ESTADÍSTICA (2012), 64, 182 y 183, pp. 85-106

REVISIÓN DE LA ESTIMACIÓN ROBUSTA EN MODELOS SEMIPARAMÉTRICOS DE SUPERVIVENCIA

ENRIQUE E. ÁLVAREZ
Universidad Nacional de La Plata y CONICET
ealvarez@mate.unlp.edu.ar

JULIETA FERRARIO
Universidad Nacional de La Plata y CONICET
jferrario@mate.unlp.edu.ar

RESUMEN

En Análisis de Supervivencia se analizan datos referidos al tiempo final de ocurrencia de un evento, T, y asociado a éste se recogen un vector de variables explicativas independientes o "covariables", $\boldsymbol{Z}$. Lo que se desea es modelar la relación entre T y $\boldsymbol{Z}$, y el enfoque más común para esto se basa sobre la función de intensidad o tasa de riesgo, definida como

$$\lambda(t) := \lim_{\epsilon \downarrow 0} \frac{\Pr\left(T \leq t + \epsilon | T > t\right)}{\epsilon}$$

que representa el riesgo instantáneo en el tiempo t. Una generalización de los modelos para la función de riesgo incluye variables regresoras. Éstos pueden ser formados de varias maneras y los tres modelos semiparamétricos más utilizados y a los que hacemos referencia aquí son: de riesgo proporcional, de tiempo de falla acelerado y de riesgo aditivo. El objetivo de esta revisión es sintetizar las propuestas de robustificación realizadas hasta el momento para los modelos proporcional, de falla acelerado y aditivo, comentando posibles generalizaciones y extensiones.

Palabras clave

Modelos de riesgo proporcional; modelo de riesgo de falla acelerado; modelo de riesgo aditivo; estimación robusta.

ABSTRACT

Survival Analysis analyzes data referring to times until the occurrence of an event, T, which is collected together with a vector of independent variables or "covariates", $\boldsymbol{Z}$. What is desired is to model the relationship between T and $\boldsymbol{Z}$, and the most common approach is based on the intensity function or hazard rate, defined as

$$\lambda(t) := \lim_{\epsilon \downarrow 0} \frac{\Pr\left(T \leq t + \epsilon | T > t\right)}{\epsilon},$$

which represents the instantaneous risk at time t. A generalization of the models for the hazard function includes covariates. These can be formed in various ways and the three most commonly used semiparametric models and are referred in the review. They are the proportional hazards, the accelerated failure time and the additive risks models. The aim of this review is to summarize the proposals made for robustification in the proportional model, the accelerated failure time model and the additive model, commenting on possible generalizations and extensions.

Keywords

Proportional hazards model; accelerated failure time model; additive hazards model; robust estimation.

1. Introducción

Para analizar datos referidos al tiempo final de ocurrencia, T, de un evento, éste se refiere como tiempo de falla. Estos tiempos suelen ser en muchos casos muy grandes, entonces en lugar de esperar hasta su falla se censuran. Dentro de los tipos de censura los que se utilizarán aquí son: censuras fijas (se observa $T \wedge \tau$, con τ fijo y finito) o censuras aleatorias (se observa $T \wedge C$, donde C es una variable aleatoria no observada independiente de T). El evento que se analiza puede ser recurrente o no, y además su tiempo de falla puede ser causado por varios motivos. Asociado a cada tiempo de falla del evento se recoge un vector de covariables $\boldsymbol{Z} \in \mathbb{R}^p$ que puede incluir variables cuantitativas, cualitativas, que dependan o no del tiempo e incluso pueden ser variables externas (independientes del proceso del evento recurrente) y/o internas (si no es externa).

Para modelar y determinar la distribución de T, $F(t)$, la que suponemos continua en todo este trabajo (es decir, $F(t) = \int_0^t f(u)\, du$) se definen dos funciones importantes y útiles en las aplicaciones de supervivencia:

- la función de supervivencia

$$S(t) = 1 - F(t);$$

- la función de riesgo o función de intensidad de falla

$$\lambda(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} \ln\left(S(t)\right).$$

Nosotros nos focalizamos sólo en los modelos basados sobre la función de intensidad. Esta función, antes ya definida, también se la puede definir a través del proceso de conteo de la ocurrencia del evento, $\{N(t) : t \geq 0\}$. La función de intensidad de un individuo para el proceso del evento es definida como

$$\lambda\left[t|H(t)\right] = \lim_{\Delta t \downarrow 0} \frac{\Pr\left[\Delta N(t) = 1|H(t)\right]}{\Delta t} \ ,$$

donde $\Delta N(t) = N(t + \Delta t^-) - N(t^-)$ y $H(t) = \{N(s) : 0 \leq s < t\}$ es la historia de los eventos antes de t del individuo. Esto es, $\lambda[t|H(t)]\Delta t$ es, para Δt chico, la probabilidad aproximada de que un evento ocurra en $(t, t + \Delta t^-]$ dado la historia del proceso antes de t.

Una generalización de los modelos para la función de intensidad del tiempo de falla T asociada con el vector $\boldsymbol{Z}$ incluye variables regresoras. Éstas pueden ser formadas de varias maneras y las tres más comunes y las cuales nos referiremos aquí, para cada individuo con covariables $\boldsymbol{Z}$, son

1. Modelo de riesgo proporcional o riesgo multiplicativo de Cox

$$\lambda(t|\boldsymbol{Z}) = \lambda_0(t) \exp\left[\boldsymbol{\beta}_0' \boldsymbol{Z}(t)\right]$$

2. Modelo de tiempo de falla acelerado

$$\lambda(t|\boldsymbol{Z}) = \lambda_0 \left\{t \exp\left[\boldsymbol{\beta}_0' \boldsymbol{Z}(t)\right]\right\} \exp\left[\boldsymbol{\beta}_0' \boldsymbol{Z}(t)\right]$$

3. Modelo de riesgo aditivo

$$\lambda(t|\boldsymbol{Z}) = \lambda_0(t) + \boldsymbol{\beta}_0' \boldsymbol{Z}(t)$$

donde $\boldsymbol{\beta}_0 \in \mathbb{R}^p$ es el vector de parámetro de regresión y $\lambda_0(t) = \lambda(t|\boldsymbol{Z} = 0)$ es la función baseline, desconocida, arbitraria y no negativa en función del tiempo. En los tres casos se desea estimar el vector de parámetro de regresión, $\boldsymbol{\beta}_0$, y la función baseline, $\lambda_0(t)$. Si bien esta función puede ser considerada en forma paramétrica (por ejemplo, los modelos exponencial, Weibull, Gamma, log-normal, etc.), aquí sólo veremos modelos en los que se la consideran en forma no paramétrica. Es decir, en esta revisión sólo veremos modelos semiparamétricos, ya que se supone un modelo paramétrico solamente para el efecto de la variable independiente $\boldsymbol{Z}$. Observemos que la función de intensidad es modificada por una proporción (en el modelo de Cox), una reescalación (en el modelo de falla acelerado) y una traslación (en el modelo aditivo).

Haremos, para cada uno de los modelos semiparamétricos anteriores, una revisión y análisis de algunos trabajos focalizándonos solamente sobre los resultados estadísticos de la estimación del parámetro de regresión, motivados por la búsqueda de estimadores menos sensibles que los métodos clásicos ante observaciones extremas y de alta palanca y/o a perturbaciones en la distribución de F, que se han estudiado hasta el momento. Si bien es más difícil estimar en modelos semiparamétricos [ya que en ellos hay al menos un parámetro de dimensión infinita, como es el caso de $\lambda_0(\cdot)$] que en modelos paramétricos, cabe destacar que los procedimientos de estimación descriptos aquí estiman al parámetro de regresión $\boldsymbol{\beta}_0$ sin la necesidad de estimar conjuntamente ni previamente a la función λ_0. Algunos procedimientos robustos basados en modelos paramétricos se pueden encontrar por ejemplo en Huber (1981), Beran (1981), Basu, Basu y Jones (2006) y Maronna, Martin y Yohai (2006), entre otros.

En la siguiente sección desarrollamos las propuestas de estimación del parámetro para el modelo de riesgo proporcional en el que exponemos, para el caso en que el evento sea simple, el conocido estimador de verosimilitud parcial (o estimador de Cox, 1972 y 1975) y dos propuestas robustas, la de Sasieni (1993a y 1993b) y la de Bednarski (1993). También desarrollamos la extensión del estimador de Cox, que realizan Huang y Chen (2003), a eventos recurrentes. En la Sección 3 desarrollamos las propuestas robustas (para eventos simples) de Salibian–Barrera y Yohai (2008), y (para eventos recurrentes) de Lin, Wei y Ying (1998) y Strawderman (2005) para el modelo de falla acelerado. Por último, en la Sección 4 para la estimación en el modelo aditivo daremos las ideas de Lin y Ying (1994), para eventos simples y Sun, Park y Sun (2006) para eventos recurrentes.

2. Modelo de Riesgo Multiplicativo

En el modelo clásico de riesgo proporcional de Cox (1972), los datos provienen de un evento simple y donde la variable explicativa es independiente del tiempo. Se observan las m ternas independientes $(t_i, \boldsymbol{Z}_i, \Delta_i)$, $i = 1 : m$, donde para cada individuo i, t_i es el tiempo de falla observado (censurado o no), $\boldsymbol{Z}_i \in \mathbb{R}^p$ su vector de covariable asociado y Δ_i vale 0 si t_i es un tiempo de falla censurado a derecha y vale 1 en caso contrario. Entonces, para el tiempo de supervivencia t_i del individuo i con vector de variables independientes $\boldsymbol{Z}_i$, la función de intensidad es de la forma

$$\lambda(t_i|\boldsymbol{Z}_i) = \lambda_0(t_i)\exp(\boldsymbol{\beta}_0'\boldsymbol{Z}_i) \;, \tag{1}$$

donde $\boldsymbol{\beta}_0$ es un $p-$vector de coeficientes de regresión desconocido y λ_0, la función de riesgo para $\boldsymbol{Z}_i = \mathbf{0}$ (conocida como baseline), es también desconocida y es una función arbitraria no negativa del tiempo.

