Rev.Acad.Canar.Cienc., XI (Núms. 1-2), 153-165 (1999)

# MODELOS DE REGRESIÓN DEFINIDOS A TROZOS CON RIESGOS PROPORCIONALES Y DOBLE CENSURA

**Arturo J. Fernández, José I. Bravo e Íñigo De Fuentes**

Departamento de Estadística, I. O. y Computación. Universidad de La Laguna
38271 Santa Cruz de Tenerife. España (e-mail: ajfernan@ull.es)

## RESUMEN

El propósito del presente artículo es describir la aplicación de los modelos de regresión con riesgos proporcionales al análisis de datos de supervivencia doblemente censurados, con covariables o factores de riesgo, mediante la modelización paramétrica de la tasa de fallo base en términos de distribuciones definidas a trozos. Presentamos las ecuaciones de verosimilitd y procedimientos de inferencia asociados. Asimismo, obtenemos una estimación consistente de la matriz de covarianza asintótica del estimador máximo-verosímil del parametro desconocido. En particular, consideramos el ajuste de las distribuciones exponencial, Weibull y de valor extremo generalizada a trozos. Puede esperarse que los modelos definidos a trozos de esta clase describirán adecuadamente muchos de los procesos de supervivencia con riesgos proporcionales que incluyan una serie de puntos en los cuales se alteren las condiciones imperantes. Además, modificamos algunos métodos para evaluar la influencia de los factores de riesgo y varias técnicas gráficas estándar para comprobar los supuestos del modelo para datos no censurados de tal forma que admita la censura doble.

## ABSTRACT

The purpose of the present paper is to describe the application of proportional hazards regression models to the analysis of doubly censored survival data, with covariates or risk factors, by means of a parametric modelling of the base-line hazard rate in terms of piecewise distributions. Likelihood equations and associated inferential procedures are presented. Likewise, a consistent estimation of the asymptotic covariance matrix of the maximum likelihood estimator of the unknown parameter is obtained. In particular, we consider the fitting of the piecewise exponential, Weibull and generalized extreme value distributions. It can be expected that piecewise models of this kind will usefully describe many proportional hazards survival processes involving changepoints at which the ruling conditions suddenly alter. In addition, several methods for evaluating the effects of the risk factors and standard graphical techniques to check model assumptions for uncensored data are modified to allow for double censoring.

*Palabras y frases claves: Censura por la derecha y por la izquierda; factores de riesgo; función de supervivencia; máxima verosimilitud; puntos de cambio.*

*Clasificación AMS:* 62A10, 62F10, 62J99, 62P10

# 1. PRELIMINARES

## 1.1 Introducción al Análisis de Supervivencia

El análisis de supervivencia engloba una amplia variedad de técnicas estadísticas orientadas al estudio del comportamiento de variables aleatorias no negativas, que representan el tiempo que transcurre desde un instante inicial hasta la ocurrencia de algún suceso de interés (a menudo, la muerte o el fallo). En general, la variable en estudio se denomina génericamente *tiempo de supervivencia, de vida o de fallo.* El tiempo que transcurre hasta la muerte o recaída de una paciente con cáncer, el que tarda en fallar una componente eléctrica o mecánica, o en germinar una semilla, o el de duración de una huelga son algunos de los muchos ejemplos que podrían citarse como variables de interés.

Los orígenes del análisis de supervivencia pueden atribuirse a los primeros trabajos sobre tablas de mortalidad de hace ya varios siglos, no obstante, su era moderna comenzó hace aproximadamente medio siglo a raíz de las aplicaciones de la teoría de la fiabilidad en ingeniería, especialmente las de tipo militar, puestas en práctica a partir de la Segunda Guerra Mundial. Hoy en día, estas técnicas no sólo se emplean con gran profusión en las ciencias biomédicas y actuariales, sino que también se han introducido en muchas otras áreas del conocimiento como son la biometría forestal, botánica, criminología, demografía, ecología, economía, fiabilidad de sistemas, física, ingeniería, procesos industriales, zoología, y un largo etcétera.

La variabilidad de los resultados que se obtienen al medir el tiempo transcurrido entre el instante inicial y el momento en que ocurre el suceso de interés y la existencia de cierta regularidad inherente a esa incertidumbre justifican la utilización de modelos probabilísticos, que proporcionan el soporte teórico esencial y las herramientas necesarias para un análisis profundo, en los problemas de supervivencia.

Una de las características distintivas del análisis de supervivencia es la presencia de complicados patrones de censura que dificultan en gran medida el estudio. Asimismo, otro aspecto fundamental es que el *mecanismo de fallo/censura* puede modelizarse preferentemente utilizando técnicas extraídas de la probabilidad y estadística. Debido a los métodos de muestreo y a ciertos factores que escapan al control experimental, en ocasiones el *tiempo de vida* no es directamente observable. En la mayoría de esos casos sólo se sabe que es mayor o menor que un determinado *tiempo de censura por la derecha o por la izquierda*, respectivamente. Si algunos individuos de la muestra nacieron, o empezaron a funcionar, antes del inicio del estudio, existirá *censura a la izquierda.* El otro tipo de información censurada se origina cuando, finalizado el experimento, hay individuos en la muestra para los que no ha ocurrido todavía el suceso de interés. Esta clase de censura se denomina *censura a la derecha.* Cuando algunos datos están censurados a la derecha y otros a la izquierda, se dice que existe *doble censura.* Los artículos de Gu y Zhang [6], Bravo et al. [2], Mykland and Ren [9], Zhang and Li [13] and Fernández et al. [3] son algunos de los muchos trabajos recientes que consideran este tipo de observaciones.

