



PUNTOS DE CAMBIO EN LA FUNCIÓN DE RIESGO

Guadalupe Yoanna Arenas Martínez, Francisco Solano Tajonar Sanabria^a, Hugo Cruz Suárez^a,
Fernando Velasco Luna^a.

^aFacultad de Ciencias Fisico Matematicas, BUAP, Puebla. guadalupe_yam@hotmail.com,
ftajonar@cfm.buap.mx, hcs@cfm.buap.mx, fvelasco@cfm.buap.mx.

RESUMEN

El estudio de los puntos de cambio en el análisis de supervivencia, son de vital importancia ya que estos ayudan a determinar, los cambios que se presentan durante algún evento, los primeros estudios realizados para determinar los puntos de cambio, solo se basan en un solo punto, pero se ha observado que en el área de la industria y la medicina, los eventos estudiados presentan más de un solo punto de cambio. En el área médica, este tipo de estudios es de vital importancia, por ejemplo en el estudio del cáncer o enfermedades crónicas, se puede llegar a estimar cuando los pacientes sufren algún cambio, de acuerdo al tratamiento al que están sometidos. En este trabajo se propone un método para poder determinar y estimar múltiples puntos de cambio en la función de riesgo, y poder aportar dicha información al investigador, y de esta manera se podrá dar un mejor diagnóstico.

1. INTRODUCCIÓN

El análisis de supervivencia es una línea de investigación de la estadística, que nace de una teoría paralela llamada "Teoría de confiabilidad", la cual designa la probabilidad de que un sistema cumpla satisfactoriamente con su respectiva función, durante determinado periodo, así un evento que interrumpa este funcionamiento se denomina "falla".

Durante los primeros estudios realizados, este tipo de análisis se empleó sin proponer el concepto de censura, el cual, toma en consideración complicaciones que pueden surgir durante el evento en un determinado periodo de tiempo.

En el área de la medicina los análisis clínicos están dados en términos de supervivencia, la cual generalmente es sobre pacientes que son sometidos a un determinado tratamiento, estos sufren cambios durante periodos de tiempo muy pequeños. Un caso particular son los tratamientos emitidos para poder radicar el cáncer, en esta situación es de utilidad emplear el análisis de supervivencia.

El considerar la censura en este tipo de estudios, ayuda a analizar mejor el tratamiento que se está empleando sobre el paciente. Otro factor de vital importancia es el análisis de los cambios que el paciente tiene cuando es sometido al tratamiento, estos se dan en determinados periodos de tiempo, y se espera que sean durante el inicio del tratamiento.

Para poder estimar cuando ocurren los cambios, proponemos un modelo constante por tramos con múltiples puntos de cambio. Este modelo ayuda a estimar dichos puntos en periodos de tiempo muy pequeños, de tal manera que en cada tramo solo se encuentre a lo más un cambio, el uso de la función de riesgo nos proporciona la posibilidad de fallar en un determinado momento,



para poder culminar con este análisis se emplea el método secuencial, el algoritmo de Nelder Mead, y una prueba de tipo Wald.

El presente trabajo tiene como objetivo principal el análisis de un modelo de regresión con puntos de cambio, dichos cambios se considera tanto en la función de riesgo como en los parámetros de regresión, este ocurre en algún momento del tiempo, el cual es desconocido.

2. TEORIA

En muchos estudios quirúrgicos, y especialmente los relacionados con el cáncer, la variable principal que queremos valorar es el tiempo que transcurre hasta un suceso dado. A este tiempo, normalmente se le llama “supervivencia”, aunque el suceso en cuestión no tiene por qué ser la defunción. Cuando el suceso que queremos evaluar ocurre en todos los pacientes, entonces disponemos de varias técnicas estadísticas que pueden aplicarse. Sin embargo, lo más habitual es que el suceso en cuestión no se presente en todos los individuos

Los datos de supervivencia se pueden dar y estudiar con dos tipos de probabilidades diferentes: supervivencia y riesgo. La “función de supervivencia” es la probabilidad de que un individuo sobreviva desde la fecha de entrada en el estudio hasta un momento determinado en el tiempo t . Pero nuestro interés está dado sobre la “función de riesgo” que es la probabilidad de que a un individuo que está siendo observado en el tiempo t le suceda el evento en ese momento.

