la función de supervivencia conjunta coincidía con la transformada de Laplace de Z

$$\phi(u)=\phi(\Lambda_{10}(t_1)+\Lambda_{20}(t_2)).$$

Ahora, en cambio, la función de supervivencia conjunta es el producto de las funciones de supervivencia marginales base por la transformada de Laplace de Z y con argumento t_1 + t_2.

En el caso en el que el frailty Z pueda modelizarse por una distribución estable positiva, es decir

$$\phi(u)=\exp(-u^{\alpha})$$

entonces la función de supervivencia conjunta será

$$S(t_1,t_2)=S_{10}(t_1)S_{20}(t_2)\exp(-(t_1+t_2)^{\alpha}).$$

Este modelo semiparamétrico podrá concretarse si conocemos la familia de distribuciones a la que pertenecen las marginales (Weibull, etc) quedándonos el correspondiente modelo paramétrico.

5.2 INTRODUCCIÓN DE COVARIABLES EN MODELOS DE AZAR NO PROPORCIONAL

En el capítulo 2 planteábamos diversas alternativas en la introducción de covariables explicativas en modelos de azar proporcional. Siguiendo con la misma filosofía exponemos a continuación algunas alternativas en cuanto a la introducción de covariables en modelos de azar no proporcional. En todos los casos que a continuación se plantean se supone que el frailty Z toma valores de manera que la función de supervivencia marginal condicionada a un valor fijo de dicho frailty está bién definida. Como regla general se supone que Z es una variable no negativa.

a) Si consideramos el caso siguiente

$$h_i(t_i,z,\psi_i(x_i))=\exp(\beta_i x_i)+zh_i(t_i)$$

la función de supervivencia conjunta condicionada quedará

$$S(t_1,t_2/z,\psi_1(x_1),\psi_2(t_2))=$$

$$=\exp[-t_1\exp(\beta_1 x_1)-z\exp\Lambda_1(t_1)-t_2\exp(\beta_2 x_2)-z\Lambda_2(t_2)]=$$

$$=\exp[-t_1\exp(\beta_1 x_1)-t_2\exp(\beta_2 x_2)-z(\Lambda_1(t_1)+\Lambda_2(t_2))]$$

y la función de supervivencia conjunta será

$$S(t_1,t_2)=\exp[-t_1\exp(\beta_1x_1)-t_2\exp(\beta_2x_2)].$$
$$.\int\exp(-z(\Lambda_1(t_1)+\Lambda_2(t_2)))dF(z)$$

y en el caso de un frailty con distribución estable positiva

$$S(t_1,t_2)=\exp[-t_1\exp(\beta_1x_1)-t_2\exp(\beta_2x_2)].\exp[-(\Lambda_1(t_1)+\Lambda_2(t_2))^{\alpha}].$$

b) Si

$$h_i(t_i,z,\psi_i(x_i)=z(\exp(\beta_ix_i)+h_i(t_i))$$

la función de supervivencia conjunta condicionada será

$$S(t_1,t_2/z,\psi_1(x_1),\psi_2(x_2))=$$
$$=\exp[-z\exp(\beta_1x_1)t_1-z\Lambda_1(t_1)-z\exp(\beta_2x_2)t_2-z\Lambda_2(t_2)]=$$

$$=\exp[-z(t_1\exp(\beta_1x_1)-t_2\exp(\beta_2x_2)-\Lambda_1(t_1)-\Lambda_2(t_2))]$$

y la función de supervivencia será

$$S(t_1,t_2)=$$

$$=\int\exp[-z(t_1\exp(\beta_1x_1)-t_2\exp(\beta_2x_2)-\Lambda_1(t_1)-\Lambda_2(t_2))]dF(z)$$

y si Z sigue una disribución estable

$$S(t_1,t_2)=\exp[-t_1\exp(\beta_1x_1)-t_2\exp(\beta_2x_2)-\Lambda_1(t_1)-\Lambda_2(t_2)].$$

