

Comparación entre riesgos competitivos vía el estimador cópula-gráfico

Comparison between Competing Risks via the Copula-Graphic Estimator

SERGIO YÁÑEZ^{1,a}, HUGO BRANGO^{2,b}, MARIO C. JARAMILLO^{1,c},
CARLOS M. LOPERA^{1,d}

¹ESCUELA DE ESTADÍSTICA, FACULTAD DE CIENCIAS, UNIVERSIDAD NACIONAL DE COLOMBIA,
MEDELLÍN, COLOMBIA

²DEPARTAMENTO DE MATEMÁTICAS Y ESTADÍSTICA, UNIVERSIDAD DE CÓRDOBA, MONTERÍA,
COLOMBIA

Resumen

En riesgos competitivos, el problema de identificabilidad asociado a la dependencia entre los modos de falla, se puede resolver utilizando el estimador cópula-gráfico que asume la forma de la cópula conocida. En este trabajo se compara, mediante simulación, este estimador para un modelo de riesgos competitivos dependientes con el estimador tradicional que asume independencia. Se muestra que ignorar la dependencia puede causar subestimación significativa de la función de confiabilidad. Se ilustra la temática con un ejemplo y se visualiza a futuro la posible utilidad del estimador cópula-gráfico en confiabilidad industrial para predecir en experimentos acelerados.

Palabras clave: confiabilidad, cópula, identificabilidad, riesgos competitivos.

Abstract

The identifiability problem in competing risks associated to the dependence measure between the failure modes, can be solved by means of the copula-graphic estimator which assumes the copula function known. This paper compares, through a simulation study, the copula-graphic estimator for a dependent competing risks model with the traditional one under the assumption of independence. It is shown that a significant subestimation of the reliability function can be made when the dependence is not taken into account. An example illustrates this point and as a future work it is seen the possibility of using copula-graphic in industrial reliability to predict in accelerated life tests.

Key words: Competing risks, Copula, Identifiability, Reliability.

^aProfesor asociado. E-mail: syanez@unal.edu.co

^bProfesor asistente. E-mail: habrango@hotmail.com

^cProfesor asociado. E-mail: mcjarami@unal.edu.co

^dProfesor asistente. E-mail: cmlopera@unal.edu.co

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 & 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 competitivos (“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 de manera que censura así a los otros modos de falla.

Muchas son las aplicaciones de riesgos competitivos en confiabilidad. Basu & Klein (1982) obtuvieron algunos resultados en la teoría de riesgos competitivos; Nelson (1990) presenta datos para el tiempo de falla en horas de calentadores industriales, los cuales tienen dos modos de falla; Bedford & Lindqvist (2004) trataron el problema de la identificabilidad en sistemas reparables cuando se presentan los riesgos competitivos; Bedford (2005) muestra como modelar la confiabilidad en presencia de riesgos competitivos; Manotas, Yáñez, Lopera & Jaramillo (2008), estudiaron el efecto en la estimación de la confiabilidad, cuando se asume el supuesto de independencia entre los tiempos de falla que compiten, que realmente son dependientes; Meeker, Escobar & Hong (2009) usaron pruebas de vida aceleradas para predecir la distribución del tiempo de falla de un nuevo producto con dos modos de falla.

En 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 (Tsiatis 1975). Inclusive en el caso paramétrico, los datos pueden contener poca información sobre el coeficiente de asociación entre las variables y es necesario hacer algunos supuestos al respecto (Meeker et al. 2009). Para el caso no-paramétrico el estimador cópula-gráfico de Zheng & Klein (1995) resuelve el problema bajo el supuesto del conocimiento de la forma funcional de la cópula que da cuenta de la estructura de asociación entre los riesgos competitivos. Es de anotar que bajo independencia el problema de identificabilidad no existe.

Yáñez, Manotas & Lopera (2006) y Manotas et al. (2008), estudiaron, para el caso Weibull y lognormal, el efecto en la estimación de la sobrevivencia cuando se asume el supuesto de independencia entre tiempos de falla que realmente son dependientes. Encontraron que bajo el supuesto de independencia, que es la metodología tradicional utilizada en muchos campos de la ciencia (Pintilie 2006), no genera errores significativos en la estimación de la función de confiabilidad del tiempo mínimo. Puesto que la literatura reciente reporta que el supuesto de independencia en riesgos competitivos generalmente da una visión pesimista del comportamiento del sistema, ellos comentan la importancia de estudiar en más detalle la temática.

En este trabajo se estima la función de confiabilidad bajo dependencia mediante el estimador cópula-gráfico propuesto en Zheng & Klein (1995). Se hace un estudio de simulación para comparar este estimador que permite incorporar la dependencia con el que considera independencia entre los modos de falla.

En la sección 2 se presentan los conceptos básicos de riesgos competitivos y se detalla el ejemplo que se usará como ilustración. El estimador cópula-gráfico se discute en la sección 3, con una ilustración que determina la cópula a usar en este artículo. En la sección 4 se presenta el esquema de simulación utilizado. El análisis de resultados se hace en la sección 5. Finalmente, en la sección 6 se presentan las conclusiones y posible trabajo futuro, donde se muestra que ignorar la dependencia puede causar una subestimación representativa de la función de confiabilidad.

2. Riesgos competitivos

En este artículo se trabaja con dos modos de falla. El tiempo de falla de un sistema con dos modos de falla puede ser modelado como un sistema en serie o un modelo de riesgos competitivos, como se ilustra en la figura 1. Cada unidad tiene un tiempo potencial de falla. El tiempo de falla observado es el mínimo de esos tiempos potenciales individuales.



FIGURA 1: Sistema con dos componentes en serie.

Así por ejemplo, para un sistema con dos modos de falla, sean X y Y los respectivos tiempos potenciales, entonces lo que se observa en la práctica son parejas de la forma (T, δ) , donde $T = \min(X, Y)$ y δ es una variable discreta que toma valores de 0 cuando se presenta una censura a derecha, 1 cuando falla debido al primer modo de falla y 2 cuando falla debido al segundo modo de falla.