Definición 1. *El análisis de supervivencia es el conjunto de técnicas que permiten analizar, estudiar y modelar la variable de interés, que en este caso, representa una variable de tipo evento-tiempo.*

A lo largo de este trabajo, se considera que la variable aleatoria T , denota el tiempo transcurrido hasta que ocurre un evento específico. Este puede ser la muerte, la aparición de un tumor, el desarrollo de alguna enfermedad, dejar de fumar, fallo en maquinaria, etc. De forma más específica, T será una variable aleatoria (v.a.) continua no negativa, esto es, T toma valores en el intervalo $(0, \infty)$.

Para nuestro propósito denotaremos a T una variable aleatoria no negativa, que representa el tiempo de vida de los individuos en una población.

Sea $G(u)$ una función no decreciente, continua por la derecha, con límite a la izquierda y con un número finito de discontinuidades. Asumimos que $g(u) = G'(u)$ existe excepto en los puntos de discontinuidad de $G(u)$, donde los puntos de discontinuidad tienen la siguiente forma: $G(a_j) - G(a_{j-}) = g_j$ donde $\lim_{\Delta a \rightarrow 0} G(a - \Delta a) = G(a_{j-})$.

La función de densidad de $G(u)$ sobre el intervalo $(a, b]$ está dada mediante la integral de Riemman Stieltjes de está definida por:

$$\int_a^b dG(u) = \int_a^b g(u) du + \sum_{j: a < a_j \leq b} g_j \quad (1)$$

de donde podemos pensar que $dG(u) = g(u) du + G(u) - G(u-)$.

En general una función de distribución $F(t) = P(T \leq t)$ es continua por la derecha no decreciente con saltos en los puntos a_j , donde $P(T = a_j) = f_j$ con $f_j > 0$ y la función de densidad de probabilidad es $f(u) = F'(u)$ en los puntos donde $F(u)$ es continua.



$$P(a < T \leq b) = F(b) - F(a) = \int_a^b dF(u) = \int_a^b f(u) du + \sum_{j: a < a_j \leq b} f_j \quad (2)$$

Si $F(t)$ es continua no tiene puntos de cambio y si T tiene una distribución discreta entonces $F(u)$ es una función discreta con $F(u) = 0$ en los puntos continuos.

Para definir a la función de riesgo se hará mediante la integral producto, para lo cual se toma $a = u_0 < u_1 < \dots < u_m = b$ una partición de $(a, b]$ con $\Delta u_i = u_i - u_{i-1}$ y el $\max(\Delta u_i) \rightarrow 0$ cuando $m \rightarrow \infty$. La integral producto de una función $dG(u)$ es definida por

$$\prod_{(a,b]} \{1 + dG(u)\} = \lim_{m \rightarrow \infty} \prod_{i=1}^m \{1 + G(u_i) - G(u_{i-1})\} \quad (3)$$

Si $G(u)$ es continua para todo u en $(a, b]$ entonces $dG(u) = g(u) du$ y tenemos que

$$\prod_{(a,b]} \{1 + g(u) du\} = \lim_{m \rightarrow \infty} \prod_{i=1}^m \{1 + g(u_i) \Delta u_i + o(\Delta u_i)\} = \lim_{m \rightarrow \infty} \prod_{i=1}^m \{1 + g(u_i) \Delta u_i\} \quad (4)$$

donde $o(x)$ es una función de medida de una función $w(x)$ tal que $\frac{w(x)}{x} \rightarrow 0$ si $x \rightarrow 0$. Tenemos que $\log\{1 + g(u_i) \Delta u_i\} = g(u_i) \Delta u_i + o(\Delta u_i)$ para un Δu_i pequeño y tomando el logaritmo de la integral producto, en el caso continuo tenemos

$$\lim_{m \rightarrow \infty} \prod_{i=1}^m \{1 + g(u_i) \Delta u_i\} = \exp\left[\int_a^b g(u) du\right] \quad (5)$$

Si $G(u)$ tiene saltos en los puntos a_j con $j = 1, 2, \dots$, de tamaño g_j entonces

$$\prod_{(a,b]} \{1 + dG(u)\} = \prod_{(a,b]} \{1 + g(u) du\} \prod_{j: a < a_j \leq b} \{1 + g_j\}. \quad (6)$$

$h(u) = \frac{f(u)}{S(u)}$, representa la función de riesgo para T en los puntos donde $F(u)$ y $S(u)$ es continua y $h_j = P(T = a_j | T \geq a_j)$ en los valores de tiempo discreto con salto en cada a_j . La función de riesgo acumulativa está definida por la integral de Riemman Stieltjes

$$H(t) = \int_a^b dH(a) = \int_a^b h(u) du + \sum_{j: a_j \leq t} h_j \quad (7)$$