c) Si

$$h_i(t_i,z,\psi_i(x_i))=z+\exp(\beta_ix_i)h_i(t_i)$$

la función de supervivencia conjunta condicionada será

$$S(t_1,t_2/z,\Psi_1(x_1),\psi_2(x_2))=$$
$$=\exp[-zt_1-\exp(\beta_1 x_1)\Lambda_1(t_1)-zt_2-\exp(\beta_2 x_2)\Lambda_2(t_2)]=$$
$$=\exp[-\exp(\beta_1 x_1)\Lambda_1(t_1)-\exp(\beta_2 x_2)\Lambda_2(t_2)-z(t_1+t_2)]$$

y la función de supervivencia conjunta será

$$S(t_1,t_2)=$$
$$=\exp[-\exp(\beta_1 x_1)\Lambda_1(t_1)-\exp(\beta_2 x_2)\Lambda_2(t_2)].\int\exp(-z(t_1+t_2))dF(z)$$

y si Z sigue una disribución estable

$$S(t_1,t_2)=$$
$$=\exp[-\exp(\beta_1 x_1)\Lambda_1(t_1)-\exp(\beta_2 x_2)\Lambda_2(t_2)].\exp(-(t_1+t_2)^{\alpha}).$$

d) Si

$$h_i(t_i,z,\psi_i(x_i))=z.\exp(\beta_i x_i)+h_i(t_i)$$

la función de supervivencia conjunta condicionada será

$$S(t_1,t_2/z,\psi_1(x_1),\psi_2(x_2))=$$
$$=\exp[-zt_1\exp(\beta_1 x_1)-\Lambda(t_1)-zt_2\exp(\beta_2 x_2)-\Lambda_2(t_2)]=$$
$$=\exp[-\Lambda_1(t_1)-\Lambda_2(t_2)-z(t_1\exp(\beta_1 x_1)+t_2\exp(\beta_2 x_2))]$$

y la función de supervivencia será

$$S(t_1,t_2)=$$
$$=\exp[-(\Lambda_1(t_1)+\Lambda_2(t_2))]\int\exp[-z(t_1\exp(\beta_1x_1)+t_2\exp(\beta_2x_2))]dF(z)$$

y si Z sigue una distribución estable

$$S(t_1,t_2)=$$
$$=\exp[-(\Lambda_1(t_1)+\Lambda_2(t_2))]\exp[-(t_1\exp(\beta_1x_1)+t_2\exp(\beta_2x_2))^\alpha].$$

e) Si

$$h_i(t_i,z,\psi_i(x_i))=\exp(\beta_ix_i).(z+h_i(t_i))$$

la función de supervivencia conjunta condicionada será

$$S(t_1,t_2/z,\psi_1(x_1),\psi_2(x_2))=$$
$$=\exp[-zt_1\exp(\beta_1x_1)-\exp(\beta_1x_1)\Lambda_1(t_1)-$$
$$-zt_2\exp(\beta_2x_2)-\exp(\beta_2x_2)\Lambda_2(t_2)]=$$

$$=\exp[-\exp(\beta_1x_1)\Lambda_1(t_1)-\exp(\beta_2x_2)\Lambda_2(t_2)-$$
$$-z(t_1\exp(\beta_1x_1)+t_2\exp(\beta_2x_2))]$$

y la función de supervivencia conjunta será

$$S(t_1,t_2)=\exp[-\exp(\beta_1x_1)\Lambda_1(t_1)-\exp(\beta_2x_2)\Lambda_2(t_2)].$$
$$.\int\exp[-z(t_1\exp(\beta_1x_1)+t_2\exp(\beta_2x_2))]dF(z)$$

y si Z sigue una distribución estable

$$S(t_1,t_2)=$$
$$=\exp[-\exp(\beta_1x_1)\Lambda_1(t_1)-\exp(\beta_2x_2)\Lambda_2(t_2)].$$
$$.\exp[-(t_1\exp(\beta_1x_1)+t_2\exp(\beta_2x_2))^\alpha].$$

