

AJUSTE DEL ESTIMADOR DE TABLA DE VIDA PARA ENFRENTAR SITUACIONES DE RIESGOS COMPETITIVOS DEPENDIENTES

Mayelin Mirabal Sosa*¹, Armando Seuc Jo** y Rolando Uranga Piña***

*Instituto Finlay. Investigación y Producción de Vacunas

**Instituto Nacional de Angiología y Cirugía Vascular

***Centro Nacional Coordinador de Ensayos Clínicos

RESUMEN

Los desarrollos en el marco de un problema clásico de “riesgos competitivos” en su mayoría han consistido en, a partir del supuesto de independencia entre las causas competitivas, hacer estimaciones de diversas tasas. En este artículo se proponen ajustes de los estimadores tipos Tabla de Vida (Clásico y Tasa Neta) en una situación de riesgos competitivos para cuando no se cumple el supuesto de independencia. Se demuestra que el estimador propuesto es pseudo-máximo verosímil. Se comparan, mediante simulación, los estimadores propuestos en el trabajo y los estimadores existentes lo que corrobora el hecho de que el estimador ajustado brinda estimaciones más cercanas a la tasa real que las obtenidas al usar los estimadores clásicos.

ABSTRACT

The majority of the developments in a classic problem of competing risks have consisted in, based on the assumption of independence between all event types, make estimations to several rates. In this paper, we propose an adjusted Life-Table estimator in the context of non-independent competing risks. We proof that the proposed estimator is maximum pseudo-likelihood. We compare the adjusted and classic estimators by simulation study. It confirmed that de adjusted estimator provides closer estimations to the real rate than the classic estimator.

KEY WORDS: Life Table, Independence, Net Rate, Likelihood.

MSC 62P10

1 INTRODUCCIÓN

Las situaciones de riesgos competitivos constituyen un problema estadístico y práctico de especial importancia porque el cambio (aumento o disminución) en la tasa de ocurrencia del evento por una de las causas puede influir de disímiles maneras en el comportamiento de las tasas de ocurrencia por otras causas. Por ejemplo, el incremento observado en los últimos 10 años en las tasas de mortalidad por cáncer en algunos países europeos puede deberse a que las tasas de mortalidad por enfermedades del corazón han disminuido en el mismo periodo, lo que posibilita que los sobrevivientes de estas últimas tengan ahora la posibilidad - y de hecho lo hagan - de morir por cáncer. Ver Llorca-Delgado [2004].

Al parecer fue Chiang [1961], quien primero intentó estructurar la situación de riesgos competitivos mediante la definición de tres tasas de interés diferentes:

- a) Tasa Global de ocurrencia del evento, definida como la probabilidad global de ocurrencia del evento por cualquiera de las causas.
- b) Tasa Cruda de ocurrencia del evento por una determinada causa, definida como la probabilidad de ocurrencia del evento por dicha causa en presencia de todas las causas restantes.
- c) Tasa Neta de ocurrencia del evento por una determinada causa, definida como la probabilidad de ocurrencia del evento por dicha causa en una población hipotética en la cual sólo estuviera activa esa causa.

¹ mmirabal@finlay.edu.cu

Gran parte de los desarrollos estadísticos alrededor de este tema han consistido en, a partir de ciertos supuestos, obtener estimaciones de las distintas tasas definidas previamente. Por ejemplo, han sido desarrolladas distintas variantes de los métodos de Kaplan-Meier y de Tabla de Vida las que han estado basadas fundamentalmente en los supuestos de independencia de las censuras o de las causas competitivas. Ver Fiocco-Putter [2005], Zeng [2004], Satagopan-Ben Porat [2004].

El uso inapropiado de estos desarrollos en el marco de los estudios de riesgo competitivo, ha introducido cierto grado de confusión en el tema. Las razones principales de esta confusión son:

1. La selección inadecuada del estimador, pues se han usado indiscriminadamente sin considerar las circunstancias bajo las que cada cual es más apropiado.
2. La utilización de los estimadores de la Tasa Neta aún cuando son violados los supuestos de independencia que los sustentan.

