



UNIVERSIDAD NACIONAL DE COLOMBIA

Una comparación de pruebas de igualdad de dos riesgos competitivos

Liliana Carolina Molina Blanco

Universidad Nacional de Colombia
Facultad de Ciencias, Escuela de Estadística
Medellín, Colombia
2016

Una comparación de pruebas de igualdad de dos riesgos competitivos

Liliana Carolina Molina Blanco

Tesis presentada como requisito parcial para optar al título de Magíster en Ciencias Estadística

Director:

Carlos Mario Lopera Gómez
Doctor en Ciencias Estadísticas

Línea de Investigación:

Confiabilidad

Grupo de Investigación:

Estadística Industrial

Universidad Nacional de Colombia
Facultad de Ciencias, Escuela de Estadística
Medellín, Colombia
2016

El cielo es el límite.

Agradecimientos

Gracias Dios por tu amor incondicional, por la vida y las fuerzas que me das para seguir adelante, ser mi compañía y guía en cada sueño y retos que he asumido, porque colocaste las personas indicadas para darme palabras de aliento durante mis estudios.

A mi madre Francisca Blanco Julio, por ser un apoyo incondicional en cada una de las decisiones de mi vida, por creer en mi, por tus consejos y ayuda incondicional.

A mi asesor Carlos Mario Lorepa Gómez, por su acompañamiento y disposición para ayudarme en este proceso tan importante en mi formación académica.

“El fracaso es una gran oportunidad para empezar otra vez con más inteligencia”.

Henry Ford.

Resumen

Los estudios de confiabilidad y supervivencia buscan analizar por medio de un conjunto de técnicas la variable “tiempo hasta que ocurre un evento”, tales como el tiempo hasta la muerte o curación, la probabilidad de falla en cada instante de tiempo, el riesgo de falla, etc. El análisis de los modelos de riesgos competitivos es apropiado para estudiar el comportamiento de una unidad o sujeto que puede fallar por diferentes causas, donde se observa tanto el tiempo hasta la falla, como el tipo de falla.

En este trabajo se abordó la problemática de dos riesgos que están compitiendo para causar la falla de un sujeto; en particular determinar si los riesgos o probabilidad de falla asociada a cada tipo de falla son igualmente importantes o si un riesgo es más serio que el otro. Para este fin se hizo un estudio de la prueba de hipótesis para la igualdad de las dos funciones de incidencia acumulada asociadas a los riesgos. Se realizó un estudio de simulación donde se comparan algunos de los procedimientos de prueba que han sido propuestos para este fin; y así, poder determinar el comportamiento de estos procedimientos de prueba bajo varios escenarios que permitan evaluar el desempeño de los mismos. Se incluyen los procedimientos de prueba usando datos reales de pacientes con linfoma, presentados en Pintilie (2006) y Petersen et al. (2004).

Palabras clave: Función de incidencia acumulada, tasa de riesgo de causa-específica, bootstrap, aproximación de simetrización aleatoria, remuestreo y supremo generalizado.

Abstract

In survival and reliability studies the analysis is intended by means of a set of variable techniques studying “time until an event occurs”, for example, the time to death or cure, the probability of failure at each instant, risk of failure, etc. The analysis of competing risk models is appropriate to study the behavior of a unit that can fail for different causes, where the time and time to failure are observed.

In this paper, it is tackled the problematic of the risks that are competing to cause the failure from the subject; in particular whether the risks or likelihood of failure associated with each type of failure are equally important or whether a risk is more serious than the other. For this purpose will be made a study of hypothesis tests for equality of cumulative incidence functions of associated with risks. A comparative study of some of the test procedures that have been proposed for this purpose, and thus able to determine the behavior of the different tests in various scenarios to evaluate the performance of the same will be made. Test procedures are included using real data of patients with lymphoma, presented in Pintilie (2006) y Petersen et al. (2004)

Keywords: Cumulative incidence function, cause-specific hazard rates, quantile, bootstrap, random symmetrization approximation, resampling and generalized supremum.

Contenido

Resumen	VII
Lista de figuras	XIII
Lista de tablas	XV
1. Introducción	1
2. Marco teórico	5
2.1. Reseña histórica	5
2.2. Modelos de riesgos competitivos	6
2.2.1. Tasa de riesgos de causa-específica.	7
2.2.2. Subdistribución o Función de incidencia acumulada (CIF).	8
2.3. Pruebas para la comparación de dos riesgos en la misma población	9
2.3.1. Prueba del supremo generalizado para la igualdad de dos CIF's	10
2.3.2. Pruebas para la igualdad de dos CIF's por métodos de remuestreo	12
2.3.3. Método bootstrap alternativo	14
3. Estudio de simulación	17
4. Análisis de resultados	23
4.1. Tablas de resultados del estudio de simulación	23
4.2. Análisis gráfico del nivel de significancia alcanzado	25
4.3. Análisis gráfico de potencia empírica	28
5. Ejemplo: Pacientes con linfoma	37
6. Conclusiones y trabajo futuro	41
Apéndices	43
A. Anexos: Códigos en R de la simulación	45
Bibliografía	65

Lista de Figuras

4-1. Significancia alcanzada bajo independencia	26
4-2. Significancia alcanzada bajo dependencia	27
4-3. Potencia empírica bajo independencia y tamaño de muestra $n = 50$. . .	29
4-4. Potencia empírica bajo dependencia y tamaño de muestra $n = 50$. . .	30
4-5. Potencia empírica bajo independencia y tamaño de muestra $n = 100$. .	31
4-6. Potencia empírica bajo dependencia y tamaño de muestra $n = 100$. . .	32
4-7. Potencia empírica bajo independencia y tamaño de muestra $n = 250$. .	33
4-8. Potencia empírica bajo dependencia y tamaño de muestra $n = 250$. . .	34
5-1. CIF para pacientes con recaída y muerte por recaída	38