f) Si

$$h_i(t_i,z,\psi_i(x_i))=z+\exp(\beta_ix_i)+h_i(t_i)$$

la función de supervivencia conjunta condicionada será

$$S(t_1,t_2/z,\psi_1(x_1),\psi_2(x_2))=$$
$$=\exp[-zt_1-t_1\exp(\beta_1x_1)-\Lambda_1(t_1)-zt_2-t_2\exp(\beta_2x_2)-\Lambda_2(t_2)]=$$

$$=\exp[-t_1\exp(\beta_1x_1)-\Lambda_1(t_1)-t_2\exp(\beta_2x_2)-\Lambda_2(t_2)-z(t_1+t_2)]$$

y la función de supervivencia será

$$S(t_1,t_2)=\exp[-t_1\exp(\beta_1x_1)-\Lambda_1(t_1)-t_2\exp(\beta_2x_2)-\Lambda_2(t_2)].$$
$$.\int\exp(-z(t_1+t_2))dF(z)$$

y si Z sigue una distribución estable

$$S(t_1,t_2)=$$
$$=\exp[-t_1\exp(\beta_1x_1)-t_2\exp(\beta_2x_2)-\Lambda_1(t_1)-\Lambda_2(t_2)]\exp(-(t_1+t_2)^\alpha).$$

5.3 DATOS AGRUPADOS

Una situación bastante frecuente en la práctica es aquella en la que los datos se presentan agrupados. Estas agrupaciones pueden considerarse como intervalos de medición. En el caso multivariante se considera que en cada distribución marginal el tiempo de fallo viene agrupado en general vía un modelo de azar proporcional y se parte de la función de azar directamente. Por consiguiente, se trata de estimar los parámetros de las covariables, esto es, un análisis de regresión. Un estudio de análisis de regresión para datos agrupados en el caso univariante puede verse en Lawless (1982) y en Prentice & Gloecker (1978). Extensiones multivariantes se encuentran en Wei, Lin & Weissfeld (1989), Lee,Wei & Amato (1992) y en Guo & Lin (1993). La situación es la siguiente: consideremos una variable bivariante (m=2) donde se observan n unidades, cada unidad es una agrupación.

$$\lambda_i(t_i, z_i) = \exp(\beta_i z_i)\lambda_{i0}(t_i) \qquad i:1,2$$
$$\beta_i = (\beta_{i1}, \ldots, \beta_{ip}).$$

Estos parámetros de regresión son desconocidos. El azar base marginal podrá ser igual en cada una de las funciones de azar marginales o no. También es posible generalizarlo al caso de ser $z_i = z_i(t_i)$, es decir, tiempo dependientes. Si j:1,...,n denomina a la agrupación, entonces

$$(y_{ij}, \delta_{ij}, z_{ij})$$

son los tiempos censurados, indicador de censura (1 si y_{ij} no es censurado y 0 si sí lo es) y vector de covariables, respectivamente. Podemos suponer que entre unidades o agrupaciones las observaciones son independientes y que condicionalmente a las covariables las censuras y los mecanismos de fallo son independientes. Si hacemos una partición del eje de tiempos en intervalos, para una unidad dada esta toma un tiempo t_k que corresponde al intervalo k.

Ejemplos de estos casos son aquellos en los que los individuos son controlados periódicamente para determinar si el fallo ha ocurrido en tal intervalo. Por ejemplo, semanalmente, mensualmente,etc. Así, sólo se tiene información de si ha ocurrido fallo en tal intervalo, pero no en qué momento preciso. Según esto el enfoque sería el siguiente. El eje de tiempos se particiona así

$$0=a_0<a_1<a_2<\ldots\ldots<a_{k+1}=\infty$$

$$I_r=[a_{r-1},a_r) \qquad r:1,\ldots,k+1$$

$$D_r=\textit{conjunto de individuos que fallan en } I_r$$

$$d_r=\textit{cardinal de } D_r.$$

Si se toma una muestra de n individuos y se observan los tiempos de fallo, en lugar de tener

$$(T_{1i},T_{2i}) \qquad i:1,\ldots,n \qquad [5.1]$$

se tendrán los valores t_k correspondientes a cada intervalo. Es decir, varios valores i toman un primer valor, otros varios toman otro valor y así sucesivamente. Luego en realidad se tendrán menos valores de los indicados en [5.1].