Para estimar $\boldsymbol{\beta}_0$ en (1), Cox (1975) por la falta de conocimiento de $\lambda_0(\cdot)$, propuso un método de estimación denominado *verosimilitud parcial* (VP). La función de verosimilitud parcial no es una verosimilitud en el sentido usual sino que fue motivada condicionando sobre los tiempos de fallas observados t_i, donde la probabilidad condicional de que halla una falla en t_i dado los casos que se encuentran en riesgo de fallar en el tiempo t_i (es decir, aquellos para los cuales $t_k \geq t_i$) es

$$\frac{\lambda(t_i|\boldsymbol{Z}_i)}{\sum_{t_k \geq t_i} \lambda(t_i|\boldsymbol{Z}_k)} \ .$$

Luego la función de VP es

$$L_P(\boldsymbol{\beta}) = \prod_{i=1}^{m} \left[\frac{\lambda(t_i|\boldsymbol{Z}_i)}{\sum_{t_k \geq t_i} \lambda(t_i|\boldsymbol{Z}_k)} \right]^{\Delta_i} = \prod_{i=1}^{m} \left[\frac{\exp(\boldsymbol{\beta}'\boldsymbol{Z}_i)}{\sum_{t_k \geq t_i} \exp(\boldsymbol{\beta}'\boldsymbol{Z}_k)} \right]^{\Delta_i} ,$$

donde se canceló el parámetro de "nuisance" $\lambda_0(t)$. Esta función es tratada como la verosimilitud usual. Notar que el numerador depende únicamente de la información del individuo quien experimentó el evento, mientras que el denominador utiliza información sobre todos los individuos quienes no han experimentado el evento (incluyendo algunos individuos quienes serían luego censurados).

Entonces, el estimador de VP de $\boldsymbol{\beta}_0$ es solución de la siguiente ecuación de score

$$\boldsymbol{U}_P(\boldsymbol{\beta}) := \sum_{i=1}^{m} \Delta_i \left\{ \boldsymbol{Z}_i - \frac{\sum_{t_k \geq t_i} \boldsymbol{Z}_k \exp(\boldsymbol{\beta}'\boldsymbol{Z}_k)}{\sum_{t_k \geq t_i} \exp(\boldsymbol{\beta}'\boldsymbol{Z}_k)} \right\} = \mathbf{0}. \tag{2}$$

Bajo ciertas condiciones de regularidad, el estimador es consistente y asintóticamente normal.

Andersen y Gill (1982) extienden el modelo de Cox para eventos recurrentes y para variables explicativas externas que dependen del tiempo. Consideran el proceso de conteo multivariado de m componentes $N = (N_1, \ldots, N_m)$ de la vida de m individuos, donde N_i es la cantidad de eventos observados en la vida del i-ésimo individuo, $i = 1 : m$ sobre el intervalo $[0, 1]$. Es decir que los tiempos de falla serán censurados por derecha. Ellos trabajan en este intervalo de tiempo por simplicidad pero hacen una discusión en su trabajo (Sección 4) de cómo extender los resultados a $[0, \infty)$. Asumen que N tiene un proceso de intensidad aleatorio $\lambda = (\lambda_1, \ldots, \lambda_m)$ tal que

$$\lambda_i(t|\boldsymbol{Z}_i) = Y_i(t)\lambda_0(t) \exp\left[\boldsymbol{\beta}_0'\boldsymbol{Z}_i(t)\right], \tag{3}$$

donde $Y_i(t)$ es el proceso $\{0, 1\}$ continuo a derecha, que indica con 1 cuando el $i-$ésimo individuo está bajo observación en el tiempo t, y que el vector de dimensión p de procesos de variables independientes $\boldsymbol{Z}_i = (Z_{i1}, \ldots, Z_{ip})$ es predecible y localmente acotado.

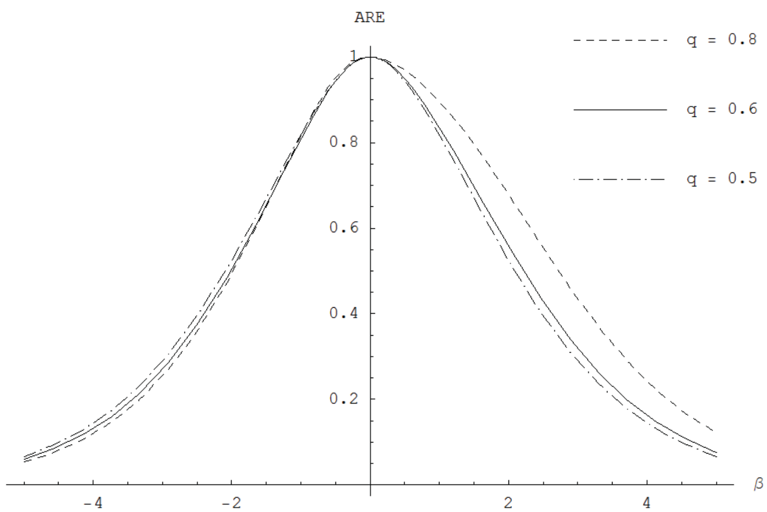
El estimador de VP de $\boldsymbol{\beta}_0$ en (3) se define como la solución de la ecuación

$$\boldsymbol{U}_{AG}(\boldsymbol{\beta}) = \sum_{i=1}^{m} \int_0^1 \boldsymbol{Z}_i(s)\, dN_i(s) - \int_0^1 \frac{\sum_{i=1}^{m} \boldsymbol{Z}_i(s) Y_i(s) \exp\left[\boldsymbol{\beta}'\boldsymbol{Z}_i(s)\right]}{\sum_{i=1}^{m} Y_i(s) \exp\left[\boldsymbol{\beta}'\boldsymbol{Z}_i(s)\right]}\, d\bar{N}(s) = \mathbf{0},$$

donde $\bar{N} = \sum_{j=1}^{m} N_j$, que, bajo ciertas condiciones de regularidad y utilizando resultados de martingalas locales, es consistente y asintóticamente normal. Observar que en el caso especial en que las covariables $\boldsymbol{Z}_i$ sean independientes del tiempo, las funciones $\boldsymbol{U}_P(\boldsymbol{\beta})$ y $\boldsymbol{U}_{AG}(\boldsymbol{\beta})$ coinciden.

Pero el estimador de VP, $\widehat{\boldsymbol{\beta}}$, pierde eficiencia con respecto al estimador de máxima verosimilitud cuando $\boldsymbol{\beta}_0$ se aleja del origen (por ejemplo, para $p = 1$, ver la Figura 1), y tiene función de influencia no acotada (Reid y Crépeau, 1985) haciéndose sensible a observaciones extremas.

Figure 1. Para un problema de dos muestras ($Z = 0, 1$), con ausencia de censura y $\lambda(t|Z) = \lambda_0 \exp(\beta_0 Z)$ (λ_0 es constante, independiente de t), Efron (1977, ecuación 4.11) calculó la eficiencia relativa asintótica (ARE) del estimador de VP comparada con el estimador de máxima verosimilitud. Ésta es su gráfica donde q representa la proporción de 1 en la muestra.



Uno de los autores que buscaron estimadores robustos a partir del estimador de VP, fue Sasieni (1993a). Él propone una familia de estimadores que maximizan la VP "pesada" (nos referiremos a ellos como "estimadores pesados"), es decir, le asigna, a través de una función de peso, más importancia a lo que sucede en algunos tiempos de fallas que en otros. Esta función es $w(t, \mathbb{P}_m)$, donde $\mathbb{P}_m$ es la medida empírica del tiempo de falla con masa $1/m$ en cada punto, tal que $w(\cdot, \mathbb{P}_m)$ sea predecible, no negativa y localmente acotada (entre otras condiciones de regularidad). Formalmente, Sasieni propone el estimador pesado de $\boldsymbol{\beta}_0$ como solución de

$$\boldsymbol{U}_S(\boldsymbol{\beta}; w) := \sum_{i=1}^{m} w(t_i, \mathbb{P}_m) \left\{ \boldsymbol{Z}_i(t_i) - \frac{\sum_{j=1}^{m} Y_j(t_i) \boldsymbol{Z}_j(t_i) \exp\left[\boldsymbol{\beta}' \boldsymbol{Z}_j(t_i)\right]}{\sum_{j=1}^{m} Y_j(t_i) \exp\left[\boldsymbol{\beta}' \boldsymbol{Z}_j(t_i)\right]} \right\} \Delta_i = \mathbf{0} \, .$$

Este estimador resulta ser, bajo ciertas condiciones, consistente de tasa $\sqrt{m}$ y asintóticamente normal.