Lista de Tablas

3-1. Factores tenidos en cuenta en la simulación	18
3-2. Simulación de la prueba del supremo generalizado para $n = 35$	20
4-1. Tasas de error tipo I y potencias empíricas con $\alpha = 0.05$ y $n = 50$	23
4-2. Tasas de error tipo I y potencias empíricas con $\alpha = 0.05$ y $n = 100$	24
4-3. Tasas de error tipo I y potencias empíricas con $\alpha = 0.05$ y $n = 250$	24
5-1. Muestra de recaída y muerte de pacientes con linfoma folicular	37
5-2. Estadísticos de prueba (Valor-p) para los datos de pacientes con linfoma	38

1. Introducción

En diferentes campos de investigación es de interés estudiar las causas por las cuales un sujeto o componente pueden fallar; es común encontrar casos en los que el sujeto pueda estar expuesto a presentar más de un modo de falla. Por ejemplo; en el campo de la medicina se hace necesario desarrollar tratamientos adecuados para una enfermedad que requiere análisis de los resultados específicos ésta. Los criterios de valoración o control de la enfermedad, pueden ser imposibles de observar debido a la aparición previa de un tipo de evento diferente (como la muerte por otra causa), a este evento se le denomina riesgo competitivo (Pintilie 2011). En el campo de la industria existen sistemas o 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 objetivo de mejorar la confiabilidad (Meeker & Escobar 1998). Yáñez et al. (2011) mencionan que el tiempo de falla de un sistema con diversos modos de falla puede ser modelado considerando un sistema en serie o un modelo de riesgos competitivos, donde cada modo de falla es un componente del sistema en serie, y cada componente tiene un tiempo de vida aleatorio; por ende, 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.

En muchas situaciones prácticas es común analizar únicamente el evento de interés sin tener en cuenta los riesgos que están compitiendo; los procedimientos de análisis en este contexto solo responden preguntas en las que el objeto de estudio es estimar el efecto “puro” debido a una sola causa de falla (Pintilie 2006). En otros campos de investigación se desean estudiar los riesgos que están actuando simultáneamente en una misma población, para determinar si los diversos riesgos que se consideran son igual de graves o si uno de los riesgos es más serio que el otro, para tal fin es necesario estudiar las diferentes pruebas que se han diseñado para analizar este tipo de situaciones, con fin de determinar el comportamiento de éstas.

Pintilie (2011) menciona que las técnicas habituales de análisis del tiempo hasta el evento aplicadas en presencia de riesgos competitivos producen resultados sesgados o no interpretables. Por ejemplo, en el caso de tener dos tipos de eventos; el evento de

interés y el conjunto de eventos de riesgos competitivos, se debe tener cuidado en el momento de analizar la probabilidad de falla, dado que esta puede ser sobreestimada cuando se aplican las técnicas clásicas; esto llevaría a dar recomendaciones erróneas de la probabilidad de falla cuando los riesgos competitivos están presentes, ya que en este caso se estará indicando que el sujeto va a fallar antes del tiempo en el que realmente puede fallar. La estimación de la probabilidad del evento debe calcularse con el empleo de técnicas específicas como la función de incidencia acumulada introducida por Kalbfleisch & Prentice (1980). Cabe mencionar que, con el empleo de técnicas específicas para el análisis de los riesgos competitivos se asegurará que los resultados no estén sesgados y se puedan interpretar correctamente.

Recientemente en la literatura se han estudiado ampliamente las pruebas para la comparación de riesgos independientes. Sin embargo en situaciones reales es frecuente que los riesgos sean dependientes y los tiempos de falla estén sujetos a censuras a derecha. En este contexto general, Aly et al. (1994) propusieron pruebas de distribución asintóticamente libre para la comparación de las funciones de incidencia acumulada o equivalentemente de las tasas de riesgo de causa específica. La propuesta de Aly et al. (1994) fue ampliada en la investigación de la prueba del supremo generalizado para la igualdad de tasas de riesgo de causa-específica propuesta por Kochar et al. (2002), donde se proponen tres clases de pruebas de hipótesis y por medio de un estudio de simulación se establece que una de las pruebas es menos sensible a la elección de la función de peso y es adecuada cuando no se tiene información suficiente en cuanto a la naturaleza de los datos, por tanto, ésta sería una buena prueba para comparar dos riesgos competitivos; en general, cuando no se tiene información acerca de las características de los datos. Por otro lado, se han desarrollado varios tipos de pruebas para la comparación de funciones de incidencia acumulada a través de métodos de remuestreo propuesta por Kam et al. (2002).

En este trabajo se abordó la problemática de determinar si los riesgos o probabilidad de falla asociada a cada tipo de falla son igualmente importantes o si un riesgo es más serio que el otro. Para este fin se realizó un estudio comparativo de la prueba del supremo generalizado Kochar et al. (2002) y las pruebas basadas en métodos de remuestreo para la igualdad de las funciones de incidencia acumulada: el método bootstrap y la aproximación por simetrización aleatoria propuestas por Kam et al. (2002). En este último caso se propuso una modificación de estas pruebas, específicamente en el método bootstrap; luego, a través de un estudio de simulación se evaluó el comportamiento de las diferentes pruebas en varios escenarios con el fin de estudiar el desempeño de las mismas.

