REVISTA INGENIERÍA UC. Vol. 15, Nº 1, 86-96, 2008

Modelos de estimaciones no paramétricas del análisis de supervivencia con eventos recurrentes

Carlos Martínez (1), Rafael Borges (2)

(1) Escuela de Ingeniería Industrial , Facultad de Ingeniería , Universidad de Carabobo, Valencia, Venezuela

(2) Escuela de Estadística, Facultad de Ciencias Económicas y Sociales

Universidad de Los Andes, Mérida, Venezuela.

Email: cmartin@uc.edu.ve, borgesr@ula.ve

Resumen

El desarrollo del análisis de supervivencia con fenómenos de naturaleza recurrente ha tenido importantes avances en las últimas cuatro décadas. Tradicionalmente, las investigaciones en este campo estaban centradas en el estudio de una única ocurrencia del evento por unidad de estudio (análisis clásico de supervivencia). Recientemente se han propuesto varios modelos para estudiar fenómenos con eventos recurrentes. Los estudios de supervivencia han sido ampliamente empleados en estudios médicos para modelar la aparición de enfermedades recurrentes, y en el área de Ingeniería en la modelación de tiempos de fallas en maquinarias y equipos. Entre los modelos más conocidos, se tienen los modelos de estratificados tipo Cox, clasificados como de riesgo multiplicativo, de riesgo aditivo y de riesgo aditivo multiplicativo, algunos de estos modelos consideran covariables que varían en el tiempo, otro conjunto del campo recurrente, son los modelos no paramétricos tipo Kaplan-Meier. En este artículo nos centramos en el estudio de éstos últimos para estimar las funciones del análisis de supervivencia para el caso recurrente. El modelo GPLE (*Generalized Product-Limit Estimator*), es un modelo que generaliza el estimador clásico conocido como estimador limite-producto de Kaplan-Meier a casos con eventos de carácter recurrente. El GPLE es un modelo que asume independencia entre los tiempos de interocurrencias del evento. Nuestro objetivo principal en este trabajo consiste en proponer modelos de supervivencia no paramétricos a casos con eventos recurrentes, generalizando los modelos clásicos del análisis.

Palabras clave: Análisis de supervivencia, eventos recurrentes, modelos no paramétricos.

Non parametric estimation models from survival analysis with recurrent events

Abstract

The development of survival analysis with recurrent events has had significant progress in the last four decades. Traditionally, survival analysis had focused on the study of a single occurrence of the event per unit, known as survival classical analysis. However, recent research has been proposed several models to study phenomena where the events are the recurring type. These have been applied in medical studies for recurrent diseases and in the modeling of machinery and equipment failure. Among, the most popular models are the stratified Cox models, classified as risk multiplicative, additive risk and risk additive multiplicative models, considered co-variable over time. Another subset from recurrent field , are non-parametric models proposed by Kaplan-Meier. In this paper we focus on the study of the latter, nonparametric model used to estimate the functions of the survival analysis for recurrent events. The GPLE model, (*Generalized Product-Limit Estimator*) is widely known as the classic estimator for the limit product Kaplan-Meier estimator, in cases with recurrent events. The GPLE is a model that assumes independence between the ocurrence times of the event. Our main goal in this paper is to propose nonparametric models of survival in cases with recurring events, generalizing the classical analysis.

Keywords: survival analysis, recurrent events, nonparametric models.

1. INTRODUCCIÓN

Hasta la década de los años ochenta, los estudios de supervivencia se habían realizado considerando sólo una única ocurrencia del evento por unidad de estudio. Hoy en día, estos modelos son conocidos como modelos clásicos; entre los más conocidos, se encuentran: el modelo actuarial de Cutler-Ederer[1], el modelo de Kaplan-Meier[2] y el modelo de riesgos proporcionales de Cox [3]. La condición de recurrencia de los eventos en las unidades bajo estudio no se había considerado, pero en la década de los años ochenta introdujeron esta condición en los modelos propuestos por [4-6]. Los fenómenos con patrones de recurrencias están presentes en muchas áreas de la ciencia, en el ámbito biomédico suelen ocurrir numerosos fenómenos de este tipo; por ejemplo, pacientes enfermos del corazón con ataques cardiacos, aparición de enfermedades infecciosas en seres vivos, aparición de tumores, ataques de epilepsia, entre otros. En Ingeniería, se podrían presentar fallas en maquinarias y equipos, producciones de artículos defectuosos en líneas de producción, entre otros. En lo social, se podrían citar: la concepción de hijos en las parejas, matrimonios, divorcios, empleos, entre otros. En investigaciones policiales se puede hablar de diversos delitos como: robos, asaltos, homicidios, suicidios, secuestros, ataques terroristas, etc. Investigaciones recientes han tratado el problema con eventos recurrentes [7-9] y propusieron modelos de estimación de funciones de supervivencia para este tipo de fenómenos. En el desarrollo de estos trabajos se consideraron aspectos relacionados con la correlación de los tiempos entre ocurrencias del evento, problema de datos censurados y probabilidad de ocurrencia del evento en las unidades bajo estudio.

Los primeros modelos de supervivencia fueron desarrollados en la década de los años ochenta para los cual consideraron los efectos de covariables, clasificado por sus autores como modelos semiparamétricos, generalizaciones del modelo clásico de Cox adaptado al caso recurrente. El último conjunto de modelos desarrollados entran dentro de la clasificación de modelos no paramétricos de interés para esta investigación. Los más conocidos son el estimador de supervivencia no paramétrico GPLE que generaliza el estimador límite del producto de Kaplan-Meier y el estimador de fragilidad multiplicativa (*multiplicative frailty model o FRMLE*) [9], este último es útil para aquellos casos donde existe dependencia entre los tiempos de interocurrencias en la unidades bajo estudio, un estudio similar fue propuesto en [10]. Estos modelos de supervivencia en fenómenos de recurrencia no han sido ampliamente difundidos, y la bibliografía que abarca a estos estudios es escasa y muy especializada. La idea central de este trabajo consiste en proponer modelos de estimaciones no paramétricos para estimar la función de supervivencia en fenómenos con eventos de naturaleza recurrentes. Entre los objetivos que se persiguen en este trabajo, se encuentran:

- Proponer modelos de estimaciones no paramétricos que permitan modelar fenómenos de supervivencia con eventos de naturaleza recurrente, generalizando modelos clásicos propuestos en [11-14],[5],[15-17].
- Diseñar una herramienta computarizada en lenguaje R que permita realizar las estimaciones de los modelos propuestos.
- Aplicar modelos teóricos propuestos a un caso real y comparar resultados con modelos de Wang-Chang [10] y Peña *et al.* [9].