La verosimilitud parcial será

$$\prod_{i=1}^{n}\left(\frac{\exp(\beta_i z_{(i)})}{\sum_{j\in\Re_{(i)}}\exp(\beta_i z_i)}\right).$$

Un estudio de Sinha, Tanner y Hall (1994) utiliza el algoritmo EM de Montecarlo para buscar las estimaciones de máxima verosimilitud en el caso univariante.

5.4 SIMULACIÓN

Un estudio previo a la obtención de un conjunto de datos reales y su tratamiento puede hacerse a través de la simulación de los datos. Varias son las ideas que se han dado para realizar esto. Así en Johnson (1987) se presentan formas para generar distribuciones de probabilidad multivariantes por ordenador basadas en Johnson & Kotz. Una de ellas está basada en las distribuciones condicionadas. Se genera un valor x_1 de la distribución marginal X_1 . Se genera x_2 dado $X_1=x_1$. Se genera x_3 de la distribución de X_2 dado $X_2=x_2$. Así sucesivamente. Otro método está basado en buscar una transformación de la variable aleatoria como función de variables aleatorias independientes cada cual fáciles de simular.

Marshall & Olkin (1988) presentan un método basado en la mixtura de distribuciones de acuerdo con su forma de construir modelos de supervivencia multivariante. Si

$$H(x_1,x_2)=\int\int F_1^{\theta_1}(x_1)F_2^{\theta_2}(x_2)\,dG(\theta_1,\theta_2)$$

o bien

$$H(x_1,x_2)=\int\int K(F_1^{\theta_1}(x_1),F_2^{\theta_2}(x_2))\,dG(\theta_1,\theta_2)$$

se genera una observación

$$(\theta_1,\theta_2)$$

de G. Se genera una observación (y_1,y_2) de K. Se hace

$$x_i=F_i^{-1}(y_i^{\frac{1}{\theta_i}})\qquad i:1,2\qquad F_i(u)=\exp[-\phi_i^{-1}H_i(u)]. \qquad [5.2]$$

En este caso (x_1,x_2) es una observación de la distribución H. Las dificultades se presentan dependiendo de las formas de G y K.

Genest & Mackay (1986) dan un método para generar valores de una cópula arquimediana H(x,y). Se generan dos variables aleatorias independientes X y Z con ley uniforme sobre [0,1]. Se calcula

$$W=\phi'^{-1}(\phi'(x)/z). \qquad [5.3]$$

Se hace

$$Y=\phi^{-1}(\phi(W)-\phi(x)). \qquad [5.4]$$

Segal & Neuhaus (1993) hacen lo propio para el caso concreto de distribución de Weibull multivariante. Guo & Lin (1993) presentan un método de simulación con vistas a efectuar un análisis de regresión con datos de supervivencia multivariante agrupados.

5.5 OTROS PUNTOS DE INTERÉS

Como objeto de nuestra Linea de investigación presentamos a continuación algunos de los puntos de interés de inmediato estudio:

. Variaciones en los modelos frailty estudiados en esta memoria.
. Estudio del modelo frailty con hipótesis de vida acelerada concretando la distribución de dicho frailty.
. Estudio de problemas de identificabilidad.
. Estudio de modelos frailty función de otros frailties.
. Estudio de modelos con asociación negativa.
. Construcción de modelos a partir de ecuaciones diferenciales.
. Conexión con el análisis de regresión.
. Estudio de la matriz de información en algunos modelos concretos.
. Inferencia en modelos semiparamétricos definidos a partir de las funciones de azar.
. Puntos de cambio
. Modelos no paramétricos.
. Introducción de notación matricial.
. Análisis con datos reales.

