

Uso del Método “Combinación de Riesgos” para estimar la función de supervivencia en presencia de riesgos competitivos dependientes: Un estudio de simulación

Using the “Risk Combination” method to estimate the survival function in the presence of competing risks dependent: A simulation study

Osnamir E. Bru-Cordero ^{a*}
Mario C. Jaramillo-Elorza ^{b*}

Fecha de recepción: 23.10.2017

Fecha de aceptación: 19.12.2018

DOI: <https://doi.org/10.19053/01217488.v10.n1.2019.7261>

Resumen

En este trabajo se comparan distintas estructuras de dependencia para los riesgos que compiten en un modelo de riesgos competitivos dependientes trivariado, haciendo uso de las técnicas C-Vines y D-Vines cópulas, vía simulación estadística. Los Vines permiten obtener flexibilidad multivariada y son capaces de capturar todo el rango posible de dependencias entre los riesgos competitivos, las cuales son de gran interés en los mercados financieros, problemas sociales, genéticos, entre otros. Seguidamente, se estima la función de supervivencia para el tiempo mínimo, tanto para el caso independiente, por medio del estimador Kaplan Meier, como para el caso dependiente, en el que usaremos el método de combinación de riesgos, el cual es una extensión del estimador cópula gráfico. Los C-D Vines cópulas, trabajan con una variedad de cópulas bivariadas, las cuales se pueden seleccionar de manera independiente y permiten tener una amplia gama de posibilidades para la caracterización de la dependencia de los riesgos que compiten, se estudian casos particulares donde dos de los tres riesgos tienen igual dependencia y el riesgo restante es independiente a los anteriores. También se estudia el caso donde dos riesgos presentan igual dependencia y el otro con dependencia alta. Además se analiza un caso particular donde los tres riesgos presentan distinta dependencia. En todos los casos estudiados, el método de combinación de riesgos es una buena alternativa para estimar las funciones de distribución marginal y la función de supervivencia cuando se tiene dependencia entre los riesgos de un modelo de riesgos competitivos dependientes.

Palabras clave: C-Vines, D-Vines, Combinación de Riesgos, Cópula Gráfico.

a Estudiante de Doctorado, Escuela de Estadística, Universidad Nacional de Colombia, Sede Medellín.

* Correo electrónico: oebruc@umal.edu.co

b Profesor asociado, Escuela de Estadística, Universidad Nacional de Colombia, Sede Medellín.

* Correo electrónico: mcjarami@unal.edu.co

Abstract

In this paper we compare different structures of dependence for the risks that compete in a trivariate competing risk model, using the C-Vines and D-Vines copula techniques, through statistical simulation. The vines can obtain multivariate flexibility and are able to capture all the possible range of dependencies between the competing risks, which are of great interest in financial markets, social, genetic among others problems. Then, we estimated survival function for the minimum time, both for the independent case, through the Kaplan Meier estimator and for the dependent case, in which we will use the risk combination method, which is an extension of the copula graphic estimator. The C-DVines copulas work with a cascade of bivariate copulas, which can be selected independently and allow a wide range of possibilities for characterizing the dependence of competing risks, we study particular cases where two of the three risks have equal dependence and the remaining risk is independent to the previous ones. We also study the case where two risks are equally dependent and the other is highly dependent. In addition, a particular case where the three risks have different dependence is analyzed. In all the cases studied, the risk combination method is a good alternative to estimate the marginal distribution functions and the survival function when there is a dependence between the risks of a dependent competing risks model.

Key words: C-Vines, D-Vines, Risk pooling method, Copula Graphic.

1. INTRODUCCIÓN

Un sistema con tres modos de falla puede ser modelado usando un modelo de riesgos competitivos, el sistema fallará cuando se presente cualquiera de los tres modos de falla, es decir, el tiempo de falla observado del sistema es el mínimo de los posibles tiempos de falla de los respectivos modos de falla, además un modo de falla censura a los otros dos modos. La metodología tradicional asume independencia entre los tiempos de falla para la estimación de la función de supervivencia marginal, pero comúnmente este supuesto no se cumple. Para riesgos competitivos la estructura de los datos obtenidos no permite estimar la función de supervivencia ni las densidades marginales, esto es el llamado problema de identificabilidad. En el caso de dos riesgos competitivos, el estimador cópula gráfico permite estimar la función de supervivencia marginal bajo una cópula conocida y la dependencia entre los dos riesgos que compiten. Cuando se tienen modelos de riesgos competitivos con más de dos riesgos, se considera el método denominado combinación de riesgos, propuesto por [1], el cual es una extensión del estimador cópula gráfico, propuesto por [2], y es usado para la estimación de la función de supervivencia. En este trabajo se analiza la función de supervivencia para un modelo con tres riesgos bajo distintas dependencias entre los modos de falla que compiten, por medio del método antes mencionado. Para encontrar la estimación de la función de supervivencia se supone independencia entre los modos de

falla. En teoría de probabilidad las funciones llamadas cópulas son en realidad funciones de distribución que representan las relaciones de dependencia entre las variables aleatorias. Las funciones cópula bidimensionales son funciones bivariadas que unen o bien copulan dos funciones de distribución univariadas para construir funciones de distribución bivariadas. La cópula representa una forma paramétrica conveniente para modelar la estructura de dependencia en distribuciones conjuntas de variables aleatorias. Este concepto de cópula permite construir modelos que van más allá de los estándares en el análisis de dependencia entre variables estocásticas. La metodología de cópula es capaz de capturar relaciones no lineales y en particular permite relacionar eventos extremos que ocurren en la naturaleza [3].