La estructura del trabajo es la siguiente: en el Capítulo 2 se dan algunos conceptos básicos de riesgos, nociones de los modelos riesgos competitivos y se presentan las pruebas para comparar la igualdad de las funciones de incidencias acumuladas asociadas a dos riesgos competitivos, en el Capítulo 3 se presenta el estudio de simulación de los datos de riesgos competitivos usando para este fin la distribución exponencial bivariada de Block & Basu (1974) y las pruebas antes mencionadas; en el Capítulo 4 se evalúa el desempeño de las pruebas en diferentes escenarios a través del análisis de los resultados obtenidos del estudio de simulación, posteriormente en el Capítulo 5 se presenta un ejemplo con datos reales. Finalmente, en el Capítulo 6 se presentan las conclusiones más relevantes del trabajo y describe los temas relacionados con esta investigación que pueden ser de interés para trabajo futuro en el área de confiabilidad y supervivencia.

2. Marco teórico

En este capítulo se presenta una breve reseña histórica de los riesgos competitivos, algunas definiciones básicas, los modelos de riesgos de causa específica y las funciones de incidencia acumulada, estas definiciones están basadas en el texto de Pintilie (2006), además se explican las pruebas basadas en el supremo generalizado, en métodos de remuestro: bootstrap y simetrización aleatoria para comparar la igualdad de dos funciones de incidencias acumuladas y se presenta una propuesta alternativa al método bootstrap propuesto por Kam et al. (2002).

2.1. Reseña histórica

A mediados del siglo XVII John Graunt comenzó a recopilar y tabular información incluida en los boletines de mortalidad publicados semanalmente para esa época en Londres. Ese trabajo, probablemente realizado en colaboración con Sir William Petty, apareció en 1662 y contenía un resumen demográfico de las causas de muerte en Inglaterra y Gales.

En 1693, el famoso astrónomo Edmund Halley, desarrolló el concepto de tablas de vida (tablas de mortalidad) en un formato no muy diferente al utilizado actualmente en el análisis de supervivencia; sus datos se basaron en los registros de nacimientos y defunciones de la ciudad de Breslau (hoy Wroclaw, en el suroeste de Polonia). En 1760 Daniel Bernoulli aplicó el método de Halley para demostrar las ventajas de la inoculación de la viruela (variolización). Para tal fin, calculó el incremento de la función de supervivencia de los datos registrados en Halley para determinar el riesgo de morir a causa de la viruela de las personas que recibieron y no recibieron la vacunación, con el fin de determinar la eficacia de la vacunación contra dicha enfermedad. De esta manera Bernoulli fundó la teoría de los riesgos competitivos. Un resumen de su trabajo sobre los riesgos competitivos se puede encontrar en Moeschberger & David (1978).

Riesgos Competitivos

Kalbfleisch & Prentice (1980) describen los riesgos¹ competitivos como la situación en la que un individuo puede experimentar más de un tipo de evento. Gooley et al. (1999) definieron el concepto de riesgo competitivo como la situación donde un tipo de evento “impide la ocurrencia de otro evento bajo investigación o altera fundamentalmente la probabilidad de ocurrencia del otro evento”. Esto refleja estrechamente la situación con la que se enfrenta el investigador clínico o estadístico aplicado: la posibilidad de muchos tipos de falla que compiten para ser observados.

2.2. Modelos de riesgos competitivos

Pintilie (2006) describe dos enfoques matemáticos para tratar los riesgos competitivos: los riesgos competitivos como una variable aleatoria bivariada y los riesgos competitivos como tiempos de falla latentes. El primer enfoque muestra que en ausencia de riesgos competitivos, los datos de supervivencia se presentan usualmente como variables aleatorias bivariadas (T, δ) , donde δ es una indicadora del estatus del individuo que toma el valor de 1 si se observa el evento de interés, y 0 si la observación es censurada. Cuando $\delta = 1$, el primer miembro del par, T , corresponde al tiempo en que se produce el evento y cuando $\delta = 0$, T es el tiempo en el que se censuró la observación.

Esta definición se puede extender a la situación de riesgos competitivos donde son posibles $j \geq 2$ tipos de fallas o eventos. Los datos son nuevamente representados por un par (T, δ) , donde la variable de censura δ esta vez es discreta tomando el valor de 0 si la observación es censurada. En el caso de que la observación no este censurada, δ tomará el valor j , donde j es el tipo de falla o evento observado ($\delta = j$), en este caso T es el tiempo en el que el evento o falla de tipo j se produjo; de lo contrario, es el tiempo de censura (Kalbfleisch & Prentice 1980).

La segunda definición matemática de los riesgos competitivos está escrita en términos de los tiempos de eventos latentes, donde T_1, T_2, \dots, T_k son los tiempos de los eventos,

¹El concepto de riesgo en este contexto se entiende como la vulnerabilidad ante un potencial perjuicio o daño para sujeto. Cuanto mayor es la vulnerabilidad mayor es el riesgo, pero cuanto más factible es el perjuicio o daño, mayor es el peligro. Por tanto, el riesgo se refiere sólo a la teórica “posibilidad de daño” bajo determinadas circunstancias, mientras que el peligro se refiere sólo a la teórica “probabilidad de daño” bajo esas circunstancias.

uno por cada tipo de evento j . En los riesgos competitivos sólo se observa la situación donde se presenta el primer evento, por lo tanto la variable de tiempo, T , estará definida como $T = \min\{T_1, T_2, \dots, T_k\}$. La variable de censura, δ , está definida como $\delta = 0$ si la observación es censurada de lo contrario $\delta = j$, para $j = 1, 2, \dots, k$.

Se han establecido varias funciones matemáticas importantes en los riesgos competitivos, tales como: la subdistribución o función de incidencia acumulada, y la tasa de riesgos de causa-específica.