REFERENCIAS Y BIBLIOGRAFÍA

[1] Aalen, O. (1988). Heterogeneity in survival analysis. Statistics in Medicine, Vol. 7, 1121-1137.

[2] Anderson, J.E. & Louis, T.A. (1995). Survival analysis using a scale change random effects model. Journal of American Statistical Association, 90, 430.

[3] Anderson, J.E., Louis, T.A., Holm, N.V. and Harvald, B. (1992). Time dependent association measures for bivariate survival distributions. Journal of American Statistical Association, 87, 641-650.

[4] Apostol, T.M. (1972). Análisis Matemático. Ed. Reverté, S.A.

[5] Barlow, R.E. and Proschan, F. (1977). Techniques for analyzing multivariate failure data. The theory and applications of reliability. Ed. Chris P. Tsokos. Vol. 1. 373-396.

[6] Basu, A.P. (1971). Bivariate failure rate. Journal of American Statistical Association, 66, 103-104.

[7] Basu, A.P. (1988). Multivariate exponential distributions and their applications in reliability. Handbook of statistics 7. Quality control and reliability. Edited by P.R. Krishnaiah & C.R. Rao. North-Holland. 467-476.

[8] Bjerve, S. and Doksum, K. (1990). Correlation curves:measures of association as functions of covariate values. Preprint.

[9] Block, H.W. (1973). Monotone hazard and failure rates for absolutely continuous multivariate distributions. Research Report 73-20, University of Pittsburgh.

[10] Block, H.W. (1977). A family of bivariate life distributions. The theory and applications of reliability. Ed. Chris P. Tsokos. Vol. 1. 349-371.

[11] Clayton, D.G. (1978). A model for associaton in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika 65, 141-51.

[12] Clayton, D.G. & Cuzick, J. (1985). Multivariate generalizations of the proportional hazards model. Journal of the Royal Statistical Society, A, 148, 82-117.

[13] Costigan, T.M. & Klein, J.P. (1993). Multivariate survival analysis based on frailty models. Advanced in reliability. Edited by Asit P. Basu. Department of statistics University of Missouri Columbia, Missouri. 43-58.

[14] Cox, D.R. (1972). Regression models and life-tables (with discussion). Journal of the Royal Statistical Society, B, 34, 187-220.

[15] Cox, D.R. (1975). Partial likelihood. Biometrika 62, 269-76.

[16] Cox, D.R. and Oakes, D. (1984). Analysis of survival data. Chapman and Hall. London.

[17] Crowder, M. (1985). A distributional model for repeated failure time measurements. Journal of the Royal Statistical Society, B, 47, 447-452.

[18] Crowder, M. (1989). A multivariate distribution with weibull connections. Journal of the Royal Statistical Socicty, B, 51, 93-107.

[19] Cuadras, C., Echeverría, B., Mateo, J., y Sánchez, P. (1991). Fundamentos de Estadística. P.P.U. Barcelona.

[20] Daniel, W.W. (1990). Applied nonparametric statistics. 2ª ed. The duxdury advanced series in statistics and decision sciences. PWS-Kent. Boston.

[21] Dykstra, R.L., Hewett, J.E. and Thompson, W.A. (1973). Events which are almost independent. Ann. Statist. Assoc. , 67, 822-830.

[22] Downton, F. (1970).Bivariate exponential distributions in reliability theory. Journal of the Royal Statistical Society, B, 32, 408-417.

[23] Feller, W. (1971). An introduction to probability theory and its applications,2, 2nd ed. New york. Wiley.

[24] Freund, J.E. (1961). A bivariate extension of the exponential distribution. Journal of the American Statistical Association, 56, 971-977.

[25] Friday, D.S. & Patil, G.P. (1977). A bivariate exponential model with applications to reliability and computer generation of random variables. The theory and applications of reliability. Ed. Chris P. Tsokos. Vol. 1. 527-549.