En este trabajo se proponen ajustes de los estimadores tipos Tabla de Vida (Clásico y Tasa Neta) en una situación de riesgos competitivos para cuando no se cumple el supuesto de independencia entre los tiempos de ocurrencia de las censuras y los eventos. Se presenta además una función de pseudo-verosimilitud y se esboza la demostración de que el estimador ajustado es el estimador máximo pseudo-verosímil. Son presentados los resultados obtenidos de simulaciones que permiten ilustrar el proceso de estimación.

2. AJUSTE PARA LOS ESTIMADORES TIPO TABLA DE VIDA

En lo que sigue se propone un ajuste de los estimadores tipos Tabla de Vida, basado en información “a priori” acerca de la interrelación entre las causas competitivas.

2.1. El Estimador Tabla de Vida Clásico Ajustado

En el marco de un estudio de supervivencia clásico y cuando se asume que los tiempos de ocurrencia del evento no son conocidos exactamente, sino que se hallan dentro de cierto intervalo, se define el estimador de Tabla de Vida Clásico o estimador actuarial. Ver en Farley [1986]:

$$\hat{F}_{211}(t_k) = 1 - \hat{S}(t_k) = 1 - \prod_{j \leq k} \left(1 - \frac{d_j}{n_j - \frac{1}{2}w_j} \right) \quad (1)$$

donde:

d_j : número de eventos durante el intervalo $[t_{j-1}, t_j)$

w_j : número de censuras durante el intervalo $[t_{j-1}, t_j)$

n_j : número de sujetos en riesgo al inicio del intervalo $[t_{j-1}, t_j)$

Sea (J_i, σ_i) $i = 1, \dots, n$ una muestra independiente de datos de supervivencia censurados por la derecha, donde J_i es un indicador del intervalo ($J_i = j$ si al individuo i le ocurrió un evento o una censura en $[t_{j-1}, t_j)$) y σ_i es un indicador de censura ($\sigma_i = 1$ si el individuo es censurado y $\sigma_i = 0$ en otro caso). Sea $\varepsilon_{ij} = I_{(J_i=j)}$. Hagamos $\sigma = (\sigma_1, \dots, \sigma_n)$, $\varepsilon_j = (\varepsilon_{1j}, \dots, \varepsilon_{nj})$.

Es válida entonces la siguiente expresión para el número de censuras durante el intervalo $[t_{j-1}, t_j)$

usada en la fórmula (1): $w_j = \sum_{i=1}^n \sigma_i \varepsilon_{ij}$.

Definamos: $\lambda_j = \frac{1}{2} w_j = \frac{1}{2} \sum_{i=1}^n \sigma_i \varepsilon_{ij}$ (2)

Introduzcamos un nuevo vector $r = (r_1, \dots, r_n)$, cuyas componentes indican una penalización impuesta a cada individuo por el hecho de ser una censura potencial. Podría pensarse en la siguiente generalización sencilla de (2):

$$\lambda_j = \frac{1}{2} \sum_{i=1}^n \sigma_i \varepsilon_{ij} r_i \quad (3)$$

Tras esta idea subyace la condición $r_i \leq 1$ ya que el denominador de (1) no debe ser inferior al numerador. Si los r_i son interpretados como probabilidades esta restricción no representa tal dificultad. Sin embargo, como veremos, los r_i pueden ser interpretados como riesgos relativos y admiten, por tanto, valores positivos arbitrarios. La expresión (3) puede ser corregida del siguiente modo:

$$\lambda_j = \sum_{i=1}^n \sigma_i \varepsilon_{ij} \frac{r_i}{1 + r_i} \quad (4)$$

Pueden considerarse expresiones de λ_j más generales.

Sea $G : \mathfrak{R}_+^{3n} \rightarrow \mathfrak{R}_+$, entonces (2) se generaliza tomando $\lambda_j = G(\sigma, \varepsilon_j, r)$.

Es necesario imponer a la función G la restricción $G(x, y, z) \leq |x|$, donde $|x|$ denota la suma de las componentes del vector x y x, y, z representan tres vectores en \mathfrak{R}_+^n . Si se elige, por ejemplo, $G(x, y, z) = g(xy, yz)$ ($xy =$ producto escalar de x, y), se obtiene $\lambda_j = g(w_j, R_j)$, donde, como es usual, w_j es el número de censuras en el intervalo j y R_j es la penalización correspondiente al intervalo j . La función g debe satisfacer $g(u, v) \leq u$.