METODOLOGÍA

2.1 NOTACIÓN BÁSICA DEL ANÁLISIS DE SUPERVIVENCIA

La variable aleatoria principal del análisis de supervivencia es el tiempo *(T),* para medir el tiempo que transcurre desde un momento inicial predefinido hasta la aparición de un evento de interés. Esta variable tiene asociada una serie de funciones fundamentales; entre las que se encuentran: la función de densidad de probabilidades *(fdp)*, la función *(F)* o función de distribución acumulada *(fda)*, la función *(S)* o función de supervivencia, la función de riesgo instantáneo, *(h o λ)* y la función de riesgos acumulados, *(H o Λ).* A continuación se dan algunas definiciones básicas del análisis de supervivencia:

Definición 1. Sea, T la variable aleatoria que representa el tiempo de ocurrencia del evento de estudio, si F es la *fda* de la variable T, entonces: $F(t)=P(T\leq t)$; y en consecuencia, $S(t)=1-F(t)$, Así:

$$F(t)=\int_0^t f(s)\,ds \tag{1}$$

$$S(t)=\int_t^\infty f(s)\,ds \tag{2}$$

Definición 2. Sea, h la razón instantánea de ocurrencia de un evento de estudio, definida como:

$$\mathrm{h}(t)=\lim_{\Delta t\to 0}\frac{P(t\leq T<t+\Delta t/T\geq t)}{\Delta t} \tag{3}$$

De la ecuación (3), se puede deducir que:

$$\mathrm{h}(t)=\frac{f(t)}{S(t)} \tag{4}$$

$$\mathrm{h}(t)=-\frac{\partial\left[S(t)\right]}{\partial t} \tag{5}$$

$$\mathrm{h}(t)=exp\left[-H(t)\right] \tag{6}$$

Definición 3. Sea, H la función de riesgo acumulado

$$H(t)=\int_0^t \mathrm{h}(s)ds \tag{7}$$

De modo que, para funciones continuas:

$$H(t)=\int_0^t \frac{dF(s)}{S(s)} \tag{8}$$

para funciones discretas:

$$H(t)=\sum_{s\leq t}\frac{\Delta F(s)}{S(s)} \tag{9}$$

2.2 MODELOS CLÁSICOS DE SUPERVIVENCIA

Los modelos de supervivencia clásicos más utilizados son: el modelo actuarial de Cutler-Ederer, el modelo del limite-producto de Kaplan-Meier y el modelo de riesgos proporcionales de Cox. Estos modelos no consideran la condición de recurrencia del evento de estudio; por lo tanto, no pueden ser utilizados para fenómenos recurrentes, ya que fueron diseñados bajos otros supuestos y otro fin. Tanto los modelos actuariales como el modelo de Kaplan-Meier son estimadores no paramétricos del análisis de supervivencia clásico. Mientras que el modelo de Cox es un modelo clásico que considera el efecto de covariables en el análisis. A continuación, se detallan estos modelos.

2.2.1 Modelos actuariales

Los modelos actuariales del análisis de supervivencia son modelos clásicos, no paramétricos. Entre los más conocidos: Bhomer, Berkson-Gage y Cutler-Ederer [9-10], éstos son aplicables en aquellos casos donde no se dispone de los tiempos exactos de ocurrencia del evento. La información se presenta agrupada en intervalos de tiempo. Por ejemplo, los datos de mortalidad para un país determinado suelen presentarse agrupados en intervalos anuales. Para realizar un análisis de supervivencia es conveniente aplicar métodos actuariales; donde, el período de observación es dividido en subintervalos de tiempo independientes, de igual longitud $[t_0,t_j]$, para todo j=1,2,...,k; donde, t_k es el mayor de los tiempos de ocurrencia del evento en todas las unidades del estudio, t_k coincide con el mayor tiempo de observación de las unidades, que se particiona convenientemente en k subintervalos.

2.2.2 Modelo de Kaplan y Meier

Es un estimador de la función de supervivencia, en presencia de datos censurados u observaciones incompletas, conocido hoy en día como estimador "Límite-producto" [2]. Este modelo, al igual que los estimadores actuariales es un modelo clásico no paramétrico. La diferencia metodológica entre este método y el método actuarial de supervivencia es que en el primero los datos u observaciones no son agrupados en intervalos; mientras que en este modelo no se predefinen intervalos. La función de supervivencia es estimada en cada momento de ocurrencia del evento. El estimador de la función de supervivencia de Kaplan-Meier se basa en el mismo principio actuarial y se calcula como el producto de probabilidades condicionales. El modelo también supone independencia entre las observaciones y para una muestra aleatoria de tamaño n, el estimador Kaplan-Meier con tiempos de supervivencia no repetidos, está dado por (10):

$$S(t_j)=\prod_{i=1}^{j}\frac{n_i-d_i}{n_i} \tag{10}$$

donde, d_i representa el número total de ocurrencia en el *i-ésimo* momento y n_i representa el numero de unidades a riesgo, justo antes del tiempo t_j.

2.2.3 Modelo de Cox

El modelo de riesgos proporcionales propuesto por Cox [3], a diferencia de los modelos anteriores, supone que existe un conjunto de covariables $X=(x_1, x_2,\ldots,x_p)'$, que influyen sobre el comportamiento del tiempo de ocurrencia de los eventos. El modelo asume independencia entre las observaciones de cada unidad y entra en la clasificación de los modelos clásicos, semiparamétricos y multivariante. Donde, la función de riesgo condicionada, está dada por (11):

$$h(t/X) = h_o(t) \times e^{\sum_{j=1}^{q} \beta_j X_j} \tag{11}$$

y en consecuencia:

$$S(t/X) = \left[S_0(t)\right]^{\exp[\beta' X]} \tag{12}$$

donde, $h_0(t)$ es la función de riesgo base, $S_0(t)$ es la función de supervivencia base y β es el vector de parámetros desconocidos que miden el efecto de las covariables.