2.2.1. Tasa de riesgos de causa-específica.

Pintilie (2006) define la función de riesgo o tasa de riesgo como la tasa instantánea del evento para un individuo que ha llegado al tiempo t sin experimentar el evento. Cuando el evento de interés es la muerte la función de riesgo representará tasa instantánea de mortalidad para un individuo que ha sobrevivido al tiempo t . La tasa de riesgo, expresa la propensión a fallar en el siguiente intervalo pequeño de tiempo, dado que se sobrevivió al tiempo t . Esta función a menudo es de interés en el análisis de los datos de tiempo a un evento, porque ilustra cómo la tasa instantánea del evento varía con el tiempo. Matemáticamente la función de riesgo es la probabilidad condicional de que un evento ocurra dentro de un tiempo dado, digamos entre t y $t + \Delta t$, dado que no había ocurrido ningún evento hasta el tiempo t . Por lo tanto, la función de riesgo denotada por $\lambda_j(t)$, se define como:

$$\begin{aligned} \lambda_j(t) &= \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t, \delta = j | T \geq t)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t, \delta = j)}{\Delta t P(T \geq t)} \\ &= \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t, \delta = j)}{\Delta t} \times \frac{1}{1 - P(T \leq t)} \\ &= \frac{f_j(t)}{1 - F_j(t)} \end{aligned}$$

La tasa de riesgo total de un evento de cualquier tipo hasta la falla se puede encontrar sumando sobre todas las tasas de riesgo de causa específica:

$$\lambda_T(t) = \sum_{i=1}^j \lambda_i(t)$$

La función de riesgo acumulado está definida como

$$\Lambda_j(t) = \int_0^t \lambda_j(x) dx = \int_0^t \frac{f_j(x)}{S(x)} dx$$

donde $S(t)$ es la función de supervivencia, que mide la probabilidad de que un sujeto sobreviva más allá de un periodo de tiempo dado, esto es

$$S(t) = P(T > t) = \int_t^{\infty} f(u) du = 1 - F(t)$$

La función de supervivencia $S(t)$ puede ser estimada por medio del estimador de Kaplan-Meier introducido por Kaplan & Meier (1958), este está definido como el producto

$$\hat{S}(t) = \prod_{t_i \leq t} \frac{n_i - d_i}{n_i}$$

donde d_i es el número de fallas en el tiempo t_i y n_i es el número de sujetos en riesgo justo antes de t_i , con $i = 1, 2, \dots, N$ observaciones.

2.2.2. Subdistribución o Función de incidencia acumulada (CIF).

Pintilie (2006) define la CIF, o subdistribución, para un evento de tipo j ($j = 1, 2, \dots, k$) como la probabilidad conjunta

$$F_j(t) = P(T \leq t, \delta = j)$$

En otras palabras, la CIF es la probabilidad de que un evento de tipo j ocurre en o antes del tiempo t .

La función de distribución acumulada $F(t)$ es la probabilidad de que un evento de cualquier tipo ocurra en o antes del tiempo t ; ésta es igual a la suma de las CIF's, para todos tipos de eventos. Por lo tanto,

$$F(t) = Pr(T \leq t) = \sum_{j=1}^k P(T \leq t, \delta = j) = \sum_{j=1}^k F_j(t)$$

Cuando se están comparando riesgos competitivos en una misma población, las pruebas de las tasas de riesgos de causa específica son equivalentes a probar las subdistribuciones (CIF's) (ver Gray (1988), Luo & Turnbull (1999) y Pintilie (2006)).

2.3. Pruebas para la comparación de dos tipos de riesgo en la misma población

En muchas situaciones prácticas, es importante saber si en una población donde actúan simultáneamente dos riesgos, éstos se pueden considerar igual de graves o si uno de los riesgos es más serio que el otro. Para dar respuesta a este problema se han diseñado pruebas estadísticas que usan los modelos de riesgos competitivos; cabe recordar que en presencia de los riesgos competitivos, la comparación de los riesgos de causa específica entre dos grupos no es equivalente a la comparación de las funciones de subdistribución (CIF), sin embargo, comparar los riesgos de causa específica de dos riesgos en una misma población es equivalente a comparar las subdistribuciones (CIF). (Pintilie 2006)

Sea T el tiempo de vida del sujeto, asumiendo que el tiempo es continuo, con función de distribución F y función de supervivencia S , y sea δ la causa de falla, es decir, $\{\delta = j\}$ es el evento cuya falla se debe al riesgo $j = 1, 2$, la función de subdistribución está definida como (Pintilie 2006):

$$F_j(t) = P(T \leq t, \delta = j) = \int_0^t S(u) \lambda_j(u) du, j = 1, 2$$

La función de distribución acumulada de falla $F(t)$ estará dada por $F(t) = F_1(t) + F_2(t)$. Sea C un tiempo de censura independiente de T con función de supervivencia $S_C(t)$. Asuma que $S_C(t) > 0$ para todo t , entonces δ se convierte en una variable aleatoria discreta que representa el estatus del individuo que toma valores de $\delta = 0, 1, 2$, donde $\{\delta = 0\}$ es el evento en el que el sujeto fue censurado (censura a derecha) y $\{\delta = j\}$ cuando el sujeto falla debido a la causa j . Se tienen n observaciones independientes e idénticamente distribuidas (i.i.d) de la forma (X_i, δ_i) , $i = 1, 2, \dots, n$, donde $X_i = \min\{T_i, C_i\}$. De acuerdo a los datos mencionados, se formula el problema de poner a prueba las siguientes hipótesis

$$\mathbf{H}_0 : F_1(t) = F_2(t), \text{ para } t \geq 0 \quad \text{vs} \quad \mathbf{H}_1 : F_1(t) \neq F_2(t) \text{ para algún } t$$

o equivalentemente

$$\mathbf{H}_0 : \lambda_1(t) = \lambda_2(t), \text{ para } t \geq 0 \quad \text{vs} \quad \mathbf{H}_1 : \lambda_1(t) \neq \lambda_2(t), \text{ para algún } t$$

Para probar estas hipótesis, se utilizarán las siguientes pruebas:

1. Supremo generalizado.

2. Método basados en remuestreo.
 - a. Aproximación bootstrap
 - b. Aproximación por simetrización aleatoria (**RAS**)
3. Propuesta del método bootstrap alternativo usando los estimadores de las CIFs y Kaplan-Meier.