Sasieni propuso que $w(t, \mathbb{P}_m) = \widehat{S}(t, \widehat{\boldsymbol{\beta}})$, donde $\widehat{S}(t, \mathbf{0})$ es el estimador de Kaplan-Meier (KM) (Kaplan y Meier, 1958) de la función de supervivencia marginal (es decir, ignorando las variables explicativas) y a este estimador, asumiendo que las covariables son acotadas, lo llama el estimador de Wilcoxon (por la analogía que tiene con respecto al test de Wilcoxon). De esta manera, Sasieni logra dar menos peso a observaciones grandes. La eficiencia relativa asintótica del estimador de Wilcoxon con respecto al de VP (que fue calculada por Sasieni 1993a, Corolario B, página 147) para el caso en que Z sea univariado, acotado e independiente del tiempo) crece a medida que β_0 se aleja del 0 y además va aumentando a medida que se aumenta el porcentaje de censura. Esto se ve reflejado en la simulación que él realiza. Pero, por un lado, al utilizar un estimador preliminar de $\boldsymbol{\beta}_0$ la función de peso $w(t, \mathbb{P}_m)$ dejará de ser predecible y, por otro lado, el estimador de KM no es robusto. En este sentido, Reid (1981) analizó la sensibilidad de la función de influencia del estimador de KM al agregar una observación atípica.

Cuando el modelo de Cox se verifica, el estimador de VP para $w = 1$ es eficiente (Begun, Hall, Huang y Wellner, 1983; Efron, 1977), es decir, tiene mínima varianza entre los estimadores pesados; si bien tiene función de influencia no acotada. En cualquier otro caso, cuando se considera una función de peso $w(t, \mathbb{P}_m)$ que no depende de la muestra y que no es constantemente igual a 1, o bien, que depende de la muestra, los correspondientes estimadores pesados tendrán asintóticamente la misma eficiencia que el estimador VP pero serán menos eficientes que él.

En cuanto a la función de influencia, Sasieni nota que el problema de que el estimador de VP tenga función de influencia no acotada asintóticamente es por dos potenciales razones: una, por los valores de $\boldsymbol{Z}$ "extremos" y, la otra, por los individuos que más sobrevivieron, es decir, por los valores grandes de T. El primero, no se soluciona sin importar la función de peso $w(t, \mathbb{P}_m)$ que se seleccione, ya que ésta sólo depende de los tiempos de falla y de su distribución empírica. Entonces asumiendo que $\boldsymbol{Z}$ está acotado, cuando el estimador es el de Wilcoxon se obtiene un estimador con función de influencia acotada sin importar que $\mathbb{P}_m$ sea o no un miembro del modelo Cox. Una medida de probabilidad $\mathbb{P}$ sobre $(\boldsymbol{Z}, T, \Delta)$ es un miembro del modelo de Cox con parámetro $\boldsymbol{\beta}_0$, que denotamos $\mathbb{P} \in \mathcal{P}(\boldsymbol{\beta}_0)$, si existen T^u y T^c independientes condicionalmente dado $\boldsymbol{Z}$ tal que $T = \min(T^u, T^c)$, $\Delta = I(T^u \leq T^c)$ y el riesgo proporcional de T^u dado $\boldsymbol{Z}$ en t es $\lambda(t|\boldsymbol{Z}) = \lambda_0(t) \exp\left[\boldsymbol{\beta}_0' \boldsymbol{Z}(t)\right]$.

Por lo tanto, para el método propuesto por Sasieni la selección de la función de peso $w(t, \mathbb{P}_m)$ a menudo será una compensación entre eficiencia y robustez.

Por otro lado, en otro trabajo de Sasieni (1993b) amplia la familia de los estimadores anteriormente expuestas permitiendo funciones de pesos que no sólo dependan de los tiempos de falla y de su distribución empírica sino también que dependan de $\boldsymbol{Z}$. Esta familia más grande de estimadores la llamó *estimadores de clase* K y los definió de la siguiente manera: Sea K una función medible de $\mathbb{R}^p \times \mathbb{R}^+ \times \mathcal{Q} \to \mathbb{R}^p$, donde $\mathcal{Q}$ es una extensión de $\mathcal{P}(\boldsymbol{\beta}_0)$, que lo contiene y a todas las posibles distribuciones empíricas $\mathbb{P}_m$, entonces, bajo ciertas condiciones, el estimador de clase K, $\widehat{\boldsymbol{\beta}}_K$, es solución de

$$\sum_{i=1}^{m} \left\{ K(\boldsymbol{Z}_i, T_i, \mathbb{P}_m)\boldsymbol{Z}_i - \frac{\sum_{k=1}^{m} K(\boldsymbol{Z}_k, T_i, \mathbb{P}_m) Y_k(T_i) \boldsymbol{Z}_k \exp(\boldsymbol{\beta}'\boldsymbol{Z}_k)}{\sum_{k=1}^{m} Y_k(T_i)\exp(\boldsymbol{\beta}'\boldsymbol{Z}_k)} \right\} \Delta_i = \mathbf{0}.$$

El estimador $\widehat{\boldsymbol{\beta}}_K$ es consistente de tasa $\sqrt{m}$ y asintóticamente normal.

Observación 1. *Esta clase de estimadores fueron primero propuestos por Ritov y Wellner (1987) pero estos autores utilizaron funciones de peso que sólo dependían de* $\boldsymbol{Z}$ *y* T*, mientras que Sasieni además le incorporó la dependencia de distribuciones empíricas de los tiempos de falla.*

Otro trabajo en el que se modificó la ecuación de estimación de score del riesgo proporcional (2) para obtener un estimador robusto, fue el de Bednarski (1993). Las modificaciones que realizó no sólo producen estimadores consistentes y asintóticamente normales de $\boldsymbol{\beta}_0$ para el modelo de riesgo proporcional sino que también para pequeños entornos del modelo, definidos como $\{G : \|G - F\|_\infty \leq \epsilon/\sqrt{m}\}$, donde F es la función de distribución acumulada "verdadera" de $(T, \boldsymbol{Z}, C)$ del modelo de Cox, donde C es la variable de censura independiente de T dado $\boldsymbol{Z}$. La herramienta importante que utiliza Bednarski es la diferenciabilidad Fréchet, con la que logra una función de influencia acotada y la norma del supremo del funcional del estimador conduce a un punto de ruptura no nulo.

La propuesta de Bednarski fue modificar la función de score $\boldsymbol{U}_P(\boldsymbol{\beta})$ (ver (2)), introduciéndole una función de peso $A(t, \boldsymbol{z})$ con ciertas propiedades de regularidad, como sigue

$$\boldsymbol{U}_B(\boldsymbol{\beta}; A) := \sum_{j=1}^{m} A(t_j, \boldsymbol{Z}_{(j)}) \left[\boldsymbol{Z}_{(j)} - \frac{\sum_{t_k \geq t_j} A(t_j, \boldsymbol{Z}_k)\boldsymbol{Z}_k \exp(\boldsymbol{\beta}'\boldsymbol{Z}_k)}{\sum_{t_k \geq t_j} A(t_j, \boldsymbol{Z}_k)\exp(\boldsymbol{\beta}'\boldsymbol{Z}_k)} \right] \Delta_j = \mathbf{0},$$

donde $\Delta_j := I(T_j \leq C_j)$. El efecto de la función $A(t, \boldsymbol{z})$ que está a la izquierda es para pesar hacia abajo las observaciones no censuradas con valores grandes de $t\exp(\boldsymbol{\beta}'\boldsymbol{z})$ y en las sumas del cociente, $A(t, \boldsymbol{z})$ es calculada para las "observaciones artificiales" ya que se combina el tiempo t_j con variables explicativas $\boldsymbol{Z}_k$, $t_k \geq t_j$, correspondientes a tiempos distintos, así pesa hacia abajo todas las observaciones con valores relativamente grande de $\boldsymbol{\beta}'\boldsymbol{z}$ entre todas aquellas con $t_k \geq t_j$. Con esta "doble poda" logra dar consistencia al estimador.

Por ejemplo, Bednarski (1993), Minder y Bednarski (1996), Bednarski y Nowak (2003), Bednarski y Mocarska (2006) y Bednarski (2007) proponen las siguientes funciones $A(t, \boldsymbol{z})$:

$$M - \min\left[M, t\exp(\boldsymbol{\beta}'\boldsymbol{z})\right],$$

$$M - \min\left[M, \Lambda(t)\exp(\boldsymbol{\beta}'\boldsymbol{z})\right],$$

$$\exp\left[\frac{-\Lambda(t)\exp(\boldsymbol{\beta}'\boldsymbol{z})}{\alpha M}\right],$$

donde M es una constante seleccionada apropiadamente, $\Lambda(t)$ es la función de riesgo acumulado y α es un factor de escala. Con estas funciones de peso, el estimador se calcula iterativamente y se estabiliza después de la tercera o cuarta iteración.

Por otro lado, cuando los datos provienen de un evento recurrente es usual modelarlos con intervalos de tiempo entre sus recurrencias. Es decir, sean $T_{i1} < \ldots < T_{in_i}$ los tiempos continuos de fallas observados donde $n_i := N_i(C_i) = \sum_{j\geq 1}\Delta_{ij}$ (con $\Delta_{ij} := I(T_{ij} \leq C_i)$), es el número de fallas del individuo i en el intervalo de tiempo $[0, C_i]$, donde C_i es el tiempo de censura del individuo i (con $i = 1 : m$). Luego se definen $X_{ij} := T_{i(j)} - T_{i(j-1)}$ como los tiempos de espera entre arribos o intervalos de tiempo entre la $(j-1)$ y $j-$ésima ocurrencia para el individuo i, y $j = 1 : n_i + 1$, donde $T_{i0} := 0$ y $T_{i,n_i+1} := C_i$.

Dentro de este marco, Huang y Chen (2003) desarrollan un método que, si bien no es robusto, lo destacamos por ser una buena técnica para extender cualquier método de eventos simples a eventos recurrentes y es utilizado por varios autores. Entre ellos, en la Sección 4, se describirá el trabajo de Sun, Park y Sun (2006).