En el caso no paramétrico la función de supervivencia conjunta, en un problema con múltiples modos de falla, no puede ser completamente identificada en la situación usual cuando sólo se conocen los tiempos mínimos de falla [4]. Incluso en el caso paramétrico los datos pueden contener poca información sobre el coeficiente de asociación entre las variables y resulta necesario hacer algunos supuestos al respecto [5]. Para el caso no paramétrico el estimador cópula gráfico de [2], resuelve el problema bajo el supuesto del conocimiento de la forma funcional de la cópula que proporciona la estructura de asociación entre los riesgos competitivos. Es de anotar que bajo independencia el problema de identificabilidad no existe. Para

la estimación de la función de supervivencia; [1] proponen el método denominado combinación de riesgos para un modelo con más de dos riesgos competitivos cuando la cópula es Arquimediana. En éste método la variable combinada es sólo creada conceptualmente para verificar lo relacionado con la identificación. [6] realizaron estudio de dependencia entre los riesgos, el cual es un procedimiento para una iteración serial de cópulas Arquimedianas las cuales tienen la misma estructura de dependencia, además estas cópulas que utilizaron para ese estudio satisfacen las propiedades de simetría y asociatividad, las cuales son requeridas para que el método de combinación de riesgos funcione en forma adecuada. [1] realizaron una aplicación del método de combinación de riesgos en un estudio de desempleo en Alemania, donde consideraron tres riesgos competitivos con igual dependencia, en este trabajo se estudia el caso en el cual los tres riesgos que compiten presentan diferentes dependencias. [7] estudiaron una extensión del estimador cópula gráfico para un modelo con más de dos riesgos competitivos e ilustraron la metodología con datos de la confiabilidad de interruptores tipo *FL245* en Interconexión Eléctrica S.A. E.S.P. (ISA) y mostraron la utilidad de la temática en confiabilidad industrial.

En este trabajo, se estudian distintas dependencias entre los riesgos, se usa la técnica de los C-D Vines cópulas, los cuales trabajan con una variedad de cópulas bivariadas, dichas cópulas se pueden seleccionar de manera independiente y permiten tener una amplia gama de posibilidades para la caracterización de la dependencia de los riesgos que compiten. Se estudiaron tres casos de dependencia. En el primer caso dos de los tres riesgos tienen igual dependencia y el riesgo restante es independiente a los anteriores. En el segundo caso se tomaron dos riesgos con igual dependencia y el otro riesgo con una dependencia alta. En el tercer caso se tomaron los tres riesgos con distintas dependencias.

2. RIESGOS COMPETITIVOS

El tiempo de falla de un sistema con múltiples modos de falla puede ser modelado como un sistema en serie o un modelo de riesgos competitivos. El tiempo de falla observado es el

mínimo de esos tiempos potenciales individuales, donde cada unidad tiene un tiempo potencial de falla [8].

Los datos para un modelo de riesgos competitivos trivariado se presentan como una variable aleatoria bivariada de la forma (T, δ) , con T el tiempo mínimo de falla $T = \min(X, Y, Z)$, con tiempos de falla potenciales X, Y, Z y donde δ es una variable discreta, la cual toma valores de 0 si la observación es una censura, 1 cuando la falla se presenta por el primer modo de falla, 2 cuando la falla se presenta por el segundo modo de falla y 3 cuando la falla se presenta por el tercer modo de falla.

3. CONCEPTOS FUNDAMENTALES

[8] presenta algunos conceptos fundamentales para el desarrollo del presente trabajo.

La función de *subdistribución* o función de incidencia acumulada (FIA) para el modo de falla $i, i = 1, 2, 3$, viene dada por:

$$G_i = P(T \leq t, \delta = i)$$

La función de distribución total es la probabilidad que un evento de cualquier tipo ocurra en o antes del tiempo t , es decir, es la suma de las FIA para todos los tipos de eventos;

$$G(t) = P(T \leq t) = \sum_{i=1}^3 P(T \leq t, \delta = i) = \sum_{i=1}^3 G_i(t).$$

En ausencia de riesgos competitivos la función de distribución total se encuentra en el intervalo $[0, 1]$. Luego en presencia de riesgos competitivos la función de subdistribución sólo puede tomar valores hasta $P(\delta = i)$ puesto que:

$$\lim_{t \rightarrow \infty} G_i(t) = P(\delta = i)$$

Por lo tanto, la función $G_i(t)$ no es una función de distribución propia.

La función de subsobrevivencia es la probabilidad de que el modo de falla i no ocurra antes del tiempo t

$$\overline{G}_i = P(T > t, \delta = i)$$

La función de subdensidad para el modo de falla i y T continuo, se define de la siguiente manera

$$g_i(t) = \frac{dG_i(t)}{dt} = -\frac{d\bar{G}_i(t)}{dt}$$

4. PROBLEMA DE IDENTIFICABILIDAD

Para el caso no paramétrico, la distribución conjunta en un problema con múltiples modos de falla no puede ser completamente identificada, en la situación usual, cuando sólo se conocen los tiempos mínimos de falla, para un problema de este tipo la aproximación usada para especificar el modelo de riesgos competitivos es vía tiempos latentes. Para el caso trivariado se tienen tiempos de falla potenciales X , Y , y Z asociados a cada modo de falla, donde si el sistema falla a causa del primer modo de falla, el segundo y el tercero no pueden ser observados. Así $T = \min(X, Y, Z)$ determina el tiempo de falla del sistema completo.