2.3.1. Prueba del supremo generalizado para la igualdad de dos funciones de incidencia acumulada (CIF)

Kochar et al. (2002) consideraron un modelo de riesgos competitivos con dos causas de falla, proponiendo dos clases de pruebas basadas en la distribución asintóticamente libre y la tipo Renyi, para probar la igualdad de dos riesgos con posible censura. Para tal fin, plantean el estadístico del supremo que es una generalización de las pruebas propuestas por Aly et al. (1994) mediante la adopción de diferentes funciones de peso w . El estadístico de prueba es

$$C_n^*(w) = \sup_{0 \leq s < t < \infty} \frac{|L_n(t) - L_n(s)|}{S_n(\infty)} \quad (2-1)$$

Esta prueba al igual que otras (ver, Carriere & Kochar (2000) y Aly et al. (1994)) que son usadas para probar H_0 contra varias alternativas, pueden expresarse como funciones ponderadas de los estadísticos de tipo log-rank de la forma

$$L_n(t) = \int_0^t w(u) d(\widehat{\Lambda}_2 - \widehat{\Lambda}_1)(u) \quad (2-2)$$

siendo $\Lambda_j(t) = \int_0^t \lambda_j(u) du$ la función de riesgo de causa-específica acumulada para el riesgo j , cuando $j = 1, 2$ y el estimador de Nelson-Aalen (Ver, Fleming & Harrington (1991)) de Λ_j es

$$\widehat{\Lambda}_j(t) = \sum_{i: X_i \leq t} \frac{I(\delta_i = j)}{R_i} \quad (2-3)$$

donde, $R_i = \#\{k : X_k \geq X_i\}$ es el tamaño del riesgo fijado antes del tiempo X_i denotado como el tiempo X_i^- . La función de peso $w(u)$ refleja la importancia que se concede a las funciones de incidencia acumulada (CIF) en el tiempo u .

Como se mencionó anteriormente las pruebas del supremo generalizado extiende las pruebas no estudentizadas propuesta por Aly et al. (1994), este tipo de prueba tienen una distribución asintóticamente libre solo cuando se utiliza la función de peso

$\widehat{S}_T(u-)\widehat{S}_C(u-)^{1/2}$, mientras que las pruebas de Kochar et al. (2002) son más flexibles en cuanto al uso de diferentes funciones de peso, pero se debe usar una adecuada estudentización con el fin de conservar las propiedades de la distribución asintóticamente libre; los estudios de simulación de Kochar et al. (2002) mostraron que los estadísticos de prueba estudentizados tienen mayor rendimiento que los no estudentizados de Aly et al. (1994) en el sentido de que las distribuciones de muestras finitas de los estadísticos de prueba estudentizados pueden estar estrechamente aproximados por su respectiva distribución asintótica bajo la hipótesis nula (H_0). Se sabe que bajo H_0 , $n^{1/2}L_n(t)$ es un proceso martingala con varianza predecible $\sigma^2(t)$, que se puede estimar por

$$S_n^2(t) = \int_0^t \frac{w^2(u)}{\frac{\bar{Y}^2(u)}{n}} d\bar{N}(u) \quad (2-4)$$

donde $\bar{Y}(u) = \sum_{i=1}^n I(X_i \geq u)$ es el número total de elementos en riesgo hasta $u-$, y $\bar{N}(u)$ es el número total de fallas hasta el tiempo u . Para este trabajo se tomó de referencia la función de peso $w(u) = \bar{Y}(u)$; es importante mencionar que según Kochar et al. (2002) la elección de la función peso debe basarse de acuerdo al criterio que desee enfatizar el investigador; pero en sus estudios de simulación observaron que las pruebas basadas en la función de peso $w(u) = \bar{Y}(u)$ presentaban potencias más razonables en la mayoría de situaciones prácticas. Cabe resaltar que de acuerdo a sus investigaciones se encontró que la prueba C_n^* es menos sensible a la elección de la función de peso, por ende es considerada como una prueba adecuada cuando no se tiene información acerca de las características de los datos. Kochar et al. (2002) demuestran que bajo la hipótesis nula (H_0) y algunas condiciones de regularidad, $\{n^{1/2}L_n(t)/S_n(\infty)\}$ convergen débilmente a un movimiento Browniano estándar $\{W(t), t \geq 0\}$, como consecuencia, bajo hipótesis nula (H_0),

$$n^{1/2}C_n^*(w) \rightarrow \sqrt{n} \sup_{0 \leq s < t < 1} |W_n(t) - W_n(s)| \quad (2-5)$$

El rango del movimiento browniano estándar fue estudiado por Feller (1951) con función de densidad dada por

$$h(x) = 8 \sum_{k=1}^{\infty} (-1)^{k-1} k^2 \phi(kx) \quad (2-6)$$

donde ϕ es la función de densidad de una normal estándar. Los cuantiles del 95% y 99% se encuentran en 2.497 y 3.023, respectivamente. Para tomar la decisión bajo H_0 con base a la prueba C_n^* se puede obtener un cuantil $1 - \alpha$ denotelo $q_{1-\alpha}$, se rechaza H_0 si $C_n^*(w)$ supera a $q_{1-\alpha}$, esto es, $C_n^*(w) > q_{1-\alpha}$.