El estimador de Tabla de Vida Clásico Ajustado lo definimos como:

$$\hat{F}_{221}(t_k) = 1 - \hat{S}(t_k) = 1 - \prod_{j \leq k} \left(1 - \frac{d_j}{n_j - \lambda_j} \right) \quad (5)$$

donde λ_j puede estar dado por alguna de las expresiones descritas.

2.2. El estimador tabla de vida ajustado para la tasa neta

En una situación de riesgos competitivos, el estimador tipo Tabla de Vida para la Tasa Neta se define en Farley [1986] como:

$$\begin{aligned}\hat{F}_{212}^i(t_k) &= 1 - \hat{S}_i(t_k) \\ &= 1 - \prod_{j \leq k} \left(1 - \frac{d_{ji}}{n_j - \frac{1}{2}(w_j + d_j - d_{ji})} \right)\end{aligned}\quad (6)$$

donde:

d_j : número de eventos por cualquier causa ocurridos en el intervalo $[t_{j-1}, t_j)$

d_{ji} : número de eventos por la causa de interés i en el intervalo $[t_{j-1}, t_j)$

w_j : número de censuras en el intervalo $[t_{j-1}, t_j)$

n_j : número de sujetos en riesgo al evento, por la causa de interés i , al inicio del intervalo $[t_{j-1}, t_j)$

En una situación de riesgos competitivos dependientes, podemos ajustar el estimador anterior de la siguiente forma:

$$\begin{aligned}\hat{F}_{222}^l(t_k) &= 1 - \hat{S}_l(t_k) \\ &= 1 - \prod_{j \leq k} \left(1 - \frac{d_{jl}}{n_j - \lambda_{jl}} \right)\end{aligned}\quad (7)$$

con $\lambda_{jl} = G(\sigma_l, \varepsilon_j, r_l)$, $\sigma_l = (\sigma_{1l}, \dots, \sigma_{nl})$, $r_l = (r_{1l}, \dots, r_{nl})$, $\sigma_{il} = 0$ si al individuo i le ocurrió el evento l , uno en otro caso. Las componentes de r_l indican una penalización impuesta a cada individuo por el hecho de ser una censura potencial con respecto a la causa l .

Casos particulares:

1. Tomando $\lambda_{jl} = \frac{1}{2} \sum_{i=1}^n \sigma_{il} \varepsilon_{ij}$, se obtiene el estimador Tabla de Vida para la Tasa Neta definida en (6).

2. Puede definirse $\lambda_{jl} = \frac{1}{2} \sum_{i=1}^n \sigma_{il} \varepsilon_{ij} r_{il}$; en este caso debe cumplirse $r_{il} \leq 1$.

3. El caso 2 puede corregirse si se define $\lambda_{jl} = \sum_{i=1}^n \sigma_{il} \varepsilon_{ij} \frac{r_{il}}{1 + r_{il}}$, con $r_{il} \in \mathfrak{R}_+$

3. PROPIEDAD DE MÁXIMA VEROSIMILITUD DE LOS ESTIMADORES \hat{F}_{221} Y \hat{F}_{222}

En Xie-Liu [2000] y Ibrahim [2005], se demuestra que el estimador clásico tipo Tabla de Vida, \hat{F}_{211} definido en (1), maximiza la siguiente función de pseudoverosimilitud:

$$L = \prod_j S_+(t_{j-1})^{w_j/2} \cdot f(t'_{j-1})^{d_j} S_+(t''_{j-1})^{w_j/2}$$

Donde,

$$S_+(t) = P(T > t)$$

t_{j-1} : extremo izquierdo del intervalo j-ésimo

t'_{j-1} : un momento de tiempo cualquiera a la derecha de t_{j-1} y dentro del intervalo $[t_{j-1}, t_j)$

t''_{j-1} : un momento de tiempo a la derecha de t'_{j-1} y dentro del intervalo $[t_{j-1}, t_j)$

Por pseudoverosimilitud se entiende que es la función de verosimilitud bajo ciertos supuestos: la mitad de las censuras ocurren antes de todas las muertes y la otra mitad después de todas las muertes.

El estimador \hat{F}_{212} posee una propiedad similar lo cual se deduce del hecho de que tiene una estructura idéntica a la de \hat{F}_{211} .