## 1.2 Modelo de Doble Censura Aleatoria

Supongamos que $X$, *tiempo de vida*, es una variable aleatoria continua y no negativa, con función de supervivencia $S(x) = \Pr(X > x)$, que representa el tiempo de ocurrencia de un determinado suceso de interés. Consideramos además que $X$ está sujeta a un intervalo aleatorio de censura $[U, V]$, que es independiente de ella. Por tanto, $X$ es observable si y sólo si $U \leq X \leq V$, donde $U$ y $V$ son las variables de censura por la izquierda y por la derecha, respectivamente, que verifican $\Pr(U \leq V) = 1$. Si $X$ no pertenece a $[U, V]$, no podemos determinar su valor exacto. Sólo sabemos si $X$ es menor que $U$ o mayor que $V$, y únicamente observamos $U$ o $V$ en el correspondiente caso. De este modo, la información disponible sobre $X$ puede expresarse mediante un par de variables aleatorias $T$ y $\delta$, donde

$$T = \max\{\min(X, V), U\} \tag{1}$$

y

$$\delta = \begin{cases} 0 & si \quad U \leq X \leq V \quad (no\ censura), \\ 1 & si \quad X > V \quad (censura\ por\ la\ derecha), \\ 2 & si \quad X < U \quad (censura\ por\ la\ izquierda). \end{cases} \tag{2}$$

## 1.3 Modelos de Regresión con Riesgos Proporcionales

En situaciones homogéneas se considera que los *tiempos de vida* de todos los individuos son variables aleatorias independientes e idénticamente distribuidas. En otras ocasiones se supone que, además del *tiempo de vida*, se ha observado sobre cada individuo cierto número de factores de riesgo (variables regresoras o covariables) que, de alguna forma, afectan a la distribución de la supervivencia. En el intento de incorporar este tipo de información en los modelos de supervivencia ha sido fructífero trabajar en términos de la tasa de fallo. El modelo general de regresión con riesgos proporcionales, introducido por Cox [4], considera que el efecto de los factores de riesgo sobre la tasa de fallo es multiplicativo. De este modo, para un individuo con vector columna $p$-dimensional de covariables $z = (z_1, .., z_p)^t$, la tasa de fallo es $h(x; z) = \psi(z)\lambda(x)$, donde $\psi$ es alguna función de $z$ no negativa y $\lambda(x) = f_0(x)/S_0(x)$ es la tasa de fallo base, es decir, para un individuo en condiciones normales $(z = z_0)$, mientras que $S_0(x)$ y $f_0(x)$ son las funciones de supervivencia y densidad de $X$ cuando $z = z_0$, respectivamente.

Las funciones de supervivencia y de densidad de $X$, dado $z$, son respectivamente

$$S(x; z) = \Pr(X > x \mid Z = z) = \{S_0(x)\}^{\psi(z)} \quad \text{y} \quad f(x; z) = f_0(x) \cdot \psi(z) \cdot \{S_0(x)\}^{\psi(z)-1}$$

Asimismo, $S(x;z) = \exp\{-H(x;z)\}$, donde $H(x;z) = \psi(z)\Lambda(x)$ y $\Lambda(x) = -\ln S_0(x)$ son las tasas de fallo acumulativas dados $z$ y $z_0$, respectivamente.

Si suponemos que $g(x) = \ln \Lambda(x)$, se verifica que $S(x;z) = \exp[-\exp\{g(x) + \ln\psi(z)\}]$ y $f(x;z) = g'(x) \cdot \exp\{g(x) + \ln\psi(z)\} \cdot \exp[-\exp\{g(x) + \ln\psi(z)\}]$. Nótese que, dados $z_1$ y $z_2$, la diferencia $\ln\{-\ln S(x;z_1)\} - \ln\{-\ln S(x;z_2)\} = \ln\{\psi(z_1)/\psi(z_2)\}$ no depende de $x$ ya que $\ln\{-\ln S(x;z)\} = g(x) + \ln\psi(z)$.

La función $\psi(z)$ puede ser parametrizada como $\psi(z;\beta)$, donde $\beta = (\beta_1, .., \beta_p)$ es un vector fila $p$-dimensional de parámetros desconocidos. Las tres principales parametrizaciones de $\psi$ a considerar son la lineal, $\psi(z;\beta) = 1+\beta z$, la logística, $\psi(z;\beta) = \ln\{1 + \exp(\beta z)\}$, y la log-lineal, $\psi(z;\beta) = \exp(\beta z)$, que por buenas razones se ha convertido en la más popular. El propio Cox [4] hizo hincapié en esta última formulación, la cual tiene la ventaja de que el factor $\psi(z;\beta)$ es siempre positivo y que, como $\psi(0;\beta) = 1$, en condiciones normales $z = 0$. Además, si $\beta = 0$ entonces $\psi(z;0) = 1$ y las covariables no afectan al riesgo, mientras que si $\beta_j > 0$, aumenta el riesgo si lo hace $z_j$.

Desde el punto de vista teórico existen principalmente dos líneas de investigación. En la primera, no se conoce la tasa de fallo base; en la segunda, se supone un modelo completamente paramétrico para ésta. La primera línea se inició a partir del modelo de Cox [4] y ha sido ampliamente analizada por numerosos investigadores desde entonces. La segunda línea de investigación ha sido estudiada, entre otros, por Aitkin y Clayton [1], Noura y Read [10] y García et al. [5], considerando tanto observaciones no censuradas como censuradas por la derecha. Los primeros autores utilizan las distribuciones exponencial, de Weibull, de valor extremo y de valor extremo generalizada. Los otros dos trabajos se basan en las distribuciones definidas a trozos de Weibull y de valor extremo generalizado, respectivamente. Sin embargo, estos modelos definidos a trozos, cuyo fundamento es el supuesto de que los parámetros que caracterizan a la distribución base de la supervivencia varían con el paso del tiempo, aún no se han considerado en la literatura científica en el caso de doble censura.

