

## **Aplicación de los algoritmos genéticos para estimar los parámetros en un modelo de regresión de Cox**

*Application of genetics algorithms for estimating the parameters of a Cox regression model*

**Douglas Rivas<sup>1</sup>, José Luciano Maldonado<sup>2</sup>,  
Rafael Borges<sup>3</sup> y Gerardo Colmenares<sup>4</sup>**

Recibido: 03-05-07 / Aceptado: 20-06-07. Códigos JEL: C34, C45, C42

### **Resumen**

Este trabajo, enmarcado en el campo de la Computación Evolutiva, presenta el desarrollo de un Algoritmo Genético para encontrar los parámetros óptimos de un Modelo de Regresión de Cox del Análisis de Supervivencia para pacientes del servicio de Diálisis Peritoneal del Hospital Clínico Universitario de Caracas entre 1980 y 2000, realizado por Borges (2002, 2005). Se hace uso de la técnica de los Algoritmos Genéticos como método de búsqueda de una mejor estimación de los parámetros del Modelo de Cox al obtenido por los métodos clásicos de optimización. El algoritmo fue programado completamente en el lenguaje C++, bajo un diseño de programación modular tomando en cuenta cada uno de los elementos que constituyen a los Algoritmos Genéticos. Las características principales del algoritmo son: a) La población inicial, que está constituida por 10 individuos, se genera de manera aleatoria entre un rango de valores; dicho rango fue obtenido luego de realizar diversas pruebas; b) la función de ajuste se basó en el Criterio de Información de Akaike (AIC); c) la selección de los individuos a reproducirse se realizó por torneo; e) para el cruce se usó el operador multipunto y la mutación se realizó a todos los genes de una parte de los cromosomas de la población. El algoritmo desarrollado permitió obtener estimaciones de los parámetros del Modelo de Regresión de Cox, y con mejor valor de AIC, a los obtenidos utilizando los métodos clásicos.

**Palabras clave:** Algoritmo genético, modelo de regresión de Cox, criterio de información de Akaike (AIC), análisis de supervivencia.

---

1 Universidad de Los Andes. Departamento de Estadística. e-mail: drivas@ula.ve

2 Universidad de Los Andes. Instituto de Estadística Aplicada y Computación. e-mail: maldonaj@ula.ve

3 Universidad de Los Andes. Departamento de Estadística. e-mail: borgesr@ula.ve

4 Universidad de Los Andes. Instituto de Investigaciones Económicas y Sociales (IIES). e-mail: gcolmen@ula.ve

## Abstract

This paper, involved in the area of evolutive computing, presents the development of a Genetic Algorithm in finding the optimal parameters of a Cox Regression Model for patients of the Service of Peritoneal Dialysis of the “Hospital Clínico Universitario de Caracas” between 1980 and 2002, performed by Borges (2002, 2005). The technique of the genetic algorithms is used as a method for finding a better estimation of the parameters of the Cox Model than the obtained by the classical optimization methods. The algorithm was completed programmed in the language C++, using a modular programming design, considering every element of the genetic algorithms. The main characteristics of the algorithm are: a) the initial population, of 10 subjects, is generated randomly between a range of values, this range was obtained after several essays, b) the adjustment function was based in the Akaike Information Criteria (AIC), c) the selection of the subjects to be reproduced was done by tournament, d) the multipoint operator for the crossing and, the mutation was done to all the genes of one part of the chromosomes of the population. The developed algorithm was useful to obtain the estimation of the Cox Regression Model and with better AIC values than the obtained by the classical methods.

**Keywords:** Genetic Algorithms, Cox Regression Model, Akaike Information Criteria (AIC), Survival Analysis.

## 1. Introducción

### 1.1. Formulación del problema

El análisis de supervivencia es una de las áreas de la estadística de mayor desarrollo en los últimos años que es evidenciado por la gran cantidad de textos, entre los que se incluyen como una muestra, libros que van desde un nivel introductorio, como los de Collett (2003), Cox y Oakes (1984), Hosmer y Lemeshow (1999) y el texto clásico de Miller (1981), pasando por los de nivel intermedio como Kalbfleisch y Prentice (2002), Klein Y Moeschberber (2003), Lawless (2003), Lee y Wang (2003), Martinussen y Scheike (2006) y Therneau y Grambsch (2000), hasta llegar a los de nivel más avanzado como los de Andersen et. al. (1993) y Fleming y Harrington (1991). Otro aspecto a destacar es la gran cantidad de artículos científicos que constantemente se publican en las revistas estadísticas de mayor prestigio, como *The Annals of Statistics*, *Biometrics*, *Biométrica*, *Journal of the American Statistical Association*, *Journal of the Royal Statistical Society, Serie B* y *Scandinavian Journal of Statistics*, entre

otras, así como la aparición en 1995 de una revista especializada en el tópico, *Lifetime Data Analysis*, editada por Kluwer.

