

Estimación de la Confiabilidad para Modelos de Riesgos Competitivos Dependientes usando Cópulas

Estimation Reliability for Models Dependent Competing Risks Copulas using

*Hugo Alberto Brango García**

*Sérgio Yañez Canal***

*Carlos Mario Lopera****

*Mario César Jaramillo Elorza*****

Resumen

La metodología tradicional asume independencia entre los modos de falla con riesgos en competencia ("competing risks") para la estimación de la función de confiabilidad, pero estudios recientes advierten sobre los posibles errores de estimación que dicho supuesto conlleva, cuando realmente existe dependencia entre los modos de falla. El supuesto de independencia es matemáticamente conveniente, ya que ésta puede ser definida de manera única. Cuando los modos de falla son dependientes se presenta el problema de identificabilidad ya que la estructura de dependencia de estos puede ser formulada de muchas maneras. Este artículo usa cópulas, lo que, se sabe, permite resolver el problema. Se estimó, vía simulación, la confiabilidad asumiendo una cópula y se compararon los resultados, bajo distribuciones Weibull y lognormal bivariadas, con el estimador $\hat{S}^*(t)$, mostrando su buen comportamiento y su ventaja en términos prácticos.

Palabras clave: Riesgos en competencia, Identificabilidad, Cópula

Abstract

The traditional approach assumes independence between modes fails with competing risks ("Competing risks") to estimate reliability function, but recent studies on possible errors in estimation that this course entails, when really exists dependency among failure modes. The

*Estudiante Maestría en estadística, Universidad Nacional de Colombia, Sede Medellín.
Email:habrangog@unal.edu.co

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

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

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

independence assumption is mathematically convenient, because it can be defined. When the failure modes are dependent it presents the problem of identifiability as the dependence structure of these can be formulated in many ways. This article uses copulas, which is known to solve the problem. It was tested via simulation, the reliability assuming a Copula and compared the results under Weibull and lognormal bivariate distributions with estimator $\hat{S}^*(t)$, showing their good behavior and advantage in practical terms.

Key words: Competing risks, Identifiability, Copula

1. Introducción

Las causas o las diferentes maneras en que puede fallar un sistema o componente, reciben el nombre de modos de falla. Existen muchos sistemas, subsistemas y componentes que tienen más de un modo de falla; en ciertas aplicaciones y para algunos propósitos es importante distinguir entre esos diferentes modos de falla, con el objeto de mejorar la confiabilidad (Meeker y Escobar, 1998). El tiempo de falla de un sistema con varios modos de falla puede ser modelado considerando un sistema en serie o un modelo de riesgos en competencia ("competing risks"), donde cada modo de falla es una componente del sistema en serie, y cada componente tiene un tiempo de vida aleatorio, por lo tanto el sistema fallará cuando el modo de falla con el tiempo de vida más corto falle.

Si T_1 y T_2 son los tiempos de falla asociados a un problema de riesgos en competencia ("competing risks") con dos modos de falla y estamos interesados en estimar la función de sobrevivencia marginal de T_1 , a menudo es imposible medir T_1 , debido a la ocurrencia de algún otro evento con tiempo T_2 . Si solamente estamos interesados en T_1 entonces T_2 puede considerarse como censura del evento de interés. La información de la muestra observada es $T = \min(T_1, T_2)$.

Un supuesto común para estimar la función de sobrevivencia de T_1 es que el tiempo de censura T_2 es independiente de T_1 . Bajo este supuesto se sabe que los datos observados proporcionan información suficiente para determinar de manera única la distribución marginal de T_1 (Peterson, 1977). El supuesto de independencia entre T_1 y T_2 no siempre es cierto, así la estimación de la distribución marginal de T_1 se complica ya que la dependencia puede ser formulada de muchas maneras. En muchas situaciones prácticas la dependencia es una condición usual, lo que ha generado el creciente interés por su estudio.

Tsiatis (1975) mostró que las distribuciones marginales y la distribución conjunta de dos riesgos en competencia, son en general no identificables, esto es, hay muchas funciones de distribución conjuntas diferentes que comparten las mismas funciones de sub-distribución (problema de identificabilidad). Si los riesgos son independientes, sus distribuciones marginales son identificables, estas marginales son las funciones de sub-distribución.

El problema clásico de riesgos en competencia es identificar las distribuciones marginales y las distribuciones conjuntas de los tiempos de falla asociados a los modos de falla a partir de datos de la forma (T, C) , donde T es un tiempo de falla y C es la causa de falla.

Manotas et al. (2008), estudiaron el efecto en la estimación de la sobrevivencia $S(t)$, cuando se asume el supuesto de independencia entre tiempos de falla en competencia que realmente son dependientes. Para ello simulaban tiempos de falla bivariados Weibull y Lognormales, estimando la confiabilidad conjunta mediante el estimador $\hat{S}^*(t) = \hat{S}_1(t)\hat{S}_2(t)$. Los resultados

no mostraron diferencias significativas en la estimación de la función de sobrevivencia.

En este trabajo se estima vía simulación la confiabilidad $S(t)$ bajo dependencia asumiendo un modelo cópula en el contexto de riesgos en competencia. Estos resultados se comparan con el estimador de la confiabilidad $\hat{S}^*(t)$ propuesto en Yañez et al. (2005) y luego modificado por Manotas et al. (2008).