Para estudiar este modelo de riesgos competitivos, es necesario presentar algunas funciones básicas. La función de subdistribución o función de incidencia acumulada (FIA), para el modo de falla i , $i = 1, 2$, está dada por:

$$G_i(t) = P(T \leq t, \delta = i) \quad (1)$$

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

$$F(t) = P(T \leq t) = \sum_{i=1}^2 P(T \leq t, \delta = i) = \sum_{i=1}^2 G_i(t) \quad (2)$$

Nótese que cuando no hay riesgos competitivos la función de distribución total se encuentra en el intervalo $[0, 1]$. En contraste, en presencia de riesgos competitivos la FIA sólo puede tomar valores hasta $P(\delta = i)$ debido a que

$$\lim_{t \rightarrow \infty} G_i(t) = P(\delta = i) \quad (3)$$

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

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

$$\overline{G}_i(t) = P(T > t, \delta = i) \quad (4)$$

La función de subdensidad para el modo de falla i y T continuo, se define como

$$g_i(t) = \frac{dG_i(t)}{dt} = -\frac{d\overline{G}_i(t)}{dt} \quad (5)$$

2.1. El problema de identificabilidad

La aproximación tradicional para especificar el modelo de riesgos competitivos es vía tiempos de falla latentes. En este contexto surge el denominado problema de identificabilidad. Para el caso bivariado tenemos dos tiempos de falla potenciales X, Y asociados a dos modos de falla, de manera que si ocurre el primero, el segundo no puede ser observado. En este sentido, se dice que los tiempos son latentes. Se supone que la pareja (X, Y) tiene una distribución multivariada de la forma $F(x, y) = P(X \leq x, Y \leq y)$. Así $T = \min(X, Y)$ determina el tiempo de falla del sistema completo. De manera que una vez que el sistema haya fallado, el otro tiempo de falla no puede ser observado. Las funciones de distribución marginales para los dos modos de falla son $F_1(t) = P(X \leq t)$ y $F_2(t) = P(Y \leq t)$, respectivamente.

El problema de identificabilidad (Tsiatis 1975) establece que para el caso de modelos de tiempos de falla latentes, dada una función de distribución conjunta con dependencia arbitraria entre las componentes, existe siempre una función de distribución conjunta en la cual las variables son independientes, y la cual reproduce las subdensidades $g_i(t)$ de manera precisa. Así, uno no puede conocer a partir solamente de los datos observados (T, δ) , cuál de los dos modelos es correcto, puesto que ambos se ajustan muy bien a los datos. Crowder (2001) da detalles técnicos sobre este problema que él denomina “El impase de Cox-Tsiatis”. En resumen, podemos decir que en la práctica se pueden encontrar las subdistribuciones marginales, asociados con datos de riesgos competitivos, y ese conjunto es consistente con un número infinito de distribuciones conjuntas de tiempos de falla potenciales. Esto quiere decir que es imposible identificar las distribuciones marginales a partir solamente de los datos de riesgos competitivos.

Supongamos que se tienen datos de la forma (T, δ) y es necesario identificar el modelo conjunto $F(x, y)$. Algunas posibles soluciones son: i) Asuma independencia, lo cual es frecuentemente no realista (ver subsección 2.2.1); ii) Asuma una

estructura de dependencia conocida usando una cópula específica para la distribución conjunta de los modos de falla; Zheng & Klein (1995) proponen el llamado estimador cópula-gráfico, el cual es un estimador máximo verosímil y trabaja muy bien independiente de la cópula escogida (ver sección 3); iii) Especifique un modelo paramétrico. En este último caso, a pesar del fuerte supuesto, los datos de confiabilidad pueden contener poca información acerca de la asociación entre riesgos competitivos debido, a veces, a la alta censura y a la cantidad limitada de datos. Meeker et al. (2009) muestran un caso de este tipo donde el problema de identificabilidad prevalece.

En este artículo se van a comparar las soluciones i) y ii). También, como trabajo futuro se propone en la sección 6 incorporar los resultados de este artículo a la solución iii).

2.2. Metodología tradicional: un ejemplo en ingeniería

Se ilustra con un ejemplo de Meeker & Escobar (1998), la metodología tradicional en el análisis de modelos de riesgos competitivos, que supone independencia entre los dos modos de falla.

Se tienen tiempos de falla de un cierto tipo de equipos, los cuales se deben principalmente a dos modos o causas de falla, una asociada a fallas por descargas por acumulación de daños producidos por altos picos de voltaje durante tormentas eléctricas (se denota S), y la otra falla debida al desgaste normal de los equipos (se denota W). La tabla 1 presenta los tiempos y los modos de falla de 30 unidades que fueron instaladas en un ambiente de servicio estándar. Los tiempos de falla asociados al modo de falla S ocurren de forma temprana en la vida de los equipos. El modo de falla W empieza a aparecer después de los 100 kilociclos de uso.

TABLA 1: Tiempos de falla y modos de falla para los datos del ejemplo.

Kilociclos	Modo de falla	Kilociclos	Modo de falla	Kilociclos	Modo de falla
275	W	106	S	88	S
13	S	300	–	247	S
147	W	300	–	28	S
23	S	212	W	143	S
181	W	300	–	300	–
30	S	300	–	23	S
65	S	300	–	300	–
10	S	2	S	80	S
300	–	261	S	245	W
173	S	293	W	266	W

En la figura 2 se muestran los estimadores de Kaplan Meier, para cada modo de falla, considerando el otro modo de falla como censura. Se debe aclarar que cuando hay dependencia entre los modos de falla, los estimadores de Kaplan Meier no son estimadores consistentes de las distribuciones marginales.

En la figura 3 se muestra el estimador de Kaplan Meier, para el tiempo mínimo de los dos modos de falla.

Este ejemplo se usa más adelante en la subsección 5.4, bajo el supuesto de dependencia entre los dos modos de falla.