2.2.4 Otros modelos clásicos del análisis de supervivencia

A continuación se ilustra un conjunto de modelos o métodos de estimación no paramétricos de la función de supervivencia con eventos no recurrentes, conocidos como modelos clásicos (Tabla 1). Esta serie de modelos son adaptaciones o modificaciones del estimador tradicional propuesto por Kaplan-Meier [2]. Cada uno de ellos, aplicables a situaciones particulares del análisis de supervivencia, como por ejemplo, problemas con presencia o ausencia de datos censurados o situaciones donde existen empates o medidas repetidas en los tiempos de supervivencia. El primer modelo que se ilustra en la tabla es el estimador de la función de supervivencia de Kaplan-Meier. El segundo modelo corresponde al propuesto por Alshuler [11], quien utilizó una aproximación de series de Taylor de la expresión: $e^{-x} \sim 1- x$, para su propuesta. Este estimador también es conocido como estimador de Nelson-Aalen**.** También aplicable a problemas con tiempos de supervivencia repetidos. En [12] se propone otro estimador de la función de supervivencia pero sin considerar los empates en los tiempos de ocurrencia del evento. Otros autores [13, 15, 16] generalizaron este último a estudios de supervivencia donde existen tiempos de supervivencia repetidos. Los estimadores propuestos por estos autores son útiles para aquellos casos donde existen empates (d_j). De forma particular, cuando $d_j=1$, el resto de los modelos se transforma en el estimador de Prentice [12].

2.3 Modelos de supervivencia no paramétricos para eventos recurrentes

El surgimiento de los nuevos modelos del análisis de supervivencia con eventos recurrentes, ha dotado a investigadores del área de nuevas herramientas de análisis que le permiten realizar estudios en fenómenos desde tipo. En este trabajo se discuten dos modelos: Peña *et al.* [9] y Wang-Chang [10]. Para realizar las estimaciones en estos modelos, los autores de la presente sustentan los cálculos con herramientas de conteo diseñadas durante la investigación; para realizar las propuestas de esta investigación.

Tabla 1. Estimadores clásicos no paramétricos del análisis de supervivencia.

Autor	Año	Estimador de S(t)	Autor	Año	Estimador de S(t)
Kaplan-Meier	1958	$S(t_j) = \prod_{i=1}^{j} \frac{n_i - d_i}{n_i}$	*Andersen et al.*	1982	$S(t_j) = \left[\prod_{i=1}^{j} \frac{n_i - d_i + 1}{n_i + 1}\right] \frac{n_j}{n_j + 1}$
Alshuler	1970	$S(t_j) = \prod_{i=1}^{j} exp\left(-\frac{d_i}{n_i}\right)$	*Harris-Albert*	1991	$S(t_j) = \prod_{i=1}^{j} \frac{n_i + d_i - 1}{n_i + d_i}$
Prentice	1978	$S(t_j) = \prod_{i=1}^{j} \frac{n_i}{n_i + 1}$	*Moreau et al.*	1992	$S(t_j) = \prod_{i=1}^{j} \frac{n_i}{n_i + d_i}$
Prentice-Marek	1979	$S(t_j) = \prod_{i=1}^{j} \frac{n_i - d_i + 1}{n_i + 1}$	*Hosmer-Lemeshow*	1999	$S(t_j) = \left[\prod_{i=1}^{j-1} \frac{n_i - d_i + 1}{n_i + 1}\right] \frac{n_j}{n_j + 1}$

conteo diseñadas durante la investigación; para realizar las propuestas de esta investigación.

2.3.1 Modelo de Wang y Chang (Modelo WC)

En [10] se expuso un modelo de estimación de la función de supervivencia para eventos recurrentes, aplicable tanto a casos donde existen independencia entre los tiempos entre ocurrencias (p.e.: presencia de datos correlacionados) como aquellos casos con presencia de datos correlacionados. El estimador WC es un estimador que se puede definir utilizando dos procesos de conteo (13): $d^*(t)$ y $R^*(t)$, donde $d^*(t)$ representa la suma total de las proporciones de tiempos de entre ocurrencias iguales (t), y $R^*(t)$ representa el promedio de individuos que están en riesgo en el momento t.

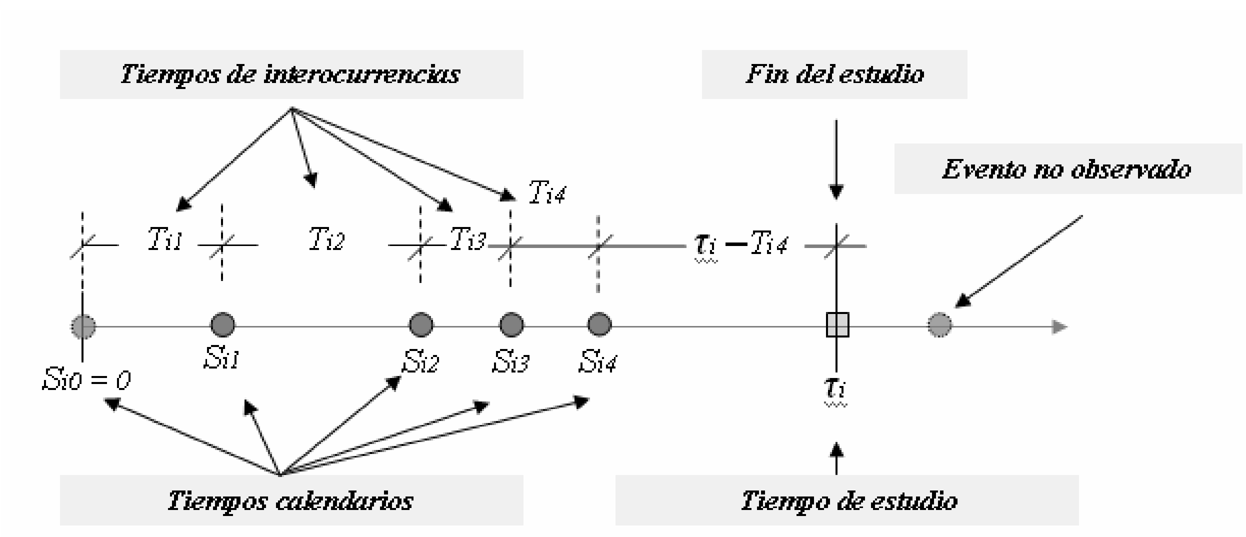
$$\hat{S}(t) = \prod_{i=1}^{n} \prod_{\{j:T_{ij} \leq t\}} \left[1 - \frac{d^*\left(T_{ij}\right)}{R^*\left(T_{ij}\right)}\right] \tag{13}$$