Dado la función de riesgo acumulativa, podemos obtener la función de supervivencia, para la cual se toma una partición del intervalo $(0, t]$, con $0 < u_0 < u_1 < \dots < u_m = t$ y está definida por

$$P(T \geq t) = \prod_{i=1}^m P(T = u_i | T \geq u_{i-1}). \quad (8)$$

Ahora para $\Delta u_i = u_i - u_{i-1}$ suficientemente pequeño $[u_{i-1}, u_i)$ puede tener a lo más un salto. En general tenemos que la función de supervivencia es

$$P(T \geq t) = \exp\left\{-\int_0^t h(u) du\right\} \prod_{[0,t]} (1 - h_j) \quad (9)$$



Existen varios modelos que son usados para el análisis de supervivencia como el el modelo Weibull, Gama, Gama generalizado, etc., pero solo nos enfocaremos en el siguiente

Modelo Exponencial: Esta es muy importante ya que una de sus principales características es que la posibilidad de fallar es independiente del pasado.

Sea T el tiempo de falla, para una variable aleatoria continua que toma valores positivos. Entonces T tiene una distribución exponencial si y solo si tiene una función de riesgo constante. Sea $h(t) = \alpha$ donde α es una constante positiva, entonces la función de distribución es $f(t) = \alpha e^{-\alpha t}$ con $t > 0$, con función de supervivencia $S(t) = e^{-\alpha t}$.

El modelo paramétrico con un punto de cambio más simple, es el que supone que en una muestra aleatoria de tamaño n , k sujetos tienen una función de riesgo común λ_1 y los $(n - k)$ restantes, tienen otra función de riesgo común λ_2 , con $\lambda_1 \neq \lambda_2$, este modelo se puede expresar como sigue

$$h(t) = \begin{cases} \lambda_1, & \text{si } 0 \leq t \leq \tau, \\ \lambda_2, & \text{si } t \geq \tau. \end{cases} \quad (10)$$

Donde, las constantes de riesgo λ_1, λ_2 y el punto de cambio τ son considerados desconocidos. Sin embargo, nuestro interés se centra principalmente en la detección de la presencia de puntos de cambio en la función de riesgo y por supuesto la estimación de todos los parámetros desconocidos involucrados en el modelo, ya que en algunos ejemplos de salud pública se sugiere que debido a la mejora de los tratamientos o de diagnósticos, puede haber dos o más puntos de cambios en la función de riesgo, es necesario considerar la función de riesgo con múltiples puntos de cambio.

El Modelo Constante por Tramos con Múltiples Puntos de Cambio.

Sea X_1, \dots, X_n , una muestra aleatoria de tiempos de vida o de falla, independientes idénticamente distribuidos, y C_1, \dots, C_n , los tiempos de censura que se suponen son independientes de X . Observamos las parejas (T_i, δ_i) , $i = 1, 2, \dots, n$, las cuales nos darán información necesaria sobre los datos obtenidos, donde $T_i = \min(X_i, C_i)$ y $\delta_i = 1$ si $X_i \leq C_i$ y cero en cualquier otro lado. Considérese el siguiente modelo para el análisis de puntos de cambio:

$$\lambda(t) = \begin{cases} \alpha_1 & 0 \leq t \leq \tau_1, \\ \alpha_2 & \tau_1 \leq t \leq \tau_2, \\ \vdots & \vdots \\ \alpha_{k+1} & t \geq \tau_k. \end{cases} \quad (11)$$

Donde $0 < \tau_1 < \dots < \tau_k$, K es el número de puntos de cambio en el modelo y los α_j es el valor de la función de riesgo entre el tiempo τ_{j-1} y τ_j , [4]. Este modelo se ha utilizado para probar la compatibilidad de los trasplantes de corazón en Stanford, California [8], también en la efectividad de los tratamientos para la prevención del cáncer [1], en el área de la industria se aplica este tipo de modelos para motores o transmisores, estos son algunos ejemplos para los cuales este modelo es aplicable.

Puesto que se ha obtenido la función de riesgo, podemos determinar de manera única a la función de densidad, la cual es una distribución exponencial por tramos y tiene la siguiente expresión



$$f(t) = \begin{cases} \alpha_1 \text{Exp}[\alpha_1 t], & 0 < t < \tau_1, \\ \alpha_2 \text{Exp}[-\alpha_1 \tau_1 - \alpha_2(t - \tau_1)], & \tau_1 \leq t < \tau_2, \\ \vdots \\ \alpha_{k+1} \text{Exp}[-\alpha_1 \tau_1 - \alpha_{k+1}(t - \tau_1)], & t \geq \tau_k. \end{cases} \quad (12)$$

En este modelo los α_i y τ_i , con $i = 1, 2, \dots$, son parámetros desconocidos, los cuales se desean determinar y estimar.