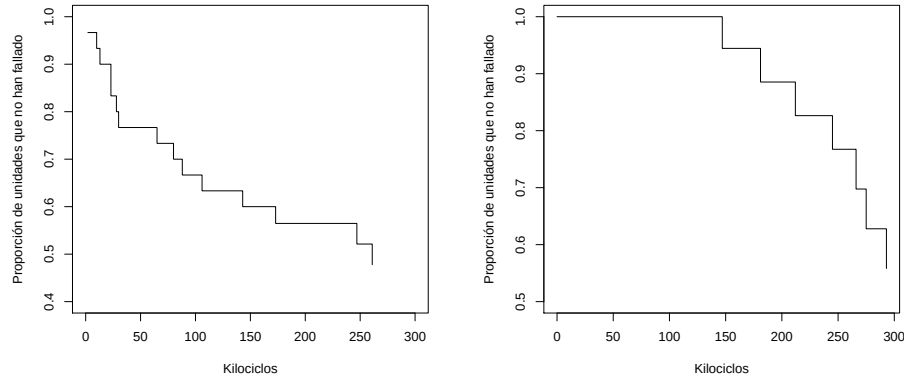


FIGURA 2: Estimación de la función de sobrevivencia de cada modo de falla.

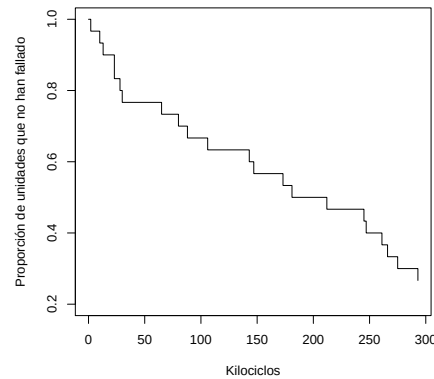


FIGURA 3: Estimación de la función de sobrevivencia del tiempo mínimo ignorando los modos de falla.

2.2.1. Estimador bajo independencia

Una estructura en serie con s componentes funciona, si y sólo si todas las componentes funcionan. Para un sistema en serie con dos componentes independientes en serie, la función de distribución es

$$\begin{aligned}
 F(t) &= P(T \leq t) = 1 - P(T > t) \\
 &= 1 - P(X > t, Y > t) \\
 &= 1 - P(X > t)P(Y > t) \\
 &= 1 - [1 - F_1(t)][1 - F_2(t)] = F_1(t) + F_2(t) - F_1(t)F_2(t)
 \end{aligned} \tag{6}$$

donde (X, Y) son los tiempos de vida asociados a las dos componentes y $F_i, i = 1, 2$ son, respectivamente, las funciones de distribución para las componentes, y $T = \min(X, Y)$ es el tiempo de falla observado del sistema.

La función de confiabilidad $S(t) = 1 - F(t)$ para el caso de independencia es

$$\begin{aligned} S(t) &= P(T > t) = P(X > t, Y > t) \\ &= S_1(t)S_2(t) \end{aligned} \quad (7)$$

donde $S_i(t), i = 1, 2$ son las funciones de confiabilidad para las dos componentes.

Para un sistema con dos componentes en serie y tiempos de falla dependientes,

$$\begin{aligned} F(t) &= P(T \leq t) = 1 - P(T > t) \\ &= 1 - P(X > t, Y > t) \end{aligned} \quad (8)$$

En este caso, la evaluación debe hacerse con respecto a la distribución conjunta de X y Y , que incluye los parámetros de dependencia.

La función de confiabilidad para este caso es

$$S(t) = P(T > t) = P(X > t, Y > t) \quad (9)$$

Para dos modos de falla en riesgos competitivos, se sabe que (Barlow & Proschan 1975)

$$\begin{aligned} S(t) &= P(X > t, Y > t) \geq P(X > t)P(Y > t) \\ &= S^*(t) = S_1(t)S_2(t) \end{aligned} \quad (10)$$

donde $S^*(t)$ es la función de confiabilidad asumiendo independencia. Por lo tanto $S^*(t)$ es pesimista en relación con $S(t)$, esto es para un mismo tiempo t , la probabilidad de sobrevivencia bajo el supuesto de independencia está siempre por debajo de la probabilidad conjunta que incluye la dependencia.

El estimador de $S^*(t)$ se denotará $\hat{S}^*(t)$, que es un estimador de la confiabilidad del tiempo mínimo de falla de un sistema con dos modos de falla que compiten. Éste se obtiene como el producto de las funciones de confiabilidad 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, así:

$$\hat{S}^*(t) = \hat{S}_1(t)\hat{S}_2(t) \quad (11)$$

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

3. Estimador cópula-gráfico

Zheng & Klein (1995) proponen un estimador no paramétrico para las distribuciones marginales, al cual denominan estimador cópula-gráfico, como una solución al problema de identificabilidad, descrito en la subsección 2.1. Para ello utilizan

cóputas como una función no paramétrica que captura la dependencia entre dos variables aleatorias. Bajo el supuesto que la cópula entre X y Y con su parámetro de asociación conocido y con los datos de riesgos competitivos disponibles, se resuelve el problema de identificabilidad.

Siguiendo a Zheng & Klein (1995), suponga que se tiene, en el contexto de riesgos competitivos, una muestra de tamaño n de tiempos de vida $T_j = \min(X_j, Y_j)$. Si $F_1(t)$ y $F_2(t)$ son las distribuciones marginales de X y Y , respectivamente, entonces, para cualquier cópula, tenemos

$$\mu_c(A_t) = P(X > t, Y > t) \quad (12)$$

donde

$$\mu_c(A_t) = \int_{I^2} \mathbf{1}_{E_t} \{F_1^{-1}(x), F_2^{-1}(y)\} dC(x, y)$$

con $E_t = \{(x, y) : x > t, y > t\}$ y $A_t = \{(x, y) : F_1(t) < x \leq 1, F_2(t) < y \leq 1\}$

$$\mu_c(B_t) = P(X \leq t, X < Y) \quad (13)$$

donde

$$\mu_c(B_t) = \int_{I^2} \mathbf{1}_{D_t} \{F_1^{-1}(x), F_2^{-1}(y)\} dC(x, y)$$

con $D_t = \{(x, y) : x < y, 0 < x < t\}$ y $B_t = \{(x, y) : 0 < x \leq F_1(t), F_2 F_1^{-1}(x) < y \leq 1\}$.