2. Método

2.1. Competing risks

En competing risks, el resultado observado comprende T , el tiempo de falla y C el tipo de causa o modo de falla. $T > 0$ es una variable aleatoria continua y C puede tomar un número fijo pequeño de valores $i = 1, 2, \dots, p$. La distribución de probabilidad es bivariada de la forma $(T; C)$ con C discreta y T continuo. Se asume que a cada falla puede asignarse una y solo una causa del conjunto de p causas.

Hay varias funciones de importancia en la teoría de competing risks: la subdistribución, subdensidad, subhazard, hazard de causa específica y la hazard de la subdistribución.

- **La CIF ó función de subdistribución**

Para un evento de tipo i ($i = 1, 2, \dots, p$) la CIF es definida como la probabilidad conjunta: $F_i(t) = P(T \leq t, C = i)$

En otras palabras la CIF es la probabilidad que un evento de tipo i ocurra hasta el tiempo t . La función de distribución conjunta es la probabilidad que un evento de cualquier tipo ocurra hasta el tiempo t . La función de distribución conjunta es igual a la suma de las CIFs para todo tipo de evento.

$$F(t) = P(T \leq t) = \sum_{i=1}^p P(T \leq t, C = i) = \sum_{i=1}^p F_i(t)$$

- **Función de Subsobrevivencia**

Es la probabilidad que un evento de tipo i no ocurra hasta el tiempo t es decir, es definida como la probabilidad de que un evento de tipo i no haya fallado hasta el tiempo t :

$$S_i(t) = P(T > t, C = i)$$

Nótese que cuando los competing risks no están presentes la función de distribución conjunta se encuentra en el intervalo $[0; 1]$. En contraste, en presencia de los competing risks la CIF puede tomar sólo el valor $P(C = i)$ porque $\lim_{t \rightarrow \infty} F_i(t) = P(C = i)$

- **Función de Subdensidad**

Para eventos de tipo i es definida como:

$$f_i(t) = \frac{\partial F_i(t)}{\partial t}$$

La función hazard cruda $h(j, t)$ normalmente no coincide con la hazard marginal, que es definida como

$$h_j(t_j) = \lim_{\delta \rightarrow 0} \frac{P(t_j \leq T_j < t_j + \delta | T_j \geq t_j)}{\delta} = -\frac{d}{dt_j} \{ \log S_j(t_j) \} \quad (1)$$

Una excepción se presenta cuando los tiempos de sobrevivencia T_j son mutuamente independientes, esto es

$$S(t_1, \dots, t_m) = \prod_{j=1}^m P(T_j > t_j) \quad (2)$$

También es posible encontrar expresiones donde las funciones hazard marginal y cruda coincidan cuando los T_j^s son estadísticamente dependientes.

2.2. Copulas

Los modelos cópula son clases de distribuciones de sobrevivencia bivariada, especificada en términos de las funciones de sobrevivencia marginal y una función cópula, que es una función de distribución continua bivariada sobre el cuadrado unitario $[0,1]$ con marginales uniformes. Una característica de la clase cópula es que la determinación de las marginales a través de la cópula ayuda a modelar y entender eficazmente la estructura de dependencia, Escarela et al. (2003).

En el contexto de riesgos en competencia, es posible especificar la distribución de sobrevivencia conjunta en términos de dos distribuciones de sobrevivencia marginal y una cópula que permite relacionar la dependencia entre las variables aleatorias individuales que corresponden a cada riesgo de falla, Escarela et al. (2003). Así, sea la cópula de distribución C_α con $\alpha \in \mathbb{R}$ para dos variables aleatorias T_1 y T_2 , donde α representa el parámetro de dependencia entre estas dos variables.

Si (T_1, T_2) proviene de una cópula para algún α , entonces la función de distribución conjunta de (T_1, T_2) está dada por

$$S(t) = C_\alpha [S_1(t_1), S_2(t_2)] \quad (3)$$

donde $S_1(t_1)$ y $S_2(t_2)$ son las funciones de distribución marginal para T_1 y T_2 respectivamente y C_α es una función de distribución bivariada. Bajo independencia la es $C_I(S_1(t_1), S_2(t_2)) = S_1(t_1)S_2(t_2)$.

2.3. Identificabilidad

En el contexto de competing risks observamos $T_i = \min(X_i, Y_i)$ y $\delta_i = 1(X_i \leq Y_i)$ ($i = 1, 2, \dots, n$). Supongamos que $P(X_i = Y_i) = 0$. Entonces usando estos datos podemos estimar directamente

$$k(t) = P(X > t, Y > t)$$

$$p_1(t) = P(X \leq t, X < Y)$$

$$p_2(t) = P(Y \leq t, Y < X)$$

Sabemos que, bajo el supuesto de independencia de X y Y , la distribución marginal de X es determinada de manera única por estas probabilidades.

A continuación mostraremos un resultado más general, que si la cópula de (X, Y) es conocida, entonces las distribuciones marginales tanto de X como de Y son determinadas de manera única por los datos de competing risks.