El análisis de supervivencia permite estudiar y construir modelos para analizar el tiempo que un suceso tarda en ocurrir. En dicho análisis o proceso las diferentes variables pronóstico permiten estimar el tiempo de aparición del suceso. Entre los diferentes tipos de modelos que se pueden emplear, uno de los más extendidos es el modelo de riesgos proporcionales, también conocido como Modelo de Cox (Cox, 1972). Éste, al igual que los demás modelos estadísticos paramétricos, enfrenta el problema de estimar sus parámetros a partir de datos observados. Es bien conocido que en todos los problemas de estimación la solución no es única; por el contrario, se tienen varias alternativas que funcionan como mínimos o máximos locales de cierta función. Por lo tanto, los problemas de estimación de los parámetros de los modelos de regresión no es otra cosa que un problema de la búsqueda de aquellos parámetros que representen mejor los datos en estudio.

Para resolver distintos problemas, en los últimos años ha tomado auge una técnica de la computación evolutiva conocida como algoritmos genéticos. Según Fleming (2002), los algoritmos genéticos son mecanismos de optimización de funciones basados en leyes biológicas de selección natural que presentan ventajas con respecto a los métodos tradicionales de optimización. En especial, estos métodos iterativos convergen al óptimo global de la función objetivo de análisis, sin importar su grado de complejidad y su dominio, evitando procedimientos adicionales que en ocasiones se desvían hacia óptimos locales. Una introducción a las técnicas de computación evolutiva puede verse en el texto editado por Aguilar y Rivas (2001).

Este trabajo propone un método alternativo para estimar los parámetros del Modelo de Regresión de Cox haciendo uso de un Algoritmo Genético. Específicamente, el algoritmo calcula los valores óptimos de los parámetros del Modelo de Regresión de Cox desarrollado por Borges (2002) en el Análisis de Supervivencia para pacientes del servicio de Diálisis Peritoneal del Hospital Clínico Universitario de Caracas entre 1980 y 2000. Es necesario aclarar que el método propuesto podría extenderse a otras aplicaciones.

En la literatura existen métodos y aplicaciones referentes a las técnicas de computación evolutiva en el área de las redes neuronales artificiales para la estimación de modelos de supervivencia (véase por ejemplo, Faraggi y Simon, 1995, Ravdin y Clark, 1992, Ravdin y De Laurentis, 1993 y Ripley, 1998)

### **1.2. Justificación: importancia y aplicaciones esperadas**

El análisis de supervivencia comprende un conjunto de técnicas de gran importancia en diversas aplicaciones prácticas, pues permite estudiar la variable tiempo “hasta que ocurre un evento” y su dependencia de otras posibles variables explicatorias. Debido a su amplia aplicación, y muy especialmente al amplio uso del modelo de regresión de Cox como técnica para estimar la variable tiempo en función de diversas variables “explicativas”, resulta interesante construir un algoritmo genético, el cual pueda estimar los parámetros de un modelo de Cox para una serie de datos. Un algoritmo de tales características resultaría en una herramienta de gran utilidad en la medicina, la ingeniería y en la estadística, pues además de presentarse como un método alternativo que selecciona los valores óptimos de los parámetros de un Modelo de Regresión de Cox, el éxito del mismo motivará futuras aplicaciones de los Algoritmos Genéticos en el campo de la estadística.