Huang y Chen argumentan (al igual que el trabajo de Wang y Chang (1999) en el que estiman la función de supervivencia marginal de los tiempos entre dos eventos sucesivos) que, bajos ciertas suposiciones, los intervalos de tiempo observados completos, X_{ij} $(j = 1 : n_i)$, es decir, aquellos intervalos de tiempo no censurados, están idénticamente distribuidos. Entonces, la intercambiabilidad de los intervalos de tiempo observados completos sugiere que un subconjunto de los datos observados puede ser tratado como datos observados de supervivencia en clases. Además, el procedimiento de regresión de Cox estándar puede ser aplicado a los datos del primer intervalo de tiempo de cada individuo, y este primer intervalo puede ser reemplazado por una selección aleatoria de la misma clase. Si bien con esto se logrará una estimación más eficiente, no obstante es una aproximación muy costosa computacionalmente. Entonces, Huang y Chen, proponen estimar $\boldsymbol{\beta}_0$ utilizando la siguiente función de score

$$\boldsymbol{U}_{HCh}(\boldsymbol{\beta}) := \int_0^\tau \left\{ d\widehat{K}_1(s) - \frac{\widehat{E}_{ij}\left[\boldsymbol{Z}_i\exp(\boldsymbol{\beta}'\boldsymbol{Z}_i)\tilde{Y}_{ij}(s)\right]}{\widehat{E}_{ij}\left[\exp(\boldsymbol{\beta}'\boldsymbol{Z}_i)\tilde{Y}_{ij}(s)\right]}\, d\widehat{K}_0(s) \right\},$$

donde τ es una constante (que por razones técnicas tiene cierta propiedad), $\tilde{Y}_{ij}(s) := I(X_{ij} \geq s)$, $\widehat{K}_0(s) = \widehat{E}_{ij}\left[\Delta_i\, I(X_{ij} \leq s)\right]$, $\widehat{K}_1(s) = \widehat{E}_{ij}\left[\boldsymbol{Z}_i \Delta_i\, I(X_{ij} \leq s)\right]$, $\widehat{E}_{ij} := \widehat{E}_i\, \widehat{E}_j$ y $\widehat{E}_j$ representa el promedio empírico sobre $j = 1 : n_i^*$, con $n_i^* := \max(n_i, 0)$. Con esto logran un estimador consistente, asintóticamente normal y más eficiente que el que sólo utiliza el primer intervalo.

Nosotros hemos realizado una simulación para comparar el estimador de Huang y Chen utilizando sólo el primer intervalo de tiempo con el que utiliza todos los intervalos de tiempo. Lo que obtuvimos, en varios casos analizados, fue que al utilizar todos los intervalos de tiempo observados no censurados se logra una mejor estimación que la que sólo utiliza el primer intervalo de tiempo observado (en los casos analizados obtuvimos una diferencia entre 0.48 y 0.92 entre las estimaciones de los coeficientes).

3. Modelo de Tiempo de Falla Acelerado

En análisis de supervivencia los modelos de regresión de tiempo de falla acelerado son una útil alternativa al modelo de riesgo multiplicativo en algunos contextos. Ellos son ejemplos de modelos transformados del tiempo que puede ser utilizado tanto en el marco de eventos simples como en eventos recurrentes. Este modelo es un caso particular del modelo de tiempos transformados ya que el efecto de $\boldsymbol{Z}$ es transformar la escala del tiempo t a $\exp(\boldsymbol{\beta}_0'\boldsymbol{Z})t$. Además, este modelo es log-lineal para T, ya que

$$\log T = \boldsymbol{\beta}_0'\boldsymbol{Z} + U, \tag{4}$$

donde U es la variable error. Luego $T = \exp(\boldsymbol{\beta}_0'\boldsymbol{Z})\tilde{T}$ donde $\tilde{T} = e^U > 0$ tienen función de riesgo $\lambda_0(\tilde{t})$, que es independiente de $\boldsymbol{\beta}_0$. Entonces la función de riesgo de T es de la forma

$$\lambda(t|\boldsymbol{Z}) = \lambda_0\left[\exp(\boldsymbol{\beta}_0'\boldsymbol{Z})\right]\exp(\boldsymbol{\beta}_0'\boldsymbol{Z}). \tag{5}$$

La ecuación (4) permite aplicar procedimientos robustos de modelos de regresión para hallar los estimadores. Por ejemplo, el trabajo de Salibian–Barrera y Yohai (2008). Ellos proponen una clase de estimadores robustos de alto punto de ruptura cuando la respuesta contiene observaciones censuradas.

Consideran el modelo de regresión lineal

$$y_i = \boldsymbol{\beta}_0'\boldsymbol{x}_i + u_i,\ i = 1 : m,$$

donde $\boldsymbol{x}_i \in \mathbb{R}^p$ es el vector de covariables, los errores u_i son independientes, idénticamente distribuidas (con distribución F, simétrica) e independientes de las covariables $\boldsymbol{x}_i$ y $\boldsymbol{\beta}_0$ es el vector de coeficientes desconocidos. Consideran censuras aleatorias a derecha, es decir, se observa $y_i^* = \min(y_i, c_i)$ donde c_i son las variables aleatorias de censura no observadas, independientes, idénticamente distribuidas e independientes de los errores u_i. Entonces se observa $(y_i^*, \boldsymbol{x}_i, \delta_i)$ con $\delta_i = I(y_i \leq c_i)$. Ellos extienden la aproximación de Buckley y James (1979) y Ritov (1990) para el caso de respuestas censuradas con una función de pérdida acotada. Proponen el M-estimador de regresión para observaciones censuradas definido por

$$\widehat{\boldsymbol{\beta}}_m = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \frac{1}{m}\sum_{i=1}^{m} \mathbb{E}_{F_{\boldsymbol{\beta}}}\left[\rho(u)|\boldsymbol{w}_i(\boldsymbol{\beta})\right],$$

donde la función $\rho : \mathbb{R} \to \mathbb{R}^+$ tiene ciertas propiedades de regularidad, $F_{\boldsymbol{\beta}}$ es la distribución de los residuos $r(\boldsymbol{\beta}) = y - \boldsymbol{\beta}'\boldsymbol{x}$, $\boldsymbol{w}_i(\boldsymbol{\beta}) = (r_i^*(\boldsymbol{\beta}), \delta_i)$, con $r_i^*(\boldsymbol{\beta}) = y_i^* - \boldsymbol{\beta}'\boldsymbol{x}_i$ los residuos censurados (con variable de censura $c_i - \boldsymbol{\beta}'\boldsymbol{x}_i$, ya que $r_i^*(\boldsymbol{\beta}) = \min[r_i(\boldsymbol{\beta}), c_i - \boldsymbol{\beta}'\boldsymbol{x}_i])$ y

$$\mathbb{E}_{F_{\boldsymbol{\beta}}}\left[\rho(u)|\boldsymbol{w}_i(\boldsymbol{\beta})\right] = \begin{cases} \rho\left[r_i^*(\boldsymbol{\beta})\right] & \text{si } \delta_i = 1 \\ \dfrac{\int_{r_i^*(\boldsymbol{\beta})}^{\infty} \rho(u)\, dF_{\boldsymbol{\beta}}(u)}{1 - F_{\boldsymbol{\beta}}\left[r_i^*(\boldsymbol{\beta})\right]} & \text{si } \delta_i = 0 \end{cases}$$

Como $F_{\boldsymbol{\beta}}$ es desconocida, se la puede estimar con el estimador de KM, $F_{m\boldsymbol{\beta}}^*$ basado sobre $r_i^*(\boldsymbol{\beta})$. Luego, para garantizar la consistencia del estimador definido por

$$\widehat{\boldsymbol{\beta}}_m = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \frac{1}{m}\sum_{i=1}^{m} \mathbb{E}_{F_{m\boldsymbol{\beta}}^*}\left[\rho(u)|\boldsymbol{w}_i(\boldsymbol{\beta})\right] \tag{6}$$

se requiere que $F_{m\boldsymbol{\beta}}^*$ sea consistente para $F_{\boldsymbol{\beta}}$ para todo $\boldsymbol{\beta} \in \mathbb{R}^p$. Para esto dan algunas condiciones sobre las funciones de distribución de los errores y de las censuras. Pero, bajo estas condiciones, el estimador de KM resulta ser consistente si $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ y para $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$ en general no lo es.

Por otra parte, el estimador $\widehat{\boldsymbol{\beta}}_m$ definido como solución de

$$\frac{1}{m}\sum_{i=1}^{m} \mathbb{E}_{F_{m\boldsymbol{\beta}}^*}\left[\psi(u)|\boldsymbol{w}_i(\boldsymbol{\beta})\right]\boldsymbol{x}_i = \boldsymbol{0}, \tag{7}$$

con $\psi(u) = \partial\rho(u)/\partial u$, es Fisher consistente. Notar que las ecuaciones (6) y (7) no son equivalentes como sucede en el caso de regresión no censurada, ya que $F_{m\boldsymbol{\beta}}^*$ depende de $\boldsymbol{\beta}$, no se puede obtener (7) derivando (6).