Teorema 1. Suponga que las funciones de distribución marginal de (X, Y) son continuas y estrictamente crecientes en $(0, \infty)$. Supongamos que la cópula C es conocida y $\mu_c(E) > 0$ para cualquier conjunto abierto E en $[0, 1] \times [0, 1]$. Entonces F y G , las funciones de distribución marginal de X y Y , son determinadas de manera única por $\{k(t), p_1(t), p_2(t), t > 0\}$

Teorema 2. Sea $u(x, y)$ la función densidad de C . Si $u(x, y) > 0$ para cualquier $(x, y) \in [0, 1] \times [0, 1]$, entonces el resultado del teorema 1 se mantiene.

Corolario 2. En el teorema 1, si hay tiempos t_1 y t_2 tales que $F(t_1) = 1$ y $G(t_2) = 1$, y tanto F y G son estrictamente crecientes en $(0, t_1)$ y $(0, t_2)$, respectivamente, entonces F y G son determinadas de manera única sobre $(0, \min(t_1, t_2))$.

2.4. El estimador Cópula-Gráfico

Si F y G son las distribuciones marginales de X y Y , entonces para cualquier cópula, tenemos

$$\mu_c(A_t) = P(X > t, Y > t) = k(t) \quad (4)$$

$$\mu_c(B_t) = P(X \leq t, X < Y) = p_1(t) \quad (5)$$

donde $A_t = \{(x, y) \mid F(t) < x \leq 1, G(t) < y \leq 1\}$, $B_t = \{(x, y) \mid 0 < x \leq F(t), GF^{-1}(t) < y \leq 1\}$

De el teorema 1 se observa que estas dos relaciones determinan de manera única a F y G y se encuentran estimadores \hat{F} y \hat{G} de F y G que preservan estas propiedades sobre una rejilla seleccionada de m puntos $0 < t_1 < t_2 < \dots < t_n < \max\{T_i, i = 1, \dots, n\}$. Para construir nuestro estimador sea

$$\hat{A}_t = \{(x, y) \mid \hat{F}(t) < x \leq 1, \hat{G}(t) < y \leq 1\}, \hat{B}_t = \{(x, y) \mid 0 < x \leq \hat{F}(t), \hat{G}\hat{F}^{-1}(t) < y \leq 1\}$$

y
 est. $P(X > t, Y > t) = \frac{1}{n} \sum (T_i > t)$, est. $P(X \leq t, X < Y) = \frac{1}{n} \sum (T_i \leq t, \delta_i = 1)$ las estimaciones empíricas de $P(X > t, Y > t)$ y $P(X \leq t, X < Y)$.

Se encuentra $\hat{F}(t_i)$ como la raíz de (3) sujeto a la definición de $\hat{G}(t_i)$ como una raíz de $\hat{F}(t_i)$ resolviendo la ecuación (2). Siendo \hat{F} y \hat{G} líneas rectas en cada intervalo (t_i, t_{i+1})

$$\mu_c(\hat{A}_{t_i}) - \text{est.}P(X > t_i, Y > t_i) = 0 \quad (6)$$

$$\mu_c(\hat{B}_{t_i}) - \text{est.}P(X \leq t, X < Y) = 0 \quad (7)$$

El algoritmo para construir \hat{F} y \hat{G} basado sobre una bisección del algoritmo raíz, es como sigue

Paso 1. Para $i = 1$, dado un punto inicial $\hat{F}(t_1)$, digamos $\hat{F}^{(1)}(t_1)$, encontrar $\hat{G}^{(1)}(t_1)$ resolviendo (2).

Paso 2. Si el punto $(\widehat{F}^{(1)}(t_1), \widehat{G}^{(1)}(t_1))$ satisface (3), vaya al Paso 4. Por otra parte use (3) para decidir si el siguiente valor, $\widehat{F}^{(2)}(t_1)$ es mayor o menor que $\widehat{F}^{(1)}(t_1)$. Use el punto medio de los intervalos $(\widehat{F}^{(1)}(t_1), 1)$ o $(0, \widehat{F}^{(1)}(t_1))$ respectivamente, como el valor de $\widehat{F}^{(2)}(t_1)$.

Paso 3. Repita los pasos 1 y 2 usando la estimación actual de $\widehat{F}(t_1)$ y $\widehat{G}(t_1)$. Después de i pasos el nuevo valor de $\widehat{F}(t_1)$ es el punto medio de los intervalos $(a_i, \widehat{F}^{(i)}(t_1))$ o $(\widehat{F}^{(i)}(t_1), b_i)$, donde

$$a_i = \max \left\{ \widehat{F}^{(k)}(t_1) \mid k < i, \widehat{F}^{(k)}(t_1) < \widehat{F}^{(i)}(t_1) \right\}, b_i = \min \left\{ \widehat{F}^{(k)}(t_1) \mid k < i, \widehat{F}^{(k)}(t_1) > \widehat{F}^{(i)}(t_1) \right\}.$$

Este proceso continúa hasta que encontremos $\widehat{F}(t_1)$ y $\widehat{G}(t_1)$ que satisfacen (2) y (3). Esto indica que la convergencia del algoritmo es garantizada.