## **2. Metodología de la investigación**

Para el desarrollo de este proyecto se siguió una metodología basada principalmente en dos aspectos. En primer lugar, la revisión exhaustiva de los fundamentos teóricos de los Algoritmos Genéticos y del Análisis de Supervivencia, específicamente del Modelo de Regresión de Cox; y en segundo lugar, la definición de las actividades inherentes a la construcción del algoritmo. Durante la revisión bibliográfica se analizaron los datos estudiados por Borges (2002, 2005). En dicho estudio se realizó un Análisis de Supervivencia a pacientes del servicio de Diálisis Peri-

toneal del Hospital Clínico Universitario de Caracas entre 1980 y 2000. Una vez realizada la revisión bibliográfica se llevó a cabo la construcción y codificación del algoritmo genético en un lenguaje de programación de alto nivel. Para la construcción del algoritmo se tomaron en cuenta los pasos recomendados en la bibliografía revisada. Las herramientas de software utilizadas fueron Microsoft Windows XP como sistema operativo, C++ como Lenguaje de Programación y Microsoft Excel de herramienta de apoyo. Luego de diseñado el algoritmo, se realizaron varias pruebas, variando en cada caso los parámetros de entrada y el número de iteraciones, con el objeto de determinar el número de iteraciones necesarias para la convergencia de las distintas soluciones. Dichas soluciones se compararon con las obtenidas por Borges (2002, 2005). Finalmente, en función de la experiencia acumulada durante la ejecución del proyecto, y en base al análisis de los resultados obtenidos con el sistema desarrollado, se extrajeron conclusiones y algunas recomendaciones para trabajos futuros.

### **3. Análisis del caso de estudio**

El Algoritmo Genético que se construye en esta investigación estima los parámetros del Modelo de Regresión de Cox del Análisis de Supervivencia para pacientes del servicio de Diálisis Peritoneal del Hospital Clínico Universitario de Caracas entre 1980 y 2000, realizado por Borges (2002, 2005).

Para dicho análisis se registró un conjunto de variables para 246 individuos en Diálisis Peritoneal, que ingresaron al servicio entre el 2 de junio de 1980 y el 6 de diciembre de 1996, y salieron entre el 15 de enero de 1981 y el 31 de octubre de 2000.

Luego de realizar un estudio exhaustivo, probando la significancia de cada una de las variables, se incluyeron solamente las siguientes variables significativas al 10: Diabetes, edad y el índice Quetelet o índice de masa corporal (quettelet), lo cual se puede apreciar en el cuadro 1.

**Cuadro 1.** Estimación de los coeficientes para el modelo definitivo de Cox para los datos según meses.

Covariable	coef	exp(coef)	ee(coef)	Z	p
diabetes	0,5492	1,732	0,3208	1,71	0,087
edad	0,0315	1,032	0,0097	3,25	0,0011
quetellet	-0,0969	0,908	0,0389	-2,49	0,013

Fuente: Elaboración propia

Donde:

coef es el coeficiente estimado mediante el modelo.

exp(coef) es el exponencial del coeficiente y se interpreta como el riesgo.

ee(coef) es el error estándar del coeficiente.

Z es el estadístico de contraste para la significación del coeficiente.

P es el p-valor de probabilidad de la significación del coeficiente.

Por lo tanto el Modelo de Regresión de Cox tiene la siguiente forma:

$$\lambda(t, z(t)) = \lambda_0(t) e^{\beta_1 Z_1(t) + \beta_2 Z_2(t) + \beta_3 Z_3(t)} \tag{1}$$

Donde

$\beta_1$ , es el parámetro asociado con la variable diabetes ( $Z_1$ )

$\beta_2$ , es el parámetro asociado con la variable edad ( $Z_2$ )

$\beta_3$ , es el parámetro asociado con la variable quetellet ( $Z_3$ )

Los riesgos asociados a cada uno de los parámetros y sus respectivos intervalos de confianza al 95% se muestran en el cuadro 2.

**Cuadro 2.** Exponencial de los coeficientes para el modelo definitivo de Cox para los datos según meses.

Covariable	Exp(coef)	exp(-coef)	LCI(95%)	LCS(95%)
diabetes	1,732	0,577	0,924	3,248
edad	1,032	0,969	1,013	1,052
quetellet	0,908	1,102	0,841	0,979

Fuente: Elaboración propia

## 4. El algoritmo genético

A continuación se establece cada uno de los pasos realizados para la construcción del Algoritmo.