2.3.2. Pruebas para determinar la igualdad de dos funciones de incidencia acumulada a través de métodos de remuestreo

Kam et al. (2002) diseñaron estadísticos de prueba basados en vector de procesos relacionados con las funciones de incidencia acumulada para determinar la igualdad de las funciones de incidencia acumulada. Como las distribuciones asintóticas parecen muy complicadas y dependen de la distribución subyacente de los datos, se utilizan dos técnicas de remuestreo (método bootstrap y método de simetrización al azar), para aproximar los valores críticos de las pruebas. Sin hacer ninguna hipótesis sobre la naturaleza de dependencia entre los riesgos, las pruebas permiten comparar dos o más riesgos ($j \geq 2$) simultáneamente bajo el modelo de censura aleatoria. Para probar H_0 vs H_1 consideran el estadístico de Cramér-Von Mises dado por:

$$D_n = n \int_0^\infty (\widehat{F}_{n,2}(t) - \widehat{F}_{n,1}(t))^2 \widehat{F}_n(dt) \quad (2-7)$$

donde $\widehat{F}_n(t)$ es el análogo muestral de $\widehat{F}(t) = P(T \leq t)$ y $\widehat{F}_{n,j}$ es la contraparte empírica de la función de sub-distribución (\widehat{F}_j); definida por

$$\widehat{F}_{n,j}(t) = n^{-1} \sum_{i=1}^n I(X_i \leq t, \delta_i = j), \quad j = 1, 2. \quad (2-8)$$

Se rechaza la hipótesis nula (H_0) cuando el valor D_n es demasiado grande. Kam et al. (2002) reescriben el estadístico (2-6) como

$$D_n = \int_0^\infty (W_n(t))^2 \widehat{F}_n(dt) \quad (2-9)$$

donde $W_n(t) = W_{n,2}(t) - W_{n,1}(t)$ siendo $W_{n,j}(t) = n^{\frac{1}{2}}(\widehat{F}_{n,j} - \widehat{F}_j)$; $j = 1, 2$. Donde \widehat{F}_j es el estimador de la función de sub-distribución definida como

$$\widehat{F}_j(t) = \int_0^t \widehat{S}(u-) \widehat{\lambda}_j(u) du$$

Los autores plantean que W_n puede ser usada como una herramienta para obtener la distribución asintótica del estadístico bajo H_0 y además demuestran que W_n converge en distribución a un vector de procesos gaussianos (Kam et al. 2002), los límites de la distribución y sus valores críticos no son analíticamente tratables por tal razón se optaron por utilizar los siguientes métodos de remuestreo:

1. Aproximación bootstrap
2. Aproximación por simetrización aleatoria (**RAS**)

Aproximación bootstrap

Sea $\{(X_i^*, \delta_i^*); i = 1, 2, \dots, n\}$ una muestra bootstrap que se extrae con reemplazo de la muestra original. Se denota las versiones bootstrap de $\widehat{F}_{n,j}$ y \widehat{F}_n por $\widehat{F}_{n,j}^*$ y \widehat{F}_n^* respectivamente. Entonces la versión bootstrap de W_n^* está dada por $W_n^*(t) = W_{n,2}^*(t) - W_{n,1}^*(t)$ donde $W_{n,j}^*(t) = n^{\frac{1}{2}}(\widehat{F}_{n,j}^* - \widehat{F}_{n,j})$. Por otra parte, el estadístico bootstrap está dado por

$$D_n^* = \int_0^\infty (W_n^*(t))^2 \widehat{F}_n^*(dt)$$

En la literatura se han encontrado investigaciones en las que utilizan técnicas basadas en la aproximación de bootstrap para construir pruebas relacionadas a los riesgos, entre estos, Burke & Yuen (1995) construyeron una prueba de bondad de ajuste para el modelo de riesgos proporcionales. Yuen & Burke (1997) utilizaron un enfoque similar para las pruebas de un modelo de riesgo aditivo semiparamétrico. El procedimiento de la aproximación bootstrap para D_n es presentado a continuación:

- Se extraen M muestras bootstrap de la forma $\{(X_i^*, \delta_i^*); i = 1, 2, \dots, n\}$ con reemplazo de la muestra original.
- A cada conjunto de datos remuestreado se le calcula el estadístico de prueba, obteniendo $D_{n_1}^*, D_{n_2}^*, \dots, D_{n_M}^*$
- Se obtienen la distribución de los D_n^* 's y se calcula el cuantil $1 - \alpha$, denotado por $q_{1-\alpha}$. Se rechaza H_0 si $D_n > q_{1-\alpha}$

Aproximación por simetrización aleatoria (RAS)

Sea $Z = \{Z_i; i = 1, 2, \dots, n\}$ una muestra aleatoria independiente e idénticamente distribuida que toman signos ± 1 con igual probabilidad. Asuma que estas variables de permutaciones son independientes de la muestra. La versión **RAS** de $\widehat{F}_{n,j}$ está dada por

$$F_{n,j}^Z(t) = n^{-1} \sum_{i=1}^n Z_i I(T \leq t, \delta_i = j)$$

esto lleva al proceso $W_{n,1,2}^Z(t) = n^{\frac{1}{2}}(F_{n,2}^Z(t) - F_{n,1}^Z(t))$. Entonces la versión **RAS** de W_n está dada por $W_n^Z = W_{n,1,2}^Z(t)$ y el estadístico de prueba **RAS** es

$$D_n^Z = \int_0^\infty (W_n^Z(t))^2 \widehat{F}_n^Z(dt)$$

El procedimiento de la prueba **RAS** para D_n es presentado a continuación:

- Denote al valor observado de D_n dado en la ecuación (2-7) como $D_n^{Z_0}$

- Genere M conjuntos de Z 's, denotados $\{Z^1, Z^2, \dots, Z^M, \}$
- Se obtienen $D_n^{Z^1}, D_n^{Z^2}, \dots, D_n^{Z^M}$
- Denote como D_{M+1} al conjunto formado por $(D_n^{Z^0}, D_n^{Z^1}, \dots, D_n^{Z^M})$
- Se rechaza la hipótesis nula H_0 cada vez que $\hat{p} \leq \alpha$, donde el valor p estimado es

$$\hat{p} = \frac{a}{M+1}$$

siendo a el número de valores del conjunto D_{M+1} que son mayores o iguales a $D_n^{Z^0}$.