Paso 4. Repita los Pasos 1-3 para $i = 2, \dots, m$. Cualquier función creciente y continua a derecha tal que la función $\widehat{G}\widehat{F}^{-1}(t)$ es una línea recta sobre cada intervalo $[\widehat{F}(t_{i-1}), \widehat{F}(t_i)]$ proporcionará un estimador consistente. Para $t > \max \{T_i\}$, definimos $\widehat{F}(t)$ y $\widehat{G}(t)$ que serán $\widehat{F}(t_m)$ y $\widehat{G}(t_m)$ respectivamente.

2.5. Estimador $\widehat{S}^*(t)$

El estimador $\widehat{S}^*(t)$ es un estimador de la función de confiabilidad del tiempo mínimo de falla de un sistema con dos modos de falla que compiten por ser la falla, obtenido estimando los parámetros asociados a cada modo de falla individual, para estimar los parámetros y la distribución asociada a los tiempos de falla, los tiempos del otro modo de falla son tiempos de censura. Así,

$$\widehat{S}^*(t) = \widehat{S}_1(t)\widehat{S}_2(t) \quad (8)$$

donde $\widehat{S}_1(t)$ y $\widehat{S}_2(t)$ son las funciones de confiabilidad marginales estimadas para cada modo de falla.

3. Estudio de simulación

En esta sección se lleva a cabo un estudio de simulación para comparar la confiabilidad estimada mediante dos métodos distintos para poblaciones con tiempos de falla Weibull y Lognormales.

Para simular los tiempos de falla Weibull bivariados se usó el algoritmo cópula compuesta de Gumbel-Hougaard (Frees y Valdes, 1997). La función de confiabilidad conjunta de la Weibull bivariada utilizada es (Lu y Bhattacharyya, 1993)

$$S(t_1, t_2) = \exp \left\{ - \left[\left(\frac{t_1}{\theta_1} \right)^{\frac{\beta_1}{\delta}} + \left(\frac{t_2}{\theta_2} \right)^{\frac{\beta_2}{\delta}} \right]^{\delta} \right\} \quad (9)$$

donde $T_1 > 0$ y $T_2 > 0$ son tiempos de falla Weibull $\beta_1 > 0$ y $\beta_2 > 0$, los parámetros de forma y $\theta_1 > 0$ y $\theta_2 > 0$, los parámetros de escala asociados a T_1 y T_2 respectivamente, $0 \leq \delta \leq 1$, es el parámetro de dependencia entre T_1 y T_2 . Cuando el parámetro de dependencia

δ entre tiempos Weibull es 1, entonces hay independencia entre T_1 y T_2 . A medida que δ disminuye, la dependencia entre T_1 y T_2 aumenta.

En este trabajo se usa como parámetro de dependencia entre tiempos de falla Weibull, T_1 y T_2 , a $\lambda = 1 - \delta$, de manera que $0 \leq \lambda \leq 1$. Cuando λ es cero, entonces hay independencia entre T_1 y T_2 . A medida que λ , aumenta, la dependencia entre T_1 y T_2 aumenta. Se consideran 30 escenarios con tiempos de falla weibull, cada uno de los cuales corresponde a la llamada función de confiabilidad verdadera, $S(t)$. Los parámetros de escala de la distribución Weibull bivariada se fijan en $\theta_1 = \theta_2 = 1$ ya que el parámetro de dependencia λ , no depende del parámetro de escala (Lu y Bhattacharyya, 1993). Los escenarios presentan situaciones donde compiten distribuciones Weibull con tasas de falla decrecientes y crecientes. Cuando el parámetro $\beta < 1$, la tasa de falla es decreciente, y cuando $\beta > 1$, la tasa de falla es decreciente. Los escenarios toman como base las distintas combinaciones posibles de dicha tasa.

Los casos de tiempos Weibull a estudiar son:

1. Tiempos de falla con parámetros de forma β_i diferentes. Modo de falla con tasa de falla decreciente vs. Modo de falla con tasa de falla creciente: $\beta_1 = 0,5$ y $\beta_2 = 2$.
2. Tiempos de falla con parámetros de forma β_i diferentes. Modo de falla con tasa de falla decreciente vs. Modo de falla con tasa de falla constante: $\beta_1 = 0,5$ y $\beta_2 = 1$.
3. Tiempos de falla con parámetros de forma β_i iguales. Modo de falla con tasa de falla decreciente vs. Modo de falla con tasa de falla decreciente: $\beta_1 = 0,5$ y $\beta_2 = 0,5$.
4. Tiempos de falla con parámetros de forma β_i iguales. Modo de falla con tasa de falla creciente vs. Modo de falla con tasa de falla creciente: $\beta_1 = 2$ y $\beta_2 = 2$.
5. Tiempos de falla con parámetros de forma β_i diferentes. Modo de falla con tasa de falla creciente vs. Modo de falla con tasa de falla constante: $\beta_1 = 2$ y $\beta_2 = 1$.

En cada uno de estos casos se toman valores del parámetro de dependencia: $\lambda = 0.0, 0.2, 0.4, 0.6, 0.8, 0.9$. De manera que en total se tienen 30 escenarios distintos para para la distribución Weibull bivariada.