### 4.1. Elementos del algoritmo genético

A continuación se describe cada uno de los elementos usados en la construcción del algoritmo, basados en la teoría anteriormente señalada:

#### 4.1.1. Selección de la variable de entrada

En este caso la variable de entrada o cromosoma está conformado por los tres parámetros del Modelo de Regresión de Cox; es decir, los parámetros de las variables diabetes ( $\beta_1$ ), edad ( $\beta_2$ ) y el índice de Quetelet ( $\beta_3$ ). Por lo tanto, el cromosoma es un vector fila con tres elementos:

$$\text{cromosoma} = [\beta_1, \beta_2, \beta_3] \quad (2)$$

Cuyo rango de valores está entre -5 y 5. Valores fuera de este rango hacen que la función de ajuste tienda a infinito.

#### 4.1.2. Representación o codificación de la variable de entrada

Aunque el rango de valores de los parámetros es un subconjunto de los números reales, las soluciones se codificaron como cadenas binarias de secuencias de 1's y 0's, cuya longitud depende del valor del parámetro más grande en valor absoluto.

#### 4.1.3. Tamaño de la población

Como lo establece Mitchell (1996), no existe un método para determinar el número óptimo de individuos, en este caso, se realizaron diversas corridas con diferentes tamaños de población inicial, y se obtuvo que 10 individuos eran suficientes para realizar el estudio, ya que un número mayor de individuos no alteraba los resultados obtenidos con 10 individuos.

#### 4.1.4. Elección de la población inicial

La población inicial se generó aleatoriamente entre el rango de valores de los parámetros fijados anteriormente.

#### 4.1.5. Función de aptitud o de adaptación

La función de aptitud utilizada está basada en el Criterio de Información de Akaike (AIC; Akaike, 1973, 1974), el cual es la maximización del logaritmo de la verosimilitud de la función de riesgo proporcional de Cox. Por lo tanto, se usó como función de ajuste la siguiente función:

$$AIC = -2\log[L(\beta)] + \alpha$$

Donde

$$L(\beta) = \prod_{i=1}^k \frac{\exp\{Z_i(t_i)\beta\}}{\sum_{l \in R(t_i)} \exp\{Z_l(t_i)\beta\}} \quad (3)$$

$t_1 \leq \dots \leq t_k$  son los tiempos de fallas observados, asumidos distintos en la muestra;  $Z_i(t_i)$  es el vector de covariables en el tiempo  $t_i$  para el sujeto que falla en el tiempo  $t_i$ ; y  $Z_l(t_i)$  es el vector de covariables correspondientes para el  $l$ -ésimo miembro de  $R(t_i)$ , el conjunto de individuos en riesgo en el tiempo  $t_i$ , (Cox y Oakes, 1984).

Es importante destacar que el AIC no es el único criterio que pudiera ser utilizado, se ha usado este criterio por su facilidad de incorporación al contexto del modelo de Cox. Otros Criterios que pudieran utilizarse son el Criterio de Información Bayesiana (BIC; Schwarz, 1978), el cual pudiera inclusive funcionar de manera más eficiente debido a la baja dimensionalidad del modelo planteado en el presente trabajo, el Criterio de Inflación de Riesgo (RIC; George y Foster, 1994), el Criterio de Inflación la Covarianza (CIC; Tibshirani y Knight, 1999), el  $C_p$  de Mallows (Mallows, 1973), la Selección de Modelos Adaptativa (Shen, Huang y Ye, 2002) y el Criterio de Información Enfocada (FIC; Claeskens y Hjort, 2003), el cual ha sido utilizado en el contexto de la selección de variables en el modelo de regresión logística (Claeskens Croux y Van Kerckhoven, 2006) y ofrece posibilidades de desarrollo interesantes en el área de los modelos de supervivencia. Dos artículos recientes referidos sobre selección de modelos son el de George (2002) y el de Kadane y Lazar (2004).

#### 4.1.6. Selección

Una vez evaluados los individuos de la población, se usó el método de selección por torneos; es decir, se seleccionaban la mitad de los individuos que tenían mayor porcentaje de contribución a la función de ajuste. En este caso, como la función de ajuste está basada en el criterio AIC, se seleccionaron los individuos que tenían menor AIC.

#### 4.1.7. Operadores

Se usaron los dos operadores genéticos principales: cruce y mutación.