Nuestro artículo se incluye dentro de la segunda línea de investigación y estudia el modelo de regresión con riesgos proporcionales y datos doblemente censurados, considerando un modelo parámetrico general para la distribución base de la supervivencia (sección 2) o un modelo definido a trozos (sección 3). Además, en la cuarta sección, analizamos algunos procedimientos gráficos para evaluar la validez de los modelos propuestos.

## 2. FORMULACIÓN PARAMÉTRICA

Supongamos que, mediante estudios preliminares, sabemos que la función de supervivencia base de $X$ está dada por un modelo paramétrico conocido $S_0(x;\theta)$, donde $\theta = (\theta_1, ...., \theta_q)$ es un vector de parámetros desconocidos que toma valores en cierto espacio paramétrico $\Theta$. Considerando el modelo de riesgos proporcionales con multiplicador $\psi(z;\beta)$, nos proponemos estimar el vector $\phi = (\theta, \beta)$ a partir de una muestra de $N$ observaciones independientes $(t_i, \delta_i, z_i)$, $i = 1, .., N$, donde $t_i$ y $\delta_i$ están definidos como en (1) y (2), respectivamente, y $z_i = (z_{i1}, .., z_{ip})^t$ es el correspondiente vector de covariables.

En muchas ocasiones podemos considerar que una parametrización interesante para la tasa de fallo base es $\lambda(t;\theta) = \exp\{\theta y(t)\}$, donde $y(t)$ es un vector columna de $q$ funciones. Si $q = 1$ e $y_1(t) = 1$, obtenemos la distribución exponencial; si $q = 2$, $y_1(t) = 1$ e $y_2(t) = \ln t$, aparece la distribución de Weibull; y si $q = 2$, $y_1(t) = 1$ e $y_2(t) = t$, se obtiene la distribución de Gompertz. En este caso, también es aconsejable utilizar la parametrización log-lineal para $\psi$, con lo cual $h(t; z, \phi) = \exp\{\theta y(t) + \beta z\}$. De forma análoga, podríamos establecer una formulación log-lineal para la tasa de fallo acumulativa base, i. e., $\Lambda(t;\theta) = \exp\{\theta y(t)\}$. Como casos particulares aparecen las distribuciones exponencial, de Weibull y de valor extremo. Por ejemplo, si la distribución base de la supervivencia es de Weibull, $\Lambda(t;\theta) = \exp\{\varepsilon + \alpha \ln t\}$, y $\psi(z;\beta) = \exp(\beta z)$, la distribución de la variable aleatoria $\exp\{\varepsilon + \beta z\} X^\alpha$ es exponencial unitaria. Por tanto, el modelo log-lineal de regresión sería $\ln X = -\varepsilon/\alpha - \beta Z/\alpha + W/\alpha$, donde el error $W$ tiene una distribución estándar de valor extremo con densidad $f(w) = \exp\{w - \exp(w)\}$, $-\infty < w < +\infty$. Como se ve, el modelo de regresión es lineal en $\ln X$, esto es, log-lineal en $X$. En el caso exponencial, el modelo sería similar salvo que ahora $\alpha = 1$.

### 2.1 Ecuaciones de Verosimilitud

Supuesto que las distribuciones de las variables de censura $U$ y $V$ no dependen del parámetro $\theta$ y que el mecanismo de censura es independiente de los factores de riesgo, la función de verosimilitud es proporcional a

$$L(\phi) = \prod_{i \in I_0} h(t_i; z_i, \phi) \exp\{-H(t_i; z_i, \phi)\} \prod_{j \in I_1} \exp\{-H(t_j; z_j, \phi)\} \prod_{k \in I_2} [1 - \exp\{-H(t_k; z_k, \phi)\}],$$

donde $I_j = \{i \in I : \delta_i = j\}$, $j = 0, 1, 2$, son los conjuntos de índices de los sujetos no censurados, censurados por la derecha y censurados por la izquierda, respectivamente, e $I = \{1, .., N\}$. Por consiguiente, la log-verosimilitud vendrá dada por:

$$l(\phi) = \sum_{i\in I_0} \{\ln \psi(z_i;\beta) + \ln \lambda(t_i;\theta) - \psi(z_i;\beta)\Lambda(t_i;\theta)\} - \sum_{j\in I_1} \psi(z_j;\beta)\Lambda(t_j;\theta) + \sum_{k\in I_2} \ln\left[1 - \exp\{-\psi(z_k;\beta)\Lambda(t_k;\theta)\}\right].$$

Por simplicidad, consideramos que $\xi(i)$ es el valor de la función $\xi$ para el sujeto $i$-ésimo, donde $\xi$ puede ser $\lambda$, $\Lambda$, $H, \psi$ o $g$, y que los subíndices $r$ y $s$ indican respectivamente $\partial/\partial\theta_r$ y $\partial/\partial\beta_s$ de la correspondiente función. De este modo, para $r = 1, .., q$, obtenemos que

$$\frac{\partial}{\partial\theta_r} l(\phi) = \sum_{i\in I_0} \left\{ \frac{\lambda_r(i)}{\lambda(i)} - \psi(i)\Lambda_r(i) \right\} - \sum_{j\in I_1} \psi(j)\Lambda_r(j) + \sum_{k\in I_2} \frac{\psi(k)\Lambda_r(k)\exp\{-\psi(k)\Lambda(k)\}}{1 - \exp\{-\psi(k)\Lambda(k)\}},$$

que también podemos expresar como

$$\frac{\partial}{\partial\theta_r} l(\phi) = \sum_{i\in I_0} \frac{\lambda_r(i)}{\lambda(i)} + \sum_{j\in I} \{w(j) - 1\} \frac{\Lambda_r(j)}{\Lambda(j)} H(j), \tag{3}$$

o bien como

$$\frac{\partial}{\partial\theta_r} l(\phi) = \sum_{i\in I_0} \left\{ \frac{g'_r(i)}{g'(i)} + g_r(i) \right\} + \sum_{j\in I} \{w(j) - 1\} g_r(j) H(j), \tag{4}$$

donde $w(j) = \mathrm{I}(\delta_j = 2)/[1 - \exp\{-H(j)\}]$ y $g'(i) = \partial g(t_i)/\partial t$.

