

UN FRAILTY ADITIVO EN MODELOS DE SUPERVIVENCIA BIVARIANTE*

ESTEBAN NAVARRETE ÁLVAREZ, JULIA GARCIA-LEAL, ANA MARÍA LARA-PORRAS

June 3, 2004

RESUMEN

Los modelos de supervivencia bivalente basados en la presencia de una covariable aleatoria no observada, frailty, suelen asumir hipótesis de riesgo proporcional. Construimos un nuevo modelo de supervivencia bivalente basado en la presencia de un frailty que actúa aditivamente en la función de riesgo.

ABSTRACT

Bivariate survival models based on the presence of an unobserved random covariate, frailty, tend to assume proportional hazard hypothesis. A bivariate survival model is constructed. It is based on a frailty model that acts additively on the hazard function.

PALABRAS CLAVE: Frailty; dependencia; riesgo proporcional.

INTRODUCCIÓN

La construcción de modelos para tiempos de supervivencia bivalente ha sido enfocada en diferentes formas dependiendo, entre otras cuestiones, del tipo de datos a estudiar. Así, por ejemplo, pueden estudiarse pares de datos de tiempo de supervivencia correspondientes a individuos relacionados, datos familiares en un sentido amplio. La asociación entre estos pares de tiempos está inducida, en los individuos de ciertos grupos, a través de una común covariable aleatoria no observada. Dicha covariable, denominada frailty, suele ponerse que actúa multiplicativamente en las funciones de riesgo marginales. Hougaard (1986a,b) y Oakes (1989) han estudiado de esta forma estos modelos. Hougaard (1986a) estudió un modelo con un frailty con una distribución estable positiva y con distribuciones marginales Weibull introduciendo covariables explicativas siempre bajo hipótesis de riesgo proporcional. A continuación se resume en términos generales este proceso.

Sea (T_1, T_2) una variable aleatoria que expresa los tiempos de supervivencia de dos individuos. Existe un factor de riesgo común, una covariable aleatoria o frailty Z , por ejemplo el factor genérico entre los individuos (padre-hijo, hermanos, etc). En casi todos

* Departamento de estadística e Investigación Operativa, Universidad de Granada, Facultad de Ciencias, Campus de Fuentenueva, s/n.18071 Granada, España; E.mail:Estebang@goliat.ugr.es

los métodos de construcción de modelos de supervivencia la hipótesis considerada es la de riesgo proporcional. Es decir, las componentes de la función de riesgo marginal puede escribirse así

$$h_i(t_i/Z) = Zh_{io}(t_i), \quad i : 1, 2$$

por lo que la función de supervivencia marginal condicionada a un valor fijo del frailty resulta

$$S_i(t_i/Z) = \exp \left[-Z \int_0^{t_i} h_{io}(t_i) dt_i \right] = \exp \left[-Z \Lambda_{io}(t_i) \right] = [S_{io}(t_i)]^Z$$

donde $\Lambda_{io}(t_i)$, $i : 1, 2$ indica la función de riesgo marginal acumulada de la variable T_i . Hemos supuesto un valor fijo de Z , $Z = z$, por lo que los tiempos de supervivencia son independientes, es decir

$$\begin{aligned} S(t_1, t_2/Z) &= P(T_1 \geq t_1, T_2 \geq t_2/Z) \\ &= \exp \left[-Z(\Lambda_{10}(t_1) + \Lambda_{20}(t_2)) \right]. \end{aligned}$$

Si se realiza la mixtura con la distribución del frailty Z resulta la función de supervivencia conjunta

$$S(t_1, t_2) = \int S(t_1, t_2/Z) dF(z) = E[\exp(-Z(\Lambda_{10}(t_1) + \Lambda_{20}(t_2)))].$$

Sea

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

una variable auxiliar, por lo tanto

$$\phi(u) = \int \exp(-uZ) dF(z) = S(t_1, t_2)$$

puede entenderse como la transformada de Laplace del frailty Z .