Dada la tripleta (X, Y, Z) con distribución multivariada:

$$F(x, y, z) = P(X \leq x, Y \leq y, Z \leq z)$$

y funciones de distribución marginales para cada modo de falla, los cuales tienen la siguiente forma $F_1(t) = P(X \leq t)$, $F_2(t) = P(Y \leq t)$ y $F_3(t) = P(Z \leq t)$. Uno de los intereses en riesgos competitivos es identificar las distribuciones marginales de las variables a partir de los datos de riesgos competitivos. En el caso donde los tiempos de falla son independientes, las distribuciones marginales son identificables, y corresponden a las funciones de distribución. Si este supuesto no se tiene, entonces a partir únicamente de los datos, en riesgos competitivos, no es posible identificar las distribuciones marginales y la distribución conjunta de los tiempos, existen muchas funciones de distribución conjunta diferentes que comparten las mismas funciones de subdistribución, lo cual es llamado el problema de identificabilidad [4].

Para este problema en particular existen varias soluciones:

1. Asumir independencia.

2. Asumir una estructura de dependencia conocida usando una cópula particular para la distribución conjunta de los modos de falla, [2] proponen el algoritmo cópula gráfico.
3. Especificar un modelo paramétrico.

5. ESTIMACIÓN DE LA FUNCIÓN DE SOBREVIVENCIA BAJO INDEPENDENCIA

Según [9] para un sistema en serie con tres componentes independientes, la función de distribución para el tiempo mínimo T está dada por:

$$F(t) = 1 - [1 - F_1(t)][1 - F_2(t)][1 - F_3(t)]$$

donde (X, Y, Z) son los tiempos de vida asociados a cada una de las tres componentes, $T = \min(X, Y, Z)$ es el tiempo observado y F_i , $i = 1, 2, 3$, las respectivas funciones de distribución para los componentes. La función de sobrevivencia $S(t) = 1 - F(t)$, para el caso de modos de falla independientes es:

$$S(t) = S_1(t)S_2(t)S_3(t).$$

donde $S_i(t)$, $i = 1, 2, 3$ son las funciones de sobrevivencia de las tres componentes. Para un sistema con tres componentes en serie y tiempos de falla dependientes, $F(t)$ tiene la siguiente forma:

$$F(t) = 1 - P(X > t, Y > t, Z > t)$$

Para este caso, la evaluación se debe hacer con respecto a la distribución conjunta de X , Y y Z , en la cual se debe incluir el parámetro de dependencia. La función de sobrevivencia para este caso es:

$$S(t) = P(X > t, Y > t, Z > t).$$

El estimador de la sobrevivencia del tiempo mínimo de falla de un sistema con tres modos de falla independientes que compiten $S(t)$, se denotará $\hat{S}^*(t)$, Para obtener este estimador basta multiplicar las funciones de sobrevivencia estimadas bajo una distribución específica, donde para cada función marginal se estiman los parámetros, considerando los tiempos del

otro modo de falla como tiempos de censura o censurados.

$$\widehat{S}^*(t) = \widehat{S}_1(t)\widehat{S}_2(t)\widehat{S}_3(t)$$

donde $\widehat{S}_1(t)$, $\widehat{S}_2(t)$ y $\widehat{S}_3(t)$, corresponden a las funciones de supervivencia marginales estimadas para cada modo de falla.

6. CÓPULAS

Una cópula se define como una distribución multivariada, con distribuciones marginales uniformes en $[0, 1]$. Las cópulas permiten representar de manera separada la estructura de dependencia y las distribuciones marginales de una distribución multivariada. La metodología de cópulas se ha convertido en una herramienta importante a la hora de modelar o generar datos multivariados.

[3] contiene una gran cantidad de información acerca del modelado de dependencias con cópulas. Combinando varias distribuciones marginales con diferentes cópulas, es posible modelar distribuciones multivariadas con una amplia gama de marginales y tipos de dependencia. A continuación se muestra el teorema fundamental para cópulas conocido como el Teorema de Sklar.

Teorema 1 (Teorema de Sklar). Sean Y_1, \dots, Y_p variables aleatorias con función de distribución conjunta F , con marginales F_1, \dots, F_p respectivamente. Entonces existe una cópula C tal que satisface:

$$F(y_1, \dots, y_p) = C(F_1(y_1), \dots, F_p(y_p))$$

Si F_1, \dots, F_p son continuas, entonces C es única. Inversamente, si C es una cópula y F_1, \dots, F_p y funciones de distribución, entonces F es una función de distribución conjunta con marginales F_1, \dots, F_p .

Para nuestro caso hacemos uso del teorema para los casos bivariado y trivariados.

6.1 Cópulas Arquimedianas

Un importante caso particular de cópulas son las cópulas Arquimedianas, se caracterizan por la facilidad con que pueden ser construidas y por la

gran variedad de estructuras de dependencia que pueden reproducir. En este trabajo se utilizó la cópula Frank debido a que en la implementación del algoritmo de [1], ellos la utilizaron en un estudio de desempleo en Alemania y los resultados fueron óptimos.

6.2 La cópula Frank

$$C_\theta(x, y) = -\frac{1}{\theta} \ln \left[1 + \frac{\{\exp(-\theta x) - 1\} \{\exp(-\theta y) - 1\}}{\{\exp(-\theta) - 1\}} \right], \theta \neq 0$$

La cópula Frank tiene grandes propiedades, captura el rango completo de dependencias, incluye las cópulas de cota superior de Fréchet cuando $\theta \rightarrow -\infty$, de cota inferior de Fréchet cuando $\theta \rightarrow \infty$ y de independencia cuando $\theta \rightarrow 0$. La cópula Frank proporciona cantidades cerradas y permite mayor facilidad al momento de programar [10].

6.3 τ de Kendall

El τ de Kendall, mide el grado de asociación entre dos variables aleatorias, en particular para las marginales dos a dos de un modelo de riesgos trivariado dependiente generado en nuestro caso. Está dado como sigue:

$$\tau_\theta = 1 - \frac{4}{\theta} [1 - D_1(\theta)]$$

donde $D_k(x)$ es la función Debye, la cual está definida para cualquier entero k por:

$$D_k(x) = \frac{k}{x^k} \int_0^x \frac{t^k}{\exp(t) - 1} dt$$

El τ de Kendall, cumple con la siguientes propiedades $\lim_{\theta \rightarrow -\infty} \tau_\theta = -1$, $\lim_{\theta \rightarrow \infty} \tau_\theta = 1$ y $\lim_{\theta \rightarrow 0} \tau_\theta = 0$ (independencia).