De forma análoga, para $s = 1, .., p$, tenemos que

$$\frac{\partial}{\partial\beta_s} l(\phi) = \sum_{i\in I_0} \frac{\psi_s(i)}{\psi(i)} - \sum_{j\in I} \psi_s(j)\Lambda(j) + \sum_{k\in I_2} \frac{\psi_s(k)\Lambda(k)}{1 - \exp\{-\psi(k)\Lambda(k)\}},$$

que también puede expresarse como

$$\frac{\partial}{\partial\beta_s} l(\phi) = \sum_{i\in I_0} \frac{\psi_s(i)}{\psi(i)} + \sum_{j\in I} \{w(j) - 1\} \frac{\psi_s(j)}{\psi(j)} H(j). \tag{5}$$

Los estimadores máximo-verosímiles $\hat{\theta}$ y $\hat{\beta}$ se obtendrían maximizando $l(\phi) = \ln L(\phi)$ o resolviendo, mediante el método de Newton-Raphson o algún algoritmo de relajación, el sistema formado por las $(q + p)$ ecuaciones de verosimilitud

$$\begin{cases} \dfrac{\partial}{\partial\theta_r} l(\phi) = \displaystyle\sum_{i\in I_0} \frac{\lambda_r(i)}{\lambda(i)} + \sum_{j\in I} \{w(j) - 1\} \frac{\Lambda_r(j)}{\Lambda(j)} H(j) = 0, \quad r = 1, .., q, \\ \dfrac{\partial}{\partial\beta_s} l(\phi) = \displaystyle\sum_{i\in I_0} \frac{\psi_s(i)}{\psi(i)} + \sum_{j\in I} \{w(j) - 1\} \frac{\psi_s(j)}{\psi(j)} H(j) = 0, \quad s = 1, .., p. \end{cases}$$

No obstante, el procedimiento propuesto por Richards [12] es más simple y computacionalmente preferible. Primero se hallaría la log-verosimilitud máxima para cada valor fijo de $\beta$. Obtendríamos así una log-verosimilitud de perfil , $l^*(\beta)$, que es función sólo de $\beta$ . Después

aplicaríamos el método de Newton-Raphson u otro similar para maximizar $l^*$ en $\beta$, y para finalizar calcularíamos la correspondiente estimación de $\theta$.

### 2.2 Estimación de la Matriz de Covarianzas Asintótica

Para estimar la matriz de covarianzas asintótica de $\widehat{\phi} = (\widehat{\theta}, \widehat{\beta})$ de forma consistente es conveniente utilizar la inversa de la matriz de información observada de Fisher,

$$A = -\frac{\partial^2}{\partial\phi^t \partial\phi} l(\phi) = \left[ \begin{array}{cc} A_{11} & A_{12} \\ A_{21} & A_{22} \end{array} \right]_{(q+p)\times(q+p)},$$

donde $A_{11} = -\left[\partial^2 l(\phi)/\partial\theta_r \partial\theta_{r'}\right]_{q\times q}$ y $A_{22} = -\left[\partial^2 l(\phi)/\partial\beta_s \partial\beta_{s'}\right]_{p\times p}$, mientras que $A_{12} = A_{21}^t = -\left[\partial^2 l(\phi)/\partial\theta_r \partial\beta_s\right]_{q\times p}$, y para $r, r' = 1, .., q$ y $s, s' = 1, .., p$,

$$\begin{aligned}
\frac{\partial^2}{\partial\theta_r \partial\theta_{r'}} l(\phi) &= \sum_{i\in I_0} \left\{ \frac{\lambda_{rr'}(i)}{\lambda(i)} - \frac{\lambda_r(i)}{\lambda(i)} \frac{\lambda_{r'}(i)}{\lambda(i)} \right\} \\
&\quad + \sum_{j\in I} \{w(j) - 1\} H(j) \left\{ \frac{\Lambda_{rr'}(j)}{\Lambda(j)} - w(j) \frac{\Lambda_r(j)}{\Lambda(j)} \frac{\Lambda_{r'}(j)}{\Lambda(j)} H(j) \right\}, \\
\frac{\partial^2}{\partial\theta_r \partial\beta_s} l(\phi) &= \sum_{j\in I} \{w(j) - 1\} \{1 - w(j)H(j)\} \frac{\Lambda_r(j)}{\Lambda(j)} \frac{\psi_s(j)}{\psi(j)} H(j), \\
\frac{\partial^2}{\partial\beta_s \partial\beta_{s'}} l(\phi) &= \sum_{i\in I_0} \left\{ \frac{\psi_{ss'}(i)}{\psi(i)} - \frac{\psi_s(i)}{\psi(i)} \frac{\psi_{s'}(i)}{\psi(i)} \right\} \\
&\quad + \sum_{j\in I} \{w(j) - 1\} H(j) \left\{ \frac{\psi_{ss'}(j)}{\psi(j)} - w(j) \frac{\psi_s(j)}{\psi(j)} \frac{\psi_{s'}(j)}{\psi(j)} H(j) \right\},
\end{aligned}$$

están evaluadas en $\widehat{\phi} = (\widehat{\theta}, \widehat{\beta})$. La dimensión de $A$, $q + p$, puede ser grande. Sin embargo, como

$$A^{-1} = \left[ \begin{array}{cc} A_{11}^{-1} + WV^{-1}W^t & -WV^{-1} \\ -V^{-1}W^t & V^{-1} \end{array} \right]_{(q+p)\times(q+p)},$$

donde $W = A_{11}^{-1} A_{12}$ y $V = A_{22} - A_{21}W$, sólo necesitamos invertir la matrix $A_{11}$ de dimensión $q$ y la matrix $V$ de dimensión $p$.