Prueba de Wald para el Modelo Constante por Tramos.

Para probar la hipótesis nula de que no hay puntos de cambio contra la alternativa que nos indica que hay un punto de cambio se puede probar la hipótesis nula, aunque este método con múltiples pruebas solo hay que probar una a la vez, para contrastar $H_0: \alpha_{k-1} - \alpha_k = 0$ vs $H_1: \alpha_{k-1} - \alpha_k \neq 0$ usamos la siguiente prueba de Wald:

$$X_w = \frac{(\widehat{\alpha}_{k-1} - \widehat{\alpha}_k)^2}{\text{Var}(\widehat{\alpha}_{k-1} - \widehat{\alpha}_k)} \sim \chi_1^2 \quad (13)$$

3. PARTE EXPERIMENTAL

El objetivo es determinar los puntos de cambio en la función de riesgo. Para generar los tiempos de vida se toma la función inversa de la distribución exponencial, mediante la distribución uniforme son generados los tiempos de vida. Se obtiene la función de log-verosimilitud de (12) y se toman los estimadores de máxima verosimilitud y se procede a realizar las pruebas de hipótesis.

Las pruebas de hipótesis se realizarán de una forma secuencial, entonces, la primera hipótesis a probar es

$$H_0: \alpha_1 - \alpha_2 = 0 \quad \text{vs} \quad H_1: \alpha_1 - \alpha_2 \neq 0$$

Haciendo uso del estadístico de Wald dado en (13), y con un nivel de significancia de .13, se rechaza la hipótesis, se continúa con este proceso, hasta encontrar al primer punto de cambio. De tal manera que el modelo final de la función de densidad es

$$f(t) = \begin{cases} 0.13 \exp(-0.13t) & 0 \leq t \leq 3.145 \\ 0.385 \exp(0.694 - 0.351t) & 3.145 \leq t \leq 4.147 \\ 0.351 \exp(0.662 - 0.351t) & t \geq 4.147 \end{cases}$$

4. CONCLUSIONES

El análisis de supervivencia es importante en los estudios del cáncer y trasplantes de órganos, en este tipo de padecimientos los pacientes sufren cambios durante intervalos de tiempo muy pequeños. La metodología propuesta ayuda a determinar estos cambios, lo cual es de gran ayuda para el médico encargado de dicho estudio, pues teniendo este tipo de información él puede tomar decisiones que ayuden a mejorar la calidad de vida del paciente.



Este tipo de técnica nos ayuda a determinar los puntos de cambio que hay en la función de riesgo, esta metodología se realizó con dos tipos de prueba, la de tipo Wald y de razón de verosimilitud. El modelo final encontrado utilizando la prueba de razón de verosimilitud y la de tipo Wald, arrojan el mismo resultado final, es decir, los dos pruebas son buenas, el uso de cada uno de ellas depende de los datos que se tengan para el estudio, puesto que las dos pruebas regresan el mismo modelo final podemos concluir que la metodología propuesta para poder determinar los puntos de cambio en la función de riesgo buena.

BIBLIOGRAFÍA

1. A. Noura and K. Iran, "Proportional Hazards Change Point Models in Survival Analysis", Vol 2, 1990, pp 241-253.
2. D. Kleinbaum and M Klein, "*Survival Analysis*" Springer 2005, Chapter 6, pp 257-300.
3. H. Mathews and O. Kurtis "*Numerical Methods Using Mat lab*", Prentice Hall, Chapter 3.115-120.
4. H Ivana, Jiri, Z "Semiparametric Estimation of Hazard Function for Cancer Patients",. 2007, pp.494-513.
5. J. Chen "*Parametric Statistical Change Point Analysis*", Birkhauser, 2000, Chapter 6, pp.143-154.
6. M, Goodman, Yi Li, "Survival Analysis whit Change Point Hazard Functions", pp 40-48.
7. J Lawlees, "*Statistical Models And Methods For Lifetime Data*", Wiley-Interscience, 2006, Chapter 2, pp.16-54.
8. W. Bruce and C. Larry "Likelihood Ratio Tests For A Change point With Survival Data" 1999, pp 555-565.