Estas dos relaciones determinan de manera única a $F_1(t)$ y $F_2(t)$. Por tanto, se pueden encontrar estimadores $\hat{F}_1(t)$ y $\hat{F}_2(t)$ para $F_1(t)$ y $F_2(t)$ respectivamente, que preservan estas relaciones sobre una rejilla seleccionada de m puntos $0 < t_1 < t_2 < \dots < t_m < \max\{T_j, j = 1, \dots, n\}$.

Para construir el estimador cópula-gráfico, sean

$$\begin{aligned} \hat{A}_t &= \{(x, y) : \hat{F}_1(t) < x \leq 1, \hat{F}_2(t) < y \leq 1\} \\ \hat{B}_t &= \{(x, y) : 0 < x \leq \hat{F}_1(t), \hat{F}_2 \hat{F}_1^{-1}(x) < y \leq 1\} \end{aligned}$$

y

$$\begin{aligned} \text{est. } P(X > t, Y > t) &= \frac{1}{n} \sum \mathbf{1}(T_j > t) \\ \text{est. } P(X \leq t, X < Y) &= \frac{1}{n} \sum \mathbf{1}(T_j \leq t, \delta = 1) \end{aligned}$$

las estimaciones empíricas de $P(X > t, Y > t)$ y $P(X \leq t, X < Y)$ respectivamente.

$\hat{F}_1(t_i)$ y $\hat{F}_2(t_i)$ son las soluciones simultáneas de (14) y (15), siendo $\hat{F}_1$ y $\hat{F}_2$ líneas rectas en cada intervalo (t_i, t_{i+1}) .

$$\mu_c(\hat{A}_{t_i}) - \text{est. } P(X > t_i, Y > t_i) = 0 \quad (14)$$

$$\mu_c(\hat{B}_{t_i}) - \text{est. } P(X \leq t_i, X < Y) = 0 \quad (15)$$

El estimador cópula-gráfico es de máxima verosimilitud y además bajo condiciones de independencia, coincide con el estimador de Kaplan Meier (Zheng & Klein 1995).

El algoritmo cópula-gráfico para la construcción de $\hat{F}_1$ y $\hat{F}_2$, basado en un algoritmo de bisección para encontrar raíces, se puede ver en detalle en Zheng & Klein (1995).

3.1. Ilustración

Se estimará $S(t)$ asumiendo la cópula Gama, la cual se muestra a continuación

$$C(F_1(x), F_2(y)) = F_1(x) + F_2(y) - 1 + \left\{ \left(\frac{1}{1 - F_1(x)} \right)^{\alpha-1} + \left(\frac{1}{1 - F_2(y)} \right)^{\alpha-1} - 1 \right\}^{-\frac{1}{\alpha-1}}, \quad \alpha > 1$$

donde $F_1(t)$ y $F_2(t)$ son las funciones de distribución marginales para los tiempos de falla X y Y , respectivamente, α es el parámetro de dependencia de la cópula que mide el grado de asociación entre los tiempos de falla X y Y . Este parámetro se relaciona con el coeficiente de correlación τ de Kendall de la siguiente forma: $\tau = \frac{1}{\alpha-1}$. Para estimar $F_1(t)$ y $F_2(t)$, se aplica el algoritmo cópula-gráfico.

A partir de la ecuación de la cópula, se puede encontrar la función de confiabilidad,

$$S(t) = \left\{ \left(\frac{1}{1 - F_1(t)} \right)^{\alpha-1} + \left(\frac{1}{1 - F_2(t)} \right)^{\alpha-1} - 1 \right\}^{-\frac{1}{\alpha-1}} \quad (16)$$

Remplazando los estimadores $\hat{F}_1(t)$ y $\hat{F}_2(t)$ obtenidos del algoritmo cópula-gráfico se obtiene $\hat{S}(t)_{\text{cópula}}$. Este último estimador es el que se utilizará en el estudio comparativo de la siguiente sección, puesto que la cópula Gama permite variables con rango $(0, \infty)$ como es el caso de la Weibull que es de amplio uso en confiabilidad. Además, Zheng & Klein (1995) muestran que este estimador es razonablemente robusto a la escogencia de la cópula.

La programación del algoritmo cópula-gráfico se realizó usando el paquete estadístico R (R Development Core Team 2010). El código utilizado se suministra bajo pedido a los autores.

4. Estudio comparativo entre $\hat{S}^*(t)$ y $\hat{S}(t)_{\text{cópula}}$

Yáñez et al. (2006) y Manotas et al. (2008) estudiaron, para el caso Weibull y lognormal, el efecto en la estimación de la confiabilidad $S(t)$, cuando se asume el supuesto de independencia entre los tiempos de falla que compiten, que realmente son dependientes. Ellos encontraron que el estimador $\hat{S}^*(t)$ bajo el supuesto de independencia, que es la metodología tradicional, no genera errores significativos en la estimación de la función de confiabilidad del tiempo mínimo. Puesto que la

literatura reciente reporta que el supuesto de independencia en riesgos competitivos generalmente da una visión pesimista del comportamiento del sistema, ellos comentan la importancia de estudiar en más detalle las propiedades del estimador $\hat{S}^*(t)$.

En este trabajo se estima $S(t)$ bajo dependencia mediante el estimador cópula-gráfico, denotado por $\hat{S}(t)_{\text{cópula}}$ propuesto en Zheng & Klein (1995). Se hace un estudio de simulación para comparar este estimador con el estimador de la confiabilidad $\hat{S}^*(t)$.

4.1. Esquema de simulación

Puesto que el estimador cópula-gráfico estima las funciones de distribución marginales, haremos el estudio comparativo entre las funciones de distribución, lo cual es equivalente a trabajar con las funciones de sobrevivencia.