Proposición: El estimador \hat{F}_{221} definido en (5) maximiza la siguiente función de pseudoverosimilitud.

$$L = \prod_j S(t_{j-1})^{\lambda_j} \cdot f(t'_{j-1})^{d_j} f(t''_{j-1})^{w_j - \lambda_j} \quad (8)$$

Demostración:

Supongamos, para simplificar, se tiene 2 intervalos de tiempo $[t_0, t_1)$ y $[t_1, t_2)$. Sean d_1 y d_2 el número de eventos que ocurren dentro de cada intervalo; w_1 y w_2 el número de individuos que se censuran en cada intervalo. Sea $S_+(t) = P(T > t)$.

La función de pseudoverosimilitud puede ser escrita como:

$$\begin{aligned} L &= S_+(t_0)^{\lambda_1} \cdot f(t'_0)^{d_1} \cdot S_+(t''_0)^{w_1 - \lambda_1} \cdot S_+(t_1)^{\lambda_2} \cdot f(t'_1)^{d_2} \cdot S_+(t''_1)^{w_2 - \lambda_2} \\ &= A^{\lambda_1} \cdot B^{d_1} \cdot C^{w_1 - \lambda_1} \cdot D^{\lambda_2} \cdot E^{d_2} \cdot F^{w_2 - \lambda_2} \end{aligned}$$

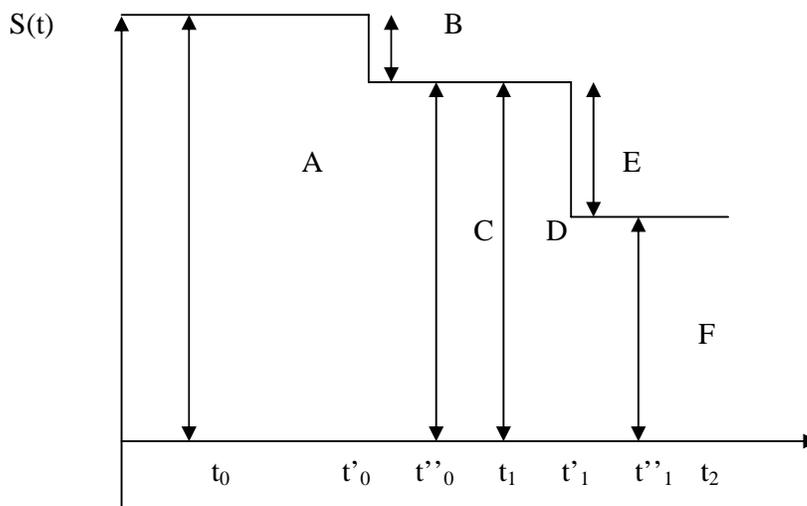


Diagrama. Configuración Óptima

Donde,

t_{j-1} : extremo izquierdo del intervalo j-ésimo donde ocurren la mitad de las censuras.

t'_{j-1} : un momento de tiempo cualquiera a la derecha de t_{j-1} y dentro del intervalo $[t_{j-1}, t_j)$ en el que ocurren todas los eventos que corresponden a ese intervalo.

t''_{j-1} : un momento de tiempo a la derecha de t'_{j-1} y dentro del intervalo $[t_{j-1}, t_j)$ donde ocurre la segunda mitad de las censuras.

El diagrama representa la configuración óptima que debe poseer S para maximizar L .

De donde se obtiene que:

$$A = 1$$

$$B + C = 1$$

$$C = D = E + F$$

Además se verifican las siguientes igualdades:

$$n_1 = n$$

$$\begin{aligned} n_2 &= n_1 - w_1 - d_1 \\ &= (d_1 + w_1 + d_2 + w_2) - w_1 - d_1 \\ &= d_2 + w_2 \end{aligned}$$

Sean:

$$E = e.D$$

$$F = f.D$$

Entonces,

$$\begin{aligned} L &= B^{d_1} C^{w_1 - \lambda_1 + \lambda_2 + d_2 + w_2 - \lambda_2} e^{d_2} f^{w_2 - \lambda_2} \\ &= B^{d_1} C^{n_1 - \lambda_1 - d_1} e^{d_2} f^{n_2 - \lambda_2 - d_2} \\ &\leq \left(\frac{d_1}{n_1 - \lambda_1} \right)^{d_1} \left(1 - \frac{d_1}{n_1 - \lambda_1} \right)^{n_1 - \lambda_1 - d_1} \left(\frac{d_2}{n_2 - \lambda_2} \right)^{d_2} \left(1 - \frac{d_2}{n_2 - \lambda_2} \right)^{n_2 - \lambda_2 - d_2} \end{aligned}$$