### 2.3 Métodos Asintóticos de Inferencia

Como sabemos, si se satisfacen las apropiadas condiciones de regularidad, el estimador de máxima verosimilitud del parámetro $\phi$ es asintóticamente la única solución de la ecuación de verosimilitud $\partial l(\phi)/\partial\phi = 0$. Mäkeläinen et al. [8] proporcionan condiciones suficientes para asegurar la existencia y unicidad de este estimador. A su vez, $\widehat{\phi}$ es un estimador consistente de $\phi$ y asintóticamente eficiente y normal de media $\phi$ y matriz de covarianzas $J(\phi)^{-1}$, donde

$J(\phi) = -E\left[\partial^2 l(\phi)/\partial\phi^t\partial\phi\right]$ es la matriz de información de Fisher (véase, por ejemplo, Lehmann [7], sec. 6.2). La matriz $J(\phi)$ no puede evaluarse sin considerar el mecanismo de censura subyacente. Sin embargo, como $A(\widehat{\phi})^{-1}$ es un estimador consistente de $J(\phi)^{-1}$, se verifica que $\widehat{\phi}$ es asintóticamente $N(\phi, A(\widehat{\phi})^{-1})$.

El método delta (Rao [11]) puede utilizarse para estimar la varianza de cualquier función real $\varphi(\phi)$ con derivadas parciales continuas. De este modo,

$$\left\{\varphi(\widehat{\phi}) - \varphi(\phi)\right\} / \sqrt{\left(\partial\varphi(\widehat{\phi})/\partial\phi\right)^t A(\widehat{\phi})^{-1} \left(\partial\varphi(\widehat{\phi})/\partial\phi\right)} \xrightarrow[N\to\infty]{d} N(0,1).$$

Por ejemplo, podemos considerar la estimación de la supervivencia y de la vida media para un valor fijo de $z$, así como la comparación de éstas entre varios grupos.

Para la comparación de varias curvas de supervivencia es interesante contrastar si $\beta = 0$. Así, un contraste de homogeneidad de $(p+1)$ curvas de supervivencia corresponde a contrastar $H_0 : \beta = 0$, considerando que $z$ es un vector de variables indicatrices para $p$ de los grupos. El estadístico de contraste basado en el gradiente de la log-verosimilitud $U_0 = [\partial l(\phi)/\partial\beta]$ evaluado en $\widehat{\theta}_0$ (estimación de máxima verosimilitud de $\theta$ cuando $\beta = 0$) y $\beta = 0$ sería $U_0 V_0^{-1} U_0^t$, donde $V_0^{-1} = V^{-1}(\widehat{\theta}_0, 0)$. La distribución de $U_0 V_0^{-1} U_0^t$ bajo $H_0$ es asintóticamente chi-cuadrado con $p$ grados de libertad. También, el estadístico de razón de verosimilitudes, $2\left\{l(\widehat{\theta}, \widehat{\beta}) - l(\widehat{\theta}_0, 0)\right\}$, y el estadístico de Wald, $\widehat{\beta} V^{-1}(\widehat{\phi})\widehat{\beta}^t$, son asintóticamente chi-cuadrados con $p$ grados de libertad si $\beta = 0$. Asimismo, es posible contrastar $H_0 : \beta_j = 0$, esto es, si el factor de riesgo $j$-ésimo influye en la supervivencia.

Debemos indicar que la teoría asintótica de la estimación de máxima verosimilitud, en la que se basan las aproximaciones normal y chi-cuadrado, requiere que se verifiquen algunas condiciones de regularidad sobre la suavidad de la función de verosimilitud. Con observaciones que no son independientes o idénticamente distribuidas, se necesita, hablando grosso modo, que la proporción de la información total en la muestra que aporta cada observación converja a cero cuando crece el tamaño muestral. No obstante, aunque se verifiquen tales condiciones, los resultados pueden ser erróneos para muestras pequeñas.

## 3. FORMULACIÓN PARÁMETRICA DEFINIDA A TROZOS

Suponemos ahora que los parámetros que caracterizan a la distribución base de la supervivencia pueden variar con el tiempo. Por esta razón, introducimos una partición del intervalo (0,∞) en $(k+1)$ subintervalos mediante los puntos de cambio $a_1, .., a_k$, de forma que la

parametrización de la distribución base se mantiene en cada subintervalo, aunque con diferentes parámetros en cada uno de ellos. Por conveniencia, definimos $a_0 = 0$ y $a_{k+1} = \infty$. En los tres subsecciones siguientes, consideramos que la distribución base de la supervivencia es, a trozos, exponencial, de Weibull y de valor extremo generalizada.

### 3.1 Distribución Exponencial a Trozos

Suponemos que $\lambda(t) = \exp(\alpha_j)$ si $t \in (a_{j-1}, a_j]$ para $j = 1, .., k+1$. Para $i = 1, .., N$, definimos $m_i$ como el índice del subintervalo al que pertenece $t_i$, esto es,

$$m_i = j \text{ si } t_i \in (a_{j-1}, a_j] \text{ para } j = 1, .., k+1.$$

y consideramos que $e_{ij}$ es el tiempo de exposición observado del individuo $i$-ésimo en el subintervalo $j$-ésimo, definido por

$$e_{ij} = \min(t_i, a_j) - a_{j-1} \text{ para } j = 1, .., m_i,$$

con lo cual $\lambda(i) = \exp(\alpha_{m_i})$ y $H(i) = \sum_{j=1}^{m_i} \sigma_{ij}$, donde $\sigma_{ij} = \exp(\alpha_j)\, e_{ij} \psi(i)$.