[26] Gamel, J.W. & McLean, I.W. (1994). A stable, multivariate extension of the log-normal survival model. Computers and biomedical research, 27, 148-155.

[27] García Leal, J. (1983). Aportaciones sobre los métodos de estimación en leyes estables multidimensionales y en procesos estables lineales. Tesis doctoral. Universidad de Granada.

[28] Genest, C. & Mackay, R.J. (1986a). Copules archimédiennes et familles de lois

bidimensionelles dont les marges sont données. The Canadian journal of statistics, vol 14, 2, 145-159.

[29] Genest, C. & Mackay, R.J. (1986b). The joy of copulas: bivariate distributions with uniform marginals. The american statistician, vol. 40, 4, 280-283.

[30] Gill, R.D. (1992a). Multivariate survival analysis. Theory probab. appl., vol. 37, 1, 18-31.

[31] Gill, R.D. (1992b). Multivariate survival analysis. Theory probab. appl., vol 37, 2, 284-301.

[32] Gumbel, E.J. (1960).Bivariate exponential distributions. Journal of the American Statistical Association, 56, 698-707.

[33] Guo, G. & Rodriguez, G. (1992). Estimating a multivariate proportional hazards model for clustered data using the EM algorithm, with an application to child survival in Guatemala. Journal of the American Statistical Association, 87, 420, 969-976.

[34] Guo, S.W. & Lin, D.Y. (1993). Regression analysis of multivariate grouped survival data. 632-639.

[35] Gutiérrez, R. and Gonzalez, A. (1991). Estadística multivariable. Vol. 1. Introdución al análisis multivariante.

[36] Harnett, and Murphy, (1987). Introducción al análisis estadístico. Addison Wesley. Iberoamericana.

[37] Heckman, J.J. and Singer, B. (1982). The identification problem in econometric models for duration data. Advanced in econometrics. Ed. W. Hildebrand, Cambrigde, U.K., Cambrigde University Press, pp 39-77.

[38] Hougaard, P. (1984). Life table methods for heterogeneous populations:distributions describing the heterogeneity. Biometrika, 71, 1, 75-83.

[39] Hougaard, P. (1986a). A class of multivariate failure time distributions. Biometrika, 73, 671-8.

[40] Hougaard, P. (1986b). Survival models for heterogeneous populations derived from stable distributions. Biometrika, 73, 387-96.

[41] Hougaard, P. (1987). Modelling multivariate survival. Scand. J. Statist., 14, 291-304.

[42] Hougaard, P. (1989). Fitting a multivariate failure time distribution. IEEE

Transactions on reliability, vol. 38, 4, 444-448.

[43] Hougaard, P., Harvald, B. & Holm, N.V. (1992). Measuring the similarities between the lifetimes of adult danish twins born between 1881-1930. Journal of the American Statistical Association, 87, 417, 17-24.

[44] Johnson, M.E. (1987). Multivariate statistical simulation. Wiley. New York.

[45] Johnson, N.L. & Kotz, S. (1975). A vector multivariate hazard rate. Journal of multivariate analysis, 5, 53-56.

[46] Kai Sun & Basu, A.P. (1993). Characterizations of a family of bivariate exponential distributions. Advanced in reliability. Edited by Asit P. Basu. 395-407.

[47] Kalbfleisch, J.D. and Prentice, R.L. (1980). The statistical analysis of failure time data. Wiley. New York.

[48] Karling, S. (1968). Total Positivity. Stanford University Press.

[49] Kendall, M.G. and Stuart, A. (1961). The advanced theory of statistical. Vol. 2. Hafner Press. New York.

[50] Klein, J.P. Keiding, N. and Kamby, C. (1989). Semiparametric Marshall-Olkin type models applied to the occurrence of metastases at multiple sites. Biometrics, 45, 1073-1086.

[51] Klein, J.P., Costigan, T.M. and Moeschberger, M.L. (1991). Assessment of risk factors and the stregth of dependence when event times are associated within groups. Ohio State University Technical Report.