Los M-estimadores con ψ monótona son sólo robustos frente a outliers de alta palanca y la principal dificultad de utilizar un ψ redescendiente en (7) es que en general esta ecuación puede tener varias soluciones con diferentes propiedades de robustez. La ecuación (6) no puede ser utilizada para obtener una solución consistente de (7). Por esta razón, ellos definen, para $\boldsymbol{\beta}, \boldsymbol{\gamma} \in \mathbb{R}^p$,

$$C_m(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \frac{1}{m} \sum_{i=1}^{m} \mathbb{E}_{F^*_{m\boldsymbol{\beta}}} \left[\rho \left(\frac{u - \boldsymbol{\gamma}' \boldsymbol{x}_i}{s_m} \right) \middle| \boldsymbol{w}_i(\boldsymbol{\beta}) \right] \tag{8}$$

donde la función ρ cumple las mismas propiedades de regularidad que antes y s_m es un estimador robusto de escala del error. Luego definen para cada $\boldsymbol{\beta} \in \mathbb{R}^p$

$$\widehat{\boldsymbol{\gamma}}_m(\boldsymbol{\beta}) = \arg \min_{\boldsymbol{\gamma} \in \mathbb{R}^p} C_m(\boldsymbol{\beta}, \boldsymbol{\gamma}).$$

Entonces definen a un estimador de $\boldsymbol{\beta}_0$ por la ecuación $\widehat{\boldsymbol{\gamma}}_m(\widehat{\boldsymbol{\beta}}_m) = \mathbf{0}$ y, alternativamente, para evitar la existencia de problemas, ellos definen a $\widehat{\boldsymbol{\beta}}_m$ como

$$\widehat{\boldsymbol{\beta}}_m = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \left[\widehat{\boldsymbol{\gamma}}_m(\boldsymbol{\beta})' A_m \widehat{\boldsymbol{\gamma}}_m(\boldsymbol{\beta}) \right], \tag{9}$$

donde $A_m = A_m(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m)$ es cualquier estimador equivariante de la matriz de covarianza de los $\boldsymbol{x}_i$. En (9) es necesario A_m para mantener la equivarianza afin del estimador. El estimador $\widehat{\boldsymbol{\beta}}_m$ puede ser considerado una extensión de los M-estimadores de Ritov (1990) para datos censurados para el caso de funciones ρ acotadas. Además, este estimador tiene las mismas propiedades asintóticas como el estimador de Ritov.

También proponen otros estimadores alternativos robustos, como los S-estimadores propuestos por Rousseeuw y Yohai (1984). Esta propuesta consiste en reemplazar a s_m en (8) por el M-escala $S_m(\boldsymbol{\beta}, \boldsymbol{\gamma})$, que es definido como solución de $C_m(\boldsymbol{\beta}, \boldsymbol{\gamma}) = b$, con $b = \mathbb{E}_F(\rho(u))$. Sea

$$\widehat{\boldsymbol{\gamma}}_m(\boldsymbol{\beta}) = \arg \min_{\boldsymbol{\gamma} \in \mathbb{R}^p} S_m(\boldsymbol{\beta}, \boldsymbol{\gamma}).$$

Notar que $\widehat{\boldsymbol{\gamma}}_m(\boldsymbol{\beta})$ es el S-estimador de regresión de los residuos $(r_i^*(\boldsymbol{\beta}), \boldsymbol{x}_i')'$, $i = 1 : m$. Entonces definen el S-estimador de regresión de las respuestas censuradas como el vector $\tilde{\boldsymbol{\beta}}_m$ tal que $\widehat{\boldsymbol{\gamma}}_m(\tilde{\boldsymbol{\beta}}_m) = \mathbf{0}$ y, alternativamente, para evitar la existencia de problemas, también se puede definir a $\tilde{\boldsymbol{\beta}}_m$ como se hizo en (9). Además, un estimador de escala de los residuos robusto s_m puede ser definido como $s_m = S_m(\tilde{\boldsymbol{\beta}}_m, \widehat{\boldsymbol{\gamma}}_m(\tilde{\boldsymbol{\beta}}_m))$. Como se necesita resolver un problema de optimización altamente complejo para calcular el estimador, ellos presentan en su trabajo un algoritmo computacional eficiente para calcularlo.

Si bien estos S-estimadores tienen un alto punto de ruptura, cuando los errores son normales no pueden alcanzar simultáneamente alta eficiencia y alto punto de ruptura. Entonces para obtener estimadores con alta eficiencia y alto punto de ruptura, Salibian–Barrera y Yohai, realizan dos propuestas: los MM-estimadores y los τ-estimadores. Ambos son extensiones a respuestas censuradas de los MM-estimadores propuestos por Yohai (1987) y de los τ-estimadores propuestos por Yohai y Zamar (1988).

Sobre eventos recurrentes, en los estudios a menudo el interés se centra en modelar la distribución del tiempo de falla entre la recurrencia de un evento (o intervalo de tiempo), o en la distribución de los tiempo de cada falla. Además existen procedimientos de inferencia basados en métodos marginales y métodos basados en intensidades. Los métodos marginales habitualmente se focalizan sobre la función de tasa acumulada o función de la media y no condicionan sobre la historia del evento completo. Un ejemplo de esto lo veremos en el trabajo que realizan Lin, Wei y Ying (1998). En cambio, los métodos de intensidad especifican como la probabilidad de recurrencia posterior dependerá de la historia del evento pasado. Aquí destacaremos el trabajo de Strawderman (2005).

Lin, Wei y Ying (1998) trabajan con los tiempos de falla recurrentes, es decir, para $i = 1 : m$ y $j = 1, 2, \ldots$, sea T_{ij} el $j-$ésimo tiempo de falla del evento para el sujeto $i-$ésimo. Asumen que los sujetos son independientes, pero no se impone ninguna estructura de dependencia sobre los tiempos de recurrencia del mismo sujeto. Definen a $N_i^*(t)$ como el número de fallas que han ocurrido sobre el sujeto i en el tiempo t en ausencia de censura, esto es $N_i^*(t) = \sum_{k\geq 1} I(T_{ik} \leq t)$. Además suponen que la función media del proceso de conteo $N_i^*(t)$ asociado al vector de variable explicativa $\boldsymbol{Z}_i \in \mathbb{R}^p$, que la suponen acotada, es de la forma

$$\mathbb{E}\left(N_i^*(t)|\boldsymbol{Z}_i\right) = \mu_0\left[\exp(\boldsymbol{\beta}_0'\boldsymbol{Z}_i)\, t\right],$$

donde $\boldsymbol{\beta}_0$ es un $p-$vector de parámetros de regresión desconocido, y μ_0 es una función continua no especificada. De acuerdo a este modelo, el número esperado de eventos en el tiempo t bajo $\boldsymbol{Z}_i = \boldsymbol{z}$ es igual al número esperado de eventos en el tiempo $t\exp(\boldsymbol{\beta}_0'\boldsymbol{z})$ bajo $\boldsymbol{Z}_i = \boldsymbol{0}$. En otras palabras, el conjunto de variables explicativas $\boldsymbol{Z}_i$ afecta la frecuencia de recurrencia sobre el tiempo expandiendo o contrayendo la escala del tiempo en aquellas ocurrencias de eventos por un factor multiplicativo de $\exp(\boldsymbol{\beta}_0'\boldsymbol{Z}_i)$ relativo a aquel de un vector de covariable cero.

Sea C_i el tiempo de censura del sujeto i, que lo asumen independiente de T_{ik} condicionado sobre $\boldsymbol{Z}_i$. Luego el proceso de conteo $N_i(t)$ de los tiempos de fallas censurados, se pueden expresar como $N_i(t) = \sum_{k\geq 1} I(T_{ik} \leq t \wedge C_i)$.

Motivados por la función de score de la verosimilitud parcial para el modelo de proceso de Poisson de intensidad proporcional (Andersen y Gill, 1982) y las funciones de estimación de rango pesado para el modelo log-lineal (4) (Prentice, 1978; Tsiatis, 1990; Wei, Ying y Ling, 1990), proponen la siguiente clase de funciones de estimaciones para $\boldsymbol{\beta}_0$

$$\boldsymbol{U}(\boldsymbol{\beta}) := \sum_{i=1}^{m} \int_0^\infty Q(t;\boldsymbol{\beta})\left[\boldsymbol{Z}_i - \bar{\boldsymbol{Z}}(t;\boldsymbol{\beta})\right]\, dN_i\left[t\exp(-\boldsymbol{\beta}'\boldsymbol{Z}_i)\right], \qquad (10)$$

donde $Q(t;\boldsymbol{\beta})$ tiene variación acotada y converge casi seguro a una función continua y

$$\bar{\boldsymbol{Z}}(t;\boldsymbol{\beta}) = \frac{\sum_{j=1}^{m} I\left[t\exp(-\boldsymbol{\beta}'\boldsymbol{Z}_j) \le C_j\right] \boldsymbol{Z}_j}{\sum_{j=1}^{m} I\left[t\exp(-\boldsymbol{\beta}'\boldsymbol{Z}_j) \le C_j\right]}.$$

Ellos se refieren a $\boldsymbol{U}(\boldsymbol{\beta})$ como la función de estimación log-rango si $Q = 1$ y como la función de estimación de Gehan si $Q(t;\boldsymbol{\beta}) = \sum_{i=1}^{m} I[t\exp(-\boldsymbol{\beta}'\boldsymbol{Z}_i) \le C_i]/m$. Para la primer función de peso, como en el caso de estimación de rango para el modelo log-lineal (4), la función de estimación $\boldsymbol{U}(\boldsymbol{\beta})$ es una función constante a trozos de $\boldsymbol{\beta}$, entonces definen el estimador $\widehat{\boldsymbol{\beta}}$ como un cero de $\boldsymbol{U}(\boldsymbol{\beta})$ o como un mínimo de $\|\boldsymbol{U}(\boldsymbol{\beta})\|_2$. Para la función de estimación de Gehan, (10) se convierte en

$$\boldsymbol{U}(\boldsymbol{\beta}) = \frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k\ge 1} \Delta_{ik}\ (\boldsymbol{Z}_i - \boldsymbol{Z}_j)\ I\left[\log\tau \log T_{ik} \ge \boldsymbol{\beta}'(\boldsymbol{Z}_i - \boldsymbol{Z}_j)\right],$$

y así obtienen $\widehat{\boldsymbol{\beta}}$ minimizando la función

$$\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k\ge 1} \Delta_{ik} \max\left[\log\tau - \log T_{ik} - \boldsymbol{\beta}'(\boldsymbol{Z}_i - \boldsymbol{Z}_j), 0\right].$$

El estimador resultante puede ser ligeramente diferente al mínimo de $\|\boldsymbol{U}(\boldsymbol{\beta})\|_2$, pero son asintóticamente equivalentes.