Como para $r = 1, .., k+1$,

$$\lambda_r(i) = \frac{\partial}{\partial \alpha_r} \lambda(i) = \mathrm{I}(m_i = r)\lambda(i) \text{ y } H_r(i) = \frac{\partial}{\partial \alpha_r} H(i) = \mathrm{I}(m_i \geq r)\sigma_{ir},$$

obtenemos a partir de la fórmula (3) que

$$\frac{\partial}{\partial \alpha_r} l(\phi) = \sum_{i \in I_0} \mathrm{I}(m_i = r) + \sum_{j \in I} \mathrm{I}(m_j \geq r) \{w(j) - 1\} \sigma_{jr} = N_{0r} + \sum_{j \in R_r} \{w(j) - 1\} \sigma_{jr},$$

donde $N_{0r} = Card\{i \in I_0 : m_i = r\}$ es el número de datos no censurados que pertenecen al subintervalo $r$-ésimo y $R_r = \{i \in I : t_i > a_{r-1}\}$ es el conjunto de índices de los individuos en riesgo al comienzo del subintervalo $r$-ésimo.

Considerando el modelo log-lineal $\psi(z; \beta) = \exp(\beta z)$, obtenemos de (5) que

$$\frac{\partial}{\partial \beta_s} l(\phi) = \sum_{i \in I_0} z_{is} + \sum_{j \in I} \{w(j) - 1\} z_{js} H(j), \quad s = 1, .., p.$$

Resolviendo el sistema de ecuaciones de verosimilitud

$$\begin{cases} N_{0r} + \sum_{j \in R_r} \{w(j) - 1\} \sigma_{jr} = 0, & r = 1, .., k+1, \\ \sum_{i \in I_0} z_{is} + \sum_{j \in I} \{w(j) - 1\} z_{js} H(j) = 0, & s = 1, .., p, \end{cases}$$

obtendríamos las estimaciones de los parámetros desconocidos.

### 3.2 Distribución Weibull a Trozos

Ahora consideramos que $g(t) = \ln \Lambda(t) = \alpha_j \ln t + \varepsilon_j$ si $t \in (a_{j-1}, a_j]$, donde $\alpha_j > 0$, para $j = 1, .., k+1$. Obsérvese que si sustituimos $\ln t$ por $t$ en la expresión anterior se obtiene la distribución de valor extremo a trozos. Suponemos además que $g$ es continua.

Por la continuidad de $g$ en los puntos de cambio $a_1, .., a_k$ se obtiene con facilidad que $\alpha_j \ln a_j + \varepsilon_j = \alpha_{j+1} \ln a_j + \varepsilon_{j+1}$ para $j = 1, .., k$, con lo cual $\varepsilon_j = \varepsilon_1 + \sum_{r=2}^{j} (\alpha_{r-1} - \alpha_r) \ln a_{r-1}$, $j = 2, .., k+1$. Luego, para $i = 1, .., N$, $g(t_i) = \varepsilon_1 + \alpha_{m_i} \ln t_i + \sum_{r=2}^{m_i} (\alpha_{r-1} - \alpha_r) \ln a_{r-1}$, donde se omite el sumatorio en $r$ si $m_i = 1$.

Supuesto el modelo log-lineal para $\psi(z; \beta)$, $H(i) = \exp\{g(t_i) + \beta z_i\}$ para $i = 1, .., N$. Por consiguiente, de la expresión (3) se tiene que

$$\frac{\partial}{\partial \varepsilon_1} l(\phi) = N_0 + \sum_{j \in I} \{w(j) - 1\} H(j), \tag{6}$$

donde $N_0$ es el número de datos no censurados, y de (5) obtenemos que

$$\frac{\partial}{\partial \beta_s} l(\phi) = \sum_{i \in I_0} z_{is} + \sum_{j \in I} \{w(j) - 1\} z_{js} H(j), \quad s = 1, .., p. \tag{7}$$

Como $g'(t_i) = \alpha_{m_i}/t_i$ y $\partial \ln g'(t_i)/\partial \alpha_r = I(m_i = r)/\alpha_r$, resulta de (4) que

$$\frac{\partial}{\partial \alpha_r} l(\phi) = \sum_{i \in I_0} \left\{ \mathrm{I}(m_i = r) \frac{1}{\alpha_r} + g_r(i) \right\} + \sum_{j \in I} \{w(j) - 1\} g_r(j) H(j), \quad r = 1, .., k+1, \tag{8}$$

donde $g_r(i) = \dfrac{\partial}{\partial \alpha_r} g(t_i) = \mathrm{I}(m_i = r) \ln t_i + \mathrm{I}(m_i > r) \ln a_r - \mathrm{I}(1 < r \leq m_i) \ln a_{r-1}$.

Las estimaciones máximo-verosímiles de los parámetros desconocidos se obtendrían resolviendo el sistema que resulta de igualar a cero las expresiones (6), (7) y (8).