En el estudio de simulación se utiliza la distribución Weibull bivariada para dos variables X y Y , la cual se puede visualizar en varios contextos, tales como los tiempos hasta la primera y segunda falla de un equipo reparable, los tiempos de falla “breakdown” de generadores duales en una planta de energía, o los tiempos de sobrevivencia en un sistema de dos órganos tal como los pulmones o riñones, en el cuerpo humano (Lu & Bhattacharyya 1990).

Para simular los tiempos de falla Weibull bivariados se usó un algoritmo cópula (Frees & Valdez 1998, Jaramillo, Lopera, Manotas & Yáñez 2008). La función de confiabilidad conjunta de la Weibull bivariada utilizada es (Lu & Bhattacharyya 1990)

$$S(x, y) = \exp \left\{ - \left[\left(\frac{x}{\theta_1} \right)^{\frac{\beta_1}{1-\lambda}} + \left(\frac{y}{\theta_2} \right)^{\frac{\beta_2}{1-\lambda}} \right]^{1-\lambda} \right\} \quad (17)$$

donde $x > 0$ y $y > 0$ son los tiempos de falla Weibull con funciones de distribución $F_1(x)$ y $F_2(y)$, $\beta_1 > 0$ y $\beta_2 > 0$ son los parámetros de forma, $\theta_1 > 0$ y $\theta_2 > 0$ son los parámetros de escala asociados a X y Y respectivamente, y $0 \leq \lambda < 1$ es el parámetro de dependencia entre X y Y . Cuando el parámetro de dependencia λ es 0, entonces hay independencia entre X y Y . A medida que λ aumenta, la dependencia entre X y Y aumenta.

Se consideran 24 escenarios con tiempos de falla Weibull, cada uno de los cuales corresponde a la llamada función de confiabilidad del tiempo mínimo de falla del sistema $S(t)$, donde $T = \min(X, Y)$. 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 de los parámetros de escala (Lu & Bhattacharyya 1990). 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 creciente. 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 versus 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 versus 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 versus 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 diferentes. Modo de falla con tasa de falla creciente versus 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 24 escenarios distintos para la distribución Weibull bivariada.

Para efectos de comparación se obtienen los intervalos de confianza empíricos basados en simulación montecarlo para $S(t)$, usando los estimadores $\hat{S}^*(t)$ y $\hat{S}(t)_{\text{cópula}}$. El procedimiento se resume como sigue:

- Para cada uno de los escenarios descritos, se generaron 1000 muestras de tamaño $n = 50$, de tiempos bivariados Weibull. Se elige este tamaño muestral, ya que en general en experimentos de confiabilidad se manejan bases de datos no muy grandes, estos estudios son costosos y los datos de falla son difíciles de obtener.
- Luego, en cada escenario se obtienen 1000 estimaciones de $S(t_p)$ usando ambos estimadores $\hat{S}^*(t_p)$ y $\hat{S}(t_p)_{\text{cópula}}$ en cada percentil t_p con $p = 0.05, 0.25, 0.50, 0.75, 0.95$. De esta forma se obtiene la distribución empírica de las estimaciones de $S(t_p)$ en cada uno de los percentiles considerados.
- Finalmente, de estas distribuciones se obtienen los límites puntuales de confianza empíricos aproximados del 95 %.

5. Análisis de resultados

En aplicaciones de confiabilidad generalmente se tienen pocas fallas y muchos datos censurados. Bajo estas características, las pruebas formales de bondad de ajuste no son útiles (Meeker et al. 2009). En lugar de ello se realizaron pruebas de hipótesis empíricas en cada percentil, utilizando como región de aceptación los intervalos de confianza empíricos, descritos en la sección anterior.

5.1. Intervalos de confianza para $S(t)$

Note que en la generación de la Weibull bivariada, el parámetro λ es una medida de dependencia entre los tiempos de falla.

La figura 4 muestra los límites de confianza aproximados para los casos Weibull en que los parámetros de forma son $\beta_1 = 0.5$; $\beta_2 = 2.0$, y parámetros de dependencia $\lambda = 0.6$ y 0.9 .

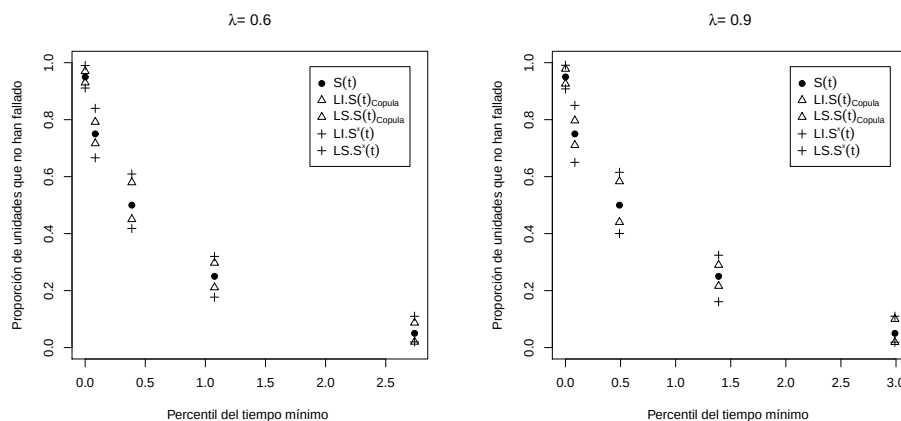


FIGURA 4: Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Weibull $\beta_1 = 0.5$ y $\beta_2 = 2$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95 para $\lambda = 0.6$ y 0.9 .

En estas figuras, en cada percentil t_p se tiene el valor verdadero de $S(t_p)$, los límites inferior y superior de confianza aproximados para $S(t_p)$ calculados con $S^*(t_p)$ (denotados como $LI.S^*(t_p)$ y $LS.S^*(t_p)$, respectivamente), y calculados con $S(t)_{\text{cópula}}$ (denotados $LI.S(t)_{\text{cópula}}$ y $LS.S(t)_{\text{cópula}}$, respectivamente). Note que las escalas en el eje de las abscisas son distintas.