Hemos utilizado dos veces la siguiente propiedad:

$$M^\alpha N^\beta \leq \left(\frac{\alpha}{\alpha + \beta} \right)^\alpha \left(\frac{\beta}{\alpha + \beta} \right)^\beta, \text{ si } M + N = 1$$

Además, el máximo de la expresión $M^\alpha N^\beta$ se alcanza para $M = \frac{\alpha}{\alpha + \beta}$ y $N = \frac{\beta}{\alpha + \beta}$

Obtenemos:

$$B = \frac{d_1}{n_1 - \lambda_1}, \quad C = 1 - \frac{d_1}{n_1 - \lambda_1}, \quad e = \frac{d_2}{n_2 - \lambda_2}, \quad f = 1 - \frac{d_2}{n_2 - \lambda_2}$$

Por tanto $\hat{S}(t_1) = C = 1 - \frac{d_1}{n_1 - \lambda_1}$

$$\hat{S}(t_2) = F = fD = fC = \left(1 - \frac{d_1}{n_1 - \lambda_1}\right) \left(1 - \frac{d_2}{n_2 - \lambda_2}\right)$$

Por último el estimador \hat{F}_{222} maximiza una función de pseudoverosimilitud similar a (8), debido a que tiene la misma estructura de \hat{F}_{221} .

4. EJEMPLO

Supongamos que existe una covariable dicotómica Z_i cuya asociación con σ_i es altamente positiva (Z_i tiende a tomar el valor 0 (1) cuando σ_i tiende a tomar el valor 0 (1)).

Definamos $r = \frac{P(E/Z=1)}{P(E/Z=0)}$, es decir, r es el cociente de la probabilidad de que ocurra el evento de

interés dado que $Z=1$ y de la probabilidad de que ocurra el evento de interés dado que $Z=0$. El parámetro r se puede interpretar como el riesgo relativo de ocurrencia del evento entre los individuos que se censuran y los que no se censuran.

Definamos además,

$$R_j = \frac{\sum_{i=1}^n (1 - \sigma_i) Z_i \varepsilon_{ij} / \sum_{i=1}^n Z_i \varepsilon_{ij}}{\sum_{i=1}^n (1 - \sigma_i) (1 - Z_i) \varepsilon_{ij} / \sum_{i=1}^n (1 - Z_i) \varepsilon_{ij}} \quad (9)$$

Asignemos al individuo i la penalización $r_i = R_j$, j tal que $\varepsilon_{ij} = 1$. Es decir, a los individuos que se censuran en un mismo intervalo les corresponde el mismo valor de r_i .

Sustituyendo en la fórmula (4) se obtiene una expresión reducida para esta de la siguiente forma:

$$\lambda_j = w_j \frac{R_j}{1 + R_j}$$

Al sustituir lo anterior en (5), se obtiene el estimador tipo Tabla de Vida Clásico Ajustado.

4.1. Simulación

Usando las facilidades de programación que brinda el lenguaje S, implementación R, versión 2.4.1, R Development Core Team [2004], se generó una muestra de tamaño 1000 como sigue:

Se generaron 1000 observaciones procedentes de la distribución exponencial con parámetro $\lambda = 1/2$, que corresponden con los tiempos latentes de ocurrencia del evento de interés. Las covariables Z_i para

cada individuo fueron generadas a partir de una distribución Bernoulli como: $Z_i \sim B\left(\frac{t}{1+t}\right)$.

Las variables indicadoras de ocurrencia del evento se generaron de dos formas diferentes.

$$P(\sigma_i = 1 / Z_i) = \begin{cases} 0.75 & \text{si } Z_i = 1 \\ 0.25 & \text{si } Z_i = 0 \end{cases} \quad (10)$$

$$P(\sigma_i = 1 / Z_i) = \begin{cases} 0.5 & \text{si } Z_i = 1 \\ 0.5 & \text{si } Z_i = 0 \end{cases} \quad (11)$$

El tiempo censurado de ocurrencia del evento se definió como:

$$X = \begin{cases} t/2 & \text{si } \sigma = 0 \\ t & \text{si } \sigma = 1 \end{cases}$$

Se dividió el eje real positivo en 5 intervalos: [0,1), [1,2), [2,3), [3,4), [4,5)

Cada observación se clasificó de acuerdo al intervalo a que pertenecía, para obtener los indicadores \mathcal{E}_{ij} .