El estimador $\widehat{\boldsymbol{\beta}}$, solución de $\boldsymbol{U}(\boldsymbol{\beta}) = \boldsymbol{0}$, resulta ser consistente y asintóticamente normal. Pero, resolver la ecuación de estimación $\boldsymbol{U}(\boldsymbol{\beta}) = \boldsymbol{0}$ puede, en general, ser arduo cuando p (la dimensión de $\boldsymbol{Z}$) es grande. Entonces, proponen $\widehat{\boldsymbol{\beta}}^{\star}$ como solución de

$$\boldsymbol{U}(\boldsymbol{\beta}) = \sum_{i=1}^{m} D_i(\widehat{\boldsymbol{\beta}})G_i,$$

donde

$$\begin{aligned} D_i(\boldsymbol{\beta}) := \int_0^{\infty} Q(t;\boldsymbol{\beta})\left[\boldsymbol{Z}_i - \bar{\boldsymbol{Z}}(t;\boldsymbol{\beta})\right] d\left\{N_i\left[t\exp(-\boldsymbol{\beta}'\boldsymbol{Z}_i)\right]\right. \\ \left. - \int_0^{t} I\left[t\exp(-\boldsymbol{\beta}'\boldsymbol{Z}_i) \le C_i\right] d\widehat{\mu}_0(s;\boldsymbol{\beta})\right\}, \end{aligned}$$

con

$$\widehat{\mu}_0(t;\boldsymbol{\beta}) = \sum_{i=1}^{m}\int_0^{t} \frac{dN_i\left[t\exp(-\boldsymbol{\beta}'\boldsymbol{Z}_i)\right]}{\sum_{j=1}^{m} I[t\exp(-\boldsymbol{\beta}'\boldsymbol{Z}_j) \le C_j]}$$

y $(G_1,\ldots,G_m)$ son variables aleatorias normales estándar independientes. También se puede obtener $\widehat{\boldsymbol{\beta}}^{\star}$ como solución de $\boldsymbol{U}(\boldsymbol{\beta})=G$, donde G es normal con media cero y matriz de covarianza $\sum_{i=1}^{m} D_i(\widehat{\boldsymbol{\beta}})D_i(\widehat{\boldsymbol{\beta}})'$. Luego $\sqrt{m}(\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\beta}}^{\star})$ tiene la misma distribución límite que $\sqrt{m}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)$ y además la matriz de covarianza de $\widehat{\boldsymbol{\beta}}$ puede ser estimada por la matriz de covarianza empírica de $\widehat{\boldsymbol{\beta}}^{\star}$.

Cuando el modelo se ajusta razonablemente a los datos, Lin, Wei y Ying proponen que el estimador de la ecuación de estimación de Gehan (solución de $\boldsymbol{U}(\boldsymbol{\beta})=\mathbf{0}$ o $\boldsymbol{U}(\boldsymbol{\beta})=G$, que se puede resolver de manera eficiente), puede ser utilizado como un estimador inicial para estimaciones con funciones de peso más generales, ya que la solución de la ecuación de estimación de Gehan será similar a las soluciones de la ecuación de estimación con pesos más generales. Además, para la mayoría de los efectos prácticos, es suficiente hacer inferencias basadas sobre la estimación de Gehan.

El trabajo de Strawderman (2005) desarrolla un nuevo modelo semiparamétrico para el efecto de las variables explicativas, independientes del tiempo, sobre la intensidad condicional de un proceso de conteo de eventos recurrentes. Su modelo es una extensión del modelo del tiempo de falla acelerado para datos de supervivencia univariado, intervalos de tiempo entre eventos, y la estimación del parámetro de regresión esta motivada por las consideraciones de eficiencia semiparamétricas.

Primero considera un sujeto con un vector de covariable $\boldsymbol{Z}$ de dimensión p independiente del tiempo, que experimenta el evento recurrente en los tiempos $0 =: T_0 < T_1 < T_2 < \cdots$. Define el j-ésimo intervalo de tiempo como $X_j = T_j - T_{j-1}$, $j \geq 1$. Sean V_1, V_2, ... variables aleatorias independientes e idénticamente distribuidas con función de distribución F_0, con ciertas propiedades de regularidad. Asume que, dado $\boldsymbol{Z}$, sus intervalos de tiempos, X_1, X_2, ... son variables aleatorias independientes donde

$$X_j = V_j \exp(-\boldsymbol{\beta}_0'\boldsymbol{Z}),$$

siendo $\boldsymbol{\beta}_0$ el vector de parámetro de regresión. Notar que $\boldsymbol{Z}$ acelera o desacelera las intervalos de tiempo V_j como sucede con los modelos de tiempos de falla acelerado.

Luego la función de riesgo de X_j, dado $\boldsymbol{Z}$ es

$$\lambda_0\left[x\exp(\boldsymbol{\beta}_0'\boldsymbol{Z})\right]\exp(\boldsymbol{\beta}_0'\boldsymbol{Z}),$$

donde λ_0 es la función de riesgo asociada a F_0. Entonces, en ausencia de censura y dado $\boldsymbol{Z}$, el proceso $N(t) := \max\left(n : \sum_{j=1}^{n} X_j \leq t\right)$ es un proceso de renovación.

Ahora supone m sujetos independientes, donde cada uno es observado en el intervalo de tiempo finito $[0, C_i]$. Entonces los datos observados son $[N_i(u \wedge C_i), \Delta_i(u), \boldsymbol{Z}_i, u \geq 0]$ para $i = 1 : m$, donde $\Delta_i(u) = I(u \leq C_i)$.

El trabajo de Strawderman fue motivado por los trabajos de Prentice (1978) (quien sugirió estimar $\boldsymbol{\beta}_0$ del modelo semiparamétrico (5) invirtiendo una clase de estadísticos de rango lineal pesado), de Tsiatis (1990) (quien estableció las propiedades asintóticas de la clase de estimadores propuesto por Prentice) y de Ritov (1990) (quien estableció una correspondencia directa entre las funciones estimadas por Prentice y Tsiatis y las basadas sobre consideraciones de eficiencia semiparamétrica para modelos de regresión lineal censurados). Strawderman propone la clase pesada de funciones de estimación, mediante la siguiente función de score

$$\bar{S}_W(\boldsymbol{\beta}) := \frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{n_i} W\left[\tilde{X}_{ij}(\boldsymbol{\beta})|\boldsymbol{\beta}\right]\left[\boldsymbol{Z}_i - \frac{\sum_{k=1}^{m}\boldsymbol{Z}_k\sum_{r=1}^{n_k+1}\tilde{I}_{kr}(\tilde{X}_{ij}|\boldsymbol{\beta})}{\sum_{k=1}^{m}\sum_{r=1}^{n_k+1}\tilde{I}_{kr}(\tilde{X}_{ij}|\boldsymbol{\beta})}\right], \quad (11)$$

donde $\tilde{X}_{ij}(\boldsymbol{\beta}) = X_{ij}\exp(\boldsymbol{\beta}'\boldsymbol{Z}_i)$ para $j = 1 : n_i + 1$, $n_i = N_i(C_i)$, $\tilde{I}_{kr}(t|\boldsymbol{\beta}) := I[\tilde{X}_{kr}(\boldsymbol{\beta}) \geq t)]$ y $W(t|\boldsymbol{\beta})$ bajo ciertas propiedades de regularidad.

Como (11) es una función de estimación basada en rango, el estimador $\widehat{\boldsymbol{\beta}}$ puede ser definido como un cero de $\bar{S}_W(\boldsymbol{\beta})$ o un mínimo de $\|\bar{S}_W(\boldsymbol{\beta})\|_2$. Sin embargo, puede existir varios mínimos porque $\bar{S}_W(\boldsymbol{\beta})$ no es necesariamente monótona. Pero, si $W(u|\boldsymbol{\beta}) = \sum_{i=1}^{m}\sum_{j=1}^{n_i+1}\tilde{I}_{ij}(u|\boldsymbol{\beta})/m$ (denominado peso de Gehan), esta dificultad desaparece. Y en este caso, (11) se reduce a

$$\bar{S}_G(\boldsymbol{\beta}) := \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\sum_{k=1}^{m}(\boldsymbol{Z}_i - \boldsymbol{Z}_k)\sum_{r=1}^{n_k+1}\tilde{I}_{kr}(\tilde{X}_{ij}|\boldsymbol{\beta}),$$

que es el gradiente de la función objetivo convexa

$$L_G(\boldsymbol{\beta}) := \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\sum_{k=1}^{m}\sum_{r=1}^{n_k+1}\max\left\{\left[\log\tilde{X}_{kr}(\boldsymbol{\beta}) - \log\tilde{X}_{ij}(\boldsymbol{\beta})\right], 0\right\},$$

Los minimizadores de $L_G(\boldsymbol{\beta})$ y $\|\bar{S}_G(\boldsymbol{\beta})\|$ son asintóticamente equivalentes (Fygenson y Ritov, 1994). Luego Strawderman toma $\widehat{\boldsymbol{\beta}}_G = \arg\min L_G(\boldsymbol{\beta})$ y resulta ser consistente de tasa $\sqrt{m}$ y asintóticamente normal.