donde, T_{ij} es el tiempo de interocurrencia del *j-ésimo* evento en la *i-ésima* unidad de investigación, S_{ij} es el tiempo calendario de la *j-ésima* ocurrencia del evento en la *i-ésima* unidad. Con $S_{i0}=0$ y $S_{ij}=T_{i1}+T_{i2}+\ldots+T_{iki}$, K_i es el número de eventos que experimenta la *i-ésima* unidad de investigación, $K_i = max\{j: S_{ij} \leq \tau_i\}$, τ_i es el tiempo total de observación de la *i-ésima* unidad de investigación y n es el número total de ocurrencias del evento (Figura 1).

2.3.2 Modelos de Peña, Strawderman y Hollander (Modelo PHS)

Peña-Strawderman-Hollander [9] presentan dos modelos de supervivencia para eventos recurrentes, uno de los modelos generaliza el estimador clásico de supervivencia de Kaplan-Meier a eventos recurrentes, útil para aquellos casos donde existe independencia entre los tiempos de interocurrencia, y otro modelo que considera la fragilidad o probabilidad de ocurrencia del evento en la unidad de estudio. En su trabajo, presentaron el estimador no paramétrico de la función de supervivencia, que generaliza el estimador limite-producto de Kaplan-Meier del caso clásico, al caso de eventos recurrentes. Para esta propuesta, los autores definieron dos procesos de conteo N e Y, que permiten realizar las estimaciones de su modelo y diseñaron una herramienta gráfica que permiten realizar estos proceso (*Apéndice A*). En el caso de una única ocurrencia del evento, los procesos quedan bien definidos por N e Y. Pero para el caso de eventos recurrentes, fue necesario definir dos escalas de tiempo: un tiempo calendario S y un tiempo de interocurrencias T. El tiempo calendario S acumula los tempos de interocurrencias T en cada una de las unidades bajo estudio. Los procesos N e Y se definen como: $N(s,t)$ e $Y(s,t)$. Donde $N(s,t)$ representa el número de eventos observados en el período calendario $[0,s]$ con tiempos de interocurrencias menores e iguales a t e $Y(s,t)$ representa el número de eventos observados en el período calendario $[0,s]$ con tiempos de interocurrencias mayores e iguales a t. De modo que, si $N_i(s,t)$ es el número de eventos observa-

Figura 1. Representación gráfica de la recurrencia de eventos en la *i-ésima* unidad de investigación.



dos en el período calendario *[0, s]*, con tiempos de interocurrencias menores e iguales a *t* para la *i-ésima* unidad de investigación, $Y_i(s,t)$ será el número de eventos observados en el período calendario *[0,s]*, con tiempos de interocurrencias mayores e iguales a *t*, en cada unidad bajo estudio.

$$N(s,t)=\sum_{i=1}^{n} N_i(s,t) \quad o \quad N(s,t)=\sum_{i=1}^{n}\sum_{j=1}^{K_i} I\{T_{ij} \le t\} \tag{14}$$

$$Y(s,t)=\sum_{i=1}^{n} Y_i(s,t) \tag{15}$$

$$Y(s,t)=\sum_{i=1}^{n}\left\{\sum_{j=1}^{K_i} I\left\{T_{ij} \ge t\right\}+I\left\{\tau_i - S_{iK_i} \ge t\right\}\right\} \tag{16}$$

De esta manera, el estimador generalizado del límite-producto de Kaplan-Meier se define como (17):

$$\hat{S}(t)=\prod_{w\le t}\left[1-\frac{N(s,\Delta w)}{Y(s,w)}\right] \tag{17}$$

El otro modelo propuesto por Peña *et al.* [9] es el modelo de fragilidad, donde se asume que existe una variable aleatoria no medible o variable latente que representa el grado de heterogeneidad entre las unidades de los grupos en estudio, la cual se asume independiente de la censura. Este es conocido como modelo de fragilidad multiplicativa, y asume que la variable de fragilidad sigue una distribución gamma con parámetros de forma y escala iguales a *α*. Donde, para cada unidad bajo estudio, existe una variable de fragilidad no observable y positiva, digamos Z_i; tal que, para cada momento *t* la función de supervivencia condicionada definida por (18):

$$S(t/Z_i = z)=[S_0(t)]^z \quad o \quad S(t/Z_i = z)=e^{-z\int_0^t \lambda_0(u)du} \tag{18}$$

donde $\lambda_0(t)$ es la función de riesgo base y $S_0(t)$ es la función de supervivencia basal. Con $Z_1,Z_2,\ldots,Z_n$ como las variables de fragilidad de las unidades, se asumen idénticamente distribuidas con una cierta distribución común H para todas. En el modelo FRMLE, H se distribuye gamma con parámetros de forma y escala iguales a *α*. De esta manera, la función de supervivencia conjunta de $(T_{i1},T_{i2},\ldots,T_{ik})$ queda definida por (19):

$$S(t)=\int_0^{\infty}\left[\prod_{j=1}^{k} S_0\left(t_j\right)\right]^{z}\frac{\alpha^{\alpha}}{\Gamma(\alpha)}z^{\alpha-1}e^{-\alpha z}dz$$

$$S(t)=\left[\frac{\alpha}{\alpha+\sum_{j=1}^{k}\Lambda_0\left(t_j\right)}\right]^{\alpha} \tag{19}$$

Simplificando (19), la función de supervivencia del modelo FRMLE se puede expresar como (20):

$$S(t)=\left[\frac{\alpha}{\alpha+\Lambda_0(t)}\right]^{\alpha} \tag{20}$$

donde $\Lambda_0(t)$ es la función marginal de riesgo acumulado, sus estimaciones son hechas a través de los métodos de verosimilitud parcial utilizados en los modelos semiparamétricos del análisis de supervivencia de Cox. Peña *et al.,* demostraron que α y Λ_0(t) pueden ser obtenidas vía maximización de la función marginal de verosimilitud de α y Λ_0(.), y sus estimaciones pueden realizarse a través de la implementación del algoritmo EM (*Expectation-Maximization*). El estimador de la función de supervivencia del modelo FRMLE de PHS es representada por (21):

$$\hat{S}(t)=\left[\frac{\hat{\alpha}}{\hat{\alpha}+\hat{\Lambda}_0(s,t)}\right]^{\alpha} \tag{21}$$