**Cruce.** Para el operador de cruce se tomaron en cuenta dos elementos: la probabilidad de que ocurriera el cruce y el número de puntos para realizar el cruce.

En cuanto a la probabilidad de cruce, se tomó 0.8 como probabilidad de cruce, ya que es común que este operador ocurra en un algoritmo genético.

Para el número de puntos de cruce se realizó un multipunto de tres puntos, tomando en cuenta que cada punto de cruce estuviese en el

rango de 0's y 1's de cada parámetro. Es decir, supongamos el siguiente cromosoma:

$$\text{cromosoma} = [011101 - 001100 - 111111]$$

Entonces los dos puntos de cruce están distribuidos de la siguiente manera:

$$\begin{array}{c} \text{cromosoma} = [011101 - 001100 - 111111] \\ \quad \quad \quad \uparrow \quad \quad \quad \uparrow \\ \quad \quad \quad \text{Puntos de cruce} \end{array}$$

**Mutación.** Para la mutación se tomaron en cuenta tres elementos: la probabilidad de que ocurra la mutación, cuántos individuos mutar y cuáles elementos de cada individuo mutar. La probabilidad de mutación seleccionada fue 0.3 y en cuanto al número de elementos a mutar se realizó de forma aleatoria, tomando en cuenta que no se podían mutar todos los individuos y por lo tanto, el de menor AIC no se selecciona para la mutación. En lo que respecta a cuáles elementos de los individuos seleccionados anteriormente mutar, se usó el criterio de mutar todos los elementos de los individuos seleccionados.

#### 4.1.8. Criterio de Parada

El criterio de Parada establecido se obtuvo por observación en las pruebas. Es decir, se realizaron varias pruebas con distintos número de iteraciones y se escogió aquel número de iteraciones donde todos los individuos tenían el mismo valor de AIC.

#### 4.2. Diseño del algoritmo genético

Una vez establecidos los elementos que componen el Algoritmo Genético, se procedió a su diseño. Para ello, se tomó en cuenta el diseño de programación modular; es decir se creó una serie de módulos que estaban encargados de realizar acciones muy específicas, que al interrelacionarlos, cumplieron con el propósito del proyecto en estudio. Entre los principales módulos creados se encuentran:

- Generación de la población inicial. Genera una población de  $n$  individuos, con tres parámetros, donde cada parámetro está en un rango de valores previamente establecidos.
- Evaluación de la Función de Ajuste. Evalúa la población en la función de ajuste o adaptación; es decir, calcula el valor AIC para cada uno de los individuos.
- Ordenamiento. Ordena los individuos de menor a mayor de acuerdo al valor AIC calculado anteriormente.
- Selección y Reproducción. Selecciona la mitad de los individuos con menor AIC y los reproduce eliminando a su vez la mitad con mayor AIC.
- Conversión o codificación. Convierte los parámetros de cada uno de los individuos en su respectiva representación binaria.
- Creación del Cromosoma. Crea los cromosomas compuestos por las cadenas binarias obtenidas anteriormente.
- Cruce. Realiza el cruce de los cromosomas obtenidos anteriormente.
- Mutación. Realiza la mutación de los individuos, asignando el valor 0 donde hay un 1 y viceversa.

Todos estos módulos, y otros no menos importantes, se interrelacionaron entre sí tomando en cuenta el número de generaciones necesarias para obtener una población cuya función de ajuste convergiera en una única solución.

## 5. Resultados

Una vez construido el Algoritmo Genético y codificado en el lenguaje C++, se procedió a realizar diversas pruebas variando algunos parámetros, tales como el número de individuos, el rango de los parámetros del modelo y el número de iteraciones. Se obtuvieron en cada uno de los casos, valores diferentes de los parámetros a estimar del Modelo de Regresión de Cox y valores diferentes del AIC.

En primer lugar, se fue variando el número de iteraciones. El método de variación fue el decreciente; es decir, se comenzó con 100 iteraciones y se fue disminuyendo de 10 en 10, comprobando que todos los valores del AIC convergieran a un único valor. Se obtuvo que para una población de 10 individuos eran suficientes 30 iteraciones. Luego, se tomaron diferentes tamaños para la población inicial pero los resultados no cambiaban significativamente. Por lo tanto, se estableció como 30 el número de iteraciones y 10 el tamaño de la población. En cuanto al rango de valores sobre el cual estaban definidos los parámetros de la población inicial se eligió el intervalo  $[-5,5]$ , valores fuera de este rango conducían a valores muy altos de AIC e incluso en algunos casos resultaba infinito.