Para este último estimador Strawderman comenta que “notablemente no se asume que el intervalo de tiempo sea acotado, es una útil consecuencia de la convexidad asociada con la función de peso de Gehan”. Además, da un algoritmo para calcular $\widehat{\boldsymbol{\beta}}_G$ y $\widehat{\Gamma}_G$, un estimador consistente de la covarianza de $\sqrt{m}(\widehat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}_0)$.

Para pesos generales, Strawderman propone un estimador $\widehat{\boldsymbol{\beta}}_W$ de un paso a partir del $\widehat{\boldsymbol{\beta}}_G$, que también resulta ser consistente de tasa $\sqrt{m}$ y asintóticamente normal.

4. Modelo Aditivo

Por último, veremos los resultados estadísticos del modelo de riesgo aditivo para la función de intensidad que es de la forma

$$\lambda(t|\boldsymbol{Z}) = \lambda_0(t) + \boldsymbol{\beta}_0'\boldsymbol{Z}(t) \tag{12}$$

donde $\boldsymbol{\beta}_0$ es el vector de parámetro de regresión y $\lambda_0(t) = \lambda(t|\boldsymbol{Z} = 0)$ es la función baseline, desconocida, arbitraria y no negativa en función del tiempo. Este modelo propuesto por Lin y Ying (1994) es una alternativa al modelo de Aalen (1980) en el cual el parámetro de regresión depende de los tiempos de falla, es decir, la función de intensidad es de la forma

$$\lambda_0(t) + \boldsymbol{\beta}_0(t)'\boldsymbol{Z}(t) = [\lambda_0(t), \boldsymbol{\beta}_0(t)'] \begin{bmatrix} 1 \\ \boldsymbol{Z}(t) \end{bmatrix}.$$

Primero plantearemos aquí algunas de las distintas funciones de verosimilitud para el modelo aditivo sólo para el caso de eventos simples.

La función de verosimilitud condicional sobre la historia del evento del sujeto i, $H_i(\tau)$, de m individuos con función de intensidad (12), donde $t_1 \leq \ldots \leq t_m$ son los tiempos de falla observados en $[0, \tau]$ (τ tiempo fijo finito) y $\boldsymbol{Z}_i$ corresponde al vector de covariable asociado a t_i, es

$$L_C(\boldsymbol{\beta}_0, \lambda_0) = \prod_{j=1}^{m} [\lambda_0(t_j) + \boldsymbol{\beta}_0'\boldsymbol{Z}_j(t_j)] \exp\left\{-\left[\Lambda_0(\tau) + \boldsymbol{\beta}_0'\boldsymbol{Z}_j^*(\tau)\right]\right\},$$

donde $\boldsymbol{Z}_j^*(t) = \int_0^t \boldsymbol{Z}_j(u)\, du$ y $\Lambda_0(t) = \int_0^t \lambda_0(u)\, du$. Luego,

$$\frac{\partial}{\partial \boldsymbol{\beta}_0} \log L_C(\boldsymbol{\beta}_0, \lambda_0) = \sum_{j=1}^{m} \left[\frac{\boldsymbol{Z}_j(t_j)}{\lambda_0(t_j) + \boldsymbol{\beta}_0'\boldsymbol{Z}_j(t_j)} - \boldsymbol{Z}_j^*(\tau)\right]. \tag{13}$$

Observemos que (13) depende del parámetro de regresión $\boldsymbol{\beta}_0$ y de la función $\lambda_0(\cdot)$, lo cual complicaría la estimación de $\boldsymbol{\beta}_0$ ya que sería necesario estimar previamente o conjuntamente la función $\lambda_0(\cdot)$.

Por otra parte, la función de verosimilitud parcial es

$$L_P(\boldsymbol{\beta}_0, \lambda_0) = \prod_{j=1}^{m} \frac{\lambda_0(t_j) + \boldsymbol{\beta}_0'\boldsymbol{Z}_j(t_j)}{\sum_{k \in R_j} [\lambda_0(t_j) + \boldsymbol{\beta}_0'\boldsymbol{Z}_k(t_j)]} \tag{14}$$

donde R_j es el conjunto de riesgo en el tiempo t_j. Entonces

$$\frac{\partial}{\partial \boldsymbol{\beta}_0} \log L_P(\boldsymbol{\beta}_0, \lambda_0) = \sum_{j=1}^{m} \left[\frac{\boldsymbol{Z}_j(t_j)}{\lambda_0(t_j) + \boldsymbol{\beta}_0'\boldsymbol{Z}_j(t_j)} - \frac{SZ_j}{r_j\lambda_0(t_j) + \boldsymbol{\beta}_0'\, SZ_j}\right], \tag{15}$$

donde r_j es el cardinal del conjunto R_j y $SZ_j := \sum_{k \in R_j} \boldsymbol{Z}_k(t_j)$.

Notar que aquí también, la función de verosimilitud parcial (14) no puede aplicarse como se hizo para el caso del modelo de riesgo multiplicativo (1), pues en este caso tampoco se eliminaría $\lambda_0(t)$ para la estimación de $\boldsymbol{\beta}_0$, (ver la función de score (15)).

Sin embargo, hay varios autores que han podido estimar $\boldsymbol{\beta}_0$ sin la necesidad de recurrir a las clásicas funciones de verosimilitud condicional y/o parcial antes descritas. Entre estos autores hemos considerado el trabajo de Lin y Ying (1994) en el cual imitaron la característica de martingala de la función de score de la verosimilitud parcial del modelo multiplicativo del parámetro $\boldsymbol{\beta}_0$ logrando construir una simple función de estimación que permite expresar en forma explícita al estimador $\widehat{\boldsymbol{\beta}}_0$ (esto no sucede en los modelos multiplicativo y de falla acelerado).

Formalmente, Lin y Ying (1994) consideran m sujetos independientes y recogen en el proceso de conteo del sujeto i, $\{N_i(t); t \geq 0\}$ el número de eventos observados hasta el tiempo t. Bajo el modelo (12), la función de intensidad para $N_i(t)$ esta dada por

$$Y_i(t)\, d\Lambda(t; \boldsymbol{Z}_i) = Y_i(t)\left[d\Lambda_0(t) + \boldsymbol{\beta}_0' \boldsymbol{Z}_i(t)\, dt\right], \tag{16}$$

donde $Y_i(t)$ indica con 1 si el sujeto i está en riesgo en el tiempo t y en en caso contrario con 0 y $\Lambda_0(t) = \int_0^t \lambda_0(u)\, du$. Ellos proponen estimar $\boldsymbol{\beta}_0$ imitando la función de score de la verosimilitud parcial del modelo multiplicativo, que bajo (16), es de la forma

$$\boldsymbol{U}(\boldsymbol{\beta}) = \sum_{i=1}^{m} \int_0^{\infty} \boldsymbol{Z}_i(t) \left\{ dN_i(t) - Y_i(t) \left[d\widehat{\Lambda}_0(\boldsymbol{\beta}, t) + \boldsymbol{\beta}' \boldsymbol{Z}_i(t)\, dt \right] \right\},$$

donde $\widehat{\Lambda}_0$ es el estimador de Λ_0 del modelo (12) definido como

$$\widehat{\Lambda}_0(\widehat{\boldsymbol{\beta}}, t) = \int_0^t \frac{\sum_{i=1}^m \left[dN_i(u) - Y_i(u) \widehat{\boldsymbol{\beta}}' \boldsymbol{Z}_i(u)\, du \right]}{\sum_{i=1}^m Y_i(u)},$$

siendo $\widehat{\boldsymbol{\beta}}$ un estimador consistente de $\boldsymbol{\beta}_0$. Luego, $\boldsymbol{U}(\boldsymbol{\beta})$ es equivalente a

$$\boldsymbol{U}_{LY}(\boldsymbol{\beta}) := \sum_{i=1}^{m} \int_0^{\infty} \left[\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}(t) \right] \left[dN_i(t) - Y_i(t) \boldsymbol{\beta}' \boldsymbol{Z}_i(t)\, dt \right],$$

donde $\bar{\boldsymbol{Z}}(t) = \sum_{j=1}^m Y_j(t) \boldsymbol{Z}_j(t) / \sum_{j=1}^m Y_j(t)$. Entonces el estimador queda definido explícitamente como

$$\widehat{\boldsymbol{\beta}} = \left\{ \sum_{i=1}^{m} \int_0^{\infty} Y_i(t) \left[\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}(t) \right]^{\otimes 2} dt \right\}^{-1} \left\{ \sum_{i=1}^{m} \int_0^{\infty} \left[\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}(t) \right] dN_i(t) \right\}.$$

este estimador resulta ser consistente de tasa $\sqrt{m}$ y asintóticamente normal, bajo ciertas condiciones de regularidad.

Nosotros hemos realizado una simulación para ver el desempeño del estimador frente a outliers y lo que hemos notado es que el estimador $\widehat{\boldsymbol{\beta}}$ se ve influenciado por estos, lo que nos motivó a una nueva propuesta de un estimador robusto en la que estamos trabajando actualmente.