Los códigos escritos en R que permitieron esta simulación pueden consultarse en el Apéndice 1.

4.2. Resultados y discusión

Considerando la variable de censura definida como en (10), es decir si asociamos la covariable con la ocurrencia de la censura, se obtienen los resultados que se muestran en la Figura No. 1.

En esta primera salida la covariable está altamente correlacionada con la censura de manera positiva, por ello, los estimadores de los riesgos relativos son en general menores que uno, por lo que las probabilidades de supervivencia ajustadas resultan mayores que las clásicas.

Note que con el ajuste propuesto se obtienen estimaciones de la verdadera supervivencia más precisas que las obtenidas usando el método clásico. Lo anterior evidencia que es posible usar información a priori para explicar la interdependencia entre las causas competitivas

Si por el contrario generamos las variables indicadoras de la censura como (11), o sea, la covariable se considera independiente de la ocurrencia del evento, entonces los estimadores clásicos y ajustados coinciden y subestiman levemente la verdadera probabilidad. Ver Figura No. 2.

En este ejemplo se ha utilizado una covariable como sustituta de la censura (información adicional requerida para realizar el ajuste). En lugar de esto puede ofrecerse como información adicional directamente los valores de los riesgos relativos observados en cada individuo (R_j) y sustituir la expresión correspondiente de los λ_j .

Otro enfoque podría ser estimar los riesgos relativos en vez de ofrecer directamente sus valores numéricos, a partir de alguna definición para ellos en términos de las probabilidades de censura y supervivencia. Por ejemplo, pueden definirse los δ -riesgos relativos como el cociente de las probabilidades:

$$P(E/C) = P(u < T < u + \delta / u < U < T)$$

$$P(E/NC) = P(u < T < u + \delta / T < U)$$

El problema se traslada entonces a estimar dichas probabilidades a partir de los datos; estos no pueden consistir en los tiempos censurados solamente. El estudio de qué tipo de datos bastaría para estimar satisfactoriamente dichas probabilidades es un problema complejo.

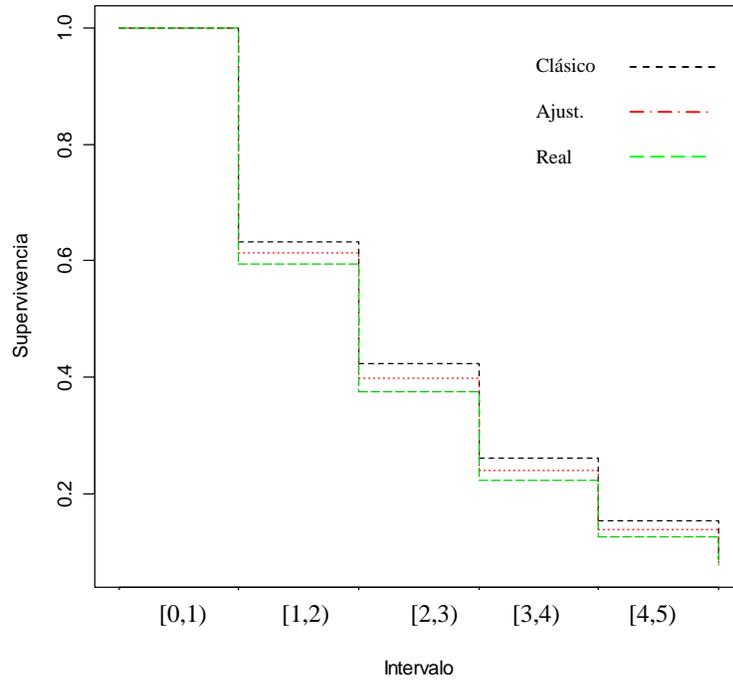


Figura 1. Comparación de las estimaciones de la supervivencia usando el estimador clásico y el ajustado con respecto a la supervivencia real, para datos simulados con censuras definidas en (10).