Una vez establecidos estos elementos del algoritmo se realizaron diversas corridas. En el cuadro 3 se muestran los resultados para 10 pruebas del algoritmo.

**Cuadro 3.** Valores estimados de los parámetros del Modelo de Regresión de Cox con sus respectivos valores de AIC, en 10 pruebas del algoritmo

Prueba	$\beta_1$	$\beta_2$	$\beta_3$	AIC
1	0,533051	0,070153	-0,584849	402,729
2	-0,052507	0,027918	-0,383819	407,276
3	-0,099685	0,04056	-0,485062	394,507
4	-0,502085	-0,022684	-0,353294	447,779
5	0,425578	-0,010896	-0,317306	430,429
6	-0,479252	0,034671	-0,520996	394,928
7	0,423431	-0,015299	-0,437233	421,427
8	-0,2721	-0,008445	-0,39926	424,742
9	-0,536755	0,05092	-0,441887	406,571
10	0,756753	-0,014112	-0,747774	428,597

Fuente: Elaboración propia

Como puede observarse, los valores obtenidos de AIC en cada una de las corridas son menores que el obtenido por el método clásico estudiado por Borges (2002), cuadro 4.

**Cuadro 4.** Valores estimados de los parámetros del Modelo de Regresión de Cox con su respectivo valor de AIC\*

$\beta_1$	$\beta_2$	$\beta_3$	AIC
0,5492	0,0315	-0,0969	507,542

Fuente: Elaboración propia

\* Nota: Este valor fue obtenido por Borges (2002)

Al comparar los resultados de los cuadros 3 y 4 se observa que el algoritmo obtiene estimaciones de los parámetros con mejor AIC que el método clásico, figura 1, ya que, todos los valores de AIC para las diez pruebas están por debajo del valor de AIC para el resultado obtenido con el método clásico.

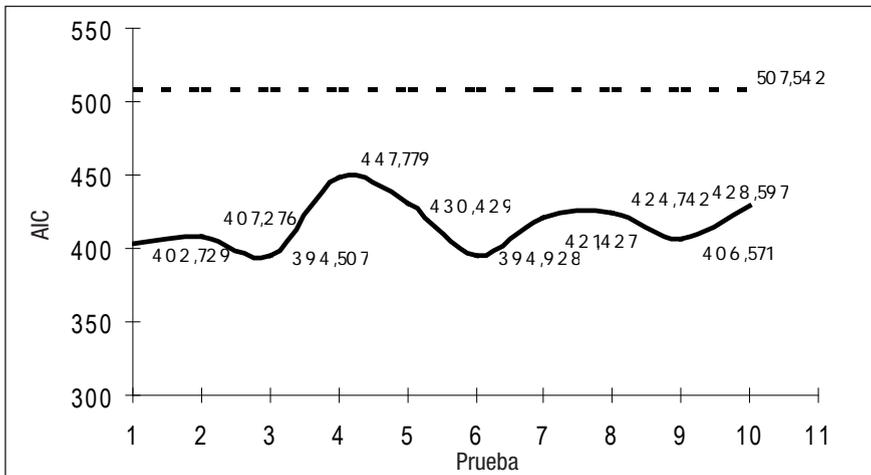


Figura.1 Comparación de los resultados obtenidos con el Algoritmo Genético (línea continua) y el método clásico (línea punteada), según el AIC.

Por otro lado, se puede ver que los resultados obtenidos varían de prueba en prueba, esto implica que las estimaciones dependen de la población inicial que se genera aleatoriamente; por lo tanto, al igual que como ocurre con los métodos clásicos, el algoritmo desarrollado no da la seguridad de obtener un óptimo global, sin embargo el algoritmo da una mejor solución que el método clásico.

Es importante resaltar que las estimaciones de los parámetros varían de prueba en prueba, y que en algunas de ellas estas estimaciones difieren significativamente de las obtenidas por el método clásico; aún así, el porcentaje de estas ocurrencias es bajo.

Además, en el cuadro 5, se muestra el riesgo asociado a cada uno de los parámetros estimados por el algoritmo.