Para simular los tiempos de falla Lognormal bivariados con coeficiente de correlación ρ_L , se generan normales bivariadas con coeficiente de correlación ρ_N , donde ρ_L está definido mediante la siguiente relación:

$$\rho_L = \frac{\exp(\rho_N \sigma_1 \sigma_2) - 1}{\{[\exp(\sigma_1^2) - 1][\exp(\sigma_2^2) - 1]\}^{\frac{1}{2}}} \quad (10)$$

y luego se sacan exponenciales a cada de los tiempos normales bivariados.

Se consideran 24 escenarios con tiempos de falla lognormales, cada uno de los cuales corresponde a la llamada función de confiabilidad verdadera, verdadera $S(t)$. Los casos de tiempos lognormales son:

1. Tiempos de falla lognormales con parámetros iguales: $\mu_1 = \mu_2 = 0$ y $\sigma_1 = \sigma_2 = 1$.

2. Tiempos de falla lognormales con parámetros μ'_i s iguales y σ'_i s diferentes: $\mu_1 = \mu_2 = 0$, $\sigma_1 = 1$ y $\sigma_2 = 2$.
3. Tiempos de falla lognormales con parámetros μ'_i s diferentes y σ'_i s iguales: $\mu_1 = 0$ y $\mu_2 = 1$, $\sigma_1 = \sigma_2 = 1$.
4. Tiempos de falla lognormales con parámetros μ'_i s y σ'_i s diferentes: $\mu_1 = 0$ y $\mu_2 = 1$, $\sigma_1 = 1$ y $\sigma_2 = 2$.

Para cada uno de estos casos hay 6 situaciones diferentes de acuerdo al parámetro de dependencia entre los tiempos de falla. Para tiempos de falla lognormales T_1 y T_2 , el parámetro de dependencia es el coeficiente de correlación ρ_N entre $\log(T_1)$ y $\log(T_2)$ y este toma los siguientes valores: $\rho_N = 0.0, 0.2, 0.4, 0.6, 0.8, 1.0$. Es decir, para cada uno de los 4 casos de tiempos lognormales hay 6 situaciones distintas que van desde independencia hasta dependencia extrema. Lo que da un total de 24 escenarios para el cuales estimaremos $\widehat{S}^*(t)$ y $\widehat{S}(t)_{\text{Cópula}}$.

Para todos los escenarios de estas dos distribuciones se estimó la confiabilidad mediante dos métodos: el primero es el estimador $\widehat{S}^*(t)$ (Manotas., et al. 2008) y el segundo aplicando el algoritmo cópula gráfico presentado en la subsección 2.4 y asumiendo un modelo cópula, el modelo cópula con el cual se trabajo es el gamma frailty (Oakes, 1989). Se seleccionó este modelo por facilidad en la aplicación del algoritmo cópula gráfico y conociendo que el modelo cópula asumido no influye en la estimación de la confiabilidad conjunta si no que esta está influenciada por la dependencia entre los tiempos de falla (Zheng and Klein, 1995). La confiabilidad estimada con el método cópula lo llamamos $\widehat{S}(t)_{\text{Cópula}}$. Las comparaciones de la confiabilidad estimada para el tiempo mínimo de falla para tiempos Weibull y Lognormales bivariados con $\widehat{S}(t)_{\text{Cópula}}$ y $\widehat{S}^*(t)$ se hizo mediante el cálculo del error cuadrático medio ECM_p en los percentiles $p = 0.05, 0.25, 0.50, 0.75, 0.95$, para determinar la dispersión de las dos estimaciones respecto a la confiabilidad verdadera en los percentiles dados. Se hicieron pruebas de bondad de ajuste donde se comparaban si las distribuciones de confiabilidad estimada con los dos métodos era estadísticamente iguales al nivel 0.05 a la confiabilidad verdadera. Por último, para los tiempos weibull se multiplicó el parámetro β_2 por un factor igual a 0.3, así como el estimador $\widehat{\beta}_2$ obteniéndose $S(t_{0,3})$ y $\widehat{S}^*(t_{0,3})$ para el caso Weibull $\beta_1 = 0,5$ y $\beta_2 = 2$.

3.1. Comparación entre $S(t)$, $\widehat{S}^*(t)$ y $\widehat{S}(t)_{\text{Cópula}}$

Para cada uno de los escenarios de estudio Lognormales y Weibull se generaron 500 muestras de tamaño 50 y se estimaron las dispersiones de las diferencias entre $S(t_p)$ y $\widehat{S}^*(t_p)$; $S(t_p)$ y $\widehat{S}(t_p)_{\text{Cópula}}$ mediante el ECM_p . En la mayoría de los casos, cuando se observan los valores de ECM_p verticalmente por percentil, estos tienden a ser a más altos a medida que aumenta la dependencia para el estimador $\widehat{S}^*(t_p)$ y las dispersiones se hacen más grandes en algunos casos en los percentiles 0.5 y 0.75. Para el estimador $\widehat{S}(t_p)_{\text{Cópula}}$ las dispersiones con respecto a la función de confiabilidad verdadera son mucho más pequeñas y no se observan crecimientos de estas dispersiones cuando se aumenta la dependencia entre los tiempos de falla. Esto indica, que en general a medida que aumenta la dependencia entre los tiempos de falla Lognormal y Weibull, la dispersión de las diferencias entre $S(t)$ y $\widehat{S}^*(t_p)$ aumenta,

mientras que la dispersión de las diferencias entre $S(t)$ y $\widehat{S}(t_p)_{\text{Copula}}$ no cambian tanto cuando la dependencia aumenta .