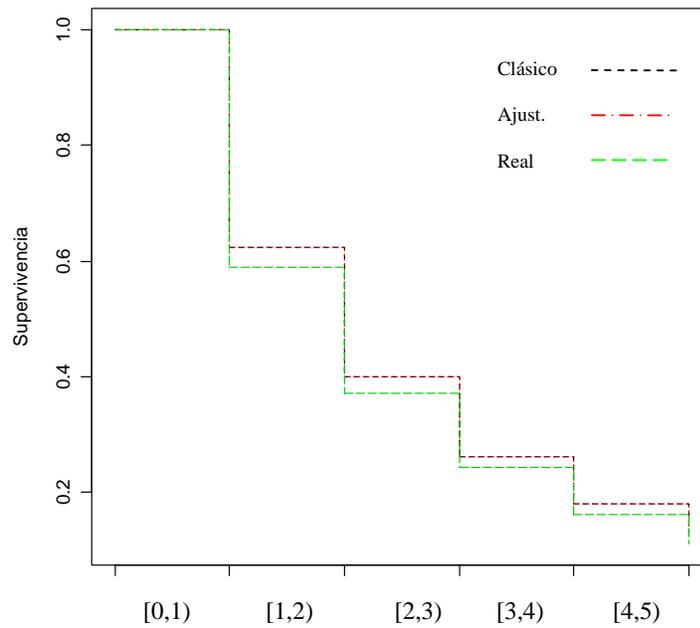


Figura 2. Comparación de las estimaciones de la supervivencia usando el estimador clásico y el ajustado con respecto a la supervivencia real para datos simulados con censuras independientes definidas en (11).

Es importante destacar que se necesita de un estrecho vínculo entre los investigadores biomédicos y los especialistas en estadística con el fin de determinar penalizaciones adecuadas para cada individuo, las que serán usadas para ajustar el estimador clásico de tabla de vida.

Received October 2007

Revised April 2008

REFERENCIAS

- [1] LLORCA J, DELGADO RM. (2004): Análisis de supervivencia en presencia de riesgos competitivos. **Gac. Sanit.** 18, 391-397.
- [2] CHIANG CL. (1961): A stochastic study of the life table and its applications: III. The follow-up study with the consideration of competing risks. **Biometrics** 17, 57-78
- [3] FIOCCO M, PUTTER H. and VAN HOUWELINGEN J.C. (2005): Reduced rank proportional hazards model for competing risk. **Biostatistics** 6, 465-478.
- [4] ZENG D. (2004): Estimating marginal survival function by adjusting for dependent censoring using many covariates. **The Annals of Statistics** 32, 1533-1555.
- [5] SATAGOPAN J M, BEN-PORAT L, BERWICK M, ROBSON M, KUTLER D and AUERBACH A D. (2004): A note on competing risks in survival data analysis. **British Journal of Cancer** 91, 1229-1235.
- [6] FARLEY T. M. M. (1986): Life-table methods for contraceptive research. **Stat Med** 5, 475-489.
- [7] XIE J. and LIU C. (2000): Adjusted Kaplan-Meier estimator and log-rank test with inverse probability of treatment weighting for survival data. **Statist. Med** 00, 1-6.
- [8] IBRAHIM JG. (2005): Applied Survival Analysis. **Presentation in the 21st Annual Summer Workshop of the American Statistical Association.**
- [9] R DEVELOPMENT CORE TEAM. (2004): **R. A language and environment for statistical computing.** R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3, Disponible en: URL: <http://www.R-project.org>.

APÉNDICE 1.