**Cuadro 5.** Riesgo asociado a cada uno de los parámetros del Modelo de Regresión de Cox

Prueba	$\text{Exp}(\beta_1)$	$\text{Exp}(\beta_2)$	$\text{Exp}(\beta_3)$
1	1,70412367	1,07267229	0,55718999
2	0,94884768	1,02831136	0,68125472
3	0,90512249	1,04139379	0,61565902
4	0,60526736	0,97757135	0,70237067
5	1,53047478	0,98916315	0,72810792
6	0,61924641	1,03527905	0,5939287
7	1,52719237	0,98481744	0,64582094
8	0,76177808	0,99159056	0,67081627
9	0,58464234	1,05223871	0,64282227
10	2,1313445	0,98598711	0,47341921

Fuente: Elaboración propia

Al comparar estos valores con los intervalos de confianza del riesgo asociado con los parámetros obtenidos por el método clásico (tabla 2), se observa que un alto porcentaje de los valores obtenidos por el algoritmo genético cae dentro del intervalo respectivo.

Todos estos resultados indican que el algoritmo genético desarrollado podría usarse como una técnica alternativa para estimar los parámetros del Modelo de Regresión de Cox del caso en estudio.

## 6. Conclusiones

- Los Algoritmos Genéticos son técnicas de gran utilidad en la optimización de funciones que son difíciles de tratar haciendo uso de las técnicas de optimización clásica.
- El Algoritmo Genético desarrollado permitió encontrar estimaciones de los parámetros del Modelo de Regresión de Cox del Análisis de Supervivencia para pacientes del servicio de Diálisis Peritoneal del Hospital Clínico Universitario de Caracas entre 1980 y 1997, mejores que los encontrados por el programa R , basándose en la comparación del AIC.
- Un alto porcentaje de los valores de los parámetros del Modelo de Regresión de Cox obtenidos por el Algoritmo Genético son cercanos a los obtenidos por el método clásico, puesto que, gran parte de los riesgos asociados a los parámetros, cae dentro del intervalo de confianza obtenido por la técnica clásica a través del programa R.

## 7. Recomendaciones

- Usar otros criterios para la función de ajuste, especialmente criterios que estudien cada uno de los parámetros por separados, tales como el test de Wald, con el objeto de optimizar cada parámetro por separado.
- Estudiar de manera más detallada la población inicial, ya que una población inicial heterogénea puede llevar al óptimo global.
- Desarrollar Algoritmos Genéticos para la optimización de los parámetros de otras funciones estadísticas como método alternativo a los métodos clásicos.
- Realizar un estudio con un número mayor de corridas, con la finalidad de estudiar las características del estimador propuesto.

## 8. Agradecimientos

Los autores agradecen las sugerencias y recomendaciones de los árbitros. Rafael Borges agradece el financiamiento recibido del Consejo de Desarrollo Científico, Humanístico y Tecnológico de la Universidad de Los Andes (CDCHT-ULA) a través del proyecto cuyo código es E-199-02-09-C.

## 9. Referencias

- Aguilar, José y Francklin Rivas (editores) (2001). *Introducción a las Técnicas de Computación Inteligente*. Editorial Meritec, Mérida, Venezuela.
- Andersen, Per Kragh, Ornulf Borgan, Richard D. Gill, y Niels Keiding. (1993). *Statistical Models Based on Counting Processes*. New York: Springer-Verlag.
- Akaike, Hirotugu (1973). Information Theory and an Extension of the Maximum Likelihood Principle, in *Second International Symposium on Information Theory*, eds. B. N. Petrox and F. Caski. Budapest: Akademiai Kiado, pp. 267–281.
- Akaike, Hirotugu (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19: pp. 716-723.
- Borges, Rafael E. (2002). *Análisis de supervivencia aplicado a un caso de diálisis renal y diálisis peritoneal en el hospital clínico universitario de caracas y hemodiálisis en el hospital de clínicas caracas 1980-2000*. Tesis de Maestría, Universidad de Los Andes, Facultad de Ciencias Económicas y Sociales. Instituto de Estadística Aplicada y Computación. Mérida, Venezuela.
- Borges, Rafael E. (2005). Análisis de supervivencia aplicado a diálisis peritoneal (DPA). *Revista Colombiana de Estadística*, 28: pp. 243-259.
- Claeskens, Gerda, Christophe Croux y Johan Van Kerckhoven, (2006). Variable Selection for Logistic Regression using a Prediction- Focussed Information Criterion. *Biometrics*, 62: pp. 972-979.
- Claeskens, Gerda y Niels L. Hjort, (2003). The Focussed Information Criterion. *Journal of the American Statistical Association*, 98: pp. 900-916.
- Collett, David (2003). *Modelling survival data in medical research*. Segunda