$\hat{\Lambda}_0(s,t)$: estimador de la función marginal de riesgo acumulado Λ_0(t)

3. GENERALIZACIÓN DE LOS MODELOS CLÁSICOS DEL ANÁLISIS DE SUPERVIVENCIA AL CASO RECURRENTE

Las propuestas se fundamentan en el trabajo de Peña *et al.* [9], cuyos autores derivaron el estimador GPLE para eventos recurrentes a partir de los estimadores clásicos Nelson-Aalen y Kaplan-Meier. Estos autores diseñaron dos procesos contadores a los que denotaron *N(s,t)* y *Y(s,t)*. Estos procesos de conteo tienen sus equivalentes en los estudios de supervivencia del análisis clásico. Para el caso de una única ocurrencia del evento por unidad de estudio, la simbología utilizada es la siguiente: d_i y n_i. Donde n_i representa los eventos ocurridos en el momento *i-ésimo* y d_i representa el número de unidades a riesgos justo antes de la ocurrencia del evento en ese instante. En esta investigación se utilizan los dos procesos contadores diseñados en [9] como base fundamental en la generalización los estimadores de Altshuler [11], Prentice [12], Prentice-Marek [13], Andersen *et al*. [5], Moreau [16] y Hosmer-Lemeshow [17] a fenómenos con eventos recurrentes, (Tabla 2).

4. APLICACIÓN Y DISCUSIÓN DE RESULTADOS

Los modelos propuestos se aplicaron a un conjunto de datos que corresponde a un estudio de motilidad del intestino delgado (*Apéndice B*) [9][7] [18]. Los datos de esta investigación están disponibles en la librería *survrec* del lenguaje R. El estudio original de los investigadores consistía en estimar el tiempo medio de supervivencia del periodo del Complejo Motor Migratorio (CMM). En esta investigación se pretende comparar los resultados de los ajuste de los modelos de Wang-Chang y Peña *et al.*, con los ajustes de los

Tabla 2. Generalización de los estimadores clásicos no paramétricos del análisis de supervivencia al caso de eventos recurrente.

Autor	*Año*	*Estimador de S(t)*	*Notación*
Kaplan-Meier	1958	$\prod_{w\le t_j}\left[1-\frac{N(s,\Delta w)}{Y(s,w)}\right]$	GPLE
Alshuler	1970	$\prod_{w\le t_j}\exp\left[-\frac{N(s,\Delta w)}{Y(s,z)}\right]$	GEAlsh
Prentice	1978	$\prod_{w\le t_j}\left[\frac{Y(s,w)}{Y(s,w)+1}\right]$	GEPren
Prentice-Marek	1979	$\prod_{w\le t_j}\left[\frac{Y(s,w)+1-N(s,\Delta w)}{Y(s,w)+1}\right]$	GEPrMa
Andersen et al.	1982	$\prod_{w\le t_j}\left[\frac{Y(s,w)+1-N(s,\Delta w)}{Y(s,w)+1}\right]\frac{Y(s,w)}{Y(s,w)+1}$	GEAnde
Harris-Albert	1991	$\prod_{w\le t_j}\left[\frac{Y(s,w)-1+N(s,\Delta w)}{Y(s,w)+N(s,\Delta w)}\right]$	GEHaAl
Moreau et al.	1992	$\prod_{w\le t_j}\left[\frac{Y(s,w)}{Y(s,w)+N(s,\Delta w)}\right]$	GEPrMo
Hosmer-Lemeshow	1999	$\prod_{w\le t_{j-1}}\left[\frac{Y(s,w)+1-N(s,\Delta w)}{Y(s,w)+1}\right]\frac{Y(s,w)}{Y(s,w)+1}$	GEHoLE

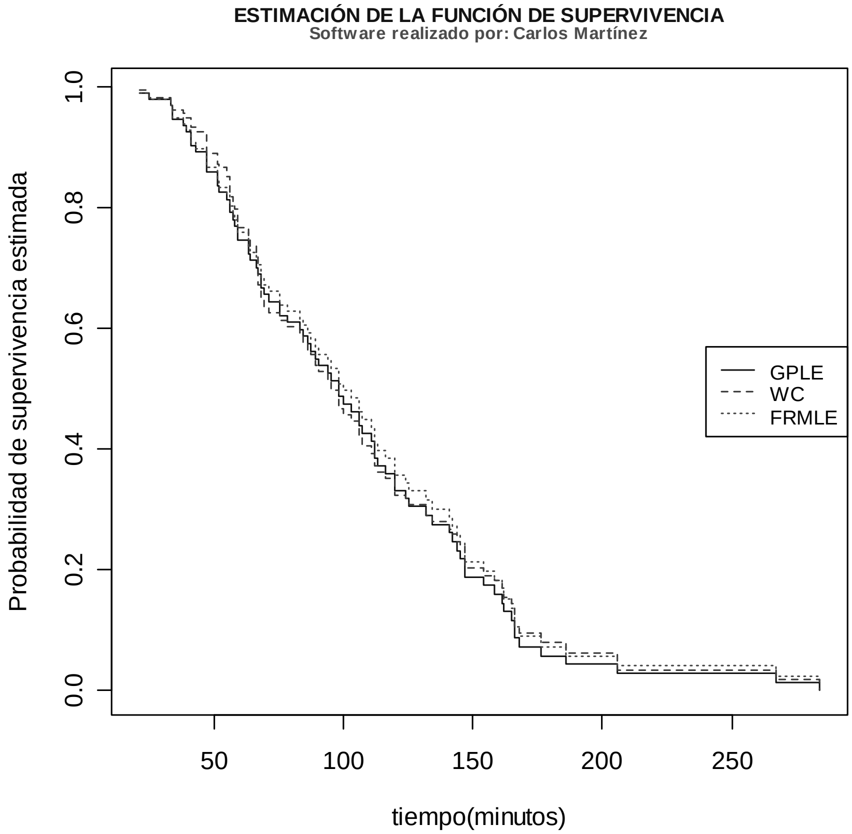


Figura 2. Estimaciones de *S(t)* de los modelos GPLE, WC y FRMLE.