ρ	ECM $_{p=0,05}$		ECM $_{p=0,25}$		ECM $_{p=0,5}$		ECM $_{p=0,75}$		ECM $_{p=0,95}$	
	$\widehat{F}^*(t_p)$	$\widehat{F}(t_p)_{\text{Copula}}$	$\widehat{F}^*(t_p)$	$\widehat{F}(t_p)_{\text{Copula}}$	$\widehat{F}^*(t_p)$	$\widehat{F}(t_p)_{\text{Copula}}$	$\widehat{F}^*(t_p)$	$\widehat{F}(t_p)_{\text{Copula}}$	$\widehat{F}^*(t_p)$	$\widehat{F}(t_p)_{\text{Copula}}$
0,0	0.00202	0.0005	0.0243	0.0038	0.0265	0.0062	0.0167	0.0046	0.00082	0.00074
0,2	0.00298	0.0008	0.0321	0.0045	0.0421	0.0057	0.0218	0.0041	0.00099	0.00072
0,4	0.00314	0.0006	0.0368	0.0033	0.0543	0.0061	0.0432	0.0056	0.00143	0.00083
0,6	0.00376	0.0006	0.0461	0.0034	0.1008	0.0039	0.0721	0.0031	0.00314	0.00062
0,8	0.00342	0.0007	0.0498	0.0033	0.1265	0.0049	0.1085	0.0050	0.00397	0.00327
1,0	0.0039	0.0005	0.0564	0.0040	0.1324	0.0073	0.1987	0.0062	0.00971	0.00114

Tab. 1: ECM $_p$ de $\widehat{F}^*(t_p)$ y $\widehat{F}(t_p)_{\text{Copula}}$ para tiempos asociados a dos modos de falla Lognormal con parámetros $\mu_1 = 0$; $\mu_2 = 1$ y $\sigma_1 = 1$; $\sigma_2 = 2$

λ	ECM $_{p=0,05}$		ECM $_{p=0,25}$		ECM $_{p=0,5}$		ECM $_{p=0,75}$		ECM $_{p=0,95}$	
	$\widehat{S}^*(t_p)$	$\widehat{S}(t_p)_{\text{Copula}}$	$\widehat{S}^*(t_p)$	$\widehat{S}(t_p)_{\text{Copula}}$	$\widehat{S}^*(t_p)$	$\widehat{S}(t_p)_{\text{Copula}}$	$\widehat{S}^*(t_p)$	$\widehat{S}(t_p)_{\text{Copula}}$	$\widehat{S}^*(t_p)$	$\widehat{S}(t_p)_{\text{Copula}}$
0,0	0.0031	0.0039	0.034	0.0131	0.0282	0.0219	0.0531	0.0135	0.0214	0.0309
0,2	0.0022	0.0041	0.049	0.0128	0.0292	0.0222	0.0312	0.0142	0.0219	0.0223
0,4	0.0021	0.0032	0.070	0.0098	0.0456	0.0114	0.0345	0.0157	0.0342	0.0346
0,6	0.0051	0.0033	0.105	0.00852	0.0987	0.0104	0.0478	0.0949	0.0987	0.0982
0,8	0.0024	0.0043	0.119	0.0086	0.1232	0.0094	0.1564	0.0965	0.0978	0.0893
0,9	0.0948	0.0044	0.184	0.0082	0.1452	0.0106	0.1908	0.1280	0.1092	0.0987

Tab. 2: ECM $_p$ de $\widehat{S}^*(t_p)$ y $\widehat{S}(t_p)_{\text{Copula}}$ para tiempos asociados a dos modos de falla Weibull con parámetros $\beta_1 = 0,5$ y $\beta_2 = 2$ y $\theta_1 = \theta_2 = 1$

3.2. Pruebas analíticas de bondad de ajuste para $S(t)$ vs. $\widehat{S}^*(t)$ y $S(t)$ vs. $\widehat{S}(t)_{\text{Copula}}$

En esta sección se muestran los p-valores de las prueba de Cramer-Von Mises para la comparación de igualdad de distribución de $S(t)$ vs. $\widehat{S}^*(t)$ y $S(t)$ vs. $\widehat{S}(t)_{\text{Copula}}$. En las tablas 3 y 4 se muestran los p-valor de las comparaciones para los casos de tiempos de falla Lognormal y Weibull respectivamente con diferente grado de dependencia entre los modos de falla para el ajuste de la función de confiabilidad para el tiempo mínimo. Se acepta la hipótesis de igualdad de distribución de confiabilidad estimada mediante $\widehat{S}(t)_{\text{Copula}}$ con la confiabilidad verdadera para todos los casos de dependencia. Mientras que para la distribución de confiabilidad estimada con el estimador $\widehat{S}^*(t)$ solo se acepta la hipótesis de igualdad distribucional cuando hay independencia o dependencia leve.