Observe que los intervalos para $S(t)$ basados en $\hat{S}^*(t_p)$ son más amplios que aquellos basados en $\hat{S}(t_p)_{\text{cópula}}$, indicando que el estimador $\hat{S}(t)_{\text{cópula}}$ tiene mayor precisión a la hora de estimar $S(t)$. La amplitud de estos intervalos se hace más evidente a medida que aumenta la dependencia λ . Al ser tan amplios los límites de confianza, la estimación de $S(t)$ con $\hat{S}^*(t)$ puede tomar diferentes comportamientos, sobreestimando o subestimando la confiabilidad verdadera.

5.2. Estimación marginal

Se compara la función de distribución marginal para el tiempo X , $S_1(t)$, con las funciones de distribución marginales estimadas, $\hat{S}_1(t)_{\text{cópula}}$ y $\hat{S}_1(t)$.

La figura 5 muestra los intervalos de confianza puntuales empíricos de la distribución marginal $S_1(t)$, basados en el método de simulación montecarlo, para tiempos de falla bivariados Weibull con parámetros de forma $\beta_1 = 0.5$; $\beta_2 = 2$ y

dependencia $\lambda = 0.6$ estimados mediante el algoritmo cópula-gráfico (a la izquierda) y mediante la distribución “marginal” estimada $S_1(t)$, que asume independencia entre los tiempos de falla X y Y (a la derecha).

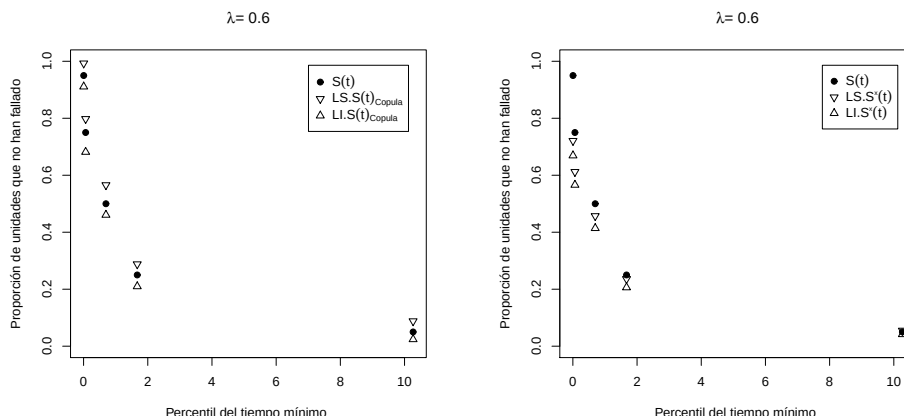


FIGURA 5: Intervalos de confianza empíricos aproximados del 95 % para la distribución marginal $S_1(t)$ en el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 2.0$ y $\lambda = 0.6$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95 .

Se observa como el estimador que emplea el algoritmo cópula-gráfico estima la marginal ya que $S_1(t_p)$ cae dentro de los límites del intervalo empírico, mientras que el otro estimador realmente no estima marginales puesto que los límites del intervalo construido no contiene la distribución marginal $S_1(t)$ como era de esperarse, ya que el algoritmo cópula-gráfico resuelve el problema de identificabilidad al asumir una cópula y la dependencia.

5.3. Predicción

En problemas de ingeniería donde se estudia la confiabilidad de algún sistema, es de interés evaluar el efecto sobre la sobrevivencia debido a un cambio en una marginal. En esta sección se realizan cambios en uno de los parámetros de forma de la distribución Weibull y el mismo cambio se le hace al estimador de ese parámetro en $\hat{S}^*(t)$, y se estima el cambio porcentual con relación a la confiabilidad verdadera en los percentiles $p = 0.05, 0.25, 0.50, 0.75, 0.95$.

Meeker et al. (2009) señalan: “Los ingenieros de diseño tienen herramientas que les permiten obtener predicciones razonables acerca de los efectos de un simple cambio geométrico o cambios de tamaño que una pieza de un producto mecánico tendrá sobre la confiabilidad”, y esto se puede traducir en cambios en los parámetros de las marginales.

Se realizó un estudio de simulación para mostrar el efecto de un cambio en el parámetro β_2 de la distribución Weibull para predecir la nueva confiabilidad de acuerdo a dicho cambio. El esquema de simulación se diseñó como una réplica

ca de lo realizado por Meeker et al. (2009). La función de confiabilidad conjunta de la Weibull $S(t)$ depende de los parámetros $(\beta_1, \beta_2, \lambda)$ (con $\theta_1 = \theta_2 = 1$, de acuerdo a lo fijado en la subsección 4.1) y se realiza un cambio en β_2 multiplicándolo por 0.3, obteniendo una función de confiabilidad nueva que denotamos por $S(t)_{\text{nuevo}} = S(t; \beta_1, 0.3\beta_2, \lambda)$. Ahora bien, el estimador bajo independencia nuevo (i.e. con el cambio en $0.3\beta_2$) se obtiene de $\hat{S}^*(t)_{\text{nuevo}} = \hat{S}_1(t)\hat{S}_2(t)_{\text{nuevo}}(t; \hat{\theta}_2, 0.3\hat{\beta}_2)$, donde $\hat{S}_2(t)_{\text{nuevo}}$ se obtiene al reemplazar el estimador $\hat{\beta}_2$ por $0.3\hat{\beta}_2$, donde $\hat{\beta}_2$ es obtenido de los datos de la distribución marginal 2 y el estimador $\hat{\theta}_2$ se deja igual; como se describió en la subsección 2.2.1. Para comparar se obtuvieron los cambios porcentuales (CP_p) en los percentiles t_p , así

$$CP_p = \frac{|S(t_p) - \hat{S}^*(t_p)|}{S(t_p)} \times 100 \quad (18)$$

que corresponde al cambio relativo entre el valor verdadero $S(t)$ y el valor estimado $\hat{S}^*(t)$ antes de hacer cambios en el parámetro β_2 , lo que se interpreta como el sesgo relativo del estimador $\hat{S}^*(t)$. El cambio porcentual nuevo $CP_p(\text{nuevo})$ en los percentiles t_p se obtiene, así

$$CP_p(\text{nuevo}) = \frac{|S(t_p)_{\text{nuevo}} - \hat{S}^*(t_p)_{\text{nuevo}}|}{S(t_p)_{\text{nuevo}}} \times 100. \quad (19)$$

En la tabla 2 se muestran los cambios porcentuales CP_p y $CP_p(\text{nuevo})$.