Tabla 2. Medidas descriptivas de las estimaciones de los modelos GPLE, WC y FRMLE de los períodos del CMM.

Cuartiles(min)	**GPLE**	**WC**	**FRMLE**
1er	142	144	145
2do	98	95	100
3er	59	63	63
A	-	-	10.18

nuevos modelos. Para obtener las estimaciones de los ajustes de los nuevos estimadores, se diseñaron rutinas en lenguaje R (apéndice C).

La Figura 2 y la Tabla 3 ilustran los resultados de las estimaciones de los modelos GPLE, WC y FRMLE. En la Tabla 3 se observan la estimaciones de las medianas de los modelos de supervivencia realizadas a través de esos modelos y cuyos resultados son: 98, 95 y 100 minutos, respectivamente. Según se puede apreciar en la tabla estas estimaciones de las medianas son bastante parecidas, considerando la escala de tiempo utilizada. Igual sucede con las estimaciones del primer y tercer cuartil. La tabla también muestra la estimación del parámetro alfa, en el modelo de fragilidad de Peña *et al.* (α=10.17623). Este parámetro mide el grado de asociación entre los tiempos de ocurrencia. Valores pequeños de alfa y cercanos a cero, significa alta dependencia entre estos tiempos y valores muy altos del parámetro, significa independencia. En este caso, α=10.17623, lo que indica la existencia de una asociación débil entre los tiempos de interocurrencias del fenómeno.

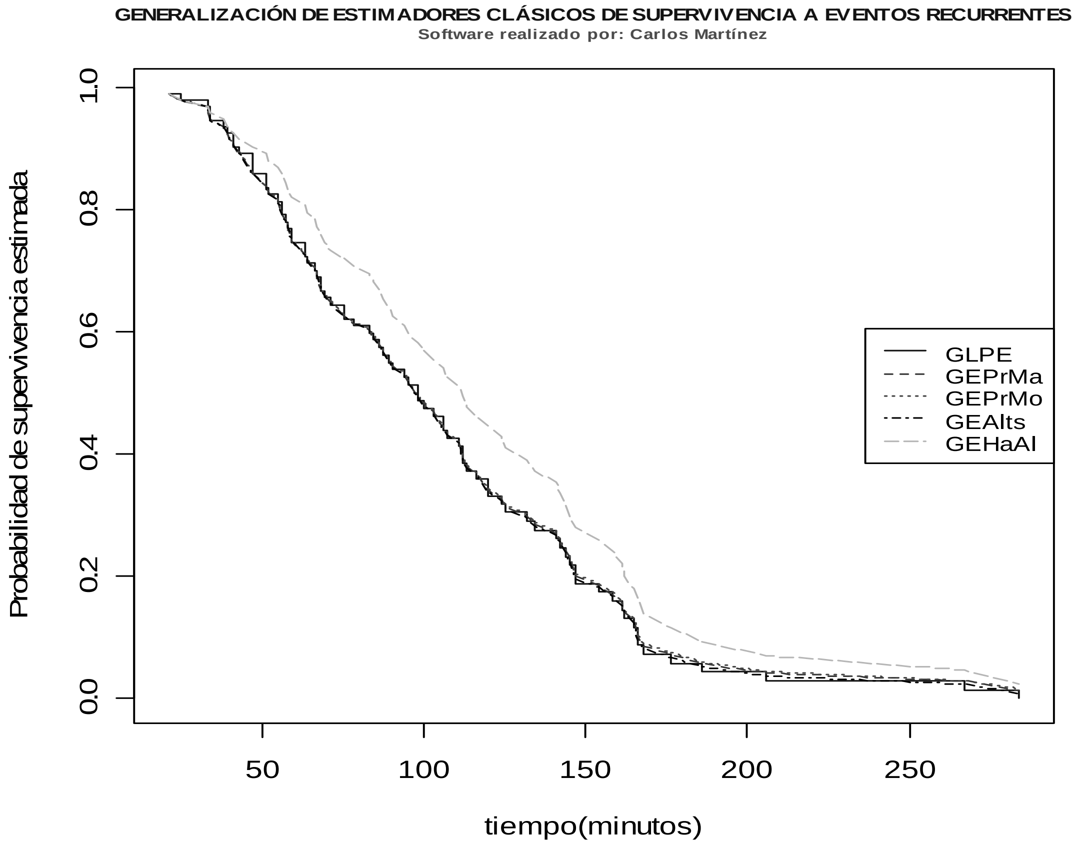


Figura 2. Estimaciones de *S(t)* de los nuevos modelos.

La Figura 3 muestra que los nuevos modelos se comportan de manera similar al estimador GPLE, propuesto por Peña *et al.* [9], con excepción del estimador tipo Harris-Albert (GEHaAl), que sobrestima de manera apreciativa la curva de supervivencia del período CMM. Esto hace suponer que en condiciones similares, las propiedades estadísticas de nuevos estimadores son también similares al estimador GPLE, excepto el estimador GEHaAl.

4. CONCLUSIONES

En este trabajo se generalizan los modelos clásicos no paramétricos del análisis de supervivencia a casos donde los eventos son de naturaleza recurrente. Los nuevos estimadores de las curvas de supervivencia fueron aplicados a un conjunto de datos provenientes de un experimento que trata sobre el estudio de motilidad del intestino delgado, experimento aplicado a 19 pacientes que originalmente fueron tratados por Aalen Husebye [18] y posteriormente utilizados por González-Peña [7] en las estimaciones de los modelos GPLE, WC y FRMLE . Los resultados de los ajustes y las estimaciones de cada modelo se obtuvieron a partir de rutinas computarizadas, diseñadas y programadas en lenguaje R, v2.3.1. Los nuevos estimadores de la función de supervivencia con eventos recurrentes, propuestos en este trabajo se comportan de manera similar a los estimadores GPLE de Peña *et al* [9], Wang-Chang [10] y el modelo de fragilidad de Peña *et al.* o FRMLE. Estos nuevos estimadores tienden a comportarse a los modelos GPLE, **WC** y FRMLE en la medida en que los tiempos de interocurrencias del evento no están correlacionados, tal y como se ilustró en el ejemplo práctico de esta investigación. Todos los modelos de supervivencia con eventos de carácter recurrentes discutidos en este trabajo solo consideran fenómenos con datos censurados por la derecha.