Kam et al. (2002) demostraron que en casi todas las secuencias de las muestras, del proceso bootstrap W_n^* y del proceso **RAS** de W_n^Z , convergen condicionalmente en distribución a un proceso gaussiano.

El enfoque **RAS** está motivado por métodos de ponderación al azar que han sido técnicas básicas para abordar la convergencia de los procesos empíricos; véase, por ejemplo, Gine & Zinn (1984) y Pollard (1984). Neuhaus & Zhu (1998) utilizaron Z_i 's para la prueba de simetría de reflexión de una distribución multivariante. Yuen, Zhu & Tang (2003) estudiaron la prueba **RAS** para un modelo de muestreo censurado semiparamétrica. Otras pruebas que usan un enfoque similar se puede encontrar en Zhu, Yuen & Tang (2002) y Zhu & Neuhaus (2000).

2.3.3. Método bootstrap alternativo usando los estimadores de las CIF's y Kaplan-Meier

Dado que Kam et al. (2002) sugieren en los métodos de remuestreo el uso de las funciones empíricas, tanto de las distribución y subdistribución de las probabilidades de falla en la construcción del estadístico de prueba; una propuesta alternativa es cambiar los estadísticos D_n por una versión modificada donde se usen las CIF's y Kaplan-Meier. Para probar H_0 vs. H_1 , se considera el estadístico de Cramér-Von Mises alternativo

$$D'_n = n \int_0^\infty (\widehat{F}_2(t) - \widehat{F}_1(t))^2 (1 - \widehat{S}(dt)) \quad (2-10)$$

Donde $\widehat{S}(t) = P(T > t)$ es la función de supervivencia estimada por medio del método de Kaplan-Meier y $\widehat{F}_j(t)$, $j = 1, 2$ son las CIF's estimadas para los dos modos de falla.

Sobre la modificación hecha al estadístico de Cramér-Von Mises, se aplica el mismo procedimiento de prueba que se usó en el método bootstrap, que consiste en encontrar la distribución bajo la hipótesis nula (H_0) del estadístico alternativo D'_n , para finalmente tomar una decisión acerca de la hipótesis de interés, es decir, se decide rechazar H_0 si el valor D'_n resulta mayor al cuantil $(1 - \alpha)$ de la distribución empírica del estadístico bajo la hipótesis nula.

3. Estudio de simulación

Para evaluar el desempeño de los estadísticos de prueba vistos en la sección 2.3 se consideraron dos causas de falla. Sea (T_1, T_2) los tiempos potenciales de falla de dos componentes de un sistema en serie. El tiempo de falla $X = \min\{T_1, T_2\}$ fue generado usando la distribución exponencial absolutamente continua de Block & Basu denotada por (ACBVE) usando las indicaciones de Leandro & Achcar (1996), Block & Basu (1974) y Friday & Patil (1977) donde la pareja (T_1, T_2) bajo esta distribución tiene una densidad dada por

$$f(t_1, t_2) = \begin{cases} \frac{\lambda\lambda_1\lambda_{02}}{\lambda_{12}} e^{-\lambda_1 t_1 - \lambda_{02} t_2} & \text{si } t_1 < t_2 \\ \frac{\lambda\lambda_2\lambda_{01}}{\lambda_{12}} e^{-\lambda_{01} t_1 - \lambda_2 t_2} & \text{si } t_1 \geq t_2 \end{cases} \quad (3-1)$$

donde $\lambda_{12} = \lambda_1 + \lambda_2$, $\lambda_{01} = \lambda_0 + \lambda_1$, $\lambda_{02} = \lambda_0 + \lambda_2$ y $\lambda = \lambda_0 + \lambda_1 + \lambda_2$.

Siendo λ_0 el parámetro que controla el grado de dependencia entre T_1 y T_2 , donde $\lambda_0 = 0.0$ corresponde a la independencia de los dos riesgos y $\lambda_0 = 1.0$ la dependencia de los dos riesgo. En este caso la tasa de riesgo de causa específica son proporcionales, y están dadas por

$$\lambda_j(t) = \frac{\lambda_j(\lambda_0 + \lambda_1 + \lambda_2)}{\lambda_1 + \lambda_2}; \quad j = 1, 2 \quad (3-2)$$

Al usar la distribución ACBVE se planteó las hipótesis H_0 versus H_1 como

$$H_0 : \lambda_1(t) = \lambda_2(t) \quad \text{vs.} \quad H_1 : \lambda_1(t) \neq \lambda_2(t)$$

Se fijó la tasa de falla del primer evento como $\lambda_1 = 1.0$ y se varió la tasa de falla del segundo evento λ_2 tomando diferentes valores, tales como, $\lambda_2 = 0.5, 1.5, 2.0, 2.5$ y en el caso de ser iguales las tasas $\lambda_2 = 1.0$, para determinar el comportamiento de las pruebas a medida que este cambia. Los valores asignados al parámetro λ_0 y λ_1 se tomaron teniendo como referencia los estudios de simulación realizados por Aly et al. (1994), Kam et al. (2002), Kochar et al. (2002) y Sankaran et al. (2010).