### 3.3 Distribución de Valor Extremo Generalizada a Trozos

En este caso, $g(t) = \ln \Lambda(t) = \alpha_j t^{\gamma_j} + \varepsilon_j$ si $t \in (a_{j-1}, a_j]$, donde $\alpha_j > 0$ y $\gamma_j > 0$, para $j = 1, .., k+1$. Por la supuesta continuidad de la función $g$ en $a_j$, $j = 1, .., k$, se cumple que $\alpha_j a_j^{\gamma_j} + \varepsilon_j = \alpha_{j+1} a_j^{\gamma_{j+1}} + \varepsilon_{j+1}$, de lo cual obtenemos que

$$\varepsilon_j = \varepsilon_1 + \alpha_1 a_1^{\gamma_1} + \sum_{r=2}^{j-1} \alpha_r (a_r^{\gamma_r} - a_{r-1}^{\gamma_r}) - \alpha_j a_{j-1}^{\gamma_j}, \quad j = 2, .., k+1,$$

donde el sumatorio en $r$ se omite si $j = 2$. De este modo, resulta que

$$g(t_i) = \varepsilon_1 + \mathrm{I}(m_i = 1) \alpha_1 t_i^{\gamma_1} + \mathrm{I}(m_i > 1) \left\{ \alpha_{m_i} (t_i^{\gamma_{m_i}} - a_{m_i-1}^{\gamma_{m_i}}) + \alpha_1 a_1^{\gamma_1} + \sum_{r=2}^{m_i - 1} \alpha_r (a_r^{\gamma_r} - a_{r-1}^{\gamma_r}) \right\}$$

para $i = 1, .., N$, donde el sumatorio en $r$ desaparece si $m_i \leq 2$.

En esta ocasión, debemos resolver el sistema de $(p + 2k + 3)$ ecuaciones

$$\begin{cases} \dfrac{\partial}{\partial \varepsilon_1} l(\phi) = 0 \; ; \; \dfrac{\partial}{\partial \beta_s} l(\phi) = 0, \; s = 1, .., p, \\ \dfrac{\partial}{\partial \alpha_r} l(\phi) = 0 \; , \; \dfrac{\partial}{\partial \gamma_r} l(\phi) = 0, \; r = 1, .., k+1. \end{cases}$$

Considerando el modelo log-lineal para el multiplicador $\psi$, obtenemos de (5) que

$$\frac{\partial}{\partial \beta_s} l(\phi) = \sum_{i \in I_0} z_{is} + \sum_{j \in I} \{w(j) - 1\} \, z_{ij} H(j), \; s = 1, .., p,$$

donde $H(j) = \exp\{g(t_j) + \beta z_j\}$ para $j = 1, .., N$, y de (3) resulta que

$$\frac{\partial}{\partial \varepsilon_1} l(\phi) = N_0 + \sum_{j \in I} \{w(j) - 1\} \, H(j).$$

Sustituyendo

$$\frac{\partial}{\partial \alpha_r} g(t_i) = \mathrm{I}(m_i > r)(a_r^{\gamma_r} - a_{r-1}^{\gamma_r}) + \mathrm{I}(m_i = r)(t_i^{\gamma_r} - a_{r-1}^{\gamma_r}) \; \text{ y } \; \frac{\partial}{\partial \alpha_r} \ln g'(t_i) = \mathrm{I}(m_i = r) \frac{1}{\alpha_r}$$

en la fórmula general (4) hallaríamos $\partial l(\phi)/\partial \alpha_r$ , $r = 1, .., k+1$.

A su vez, obtendríamos $\partial l(\phi)/\partial \gamma_r$ , $r = 1, .., k+1$, a partir de (4), utilizando que

$$\begin{aligned} \frac{\partial}{\partial \gamma_1} g(t_i) &= \alpha_1 \{\mathrm{I}(m_i = 1) t_i^{\gamma_1} \ln t_i + \mathrm{I}(m_i > 1) a_1^{\gamma_1} \ln a_1\}, \\ \frac{\partial}{\partial \gamma_r} g(t_i) &= \alpha_r \{\mathrm{I}(m_i = r)\,(t_i^{\gamma_r} \ln t_i - a_{r-1}^{\gamma_r} \ln a_{r-1}) \\ &\quad + \mathrm{I}(m_i > r)\,(a_r^{\gamma_r} \ln a_r - a_{r-1}^{\gamma_r} \ln a_{r-1})\}, \; r = 2, .., k+1, \\ \frac{\partial}{\partial \gamma_r} \ln g'(t_i) &= \mathrm{I}(m_i = r)\,(1/\gamma_r + \ln t_i), \; r = 1, .., k+1. \end{aligned}$$

Nótese que si $\gamma_j = 1$ para $j = 1, .., k+1$, aparece la distribución de valor extremo, y que si hacemos $\gamma_j \to 0^+$, obtenemos la distribución de Weibull. Por tanto, podemos utilizar el estadístico de razón de verosimilitudes para contrastar la hipótesis nula de distribución de Weibull (o de valor extremo) frente a la alternativa de una distribución de valor extremo generalizada.

## 4. PROCEDIMIENTOS GRÁFICOS

Los procedimientos gráficos pueden utilizarse tanto en la exploración preliminar de los datos como en la comprobación de la validez de los modelos considerados, e incluso para la obtención

de estimaciones aproximadas de los parámetros desconocidos. La idea básica es la construcción de gráficas que serían aproximadamente lineales si el modelo propuesto se ajustase a los datos obtenidos. De este modo, puede apreciarse a simple vista las desviaciones de la linealidad, las cuales proporcionarían evidencias de la incorrección del modelo supuesto.

Como análisis previo es aconsejable agrupar los individuos en subconjuntos con valores similares de las principales covariables y examinar gráficamente la estimación autoconsistente (Gu y Zhang [6]; Mykland y Ren [9]), $S^a(t)$, de la función de supervivencia para cada grupo. Podemos considerar apropiadas las distribuciones de Weibull (exponencial), de valor extremo o de valor extremo generalizada si la gráfica de $\ln\{-\ln S^a(t)\}$ es aproximadamente lineal a trozos en $\ln t$ (y con pendiente uno), en $t$ o en algunas potencias de $t$, respectivamente. Además, este procedimiento puede sugerirnos la localización aproximada de los puntos de cambio.