2 MODELO CON FRAILTY ADITIVO

En el anterior modelo, ampliamente tratado en la literatura, la función de riesgo marginal varía, a partir de una función de riesgo base, a través de un factor multiplicativo. Este factor puede, en múltiples ocasiones, considerarse aditivo. Es decir, una función de riesgo marginal que resulte de acumular el valor del factor de riesgo o frailty a un riesgo base. Cox and Oakes (1984) sugieren este tratamiento no multiplicativo de manera que las funciones de riesgo marginal más que proporcionarles al riesgo base sean semejantes o paralelas. Nosotros hemos considerado esta posibilidad para la construcción de modelos. Es decir, se trata de suponer que

$$h_i(t_i, z) = z + h_{io}(t_i), \quad i : 1, 2$$

es decir, una forma aditiva en la función de riesgo marginal. De forma similar al caso multiplicativo, $h_{io}(t_i)$ es el riesgo base para un individuo bajo las condiciones estándar, esto es, para $z = 0$. Para cada individuo, la variable aleatoria Z tomará una realización independiente $Z = z$

En estos supuestos la función de supervivencia marginal será

$$\begin{aligned} S_i(t_i, z) &= \exp \left[- \int_0^{t_i} (z + h_{io}(t_i)) dt_i \right] \exp \left[-zt_i - \Lambda_{io}(t_i) \right] \\ &= S_{io}(t_i) \cdot \exp(-zt_i) \end{aligned}$$

Esta función de supervivencia marginal es una verdadera función de supervivencia

Veámoslo. Puesto que $z, t_i \geq 0$ se cumple que

$$\frac{S_{io}(t_i)}{\exp(zt_i)} \in [0, 1]$$

$$t_i \rightarrow 0 \implies S_i(t_i, z) \rightarrow 1$$

$$t_i \rightarrow \infty \implies S_i(t_i, z) \rightarrow 0$$

Obsérvese que, en este caso, la función de riesgo base se alcanza para un valor del frailty igual a cero, es decir, las condiciones estándar corresponden a $Z = 0$. En este caso

multiplicativo lo era para $Z = 1$. 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

$$\begin{aligned} S(t_1, t_2/z) &= S(t_1/z)S(t_2/z) \\ &= \exp \left[- \int_0^{t_1} (z + h_{10}(t_1)) dt_1 \right] \cdot \exp \left[- \int_0^{t_2} (z + h_{20}(t_2)) dt_2 \right] \\ &= \exp \left[- \int_0^{t_1} (z + h_{10}(t_1)) dt_1 - \int_0^{t_2} (z + h_{20}(t_2)) dt_2 \right] \\ &= \exp \left[-z \left(\int_0^{t_1} dt_1 + \int_0^{t_2} dt_2 \right) - \int_0^{t_1} h_{10}(t_1) dt_1 - \int_0^{t_2} h_{20}(t_2) dt_2 \right] \\ &= \exp S_{10}(t_1) S_{20}(t_2) \exp (-z(t_1 + t_2)) \end{aligned}$$

Por consiguiente, la función de supervivencia conjunta será

$$\begin{aligned} S(t_1, t_2) &= \int S_{10}(t_1) S_{20}(t_2) \exp (-z(t_1 + t_2)) dF(z) \\ &= S_{10}(t_1) S_{20}(t_2) \int \exp (-z(t_1 + t_2)) dF(z) = S_{10}(t_1) S_{20}(t_2) \phi(u) \end{aligned}$$

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 riesgo marginal. En este caso la función de supervivencia conjunta coincidía con la transformada 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 funciones de supervivencia marginales base por la transformada de Laplace de Z y con argumento $t_1 + t_2$.

En el caso en que el frailty Z pueda modificarse por una distribución estable positiva, es decir, $\phi(u) = \exp(-u^\alpha)$, entonces la función de supervivencia conjunta será

$$S(t_1 + t_2) = S_{10}(t_1) S_{20}(t_2) \exp(-(t_1 + t_2)^\alpha). \quad [4]$$

Este modelo se considerará paramétrico si se considera la familia de distribuciones a la que pertenecen las marginales (Weibull, etc).