6.4 La cópula Frank trivariada

$$C_{\theta_{123}}^3(\mathbf{X}) = -\frac{1}{\theta_1} \left[1 + \prod_{i=1}^3 \frac{\exp(-\theta_i x_i) - 1}{\{\exp(-\theta_1) - 1\}^2} \right]$$

donde θ_1 es el parámetro de la cópula C_{12} , θ_2 es el parámetro de la cópula C_{13} y θ_3 es el parámetro de la cópula $C_{23|1}$, donde la cópula $C_{23|1}$ es una cópula condicional.

7. ESTIMADOR CÓPULA GRÁFICO

[2] aplican cópulas Arquimedianas en modelos de riesgos en competencia con datos censurados dependientes, proponiendo un estimador no paramétrico para las distribuciones marginales llamado estimador cópula gráfico, el cual da solución al problema de identificabilidad, dicho problema es resuelto con datos de riesgos competitivos y bajo el supuesto de que la cópula y su parámetro de asociación son conocidos de antemano.

En el marco de riesgos competitivos se observa $T_j = \min(X_j, Y_j)$ y $\delta_j = I_{(X_j \leq Y_j)}$ ($j = 1, \dots, n$). Suponiendo que $P(X_j = Y_j) = 0$. Haciendo uso de esta información se puede estimar directamente:

$$k(t) = P(X > t, Y > t), p_1(t) = P(X \leq t, X < Y), \\ p_2(t) = P(Y \leq t, Y < X), \text{ para } 0 \leq t < \infty$$

donde $p_1(t)$ y $p_2(t)$ corresponden a las funciones de subdistribución.

[2] afirman que bajo el supuesto de independencia entre X y Y , las distribuciones marginales son determinadas de manera única por esas probabilidades, muestran en general, que si la cópula de (X, Y) es conocida, entonces las distribuciones marginales son determinadas de manera única por los datos de riesgos competitivos.

El estimador cópula gráfico está basado en el algoritmo de búsqueda de raíces, el cual trabaja dividiendo el intervalo a la mitad y seleccionando el subintervalo que contiene la raíz de la función de variable real. Este método es sencillo y fácil para resolver ecuaciones, se basa en el teorema del valor intermedio, el cual establece que toda función continua f en el intervalo $[a, b]$ y $f(a)$ y $f(b)$ tienen signos opuestos, entonces f tiene al menos una raíz en el intervalo (a, b) .

7.1 Extensión del Estimador Cópula Gráfico: Método de Combinación de Riesgos

[1] presentaron una extensión del estimador cópula gráfico propuesto por [2], la propuesta es llamada el método de combinación de riesgos, la cual se aplica a modelos con más de dos riesgos competitivos cuando la cópula es Arquimediana.

A continuación se presentan detalles del método cuando se tienen tres riesgos en competencia.

Sean $(T_1, T_2, T_3) \in R_+^3$ tiempos de falla latentes, de un modelo de riesgos competitivos *tridimensional*. En éste caso se tienen tres tiempos potenciales X, Y y Z , asociados a tres modos de falla, de manera que cuando la falla se presenta por el primer modo, los otros dos modos de falla no pueden ser observados.

7.2 La dependencia modelada con C-Vine y D-Vine cópulas

[11] proponen los C-Vines y D-Vines para simular distribuciones multivariadas flexibles las cuales son necesarias en distintas áreas. La distribución Gaussiana es muy restrictiva y no puede detectar características como asimetría y colas pesadas, las cuales en los análisis financieros son muy comunes. El uso de las cópulas es un reto en dimensiones superiores donde las cópulas multivariadas estándar tienen o poseen estructuras inflexibles. Los vines cópulas superan estas limitaciones y son capaces de modelar patrones de dependencia complejos al tener una gran variedad de cópulas bivariadas como componentes básicos. Los vines son modelos gráficos flexibles para describir cópulas multivariadas usando una cascada de cópulas multivariadas llamadas pair-copulas. Tales construcciones, descomponen una densidad de probabilidad multivariada en cópulas bivariadas, donde cada pair-copula puede ser elegido de manera independiente de los otros. Esto permite una enorme flexibilidad en el modelamiento de la dependencia. En particular, las asimetrías y la dependencia de las colas pueden ser tenidas en cuenta, así como la independencia condicional para construir modelos más parsimoniosos. Los vines combinan las ventajas de modelización con cópulas multivariadas, es decir, separan del modelo marginal, la dependencia y la flexibilidad de la cópula bivariada.

El número de cópulas disponibles para modelar la estructura de dependencia entre más de dos variables es limitado. De hecho, la gran mayoría de las cópulas paramétricas disponibles son bivariadas. Además, el enfoque basado en una cópula multivariada puede no resultar apropiado cuando los pares de variables

presentan estructuras de dependencia diferentes. Finalmente, en muchos casos en que sí se cuenta con una extensión multivariada de la cópula, se utiliza un solo parámetro para describir la estructura de dependencia entre todos los pares de variables, lo cual es una limitación importante si las relaciones entre pares de variables son diferentes. Las construcciones con cópulas bivariadas y los vines constituyen una alternativa al enfoque basado en cópulas multivariadas. Estos modelos permiten extender las cópulas bivariadas a dimensiones superiores utilizando solamente cópulas bivariadas y densidades univariadas como bloques constructivos. El creciente interés en estos modelos se debe a la gran flexibilidad que brindan para modelar una amplia variedad de dependencias combinando cópulas bivariadas de familias diferentes.