En nuestro ejemplo, se observó que los períodos del CMM del experimento de Aalen-Husebye[18] son aproximadamente 98 minutos y que las curvas de supervivencia estimadas a través de los nuevos modelos se comportan de igual manera que los modelos de WC, GPLE y FRMLE. A excepción del modelo generalizado de Harris-Albert que sobrestima la supervivencia del fenómeno de estudio. Todos los modelos considerados en este trabajo con excepción de este último, arrojaron que la mediana estimada de los períodos del CMM está alrededor de los 98 minutos.

5. REFERENCIAS

[1] Cutler S. y Ederer F. (1958), Maximun utililization on the life table method in analysing survival, Biostatistics, Journal of Choronic Diseases, 8, 699-712

[2] Kaplan, E. L. y Meier, P. (1958), Nonparametric estimation from incomplete observations, Journal of the American Statistical Association, 53, 457-481.

[3] Cox, D. R. (1972), Regression models and life-tables, Journal of the Royal Statistical Society, Series B, 34, 187-220.

[4] Prentice, R., Williams, B. y Peterson, A. (1981), On the regression analysis of multivariate failure time data, Biometrika, 68(2), 373-379.

[5] Andersen, P. y Gill, R. (1982), Cox's regression model for counting processes: A large sample study, Annals of Statistics, 10, 1100-1120.

[6] Wei, L., Lin, D. y Weissfeld, L. (1989), Regression analysis of multivariate incomplete failure time data by modeling marginal distributions, Journal of the American Statistical Association, 84, 1065-1073.

[7] González, J. y Peña, E. (2004), Estimación no parametrica de la función de supervivencia para datos con eventos recurrentes, Revista Española de Salud Pública, 78, 189-199.

[8] González, J. y Peña, E. (2003), Bootstraping median survival with recurrent event data, IX conferencia Española de Biometría, A Coruña, Mayo 28-30.

[9] Peña, E., Strawderman, R. y Hollander, M. (2001), Nonparametric estimation with recurrent event date, JASA, 99, 1299-1315.

[10] Wang, M. y Chang, S. (1999), Nonparametric estimation of a recurrent survival function, Journal of the American Statistical Association, 94, 146-153.

[11] Alshuler, B. (1970), Theory for measurement of competing risks in animal experiments, Math. Biosciences, 6, 1-11.

[12] Prentice, R. (1978), Linear rank tests with right censored data, Biometrika, 65(1), 167-179.

[13] Prentice, R., Marek, P. (1979), Qualitative discrepancy between censored data rank test, Biometrics, 35, 861-867.

[14] Fleming, T., O'Fallon, J., O'Brien, P. y Harrinton, D. (1980), Modified Kolmogorov-Smirnov test procedures with application to arbitrarily right-censored data, Biometrics, 36, 607-625.

[15] Harris, A. (1991), Survivorship analysis for clinical studies., Marcel Dekker, New York, USA.

[16] Moreau, T., Maccario, J., Lellouch, J., Huber, C. (1992), Weighted log rank statistic for comparing two distributions, Biometrika, 79(1), 195-198.

[17] Hosmer, D., Lemeshow, S. (1999), Applied survival analysis: regression modeling of time to event data, John Wiley and Sons, Inc., New york, USA.

[18] Aalen, O. y Husebye, E. (1991), Statistical analysis of repeated events forming renewal processes, Sta. Med., 10, 1227-1240.

6. APENDICE A

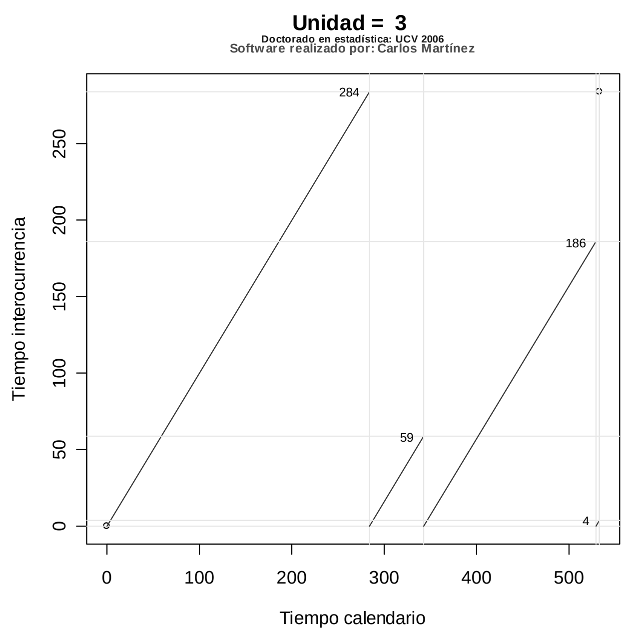


Figura 3. Tiempos de interocurrencias del evento de la unidad 3, en el experimento de Aalen-Husebye [18].

i = 3 Este subíndice indica el número de la unidad de investigación (unidad =3)

K_3(s=350) = 2, Indica que para un tiempo calendario igual a 350 minutos, la unidad ha experimentado dos eventos.

N_3(s=350,t=150) = 1, Indica que para un tiempo calendario igual 350 minutos, la unidad ha experimentado un evento con tiempo de interocurrencia menor o igual a 150 minutos.

Y_3(s=350,t=150) = 1, Indica que en un tiempo calendario igual a 350 minutos, la unidad estuvo en riesgo sólo en una ocasión con un tiempo de interocurrencia mayor o igual a 150 minutos.