Por último en el caso de eventos recurrentes destacamos el trabajo de Sun, Park y Sun (2006) en el que ajustan un modelo de riesgo aditivo $\lambda(t|\boldsymbol{Z}_i) = \lambda_0(t) + \boldsymbol{\beta}_0'\boldsymbol{Z}_i$ utilizando los intervalos de tiempo entre las ocurrencias de los tiempo de falla y toman a las covariables independientes del tiempo. Ellos extienden la idea de Lin y Ying (1994) a eventos recurrentes aplicando los mismos argumentos y suposiciones que realizan Huang y Chen (2003) además de introducirle una función de peso a la función de score de estimación.

Siguiendo los mismos argumentos y las mismas notaciones que utilizamos al describir el trabajo de Huang y Chen (2003), Sun, Park y Sun proponen la ecuación de estimación $\boldsymbol{U}_{SPS}(\boldsymbol{\beta}) = \boldsymbol{0}$ para estimar el parámetro de regresión $\boldsymbol{\beta}_0$, donde

$$\begin{aligned}\boldsymbol{U}_{SPS}(\boldsymbol{\beta}) := \int_0^\tau Q(s) \Bigg\{ & d\widehat{K}_1(s) - \frac{\widehat{G}_1(s)}{\widehat{G}_0(s)}\, d\widehat{K}_0(s) - \\ & \left[\widehat{E}_{ij}\left[\boldsymbol{Z}_i'\boldsymbol{Z}_i\, I(X_{ij} \leq s)\right] - \frac{\widehat{G}_1(s)'\widehat{G}_1(s)}{\widehat{G}_0(s)}\right] \boldsymbol{\beta}\, ds \Bigg\}\end{aligned}$$

siendo $Q(s)$ un proceso de peso con ciertas propiedades de regularidad, $\tau \in (0, \infty)$ es una constante pre-especificada (en la práctica, τ es usualmente tomada como el tiempo de seguimiento más largo), $\widehat{G}_0(t) = \widehat{E}_{ij}\left[I(X_{ij} \geq t)\right]$ y $\widehat{G}_1(t) = \widehat{E}_{ij}\left[\boldsymbol{Z}_i\, I(X_{ij} \geq t)\right]$.

Luego el estimador de $\boldsymbol{\beta}_0$ también se puede expresar de forma explícita como

$$\begin{aligned}\widehat{\boldsymbol{\beta}} = & \left\{ \int_0^\tau Q(s) \left[\widehat{E}_{ij}\left[\boldsymbol{Z}_i'\boldsymbol{Z}_i I(X_{ij} \leq s)\right] - \frac{\widehat{G}_1(s)'\widehat{G}_1(s)}{\widehat{G}_0(s)}\right] ds \right\}^{-1} \times \\ & \left\{ \int_0^\tau Q(s) \left[d\widehat{K}_1(s) - \frac{\widehat{G}_1(s)}{\widehat{G}_0(s)} d\widehat{K}_0(s)\right] \right\},\end{aligned}$$

que es consistente de tasa $\sqrt{m}$ y asintóticamente normal.

Ellos, al igual que Huang y Chen, también proponen un estimador que sólo utiliza el primer intervalo de tiempo de cada individuo pero es menos eficiente que aquel que utiliza todos los intervalos (al igual que pasa en Huang y Chen). Además comentan que "un problema que necesita ser estudiado a futuro es la selección de un proceso de peso $Q(t)$ que da el estimador más eficiente de $\boldsymbol{\beta}_0$ para una situación particular". Para el estudio de simulación que realizaron, obtuvieron resultados similares para $Q = 1$ y $Q(t) = \sum_{i=1}^{m} I(C_i \geq t)/m$.

Agradecimientos

Estamos en deuda con los árbitros cuyos comentarios han sido de mucha ayuda para mejorar este artículo.

Referencias

AALEN, O. O. (1980). "A model for non–parametric regression analysis of counting processes". *Lecture Notes on Mathematics Statistics and Probability.* **2**: 1–25.

ANDERSEN, P. K. and GILL, R. D. (1982). "Cox's regression model for counting processes: a large sample study". *The Annals of Statistics.* **10 (4)**: 1110–1120.

BASU, S., BASU, A. and JONES, M. C. (2006). "Robust and efficient parametric estimation for censored survival data". *Annals of the Institute of Statistical Mathematics.* **58**: 341–355.

BEDNARSKI, T. (1993). "Robust estimation in Cox's regression model". *Scandinavian Journal of Statistics.* **20**: 213–225.

BEDNARSKI, T. and NOWAK, M. (2003). "Robustness and efficiency of Sasieni–type estimators in the Cox model". *Journal of Statistical Planning and Inference.* **115 (1)**: 261–272.

BEDNARSKI, T. and MOCARSKA, E. (2006). "On robust model selection within the Cox model". *Ecometrics Journal.* **9**: 179–290.

BEDNARSKI, T. (2007). "On a robust modification of Breslow's cumulated hazard estimator". *Computational Statistics and Data Analysis.* **52**: 234–238.

BEGUN, J. M., HALL, W. J., HUANG, W. M. and WELLNER, J. A. (1983). "Information and Asymptotic Efficiency in Paramtric–Nonparametric Models". *The Annal of Statistics.* **11 (2)**: 432–452.

BERAN, R. (1981). "Efficient Robust Estimates in Parametric Models". *Zeitschrift fr Wahrscheinlichkeitstheorie und Verwande Gebietez.* **55**: 91–108.

BUCKLEY, J. and JAMES, I. (1979). "Linear regression with censored data". *Biometrika.* **66**: 429–436.

COX, D. R. (1972). "Regression Models and Life–Tables" (with discussion). *Journal of the Royal Statistical Society.* Series B (Methodological). **34 (2)**: 187–220.

COX, D. R. (1975). "Partial Likelihood". *Biometrika.* **62**: 262–276.

EFRON, B. (1977). "The efficiency of Cox's likelihood function for censored data". *Journal of the American Statistical Association.* **72**: 359, 557–565.

FYGENSON, T. R. and RITOV, Y. (1994). "Monotone estimating equations for censored data". *The Annal of Statistics.* **22**: 732–746.

HUANG, Y. and CHEN, Y. Q. (2003). "Marginal regression of gaps between recurrent events". *Lifetime Data Analysis.* **9**: 293–303.

HUBER, P. J. (1981). *Robust Statistics.* Wiley, New York.

KAPLAN, E. L. and MEIER, P. (1958). "Nonparametric estimation from incomplete observations". *Journal of the American Statistical Association.* **53**: 457–481.

LIN, D. Y. and YING, Z. (1994). "Semiparametric analysis of the additive risk model". *Biometrika.* **81 (1)**: 61–71.

LIN, D. Y., WEI, L. J., and YING, Z. (1998). "Accelerate failure time models for counting processes". *Biometrika.* **85 (3)**: 605–618.

MARONNA, R. A., MARTIN, R. D., and YOHAI, V. J. (2006). *Robust Statistics: Theory and Practice.* Wiley, New York.

MINDER, C. E. and BEDNARSKI, T. (1996). "A robust method for proportional hazards regression". *Statistics in Medicine.* **15**: 1033–1047.

PRENTICE, R. L. (1978). "Linear rank tests with right censored data". *Biometrika.* **65**: 167–179.

REID, N. (1981). "Influence functions for censored data". *The Annals of Statistics.* **9 (1)**: 78–92.

REID, N. and CRÉPEAU, H. (1985). "Influence functions for proportional hazards regression". *Biometrika.* **72 (1)**: 1–9.

RITOV, Y. and WELLNER, J. A. (1987). "Censoring, Martingala and The Cox Model". *Technical Report, University of Washington, Department of Statistics.* **108**.

RITOV, Y. (1990). "Estimation in a linear regression model with censored data". *The Annals of Statistics.* **18 (1)**: 303–328.

ROUSSEEUW, P. J. and YOHAI, V. J. (1984). "Robust regression by means of S–estimators". *Robust and Nonlinear Time Series Analysis (J. Franke, W. Hardle and R. D. Martin, eds.) Lecture Notes in Statist.* **26**: 256–276.

SALIBIAN–BARRERA, M. and YOHAI, V. J. (2008). "High breakdown point robust regression with censored data". *Annals of Statistics.* **36**: 118–146.

SASIENI, P. (1993a). "Maximum Weighted Partial Likelihood Estimators for the Cox Model". *Journal of the American Statistical Association.* **88 (421)**: 144–152.

SASIENI, P. (1993b). "Some new estimators for Cox regression". *The Annals of Statistics.* **21 (4)**: 1721–1759.

STRAWDERMAN, R. L. (2005). "The accelerated gap times model". *Biometrika.* **92**: 647–666.

SUN, L., PARK, D. and SUN, J. (2006). "The additive hazards model for recurrent gap times". *Statistica Sinica.* **16**: 919–932.

TSIATIS, A. A. (1990). "Estimating regression parameters using linear rank tests for censored data". *The Annals of Statistics.* **18 (19)**: 354–372.

WANG, M. C. and CHANG, S. H. (1999). "Nonparametric estimation of a recurrent survival function". *Journal of the American Statistical Association.* **94**: 146–153.

WEI, L. J., YING, Z. and LING, D. Y. (1990). "Linear regression analysis of censored survival data based on rank tests". *Biometrika.* **7**: 845–851.

YOHAI, V. J. (1987). "High breakdown–point and high efficiency robust estimates for regression". *Annals of Statistics.* **15**: 642–656.

YOHAI, V. J. and ZAMAR, R. H. (1988). "High breakdown point and high efficiency estimates of regression by means of the minimization of an efficient scale". *Journal of the American Statistical Association.* **83**: 406–413.

Received November 2012
Revised September 2013