7.3 Metodología de los vines

Los vines son una representación gráfica para especificar las llamadas construcciones pair-copulas (PCCs) introducidas por [12]. En particular, las PCC en tres dimensiones, estarían dadas como sigue:

Sea $X = (x_1, x_2, x_3)^T \sim F$ con funciones de distribución marginales F_1, F_2 y F_3 y la respectivas funciones de densidad f_1, f_2 y f_3 , luego la función de densidad trivariada se define como:

$$f(x_1, x_2, x_3) = f_1(x_1)f(x_2|x_1)f(x_3|x_1, x_2)$$

Para realizar la selección de la estructura de los C-Vines y D-Vines; se deben seleccionar los pares de variables que serán modeladas explícitamente con cópulas. Un C-Vine resulta apropiado si una variables domina las interacciones entre las variables. Un D-Vine permite una selección más flexible de dependencias.

8. ESTUDIO DE SIMULACIÓN

Se presentan los resultados obtenidos al comparar la función de sobrevivencia verdadera del tiempo mínimo de falla $S(t)$, con la función de sobrevivencia estimada asumiendo independencia entre los modos de falla ($SI(t)$) y la función de sobrevivencia bajo dependencia ($SD(t)$) del tiempo mínimo de falla, la cual fue estimada haciendo uso del método de combinación de riesgos, para un modelo de riesgos competitivos

trivariado. Se analizan distintas dependencias entre los riesgos, dichas estructuras son generadas mediante la técnica C-DVines.

8.1 Metodología y Escenarios de simulación

En esta sección se describen los escenarios y la metodología de simulación utilizados para el análisis de la función de sobrevivencia en un modelo de riesgos competitivos dependientes trivariados bajo distintas estructuras de dependencia entre los modos de falla.

Inicialmente, se generan muestras de tamaño $n = 50$ y $n = 100$ de tiempos de falla trivariados, con la dependencia deseada entre los riesgos del modelo en riesgos competitivos dependientes y marginales Weibull se tomaron estos tamaños de muestra, puesto que en confiabilidad siempre es de gran interés estudiar problemas con tamaños de muestra pequeños debido a los costos y no es común estudiar problemas con tamaños de muestra demasiado grande.

Las dependencias estudiadas entre los modos de falla se relacionan en la siguiente tabla:

Dependencia	τ_{12}	τ_{13}	$\tau_{23 1}$
Escenario 1	0.5	0.5	0.05
Escenario 2	0.5	0.5	0.9
Escenario 3	0.2	0.5	0.8

Tabla 1. Estructuras de dependencia del modelo de riesgos competitivos dependientes trivariado

Las dependencias propuestas se generaron mediante las técnicas: C-Vines; la cual es apropiada para dependencias donde una de las variables domina toda la estructura de dependencia y D-Vines; permiten una selección más flexible de las dependencias. En el escenario 1 se tomaron dos riesgos con igual dependencia y el otro riesgo independiente a los otros dos. En el escenario 2 se tomaron dos riesgos con igual dependencia y el otro riesgo con una dependencia alta. En el escenario 3 se tomaron los tres riesgos con distintas dependencias.

Seguidamente se encuentra el tiempo mínimo, es decir, $T = \min(T_1, T_2, T_3)$. Se rotulan con 1 si la falla ocurre por el modo de falla 1; 2 si la falla ocurre por el modo 2 y 3 si la falla se presenta por el modo 3.

Para obtener los tiempos de falla Weibull, se hace uso de lo propuesto por [13] en el caso bivariado. Sean T_1 y T_2 tiempos de falla Weibull, una función de sobrevivencia conjunta para la Weibull bivariada se define como sigue:

$$S(t_1, t_2) = e^{-\left[\left(\frac{t_1}{\eta_1}\right)^{\beta_1/(1-\tau)} + \left(\frac{t_2}{\eta_2}\right)^{\beta_2/(1-\tau)}\right]^{(1-\tau)}}$$

donde β_1, β_2 son los parámetros de forma y η_1, η_2 son los parámetros de escala asociados a T_1 y T_2 respectivamente, además $0 \leq \tau \leq 1$, es el parámetro de dependencia entre T_1 y T_2 .

En los escenarios Weibull estudiados se consideraron diferentes formas de la función Hazard de los tiempos de falla. Cuando el parámetro $\beta < 1$, la tasa de falla es decreciente, cuando $\beta > 1$, la tasa es creciente y cuando $\beta = 1$, la tasa de falla es constante.

Parametros	n	β_1	β_2	β_3	η_1	η_2	η_3
Caso 1	50	2	2	2	1	1	1
Caso 2	50	1.5	1	1	1	1	1
Caso 3	50	0.5	1	1	1	1	1
Caso 4	50	1.2	1	0.5	1	1	1
Caso 5	100	2	2	2	1	1	1
Caso 6	100	1.5	1	1	1	1	1
Caso 7	100	0.5	1	1	1	1	1
Caso 8	100	1.2	1	0.5	1	1	1

Tabla 2. Parámetros de la distribución Weibull y tamaños de muestra

Seguidamente, se estimaron los parámetros para las distribuciones del tiempo de falla para cada riesgo, con estos estimadores se obtienen las funciones de sobrevivencia del tiempo mínimo de falla asumiendo independencia. Luego se estiman las distribuciones marginales para cada modo de falla mediante el método de combinación de riesgos, el cual nos permite estimar la función de sobrevivencia bajo dependencia entre los riesgos del modelo trivariado.

Mediante simulación de Monte Carlo se construyen intervalos de confianza empíricos aproximados al 95 % para los percentiles t_p , con $p=0.05, 0.25, 0.50, 0.75, 0.95$, tanto para la función de sobrevivencia estimada bajo independencia y bajo dependencia.