- Edición. Boca Ratón, Florida: Chapman & Hall.
- Cox, David R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society: Series B*, 34: pp. 187-220.
- Cox, David R. y David Oakes (1984). *Analysis of Survival Data*. Chapman and Hall/CRC, New York.
- Faraggi, David y Richard Simon (1995). A Neural Network Model for Survival Data, *Statistics in Medicine*, 14: pp. 73-82.
- Fleming, Thomas R. y David P. Harrington (1991). *Counting Processes and Survival Analysis*. New York: John Wiley & Sons, Inc.
- Fleming, Peter J. y Robin C. Purshouse (2002). *Evolutionary algorithms in control systems engineering a survey*. Control Engineering Practice. 10: pp 1223-1241.
- George, Edward I. (2000). The Variable Selection Procedure. *Journal of the American Statistical Association*, 95: pp. 1304-1308.
- George, Edward I., y Dean P. Foster (1994). The Risk Inflation Criterion for Multiple Regression. *The Annals of Statistics*, 22: pp. 947-1975.
- Hosmer, David W. y Stanley Lemeshow (1999). *Applied survival analysis: Regression modeling of time to event data*. New York: John Wiley & Sons, Inc.
- Kadane, Joseph B. y Nicole A. Lazar. (2004). Methods and Criteria for Model Selection. *Journal of the American Statistical Association*, 99: pp. 279-290.
- Kalbfleisch, John D. y Ross L. Prentice (2002). *The statistical analysis of failure time data*. (Segunda Edición). New York: John Wiley & Sons, Inc.
- Klein, John P. y Melvin L. Moeschberger (2003). *Survival analysis: Techniques for censored and truncated data*. (Segunda Edición). New York: Springer-Verlag.
- Lawless, Jerald F. (2003). *Statistical models and methods for lifetime data*. (Segunda Edición). New York: John Wiley & Sons, Inc.
- Lee, Elisa T. y Wang, John W. (2003). *Statistical methods for survival data analysis*. (Tercera Edición). New York: John Wiley & Sons, Inc.
- Mallows, Colin L. (1973). Some Comments on Cp . *Technometrics*, 15: pp 661-675.
- Martinussen, Torben y Thomas H. Scheike (2006). *Dynamic Regression Models for Survival Data*. New York: Springer-Verlag.

- Miller, Rupert G. (1981). *Survival analysis*. New York: John Wiley & Sons, Inc.
- Mitchell, Melanie (1996). *An Introduction to Genetic Algorithms*. MIT Press.
- Ravdin, Peter M y Gary C. Clark (1992). A Practical Application of Neural Network Analysis for Predicting Outcome of Individual Breast Cancer Patients. *Breast Cancer Research Treatment*, 22: pp. 285-293.
- Ravdin, Peter M. y De Laurentis M. (1993). A Technique for using neural Network Analysis to Perform Survival Analysis of Censored Data. *Cancer Letters*, 77: pp.127-138.
- Ripley, Ruth (1998). *Neural Network Models for Breast Cancer Prognosis*. Tesis Doctoral presentada en el Departamento de Ciencias de Ingeniería de la Universidad de Oxford, Inglaterra.
- Therneau, Terry M. y Patricia M Grambsch. (2000). *Modeling Survival Data: Extending the Cox Model*. New York: Springer-Verlag.
- Tibshirani, Robert y Keith Knight (1999). The Covariance Inflation Criterion for Model Selection. *Journal of the Royal Statistical Society, Series B*, 61: pp. 529-546.
- Shen, Xiaotong, Huang Hsin-Cheng y Jimmy Ye (2002). Adaptive Model Selection. *Journal of the American Statistical Association*, 97: pp. 210-221.
- Schwarz, Gideon (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6: pp. 461-464.