TABLA 2: Cambios porcentuales CP_p y $CP_p(\text{nuevo})$ para $S(t)$ y $S(t)_{\text{nuevo}}$ en tiempos de falla Weibull con parámetros $\beta_1 = 0.5$; $\beta_2 = 2$ y $\theta_1 = \theta_2 = 1$.

λ	$p = 0.05$		$p = 0.25$		$p = 0.5$	
	$CP_p(\text{nuevo})$	CP_p	$CP_p(\text{nuevo})$	CP_p	$CP_p(\text{nuevo})$	CP_p
0.0	1.14279	1.2891	1.12186	1.1097	1.0822	1.0714
0.2	71.8997	49.3511	16.3558	15.2779	7.4906	3.7173
0.4	147.1947	79.6577	34.1554	31.4991	15.9938	5.8585
0.6	226.0480	97.6448	51.7345	49.1091	20.1146	7.5282
0.8	348.2726	99.9365	78.7846	74.7928	22.5856	11.5464
0.9	379.5557	99.9736	83.7510	81.8147	31.7949	12.1519
0.98	389.0409	99.9731	93.4373	82.6682	35.2286	15.7612

λ	$p = 0.75$		$p = 0.95$	
	$CP_p(\text{nuevo})$	CP_p	$CP_p(\text{nuevo})$	CP_p
0.0	0.4668	0.3255	1.0234	0.9419
0.2	0.9193	0.8953	2.3615	2.1102
0.4	1.3844	1.2676	5.3259	2.8302
0.6	1.8549	1.5949	3.6267	3.5698
0.8	1.9075	1.8547	6.5551	3.7108
0.9	3.0218	2.9878	8.3761	4.0974
0.98	3.8366	3.7381	11.8975	6.0977

Analizando por columna para cada percentil, se observa que a medida que aumenta la dependencia λ entre los tiempos de falla Weibull, el cambio porcentual

CP_p es siempre menor que el cambio porcentual $CP_p(\text{nuevo})$. En el percentil $t_{0.05}$ la diferencia entre los dos llega a ser hasta de casi cuatro veces. De lo cual se puede decir que en la práctica el estimador $\hat{S}^*(t)_{\text{nuevo}}$, que es el estimador de la confiabilidad bajo independencia, no logra captar correctamente el cambio en el parámetro β_2 , luego para efectos de predicción en este contexto es muy mal estimador. En esta situación lo que ocurre es que $\hat{S}_2(t)_{\text{nuevo}}$ se calcula cambiando $\hat{\beta}_2$ obtenido de los datos simulados para la marginal del modo de falla 2, por $0.3\hat{\beta}_2$, pues al no tener datos nuevos, se tienen una estimación inapropiada de $S_2(t)_{\text{nuevo}}$. Por ello afirmamos que $\hat{S}^*(t)_{\text{nuevo}}$ no es útil para predecir cuando hay cambios en los parámetros.

Obsérvese además, que el estimador $\hat{S}(t)_{\text{cópula}}$ no sirve para fines de predicción, ya que este estimador es no paramétrico y al hacer cambios en los parámetros de las marginales no hay forma de incluir dicho cambio en el estimador. Pero se puede pensar que con las estimaciones de las distribuciones marginales obtenidas con el estimador cópula-gráfico se puede hacer un modelamiento paramétrico, como paso previo al proceso de predicción de la confiabilidad bajo cambios en los parámetros, lo cual mejoraría el análisis exploratorio, ya que se utilizan estimaciones adecuadas de las distribuciones marginales en lugar de las estimaciones basadas en asumir independencia. Estos ajustes paramétricos permitirían hacer predicciones en experimentos acelerados como los descritos en Meeker et al. (2009).

5.4. Ejemplo utilizando los datos de la subsección 2.2.

En la figura 6, las distribuciones marginales bajo dependencia son diferentes a las obtenidas bajo independencia.

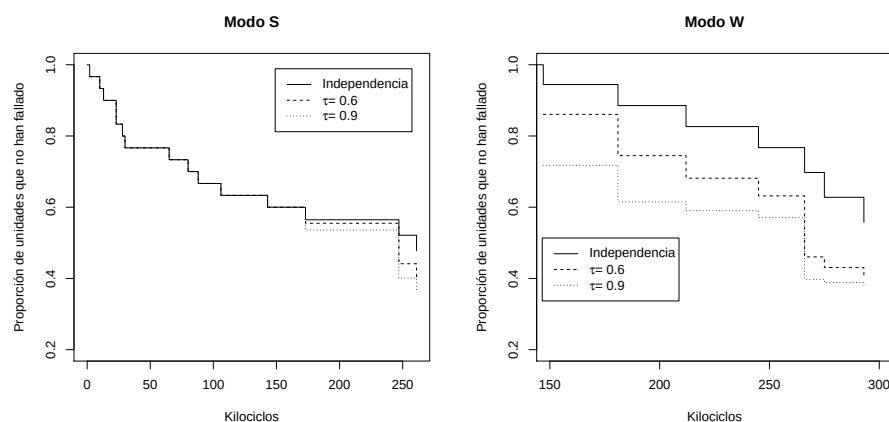


FIGURA 6: Estimación de las funciones de distribución marginales con independencia y con $\tau = 0.6$ y 0.9 .