Finalmente, se compararon los dos estimadores en cuestión, calculando la eficiencia relativa ER_p de $(\widehat{S}_I(t))$ con referencia a $(\widehat{S}_D(t))$ en los percentiles $p = 0.05, 0.25, 0.50, 0.75, 0.95$

$$ER_p = \frac{ECM(\widehat{S}_D(t_p))}{ECM(\widehat{S}_I(t_p))} = \frac{E[(\widehat{S}_D(t_p) - S(t_p))^2]}{E[(\widehat{S}_I(t_p) - S(t_p))^2]}$$

Si $ER_p < 1$ entonces $S_I(t_p)$ es menos eficiente en relación a $S_D(t_p)$ para la estimación de la función de sobrevivencia verdadera en los percentiles estudiados.

9. RESULTADOS

A continuación se muestran los resultados obtenidos en todos los escenarios de simulación analizados:

9.1 Escenario 1 de dependencia

En este escenario se simularon dos riesgos con igual dependencia y el otro riesgo se tomó independiente a los anteriores con tamaños de muestrales de 50 y 100. Los riesgos tienen dependencia $\tau_{12} = 0.5, \tau_{13} = 0.5, \tau_{23|1} = 0.05$ y cuyo parámetro de cada cópula son respectivamente; $\theta_{12} = 5.7, \theta_{13} = 5.7$ y $\theta_{23|1} = 0.45$ y se usó el método de C-Vines, el cual genera la dependencia teniendo en cuenta el primer riesgo.

n	β_1	β_2	β_3	Modo1	Modo2	Modo3
50	2	2	2	46%	28%	26%
	1.5	1	1	46%	24%	30%
	0.5	1	1	56%	26%	18%
	1.2	1	0.5	28%	24%	48%
100	2	2	2	42%	33%	25%
	1.5	1	1	39%	31%	30%
	0.5	1	1	52%	21%	27%
	1.2	1	0.5	25%	29%	46%

Tabla 3. Porcentaje de falla para el modelo Weibull con tamaños de muestra $n = 50, 100$ y parámetros de dependencia; $\tau_{12} = 0.5, \tau_{13} = 0.5$ y $\tau_{23|1} = 0.05$ y C-Vines.

En la Tabla 3, se observa que el porcentaje de falla para cada uno de los modos de falla son mayores o iguales al 15 %, lo que nos garantiza tener buenas estimaciones de la función de sobrevivencia para el tiempo mínimo cuando se realizan las estimaciones bajo dependencia e independencia entre los modos de falla del modelo de riesgos competitivos trivariado dependiente. Los porcentajes de falla en la gran mayoría de

los casos son distintos, por tanto garantiza que los modelos estudiados no se dejan influenciar por las dependencias. Lo anterior permite ilustrar claramente el problema de identificabilidad y es claro que no se están trabajando con funciones de distribución marginales sino con funciones de sub-distribución [7].

En la Tablas 4, 5, 6, 7, 8, 9,10, 11 se muestran los límites inferior y superior de la función de sobrevivencia estimada bajo dependencia $LI.S_D(t)$ y $LS.S_D(t)$ respectivamente. Los límites inferior y superior de la función de sobrevivencia estimada bajo independencia denotados por $LI.S_I(t)$, $LS.S_I(t)$, y la precisión de cada intervalo de confianza empírico aproximado para la sobrevivencia verdadera en los percentiles $p = 0.05, 0.25, 0.50, 0.75, 0.95$ y tamaño de muestral 50.

Percentil	$LI.S_D(t)$	$LS.S_D(t)$	Precisión
0.05	0.92194	0.95360	0.03166
0.25	0.74373	0.77831	0.03458
0.50	0.49458	0.52236	0.02778
0.75	0.24129	0.27104	0.02975
0.95	0.05746	0.08293	0.02547

Tabla 4. Intervalos de confianzas empíricos aproximados del 95 % para $S(tp)$ bajo dependencias. Modelo Weibull con parámetros $\beta_1=2, \beta_2=2, \beta_3=2, \eta_1=1, \eta_2=1, \eta_3=1, n=50$ y $\tau_{12}=0.5, \tau_{13}=0.5$ y $\tau_{231}=0.05$

Percentil	$LI.S_I(t)$	$LS.S_I(t)$	Precisión
0.05	0.92051	0.96526	0.04475
0.25	0.72263	0.79506	0.07243
0.50	0.48063	0.52442	0.04379
0.75	0.23449	0.28312	0.04864
0.95	0.05262	0.08731	0.03466

Tabla 5. Intervalos de confianzas empíricos aproximados del 95 % para $S(tp)$ bajo independencia. Modelo Weibull con parámetros $\beta_1=2, \beta_2=2, \beta_3=2, \eta_1=1, \eta_2=1, \eta_3=1, n=50$ y $\tau_{12}=0.5, \tau_{13}=0.5$ y $\tau_{231}=0.05$

Percentil	$LI.S_D(t)$	$LS.S_D(t)$	Precisión
0.05	0.91630	0.95091	0.03461
0.25	0.73801	0.76825	0.03025
0.50	0.50586	0.56409	0.05823
0.75	0.24050	0.31619	0.07570
0.95	0.04427	0.06895	0.02468

Tabla 6. Intervalos de confianzas empíricos aproximados del 95 % para $S(tp)$ bajo dependencia. Modelo Weibull con parámetros $\beta_1=1.5, \beta_2=1, \beta_3=1, \eta_1=1, \eta_2=1, \eta_3=1, n=50$ y $\tau_{12}=0.5, \tau_{13}=0.5$ y $\tau_{231}=0.05$