[52] Koehler-K.J., Symanowski-J.T. (1995). Constructing Multivariate Distributions with Specific Marginal Distributions. Journal Of Multivariate Analysis, vol 55, iss 2, pp 261-282.

[53] Kung-Yee Liang, Self, S.G. & Yue-Cune Chang (1993). Modelling marginal hazard in multivariate failure time data. Journal of the Royal Statistical Society, B, 55, 2, 441-453.

[54] Lara, A.M. (1995). Aportaciones a modelos de supervivencia: distribuciones base con puntos de cambio y covariables dependientes del tiempo. Tesis doctoral. Universidad de Granada.

[55] Lawless, J.F. (1982). Statistical models and methods for lifetime data. Wiley series.

[56] Lee, S. and Klein, J.P. (1988). Bivariate models with a random environmental factor. Indian Journal of Productivity, Reliability and Quality Control, 13, 1-18.

[57] Lee, E.W., Wei, L.J. and Amato, D.A. (1992). Cox-type regression analysis for large numbers of small groups of correlated failure time observations. In survival analysis: State of the art, J.P. Klein and P.K. Goel (eds), 237-247. Dordrecht:Kluwer.

[58] Lehmann, E.L. (1966). Some concepts of dependence. Ann. Math. Statist. 37, 1137-1153.

[59] Leurgans, S. Tsai, W.Y. and Crowley, J. (1982). Freund's bivariate distribution and censoring. Survival analysis, Crowley and Johson, Eds., 216-229, IMS, 1982.

[60] Lin, J. S. & Wei, L.J. (1992). Linear regression analysis for multivariate failure time observations. Journal of the Royal Statistical Association, 87, 420, 1091-1097.

[61] Lindley, D. and Singpurwalla, N.A. (1986). Multivariate distributions for the reliability of a system of components sharing a common environment. Journal of applied probability, 23, 418-431.

[62] Lindeboom, M. & Van Den Berg, G.J. (1994). Heterogeneity in models for bivariate survival: the importance of the mixing distribution. Journal of the Royal Statistical Society, B, 56, 49-60.

[63] Ling, C.H. (1965). Representation of associative functions. Publ. Math. Debrecen, 12, 189-212.

[64] Maguluri, G. (1993). Semiparametric estimation of association in a bivariate survival function. The annals of statistics, vol. 21, 3, 1648-1662.

[65] Manatunga-A.K., Oakes, D. (1996). A Measure of Association for Bivariate Frailty Distributions. Journal Of Multivariate Analysis, vol 56, iss 1, pp 60-74.

[66] Marshall, A.W. (1973). Some comments on the hazard gradient. Research Report. Dept. of Statistics, University of Rochester, N.Y.

[67] Marshall, A.W. & Olkin, I. (1967). A multivariate exponential distribution. Journal of the American Statistical Association, 62, 30-44.

[68] Marshall, A.W. & Olkin, I. (1988). Families of multivariate distributions. Journal of the American Statistical Association, 83, 403, 834-841.

[69] McGilchrist, C.A. & Aisbett, C.W. (1991). Regression with frailty survial analysis. Biometrics, 47, 461-466.

[70] McGilchrist, C.A. (1994). Estimation in generalized mixed models. Journal of the Royal Statistical Society, B, 56, 61-69.

[71] Mei-Ling Ting Lee (1993). Mixtures of lifetime distributions. Advanced in reliability. Edited by Asit P. Basu. 231-245.

[72] Miller, R.G. Jr. (1981). Survival analysis. John Wiley and sons. New York.

[73] Murphy, S.A. (1994). Consistency in a proportional hazards model incorporating a random effect. The annals of statistics, vol. 22, 2, 712- 731.

[74] Navarrete, E., García, J., Lara, A., Ollero, J. y Quesada, J.M. (1995). Construcción de un modelo para datos de supervivencia bivariante bajo hipótesis mixtas en el vector función de azar. Actas del XXII Congreso Nacional de Estadística e I.O., (SEIO), Sevilla.