	$\mu_1 = \mu_2 = 0$ y $\sigma_1 = \sigma_2 = 1$		$\mu_1 = \mu_2 = 0$ y $\sigma_1 = 1; \sigma_2 = 2$		$\mu_1 = 0; \mu_2 = 1$ y $\sigma_1 = \sigma_2 = 1$		$\mu_1 = 0, \mu_2 = 1$ y $\sigma_1 = 2; \sigma_2 = 1$	
ρ	$\hat{S}^*(t_p)$	$\hat{S}(t_p)_{\text{Copula}}$	$\hat{S}^*(t_p)$	$\hat{S}(t_p)_{\text{Copula}}$	$\hat{S}^*(t_p)$	$\hat{S}(t_p)_{\text{Copula}}$	$\hat{S}^*(t_p)$	$\hat{S}(t_p)_{\text{Copula}}$
0,0	0.98039	0.98039	0.15686	0.88235	0.98039	0.98039	0.98039	0.96078
0,2	0.21568	0.23529	0.23529	0.92156	0.45098	0.15686	0.35294	0.72549
0,4	0.00000	0.13725	0.00000	0.96078	0.00000	0.82352	0.01960	0.96078
0,6	0.00000	0.21568	0.00000	0.70588	0.00000	0.82352	0.00000	0.33333
0,8	0.00000	0.96078	0.00000	0.74509	0.00000	0.98039	0.00000	0.94117
1,0	0.00000	0.94117	0.00000	0.07843	0.00000	0.94117	0.00000	0.07921

Tab. 3: P_{valor} para las pruebas analíticas de bondad de ajuste de las distribuciones de $S(t)$ vs. $\hat{S}^*(t_p)$ y $S(t)$ vs. $\hat{S}(t_p)_{\text{Copula}}$ para tiempos asociados a dos modos de falla Lognormal

	$\beta_1 = 0,5$ y $\beta_2 = 2$		$\beta_1 = 0,5$ y $\beta_2 = 1$		$\beta_1 = 2$ y $\beta_2 = 1$		$\beta_1 = \beta_2 = 2$		$\beta_1 = \beta_2 = 0,5$	
λ	$\hat{S}^*(t_p)$	$\hat{S}(t_p)_{\text{Copula}}$	$\hat{S}^*(t_p)$	$\hat{S}(t_p)_{\text{Copula}}$	$\hat{S}^*(t_p)$	$\hat{S}(t_p)_{\text{Copula}}$	$\hat{S}^*(t_p)$	$\hat{S}(t_p)_{\text{Copula}}$	$\hat{S}^*(t_p)$	$\hat{S}(t_p)_{\text{Copula}}$
0,0	0.99502	0.90049	0.99447	0.65193	0.54455	0.61386	0.79734	0.54817	0.99009	0.15841
0,2	0.10447	0.80099	0.01104	0.45856	0.08910	0.17821	0.01328	0.88704	0.04950	0.53465
0,4	0.00497	0.92039	0.00000	0.98342	0.00000	0.30693	0.00000	0.87251	0.00000	0.87128
0,6	0.00000	0.79104	0.00000	0.09452	0.00000	0.59405	0.00000	0.99004	0.00000	0.44554
0,8	0.00000	0.05472	0.00000	0.85074	0.00000	0.45544	0.00000	0.70822	0.00000	0.74257
0,9	0.00000	0.08964	0.00000	0.07960	0.00000	0.23762	0.00000	0.42288	0.00000	0.34653

Tab. 4: P_{valor} para las pruebas analíticas de bondad de ajuste de las distribuciones de $S(t)$ vs. $\hat{S}^*(t_p)$ y $S(t)$ vs. $\hat{S}(t_p)_{\text{Copula}}$ para tiempos asociados a dos modos de falla Weibull

3.3. Comparación de $S(t)$ con $\hat{S}^*(t)$ y $S(t_{0,3})$ con $\hat{S}^*(t_{0,3})$ para los casos de tiempos de falla Weibull

En la tabla 5 se muestran los cambios porcentuales (CP_p) entre $S(t)$ y $\hat{S}^*(t)$ y luego multiplicando el parámetro β_2 por 0.3 obteniéndose $S(t_{0,3})$ y el estimador $\hat{\beta}_2$ por 0.3 obteniéndose $\hat{S}^*(t_{0,3})$ para el caso Weibull $\beta_1 = 0,5$ y $\beta_2 = 2$ y parámetros de dependencia $\lambda = 0,0, 0,2, 0,4, 0,6, 0,8, 0,9$ y $0,98$ para cada uno de los percentiles del tiempo mínimo t_p con $p = 0,05, 0,25, 0,5, 0,75$ y $0,95$

λ	$p=0,05$		$p=0,25$		$p=0,5$		$p=0,75$		$p=0,95$	
	$CP_{p(0,3)}$	CP_p	$CP_{p(0,3)}$	CP_p	$CP_{p(0,3)}$	CP_p	$CP_{p(0,3)}$	CP_p	$CP_{p(0,3)}$	CP_p
0,0	0.4668	0.3255	1.0234	0.9419	1.0822	1.0714	1.12186	1.1097	1.14279	1.2891
0,2	0.9193	0.8953	2.3615	2.1102	7.4906	3.7173	16.3558	15.2779	71.8997	49.3511
0,4	1.3844	1.2676	5.3259	2.8302	15.9938	5.8585	34.1554	31.4991	147.1947	79.6577
0,6	1.8549	1.5949	3.6267	3.5698	20.1146	7.5282	51.7345	49.1091	226.0480	97.6448
0,8	1.9075	1.8547	6.5551	3.7108	22.5856	11.5464	78.7846	74.7928	348.2726	99.9365
0,90	3.0218	2.9878	8.3761	4.0974	31.7949	12.1519	83.7510	81.8147	379.5557	99.9736
0,98	3.8366	3.7381	11.8975	6.0977	35.2286	15.7612	93.4373	82.6682	389.0409	99.9731