Percentil	$LI.S_I(t)$	$LS.S_I(t)$	Precisión
0.05	0.90648	0.95629	0.04981
0.25	0.72281	0.80682	0.08401
0.50	0.45817	0.57274	0.11457
0.75	0.19825	0.30360	0.10534
0.95	0.04236	0.08662	0.04426

Tabla 7. Intervalos de confianzas empíricos aproximados del 95 % para $S(tp)$ bajo independencia. Modelo Weibull con parámetros $\beta_1=1.5, \beta_2=1, \beta_3=1, \eta_1=1, \eta_2=1, \eta_3=1, n=50$ y $\tau_{12}=0.5, \tau_{13}=0.5$ y $\tau_{231}=0.05$

Percentil	$LI.S_D(t)$	$LS.S_D(t)$	Precisión
0.05	0.91749	0.95199	0.03450
0.25	0.74329	0.77265	0.02936
0.50	0.49172	0.55792	0.06620
0.75	0.24294	0.28412	0.04118
0.95	0.04805	0.06686	0.01881

Tabla 8. Intervalos de confianzas empíricos aproximados del 95 % para $S(tp)$ bajo dependencia. Modelo Weibull con parámetros $\beta_1=0.5, \beta_2=1, \beta_3=1, \eta_1=1, \eta_2=1, \eta_3=1, n=50$ y $\tau_{12}=0.5, \tau_{13}=0.5$ y $\tau_{231}=0.05$

Percentil	$LI.S_I(t)$	$LS.S_I(t)$	Precisión
0.05	0.91824	0.95595	0.03771
0.25	0.72363	0.75330	0.02967
0.50	0.48593	0.55701	0.07108
0.75	0.22911	0.28969	0.06057
0.95	0.04975	0.07602	0.02627

Tabla 9. Intervalos de confianzas empíricos aproximados del 95 % para $S(tp)$ bajo independencia. Modelo Weibull con parámetros $\beta_1=0.5, \beta_2=1, \beta_3=1, \eta_1=1, \eta_2=1, \eta_3=1, n=50$ y $\tau_{12}=0.5, \tau_{13}=0.5$ y $\tau_{231}=0.05$

Percentil	$LI.S_D(t)$	$LS.S_D(t)$	Precisión
0.05	0.92666	0.95999	0.03333
0.25	0.73058	0.75438	0.02380
0.50	0.50191	0.53937	0.03746
0.75	0.24899	0.26009	0.01110
0.95	0.04619	0.06875	0.02256

Tabla 10. Intervalos de confianzas empíricos aproximados del 95 % para $S(tp)$ bajo dependencia. Modelo Weibull con parámetros $\beta_1=1.2, \beta_2=1, \beta_3=0.5, \eta_1=1, \eta_2=1, \eta_3=1, n=50$ y $\tau_{12}=0.5, \tau_{13}=0.5$ y $\tau_{231}=0.05$

Percentil	$LI.S_I(t)$	$LS.S_I(t)$	Precisión
0.05	0.91185	0.95107	0.03922
0.25	0.73049	0.76006	0.02957
0.50	0.48191	0.52224	0.04033
0.75	0.24449	0.26162	0.01714
0.95	0.03906	0.06842	0.02936

Tabla 11. Intervalos de confianzas empíricos aproximados del 95 % para $S(tp)$ bajo independencia. Modelo Weibull con parámetros $\beta_1=1.2, \beta_2=1, \beta_3=0.5, \eta_1=1, \eta_2=1, \eta_3=1, n=50$ y $\tau_{12}=0.5, \tau_{13}=0.5$ y $\tau_{231}=0.05$

Se evidencia que los intervalos de confianza bajo dependencia en los modos de falla resultan ser más precisos que los intervalos de confianza bajo independencia entre los modos de falla del modelo.

En las tablas se observa que los intervalos de confianza bajo dependencia entre los modos de falla del modelo trivariado siempre se encuentran contenidos en los intervalos de confianza bajo independencia de dicho modelo. En algunos casos el límite inferior o superior se encuentran cercanos al valor verdadero pero lo importante es que contengan al valor de la sobrevivencia verdadera.

Finalmente, se compararon los dos estimadores en cuestión calculando la eficiencia relativa ER_p de ($\widehat{S}_I(t)$) con referencia a ($\widehat{S}_D(t)$) en los percentiles $p = 0.05, 0.25, 0.50, 0.75, 0.95$

$$ER_p = \frac{ECM(\widehat{S}_D(t_p))}{ECM(\widehat{S}_I(t_p))} = \frac{E[(\widehat{S}_D(t_p) - S(t_p))^2]}{E[(\widehat{S}_I(t_p) - S(t_p))^2]}$$

Si $ER_p < 1$ entonces $S_I(t_p)$ es menos eficiente en relación a $S_D(t_p)$ para la estimación de la función de sobrevivencia verdadera en los percentiles estudiados.

		Percentiles				
<i>n</i>	<i>caso</i>	0.05	0.25	0.50	0.75	0.95
50	1	0.851	0.718	0.851	0.438	0.998
	2	0.851	0.782	0.732	0.405	0.992
	3	0.959	0.949	0.434	0.104	0.231
	4	0.995	0.801	0.420	0.341	0.417
100	5	0.086	0.049	0.244	0.729	0.121
	6	0.851	0.782	0.732	0.405	0.992
	7	0.959	0.949	0.434	0.104	0.231
	8	0.995	0.801	0.420	0.341	0.417

Tabla 12. Eficiencia relativa para el escenario 1; modelo Weibull con tamaños muestrales $n = 50, 100$ y parámetros de dependencia; $\tau_{12} = 0.5, \tau_{13} = 0.5$ y $\tau_{23|1} = 0.05$ y C-Vines