3 ESTIMACIÓN

Supongamos para comenzar que se conoce la expresión de la función de supervivencia bivalente $S(t_1, t_2)$. Se supone que, además, se está ante una situación paramétrica según el sentido visto anteriormente. La expresión [4] corresponde a un modelo aditivo con frailty estable positiva. Si las marginales siguen distribuciones Weibull la expresión quedará así

$$\begin{aligned} S(t_1, t_2) &= \exp \left(\in t_1^\alpha \right) \cdot \exp \left(\in t_2^\alpha \right) \cdot \exp \left[-(t_1 + t_2)^\alpha \right] \\ &= \exp \left[\in \left(t_1^\alpha + t_2^\alpha \right) - (t_1 + t_2)^\alpha \right] \end{aligned}$$

donde $\alpha \in (0, 1)$, el parámetro de asociación, proviene de la distribución del frailty

Si es igual a 1 corresponde al caso de variables independientes. Los demás parámetros corresponden a distribuciones de Weibull.

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. Los parámetros a estimar son

$$\in, \gamma, \alpha$$

es decir, hay 3 parámetros a estimar

Se necesita hacer inferencias sobre los parámetros, ya sean 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 bivalente esto ocurre cuando alguna componente o las dos del vector de datos no se conocen. La censura más usual es la censura de 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 supervivencia real. Al ir a medir ya había fallado. Existen otros tipos de censuras, por intervalos, de tipo I, tipo II, aleatoria, etc. Consideramos censuras a la derecha en lo que sigue. Se dispone de un dato expresado por el vector (t_1, t_2) .

Establecemos cuatro situaciones distintas definidas por los siguientes grupos:

Grupo 1 si t_1 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:

$I(j) = \{1 \text{ si } (t_1, t_2) \text{ está en el grupo } j$

$0 \text{ si } (t_1, t_2) \text{ no está en el grupo } j$

$j : 1, 2, 3, 4.$

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

$$g1 = \frac{a^2 S(t_1, t_2)}{at_1 at_2}$$

es decir, el valor correspondiente de la función de densidad. Un dato del grupo 2 aporta la expresión

$$g2 = \frac{-aS(t_1, t_2)}{at_1}$$

Un dato del grupo 3 aporta la expresión

$$g3 = \frac{-aS(t_1, t_2)}{at_2}$$

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

$(t_{1i}, t_{2i}), i : 1, \dots, n$

la función de verosimilitud es

$$L(\alpha, \gamma, \epsilon) = \prod_{i=1}^n \left[\left(\frac{a^2 S(t_{1i}, t_{2i})}{at_{1i} at_{2i}} \right) I(1) \cdot \left(\frac{-aS(t_{1i}, t_{2i})}{at_{1i}} \right) I(2) \cdot \left(\frac{-aS(t_{1i}, t_{2i})}{at_{2i}} \right) I(3) \cdot (S(t_{1i}, t_{2i})) I(4) \right]$$

entendiendo tales derivadas respecto t_1 ó t_2 valuadas en t_{1i} ó t_{2i} , respectivamente.

Consideramos una muestra

$(t_{1i}, t_{2i}), i : 1, \dots, n.$

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 el logaritmo de la función de verosimilitud es

$$\begin{aligned} 1g L(\alpha, \gamma, \epsilon) &= \sum_{i=1}^n 1g[(g1)^{I(1)} \cdot (g2)^{I(2)} \cdot (g3)^{I(3)} \cdot (g4)^{I(4)}] = \\ &= \sum_{i=1}^n [I(1)1g(g1) + I(2)1g(g2) + I(3)1g(g3) + I(4)1g(g4)] = \\ &= I(1) \sum_{i=1}^n 1g(g1) + I(2) \sum_{i=1}^n 1g(g2) + I(3) \sum_{i=1}^n 1g(g3) + I(4) \sum_{i=1}^n 1g(g4) \end{aligned}$$