De ahí, como se muestra en la figura 7 las funciones de distribución acumuladas del sistema (tiempo mínimo) con independencia generan una sobrestimación de las

funciones de distribución, lo cual tiene graves implicaciones en la determinación, por ejemplo, del esquema de mantenimiento preventivo de los equipos, con las consecuencias económicas que tales equivocaciones generarían.

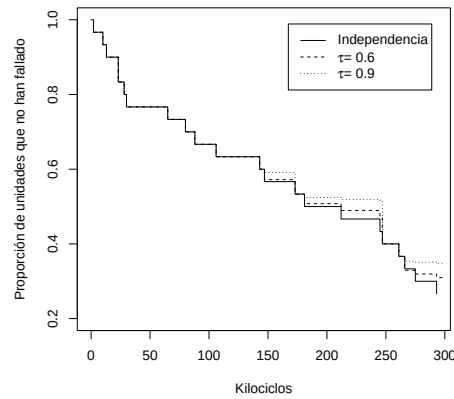


FIGURA 7: Estimación de la función de distribución del tiempo mínimo considerando independencia y con $\tau = 0.6$ y 0.9 .

6. Conclusiones y trabajo futuro

El estimador $\hat{S}(t)_{\text{cópula}}$, es mucho mejor estimador que $\hat{S}^*(t)$. Por ello recomendamos para trabajos de confiabilidad en ingeniería, utilizar el estimador cópula-gráfico, cuando no se tenga certeza sobre el supuesto de independencia, ya que tal supuesto puede generar sobrecostos y mala planeación en diseño, por ejemplo, en mantenimiento de equipos.

Para efectos de predicción en problemas de ingeniería, que implican cambios en los parámetros de las marginales para mejorar la confiabilidad conjunta, ninguno de los dos estimadores funciona. Este sería un tópico de importancia para trabajo futuro tratando de adecuar el estimador cópula-gráfico para tales efectos, y comparar con resultados como los reportados por Meeker et al. (2009). Se piensa que con las distribuciones marginales estimadas se pueden hacer ajustes paramétricos que permitan incluir la dependencia en la estimación de la predicción de la vida útil de un equipo sujeto a cambios en su diseño original, en el contexto de experimentos acelerados.

También, de acuerdo a lo sugerido por uno de los árbitros, se planteará un estudio de simulación haciendo uso de la distribución lognormal bivariada generada por cópulas, siguiendo los lineamientos de Johnson (1987) y Nelsen (2006).

Agradecimientos

Los autores agradecen a la Dirección de Investigaciones de la Universidad Nacional de Colombia, sede Medellín (DIME) por la financiación. Además, agradecen a los árbitros y al editor por sus valiosos comentarios que enriquecieron el texto del manuscrito.

Este trabajo es parte del proyecto “Riesgos competitivos en la modelación de la confiabilidad de un sistema con dos modos de falla”, código Quipu: 20101007773.

[Recibido: agosto de 2010 — Aceptado: febrero de 2011]

Referencias

- Barlow, R. & Proschan, F. (1975), *Statistical Theory of Reliability and Life Testing*, Holt, Rinehart and Winston, Inc, New York.
- Basu, A. & Klein, P. (1982), ‘Some recent results in competing risks theory’, *Proceedings on Survival Analysis. IMS monograph series* **2**, 216–229.
- Bedford, T. (2005), Competing risk modelling in reliability, in ‘Modern Statistical and Mathematical Methods in Reliability’, World Scientific Books, United Kingdom.
- Bedford, T. & Lindqvist, B. (2004), ‘The identifiability problem for repairable systems subject to competing risks’, *Advances in Applied Probability* **36**(3), 774–790.
- Crowder, M. (2001), *Classical Competing Risks*, first edn, Chapman & Hall, London.
- Frees, E. & Valdez, E. (1998), ‘Understanding relationships using copulas’, *North American Actuarial Journal* **2**(1), 1–25.
- Jaramillo, M. C., Lopera, C. M., Manotas, E. C. & Yáñez, S. (2008), ‘Generación de Tiempos de Falla Dependientes Weibull Bivariados Usando Cópulas’, *Revista Colombiana de Estadística* **31**(2), 169–181.
- Johnson, M. (1987), *Multivariate Statistical Simulation*, John Wiley & Sons, New York.
- Lu, J. & Bhattacharyya, G. (1990), ‘Some new constructions of bivariate weibull models’, *Annals of the Institute of Statistical Mathematics* **42**(3), 543–559.
- Manotas, E. C., Yáñez, S., Lopera, C. M. & Jaramillo, M. C. (2008), ‘Estudio del efecto de la dependencia en la estimación de la confiabilidad de un sistema con dos modos de falla concurrentes’, *DYNA* **75**(154), 5–21.
- Meeker, W. & Escobar, L. (1998), *Statistical Methods for Reliability Data*, John Wiley and Sons, Inc, New York.

- Meeker, W., Escobar, L. A. & Hong, Y. (2009), 'Using accelerated life tests results to predict product field reliability', *Technometrics* **51**(2), 146–161.
- Nelsen, R. B. (2006), *An Introduction to Copulas*, second edn, Springer, New York.
- Nelson, W. (1990), *Accelerated Testing: Statistical Models Test Plans and Data Analyses*, John Wiley & Sons, New York.
- Pintilie, M. (2006), *Competing Risks: A Practical Perspective*, John Wiley & Sons, United Kingdom.
- R Development Core Team (2010), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
*<http://www.R-project.org/>, accedido en 01/08/2010
- Tsiatis, A. (1975), 'A nonidentifiability aspect of the problem of competing risks', *Proceedings of the National Academy of Sciences* **72**, 20–22.
- Yáñez, S., Manotas, E. C. & Lopera, C. M. (2006), 'Estudio de la dependencia en la estimación de la confiabilidad de un sistema con dos modos de falla', *Memorias del XV Simposio de Estadística*.
- Zheng, M. & Klein, J. P. (1995), 'Estimates of marginal survival for dependent competing risks based on an assumed copula', *Biometrika* **82**(1), 12–38.