		Percentiles				
<i>n</i>	<i>caso</i>	0.05	0.25	0.50	0.75	0.95
50	1	0.673	0.210	0.081	0.252	0.989
	2	0.960	0.593	0.199	0.151	0.029
	3	0.165	0.794	0.107	0.187	0.280
	4	0.990	0.991	0.697	0.576	0.980
100	5	0.349	0.716	0.100	0.883	0.196
	6	0.065	0.123	0.290	0.228	0.130
	7	0.311	0.404	0.865	0.522	0.245
	8	0.134	0.680	0.262	0.139	0.294

Tabla 13. Eficiencia relativa para el escenario 2; modelo Weibull con tamaños muestrales $n = 50, 100$ y parámetros de dependencia; $\tau_{12} = 0.5, \tau_{13} = 0.5$ y $\tau_{23|1} = 0.9$ y C-Vines

		Percentiles				
<i>n</i>	<i>caso</i>	0.05	0.25	0.50	0.75	0.95
50	1	0.400	0.088	0.786	0.555	0.034
	2	0.023	0.021	0.603	0.979	0.999
	3	0.994	0.981	0.181	0.977	0.647
	4	0.853	0.454	0.751	0.505	0.922
100	5	0.490	0.834	0.632	0.499	0.195
	6	0.952	0.768	0.564	0.483	0.265
	7	0.487	0.218	0.291	0.466	0.462
	8	0.497	0.610	0.104	0.310	0.304

Tabla 14. Eficiencia relativa para el escenario 3; modelo Weibull con tamaños muestrales $n = 50, 100$ y parámetros de dependencia; $\tau_{12} = 0.2, \tau_{13} = 0.5$ y $\tau_{23|1} = 0.8$ y C-Vines

En las Tablas 12, 13 y 14. Se evidencia que el estimador bajo independencia es menos eficiente que el estimador bajo dependencia, puesto que los valores calculados son menores que 1, vale la pena resaltar que en algunos percentiles el valor calculado es cercano a 1 y en general no se evidencia ningún patrón en particular.

Para efectos prácticos, sólo se muestran los resultados del primer escenario, teniendo en cuenta que se trabajaron los otros dos y dichos resultados fueron similares con respecto a que el estimador bajo dependencia se comporta mejor, que el estimador bajo independencia.

CONCLUSIONES

En gran mayoría de los casos estudiados el estimador bajo dependencia $S_D(t)$ presentó mayor eficiencia frente al estimador bajo independencia

$S_{\lambda}(t)$ de $S(t)$. Por lo tanto se recomienda utilizar el estimador $S_D(t)$ cuando no se tiene certeza de que los tiempos de falla sean independientes.

Los intervalos de confianza para la función de sobrevivencia cuando se asume dependencia entre los modos de falla son más estrechos que los intervalos de confianza para la función de sobrevivencia cuando se asume independencia entre los modos de falla del modelo de riesgos trivariados con marginales Weibull.

Para un modelo de riesgos competitivos trivariados, el método de combinación riesgos (extensión del estimador cópula gráfico) es una buena técnica para estimar las funciones de distribución marginal y la función de sobrevivencia del tiempo mínimo en un modelo en riesgos competitivos, cuando se tiene dos riesgos con igual dependencia y el otro independiente a los anteriores, también cuando se tienen dos riesgos con igual dependencia y el otro riesgo presenta una alta dependencia, y por último cuando los riesgos para un caso particular, presentan distintas dependencias (generados por medio de C-DVines).

REFERENCIAS

- [1] Lo, S. and Wilke, R. A, “A copula model for dependent competing risks”, *Journal of the Royal Statistical Society*, vol. 59, no. 2, pp. 359-376, 2010.
- [2] M. Zheng, J. P. Klein, “Estimates of marginal survival for dependent competing risks based on an assumed copula” *Biometrika*, vol. 82, no 1, pp. 127-138, 1995.
- [3] Nelsen, R B, “An introduction to copulas”, *Springer*, second edn, New York, 2006.
- [4] Tsiatis, A, “A nonidentifiability aspect of the problem of competing risks”, *Proc. Natl. Acad. Sci, USA*. 72(1), pp. 20-22, 1975.
- [5] Meeker, W. Q. and Escobar, L. A. and Hong, Y, “Using Accelerated Life Tests Results to Predict Product Field Reliability”, *Technometrics*, vol. 51, no. 2, pp. 146-161, 2009.
- [6] B. Schweizer, A. Sklar, “Probabilistic Metric Spaces”, *Dover Publications*, New York, 1983.
- [7] Paz-Sabogal., M. C., Lopera-Gómez., C. M. and Yañez-Canal., S. , “Extensión del estimador cópula gráfico para un modelo con más de dos riesgos competitivos dependientes”, *Revista Ingeniería y Competitividad-Univalle* 16(1), 2014.
- [8] M. Pintilie, “Competing Risks: A practical Perspective”, *John Wiley and Sons*, 2006.
- [9] Meeker, W. Q. and Escobar, L. A, “Statistical methods for reliability data”, *New York: John Wiley and Sons*, 1998.
- [10] Escarela, G. and Carriere, J., “Fitting competing risks with an assumed copula”, *Statistical Methods in Medical Research* 12, 33-349, 2003.
- [11] Brechman, E. C. and Shepsmeier, U., “Modeling dependence with C- and D-vine copula: The r package cdvine”, *Journal of statistical software* 52, 2013.
- [12] Aas, K., Czado, C., Frigessi, A. and Bakken, H., “Pair-copula constructions of multiple dependence”, *Mathematics and Economics* 44(2), 182-198, 2009.
- [13] Lu, J. C., and Bhattacharyya, G. K., “Some new constructions of bivariate Weibull models”, *Annals of the Institute of Statistical Mathematics*, 42(3), 543-559, 1990.