Códigos generados en R: Códigos usados para generar los datos bajo el supuesto de dependencia

```
# generando el tiempo latente de ocurrencia del evento de
interés
t<-rexp(1000,0.5)

# generando la covariable
z<-rep(0,1000)
for (i in 1:1000)
z[i]<-rbinom(1,1, 1-t[i]/(10+t[i]))

# generando la variable indicadora de evento a partir de la
covariable
delta<-rep(0,1000)
for (i in 1:1000){
  if (z[i]==1) delta[i]<-rbinom( 1,1, 0.9)# delta es evento
con alta probabilidad,
  else delta[i]<-rbinom( 1,1, 0.1) # es censura con baja
probabilidad
}

# generando el tiempo censurado de ocurrencia del evento
x<-t
for (i in 1:1000){
  if (delta[i]==0) x[i]<-t[i]/2
}

# generando los eij
e<-rep(0,5000)
dim(e)<-c(1000,5)
for (i in 1:1000)
for (j in 1:5)
  if (j-1<=x[i] & x[i]<j) e[i,j]<-1

# número de individuos
n<-1000

# número de intervalos
nint<-5

# cálculo de Rj
r<-rep(0,nint)
for (j in 1:nint){
s1<-0;s2<-0;s3<-0;s4<-0
  for (i in 1:n){
    s1<-s1+delta[i]*z[i]*e[i,j]
    s2<-s2+z[i]*e[i,j]
    s3<-s3+delta[i]*(1-z[i])*e[i,j]
    s4<-s4+(1-z[i])*e[i,j]
  }
r[j]<-(s1/s2)/(s3/s4)
if (is.nan(r[j])) r[j]<-1
}
```

```

# cálculo del número de eventos en el intervalo
d<-rep(0,nint)
for (j in 1:nint)
for (i in 1:n)
    d[j]<-d[j]+delta[i]*e[i,j]

# cálculo del número de censuras en el intervalo
w<-rep(0,nint)
for (j in 1:nint)
for (i in 1:n)
    w[j]<-w[j]+(1-delta[i])*e[i,j]

# Cálculo del número en riesgo al inicio del intervalo
nr<-rep(0,nint)
nr[1]<-n
for (j in 2:nint)
nr[j]<-nr[j-1]-d[j-1]-w[j-1]

# Cálculo la función de supervivencia ajustada para cada
intervalo
SA<-rep(1,nint)
landa<-rep(1,nint)
for (j in 1:nint){
landa[j]<-round((r[j]/(1+r[j]))*w[j],1)
    SA[j]<-1-d[j]/(nr[j]-landa[j])
}
SAA<-rep(1,nint)
SAA[1]<-round(SA[1],3)
for (j in 2:nint){
for (i in 1:j)
    SAA[j]<-round(SAA[j]*SA[i],3)
}

# Cálculo la función de supervivencia clásica para cada intervalo
SC<-rep(1,nint)
for (j in 1:nint)
SC[j]<-1-d[j]/(nr[j]-(1/2)*w[j])
SCA<-rep(1,nint)
SCA[1]<-round(SC[1],3)
for (j in 2:nint){
for (i in 1:j)
    SCA[j]<-round(SCA[j]*SC[i],3)
}

# Cálculo de la función de supervivencia real para cada intervalo
dreal<-rep(0,5)
SupRealAc<-rep(0,5)
for (j in 1:5){
    for (i in 1:1000)
        if (j-1<=t[i] & t[i]<j) dreal[j]<-dreal[j]+1
}
for (j in 1:5){
    riesgo<-1000
    if (j==1) SupRealAc[1]<-1-dreal[1]/1000
}

```

```

else {
  for (k in 2:j) riesgo<-riesgo-dreal[k-1]
  SupRealAc[j]<-SupRealAc[j-1]*(1-dreal[j]/riesgo)
}
}

# dibujando la tabla
Tabla<-
data.frame(Interval=c("[0,1)","[1,2)","[2,3)","[3,4)","[4,5)"),

nrisk=nr,nevent=d,nlost=w,nlostAdj=landa,

AcumLifeTable=SCA,AcumAdjLifeTable=SAA, Real=SupRealAc)
Tabla

A<-rep(0,18)
dim(A)<-c(6,3)
A[,1]<-c(1,Tabla[,6])
A[,2]<-c(1,Tabla[,7])
A[,3]<-c(1,Tabla[,8])
Probabilidad<-A

matplot(Probabilidad, xlab="Intervalo",ylab="Supervivencia",type
= "s",lty=c(2,3,5))

```

Códigos usados para generar los datos bajo el supuesto de independencia

```

# generando el tiempo latente de ocurrencia del evento de
interés
t<-rexp(1000,0.5)

# generando la covariable
z<-rep(0,1000)
for (i in 1:1000)
z[i]<-rbinom(1,1, 1-t[i]/(10+t[i]))

# generando la variable indicadora de evento a partir de la
covariable
delta<-rep(0,1000)
for (i in 1:1000){
  delta[i]<-rbinom( 1,1, 0.9)
}

r<-rep(1,nint)
# generando el tiempo censurado de ocurrencia del evento
x<-t
for (i in 1:1000){
  if (delta[i]==0) x[i]<- runif(1, min=0, max=t[i])
}

```