Tab. 5: Cambios porcentuales CP_p y $CP_{p(0,3)}$ para $S(t)$ y $S_{0,3}(t)$ para tiempos de falla Weibull con parámetros $\beta_1 = 0,5$; $\beta_2 = 2$ y $\theta_1 = \theta_2 = 1$

A medida que aumenta la dependencia λ el cambio porcentual CP_p para $S(t)$ vs. $\hat{S}^*(t)$, y el cambio porcentual $CP_{p(0,3)}$ para $S(t_{0,3})$ vs. $\hat{S}^*(t_{0,3})$, es, en la gran mayoría de los casos $CP_{p(0,3)}$ mayor que CP_p y estas diferencias se hacen mucho más grandes en dependencia extrema y también en los cuantiles altos. En algunos casos $CP_{p(0,3)}$ es aproximadamente cuatros veces CP_p . Esto nos indica que el estimador $\hat{S}^*(t)$ de Manotas, et al. (2006) no es de máxima verosimilitud y por lo tanto no tiene la propiedad de invarianza.

4. Conclusiones

El uso de las cópulas se destaca como una herramienta útil para la estimación de función de sobrevivencia para el tiempo mínimo de falla para tiempos de falla que provengan de cual cualquier distribución en la práctica.

Los valores de ECM_p que miden la dispersión entre $F(t)$ y $\hat{F}^*(t)$ son más grandes a medida que aumenta la dependencia entre los tiempos de falla para los casos Lognormales y Weibull y también para los cuantiles altos. Los valores de ECM_p que miden la dispersión entre $F(t)$ y $\hat{F}(t)_{Cópula}$ son menores que los obtenidos para $\hat{F}^*(t)$ en la mayoría de los casos.

El estimador de la confiabilidad $\hat{F}(t)_{Cópula}$ captura la estructura de dependencia entre los tiempos asociados a dos modos de falla que se distribuyen Lognormal o Weibull, lo cual se puede extender para cualquier tipo de datos en la práctica, ya que el estimador $\hat{F}(t)_{Cópula}$ no necesita ningún supuesto distribucional de los datos, el supuesto que se hace solo es sobre la dependencia entre los tiempos de falla.

La distribución de confiabilidad estimada mediante el estimador $\hat{F}(t)_{Cópula}$ es estadísticamente igual a la confiabilidad verdadera para todos los casos de tiempos Lognormal y Weibull y no depende del grado de dependencia entre los modos de falla.

En la práctica el estimador $\hat{F}^*(t)$ no es útil ya que no se cuenta con datos, solo se podría aplicar siempre y cuando se tenga la certeza que los datos se distribuyen Weibull o Lognormal y solo se podría estimar la confiabilidad del sistema, pero no sería útil para fines predictivos.

Referencias

- [1] Crowder, M. (2001), Classical Competing Risks. *Chapman Hall/CRC*.
- [2] Escarela, G., Carriere, J., (2003) Fitting competing risks with an assumed copula. *Statistical Methods in Medical Research* **12** (4), 333-349.
- [3] Frees, E. W. & Valdez, E. A., (1998), "Understanding Relationships Using Copulas", *North American Actuarial Journal* 2(1), 1-25
- [4] Lu Jye-Chyi and Bhattacharyya, G., (1993) Some new Constructions of Bivariate Weibull Models. *Ann. Inst. Statist. Math.* **42**(3), 543-559.
- [5] Manotas, E., Yáñez, S., Lopera C.M., Jaramillo, M., (2008) Estudio del Efecto de la Dependencia en la Estimación de la Confiabilidad de un Sistema con dos Modos de Falla Concurrentes. *DYNA*, Medellín, **75**(154), 5-21.
- [6] Meeker, W. y Escobar, L.A., (1998), Statistical Methods for Reliability Data. *John Wiley & Sons*.
- [7] Oakes, D. (1989). Bivariate survival models induced by frailties. *J. Am. Statist. Assoc.* **84**, 487-493
- [8] Peterson, A. V. (1977). Expressing the Kaplan-Meier estimator as a function of empirical subsurvival functions. *J. Am. Statist. Assoc.* **73**, 11-13.
- [9] Tsiatis, A., (1975), A nonidentifiability aspect of the problem of competing risks. *Proc. Natl. Acad. Sci. USA.* **72**, 20-22.
- [10] Yáñez, S., Manotas, E., Lopera, C., (2005). Estudio de la Dependencia en la Estimación de la Confiabilidad de un Sistema con dos Modos de Falla, Memorias del XV Simposio de Estadística, Paipa.
- [11] Zheng, M., Klein, J.P., (1995). Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika* **82** (1), 12-38