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

$$\begin{aligned} \frac{a S(t_1, t_2)}{a t_1} &= [\epsilon \gamma t_1^{\gamma-1} - \alpha(t_1 + t_2)^{\alpha-1}] \exp[\epsilon(t_1^\gamma + t_2^\gamma) - (t_1 + t_2)^\alpha] \\ \frac{a S(t_1, t_2)}{a t_2} &= [\epsilon \gamma t_2^{\gamma-1} - \alpha(t_1 + t_2)^{\alpha-1}] \exp[\epsilon(t_1^\gamma + t_2^\gamma) - (t_1 + t_2)^\alpha] \\ \frac{a^2 S(t_1, t_2)}{a t_1 a t_2} &= [(\epsilon \gamma t_2^{\gamma-1} - \alpha(t_1 + t_2)^{\alpha-1})(\epsilon \gamma t_1^{\gamma-1} - \alpha(t_1 + t_2)^{\alpha-1}) - \alpha(\alpha-1)(t_1 + t_2)^{\alpha-2}] \\ &\cdot \exp[\epsilon(t_1^\gamma + t_2^\gamma) - (t_1 + t_2)^\alpha]. \end{aligned}$$

Por consiguiente la log-verosimilitud es

$$\begin{aligned} 1g L(\alpha, \gamma, \epsilon) &= I(1) [\epsilon \sum(t_1^\gamma + t_2^\gamma) - \sum(t_1 + t_2)^\alpha] \\ &+ \sum 1g[(\epsilon \gamma t_2^{\gamma-1} - \alpha(t_1 + t_2)^{\alpha-1})(\epsilon \gamma t_1^{\gamma-1} - \alpha(t_1 + t_2)^{\alpha-1}) - \alpha(\alpha-1)(t_1 + t_2)^{\alpha-2}] \\ &+ I(2) [\epsilon \sum(t_1 + t_2)^\alpha + \sum 1g(\alpha(t_1 + t_2)^{\alpha-1} - \epsilon \gamma t_1^{\gamma-1})] \\ &+ I(3) [\epsilon \sum(t_1^\gamma + t_2^\gamma) - \sum(t_1 + t_2)^\alpha + \sum 1g(\alpha(t_1 + t_2)^{\alpha-1} - \epsilon \gamma t_2^{\gamma-1})] \\ &+ I(4) [\epsilon \sum(t_1^\gamma + t_2^\gamma) - \sum(t_1 + t_2)^\alpha]. \end{aligned}$$

La solución del sistema de ecuaciones

$$\begin{aligned} \frac{a_1 g L(\alpha, \gamma, \epsilon)}{a \alpha} &= 0 \\ \frac{a_1 g L(\alpha, \gamma, \epsilon)}{a \gamma} &= 0 \\ \frac{a_1 g L(\alpha, \gamma, \epsilon)}{d \epsilon} &= 0 \end{aligned}$$

proporciona los estimadores de máxima verosimilitud $\tilde{\alpha}, \tilde{\gamma}, \tilde{\epsilon}$. La obtención de los estimadores ha de hacerse por métodos numéricos usando, por ejemplo, el método Newton Raphson.

REFERENCIAS

Cox, D.R. and Oakes, D. (1984). Analysis of survival data. Chapman and Hall. London.

Hougaard, P.(1986a). "A class of multivariate failure time distribution". Biometrika, 73,671-8.

Hougaard, P.(1986b). "Survival models for heterogeneous populations derived from stable distributions". Biometrika, 73,387-96.

Navarrete Alvarez, E.(1997). "Aportaciones a la formulación de modelos, estudio de la asociación y estimación en análisis de datos de supervivencia multivariante". Tesis doctoral. Universidad de Granada. Granada

Oakes, D. (1989). "Bivariate survival models induced by frailties". Journal of the American Statistical Association, 84,406,487-493.

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.