[75] Navarrete, E., García, J. y Ollero, J. (1997). Un modelo mixto de supervivencia bivariante con covariables en el parámetro de asociación. Actas del XXIII Congreso Nacional de Estadística e I.O., (SEIO), Valencia.

[76] Navarrete, E., García, J. y Ollero, J. (1997). Analysis of a bivariate survival model in presence of explanatory covariates. NGUS´97, IV International meeting of multidimensional data analysis. Actas, Bilbao.

[77] Nielsen, G.G., Gill, R.D., Andersen, P.K. & Sorensen, T.I.A. (1992). A counting process approach to maximum likelihood estimation in frailty models. Scand. J. Statist. 19, 25-43.

[78] Oakes, D. (1982). A model for association in bivariate survival data. Journal of the Royal Statistical Society, B, 44, 3, 414-422.

[79] Oakes, D. (1986a). Semiparametric inference in a model for association in bivariate survival data. Biometrika, 73, 353-6.

[80] Oakes, D. (1986b). A model for bivariate survival data. In modern statistical methods in chronic disease epidemiology, Moolgavkar and Prentice eds., 151-166, John Wiley and sons, New York.

[81] Oakes, D. (1989). Bivariate survival models induced by frailties. Journal of the American Statistical Association, 84, 406, 487-493.

[82] Platz, O. (1984). A Markov models for common-cause failures. Reliab. Eng., 9, 25-31.

[83] Prentice, R.L. & Cai, J. (1992). Covariance and survivor function estimation using censored multivariate time data. Biometrika, 79, 3, 495-512.

[84] Prentice, R.L. and Gloeckler, L.A. (1978). Regression analysis of grouped survival

data with applications to breast cancer data. Biometrics,34, 57-67.

[85] Prentice, R.L., Willians, B.J., Petersen, A.V. (1981). On the regresion analisys of multivariate failure time data. Biometrika, 68, 2, 373-379.

[86] Puri, P.S. & Rubin, H. (1974). On a characterization of the family of distributions with constant multivariate failure rates. Ann. Probab. 2, 738-740.

[87] Schweizer, B. & Wolff, E.F. (1981). On nonparametric measures of dependence for ramdom variables. The annals of statistics, 9, 4, 879-885.

[88] Segal, M.R. & Neuhaus, J.M. (1993). Robust inference for multivariate survival data. Statistics in Medicine, vol. 12, 1019-1031.

[89] Sklar, A. (1959). Fonctions de répartition á n dimensions et leurs marges. Publ. Inst. Statist. Math., Univ. Paris, 8, 229-231.

[90] Siegel, S. (1970). Estadística no paramétrica. Ed. Trillas.

[91] Sinha, D., Tanner, M.A. & Hall, W.J. (1994). Maximization of the marginal likelihood of grouped survival data. Biometrika, 81, 53-60.

[92] Sun, K. and Basu, A.P. (1993). Advanced in Reliability. Edited by Asit P. Basu. Department of statistics University of Missouri Columbia, Missouri. 395-409.

[93] Tsokos, M. (1987). Estadística para biología y ciencias de la salud. McGraw Hill. Interamericana.

[94] Vaupel, J.W.,Manton, K.G. and Stallard, E. (1979). The impact of heterogeneity in individual frailty and the dynamics of mortality. Demography, 16, 439-454.

[95] Vélez Ibarrola, R. y García Pérez, A. Cál. de Prob. y Est. Mat., Principios de inferencia estadística. UNED.

[96] Wassel, J.T. & Moeschberger, M.L. (1993). A bivariate survival model with modified gamma frailty for assessing the impact of interventions. Statistics in Medicine, vol. 12, 241-248.

[97] Wei, L.J., Lin, D.Y. & Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modelling marginal distributions. Journal of the American Statistical Association, 84, 408, 1065-1073.

[98] Whitmore, G.A. & Mei-Ling Ting Lee (1991). A multivariate survival distribution generated by an inverse gaussian mixture of exponentials. Technometrics, vol. 33, 1, 39-50.