7. APENDICE B

Tabla 4. Datos correspondientes al experimento de Aalen-Hussebye [18]

i	*K i*	*T_{ij} Tiempos de interocurrencias*	*Tiempo de censura*
1	*8*	*112, 145, 39, 52, 21, 34, 33, 51*	*54*
2	*2*	*206, 147*	*30*
3	*3*	*284, 59, 186*	*4*
4	*3*	*94, 98, 84*	*87*
5	*1*	*67*	*131*
6	*9*	*124, 34, 87, 75, 43, 38, 58, 142,75*	*23*
7	*5*	*116, 71, 83, 68, 125*	*111*
8	*4*	*111, 59, 47, 95*	*110*
9	*4*	*98, 161, 154, 55*	*44*
10	*2*	*166, 56*	*122*
11	*5*	*63, 90, 63, 103, 51*	*85*
12	*4*	*47, 86, 68, 144*	*72*
13	*3*	*120, 106, 176*	*6*
14	*4*	*112, 25, 57, 166*	*85*
15	*3*	*132, 267, 89*	*86*
16	*5*	*120, 47, 165, 64, 113*	*12*
17	*4*	*162, 141, 107, 69*	*39*
18	*6*	*106, 56, 158, 41, 41, 168*	*13*
19	*5*	*147, 134, 78, 66, 100*	*4*

8. APENDICE C

Código en Lenguaje R

```
require(survrec);cmmm<-choose.files(filters=Filters[c('txt'),],index=4)
nombre<-as.character(cmmm);XL<-read.table(nombre, header = TRUE,fill
= TRUE)
p<-ncol(XL);N<-nrow(XL)
cm<-survfitr(Survr(XL$ID,XL$Time,XL$event)~1,data=XL,type="pe")
xb<-c("Modelo clásico","1er Modelo PWP","2do Modelo PWP",
        "Modelo AG","Modelo WLW","Modelo GPLE","Modelo
          WC","Modelo      FRMLE","Modelo CMMM")
su<-matrix(summary(cm));nnn<-nrow(su)/4;Tevent<-matrix(su
[1:nnn,1],nnn,1)
Neventos<-matrix(su[(nnn+1):(2*nnn),1],nnn,1)
NAriesgo<-matrix(su[(2*nnn+1):(3*nnn),1],nnn,1)
SKM<-matrix(su[(3*nnn+1):(4*nnn),1],nnn,1)
NumeradorEstimadorPM<-NAriesgo-Neventos+matrix(1,nnn,1)
DenominadorEstimadorPM<-NAriesgo+matrix(1,nnn,1)
ProbEstimadorPM<-NumeradorEstimadorPM*(1/
DenominadorEstimadorPM)
Modelo<-  select.list(xb, multiple = FALSE,title="Seleccione el modelo")
p<-1;EstimadorGeneralizadoPrenticeMarek<-matrix(1,nnn,1)
for (i in 1:nnn) {EstimadorGeneralizadoPrenticeMarek[i,1]<-
p*ProbEstimadorPM[i,1]
p<-EstimadorGeneralizadoPrenticeMarek[i,1]}
NumeradorEstimadorPMoreau<-NAriesgo
DenominadorEstimadorPMoreau<-NAriesgo+Neventos
ProbEstimadorPMoreau<-NumeradorEstimadorPMoreau*(1/
DenominadorEstimadorPMoreau)
p<-1;EstimadorGeneralizadoPrenticeMoreau<-matrix(1,nnn,1)
for (i in 1:nnn) {EstimadorGeneralizadoPrenticeMoreau[i,1]<-
p*ProbEstimadorPMoreau[i,1]
p<-EstimadorGeneralizadoPrenticeMoreau[i,1]}
NumeradorEstimadorAltshuler<-Neventos
DenominadorEstimadorAltshuler<-NAriesgo
ProbEstimadorAltshuler<-exp(-NumeradorEstimadorAltshuler*(1/
DenominadorEstimadorAltshuler))
p<-1;EstimadorGeneralizadoAltshuler<-matrix(1,nnn,1)
for (i in 1:nnn) {
EstimadorGeneralizadoAltshuler[i,1]<-p*ProbEstimadorAltshuler[i,1]
p<-EstimadorGeneralizadoAltshuler[i,1]}
NumeradorEstimadorHA<-NAriesgo+Neventos-matrix(1,nnn,1)
DenominadorEstimadorHA<-NAriesgo+Neventos
ProbEstimadorHA<-(NumeradorEstimadorHA*(1/
DenominadorEstimadorHA))
p<-1;EstimadorGeneralizadoHA<-matrix(1,nnn,1)
for (i in 1:nnn) {
EstimadorGeneralizadoHA[i,1]<-p*ProbEstimadorHA[i,1]
p<-EstimadorGeneralizadoHA[i,1]}
X11()
plot(cm$time,cm$survfunc,xlab="tiempo(minutos)",ylab="Probabilidad de
supervivencia estimada",
     type="s",col = "blue", lwd=1)
title(main=list("GENERALIZACIÓN DE ESTIMADORES CLÁSICOS
DE SUPERVIVENCIA A EVENTOS RECURRENTES",       cex = 0.7,
font = 2.3,col = "blue"))
mtext("Software realizado por: Carlos Martínez",       cex = 0.6, font =
2,col = "purple",line=1)
lines(Tevent,EstimadorGeneralizadoPrenticeMarek,lty=2, col = "red")
lines(Tevent,EstimadorGeneralizadoPrenticeMoreau,lty=3, col = "brown")
lines(Tevent,EstimadorGeneralizadoAltshuler,lty=4, col = "black")
lines(Tevent,EstimadorGeneralizadoHA,lty=5, col = "green")
legend("right",c("GLPE","GEPrMa","GEPrMo","GEAlts","GEHaAl"),
col= c("blue","red","brown","black","green"), lty=c(1,2,3,4,5),cex=0.8)
failed<-matrix(cm$failed);censored<-matrix(cm$censored);time<-matrix
(cm$time)
n.event<-matrix(cm$n.event) ;AtRisk<-matrix(cm$AtRisk);m<-cm$m
GPLE1erq<-q.search(cm,q=0.25);GPLE1erq;GPLE2doq<-q.search
(cm,q=0.50);GPLE2doq
GPLE3erq<-q.search(cm,q=0.75);GPLE3erq
GPLE1erq<-q.search
(EstimadorGeneralizadoPrenticeMarek,q=0.25);GPLE1erq
GPLE2doq<-q.search
(EstimadorGeneralizadoPrenticeMarek,q=0.50);GPLE2doq
GPLE3erq<-q.search
(EstimadorGeneralizadoPrenticeMarek,q=0.75);GPLE3erq
#--------------------------
```