En muchas ocasiones el estudio de las características físicas del proceso de supervivencia nos permite obtener una estimación aceptable de los puntos de cambio. Sin embargo, la elección final de éstos debe ser optimizada analíticamente. En la mayoría de casos prácticos es suficiente considerar uno o dos puntos de cambio.

Puesto que la función de supervivencia de $X$, dado el vector de factores de riesgo $z$, es $S(x;z,\phi)$, la variable aleatoria $S(X;z,\phi)$ está uniformemente distribuida, con lo cual $H(X;z,\phi) = -\ln S(X;z,\phi)$ tiene distribución exponencial de media uno. Por consiguiente, $H(i) = H(t_i;z_i,\phi)$, $i = 1,..,N$, constituye una muestra aleatoria (doblemente censurada) de una distribución exponencial unitaria. Una vez estimado el vector $\phi$, es posible contrastar la validez del modelo ajustado analizando la variable residual $R = H(X;z,\widehat{\phi})$, cuya distribución debería ser aproximadamente exponencial unitaria, esto es, analizando los residuos estimados $r_i = \widehat{H}(i)$, $i = 1,..,N$. Si el modelo propuesto es apropiado, $r_1,..,r_N$ deben comportarse aproximadamente como un conjunto de datos (doblemente censurados) de la distribución exponencial unitaria, aunque no serán independientes. Sin embargo, como en el modelo de regresión normal, si el número de parámetros estimados no es grande en relación al número de datos no censuradas, la distribución de los residuos debería ser robusta frente a la falta de independencia y de exponencialidad exacta.

En nuestro caso, proponemos la estimación de la supervivencia residual, $\tilde{S}(r)$, como el valor esperado de la supervivencia residual empírica, $S^e(r) = N^{-1}\sum_{i=1}^{N} \mathrm{I}(R_i > r)$, condicionada a los datos observados y al modelo exponencial unitario (véase Bravo et al. [2]), esto es,

$$\tilde{S}(r) = \frac{1}{N}\left\{\sum_{i\in I} \mathrm{I}(r_i > r) + \sum_{j\in I_1} \mathrm{I}(r_j \le r)\exp(r_j - r) - \sum_{k\in I_2} \mathrm{I}(r_k > r)\frac{1-\exp(-r)}{1-\exp(-r_k)}\right\}.$$

Si el modelo es válido, la representación gráfica de $-\ln \widetilde{S}(r)$ frente a $r$ debe aproximarse a la bisectriz del primer cuadrante. Para evitar la excesiva variabilidad de $\widetilde{S}(r)$ cuando $r$ es grande, y considerando que $\widetilde{S}(r)$ es esencialmente una probabilidad binomial para cada $r$, también es aconsejable representar la transformación estabilizadora de la varianza $\operatorname{arcsen}\left\{\sqrt{\widetilde{S}(r)}\right\}$ frente a $\operatorname{arcsen}\{\exp(-r/2)\}$.

## BIBLIOGRAFÍA


[1] AITKIN, M. y CLAYTON, D. (1980). The fitting of exponential, Weibull and extreme value distributions to complex censored survival data using GLIM. *Appl. Statist.*, **29**, 156-163.

[2] BRAVO, J. I., DE FUENTES, I. y FERNÁNDEZ, A. J. (1995). A semi-parametric estimation of a survival function from incomplete and doubly censored data. *Commun. Statist.-Theory Meth.*, **24**, 2735-2752.

[3] FERNÁNDEZ, A. J., BRAVO, J. I. y DE FUENTES, I. (1997). Lifetime estimation from doubly censored data and Dirichlet process prior knowledge on the observable random vector, *Commun. Statist.-Theory Meth.*, **26**, 1541-1558.

[4] COX, D. R. (1972). Regression models and life-tables (with discussion), *J. Roy. Statist. Soc. B*, **34**, 187-220.

[5] GARCÍA, J., LARA, A. M., OLLERO, J. y PÉREZ, R. (1993). A study of a survival model with a piecewise generalized extreme value distribution. ASMDA 93: *Proc. Int. Symp. Appl. Stochastic Models and Data Analysis* (eds. J. Janssen and C. H. Skiadas), World Scientific Publ. Co., 265-274.

[6] GU, M. G. y ZHANG, C. H. (1993). Asymptotic properties of self-consistent estimators based on doubly censored data. *Ann. Statist.*, **21**, 611-624.

[7] LEHMANN, E. L. (1983). *Theory of Point Estimation*, John Wiley, New York.

[8] MÄKELÄINEN, T., SCHMIDT, K. y STYAN, G. P. H. (1981). On the existence and uniqueness of the maximum likelihood estimate of a vector-valued parameter in fixed-sized samples, *Ann. Statist.*, **9**, 758.767.

[9] MYKLAND, P. A. y REN, J.-J. (1996). Algorithms for computing self-consistent and maximum likelihood estimators with doubly censored data, *Ann. Statist.*, **24**, 1740-1764.

[10] NOURA, A. A. y READ, K. L. Q. (1990). Proportional hazards changepoint models in survival analysis, *Appl. Statist.*, **39**, 241-253.

[11] RAO, C. R. (1973). *Linear statistical inference and its applications*, John Wiley & Sons, New York.

[12] RICHARDS, F. S. G. (1961). A method of maximum-likelihood estimation, *J. Roy. Statist. Soc. B*, **23**, 469-475.

[13] ZHANG C.-H. y LI, X. (1996). Linear regression with doubly censored data, *Ann. Statist.*, **24**, 2720-2743.