Para mirar el efecto de las censuras en cada una de las pruebas se consideraron tres niveles de censura, sin censura, censura moderada (18% – 35%) y censura alta

(45% – 60%) denotadas por SC, CM y CA respectivamente. La variable censura C de acuerdo a lo planteado por Sun & Tiwari (1995), fue generada por una distribución exponencial con parámetro γ ; para lo cual se usó la probabilidad de una censura $P(C < X)$ tomando valores de 0 (sin censura), 0.2 (censura moderada) y 0.5 (censura alta), que bajo el modelo ACBVE tiene la forma

$$P(C < X) = \frac{\gamma}{\gamma + \lambda} \quad (3-3)$$

de donde se despeja γ en cada uno de los escenarios. El nivel de significancia usado en las pruebas es de $\alpha = 0.05$. Para cada conjunto de parámetros asumidos, se generaron muestras de $n = 50, 100$ y 250 . Adicionalmente se generaron 5000 muestras bootstrap, tanto para el estadístico bootstrap dado en (2-8) como para el estadístico bootstrap alternativo dado en (2-10), y se extrajeron 5000 conjuntos de datos **RAS** $\{Z_1, Z_2, \dots, Z_n\}$.

Se programaron en el software libre **R** las pruebas basadas en los métodos bootstrap, **RAS** y el método bootstrap alternativo usando los estimadores de las CIF's y Kaplan-Meier; para la prueba del supremo se usó la programación realizada por Pintilie (2006). La tabla **Tabla 3-1** presenta un resumen de los factores tenidos en cuenta en el estudio de simulación

Factor	Niveles considerados
λ_0	0.0, 1.0
λ_1	1.0
λ_2	0.5, 1.0, 1.5, 2.0 y 2.5
n	50, 100, 250
Censura	Sin censura, moderada, alta

Tabla 3-1.: Factores tenidos en cuenta en la simulación

En total se consideraron 90 escenarios de simulación, donde se evaluó el error tipo I (bajo $H_0 : \lambda_1(t) = \lambda_2(t)$) y la potencia de las diferentes pruebas presentadas (bajo $H_1 : \lambda_1(t) \neq \lambda_2(t)$). La tasa el error tipo I está dado por:

$$\hat{\alpha} = \frac{1}{N} \sum I(\text{Rechazar } H_0 | H_0 \text{ es Cierta}),$$

donde $I(\text{Rechazar } H_0 | H_0 \text{ es Cierta})$ es una indicadora de rechazo de una hipótesis cierta para una prueba particular.

La potencia alcanzada está definida como la probabilidad de rechazar una hipótesis nula falsa, esto es

$$1 - \hat{\beta} = \frac{1}{N} \sum I(\text{Rechazar } H_0 | H_0 \text{ es Falsa})$$

donde $I(\text{Rechazar } H_0 | H_0 \text{ es Falsa})$ es una indicadora de rechazo de una hipótesis falsa para una prueba particular y N es el número de simulaciones realizadas en cada una de las pruebas estudiadas, en cada uno de los escenarios de simulación considerados.

El proceso de simulación se describe a continuación:

- **Paso 1.** Se genera una muestra de tamaño n de la forma (X_i, δ_i) con un porcentaje de censura especificado
- **Paso 2.** A este conjunto de datos se le aplican las pruebas basadas en el supremo, los métodos de remuestreo: bootstrap, **RAS** y la alternativa, con un nivel de significancia fijo.
- **Paso 3.** Los pasos 1 y 2 son simulados N veces obteniendo la decisión para cada simulación.
- **Paso 4.** Se suman las indicadoras de las N simulaciones y se promedian para obtener el error tipo **I** o la potencia alcanzada esto es:

$$\hat{\alpha} = \frac{1}{N} \sum I(\text{Rechazar } H_0 | H_0 \text{ es Cierta}),$$

o

$$1 - \hat{\beta} = \frac{1}{N} \sum I(\text{Rechazar } H_0 | H_0 \text{ es Falsa})$$

A continuación se presenta un ejemplo de los pasos realizados en la simulación por medio del siguiente caso: se desea establecer de manera particular el desempeño de la prueba del supremo generalizado en una muestra de tamaño $n = 35$ con dos modos de falla donde el parámetro $\lambda_2 = 0.5$ indica que la función de riesgo de causa específica asociada a la causa 2 de falla es la mitad de la correspondiente función de riesgo de causa específica de la causa 1 de falla, y se tiene un caso de censura moderada, donde además los tiempos falla son dependientes ($\lambda_0 = 1.0$), cuando se desea probar si

$$\mathbf{H}_0 : F_1(t) = F_2(t) \quad \text{vs.} \quad \mathbf{H}_1 : F_1(t) \neq F_2(t)$$

- Se genera una muestra de tamaño $n = 35$ de la forma (X_i, δ_i) con un porcentaje de censura del 15% en los datos.
- Se realizan $N = 20$ simulaciones en las mismas condiciones definidas antes.
- En cada simulación se aplica el estadístico de prueba del supremo generalizado bajo la hipótesis nula (H_0), con un nivel de significancia $\alpha = 0.05$. Los resultados obtenidos se muestran en la **Tabla 3-2**

simulación	n	λ_0	λ_2	λ_1	C_n^*	Valor-p	Decisión*
1	35	1	0.5	1	1.7408	0.3231	0
2	35	1	0.5	1	3.0533	0.0090	1
3	35	1	0.5	1	2.5019	0.0495	1
4	35	1	0.5	1	1.2572	0.7410	0
5	35	1	0.5	1	1.7678	0.3053	0
6	35	1	0.5	1	2.0083	0.1779	0
7	35	1	0.5	1	1.4856	0.5260	0
8	35	1	0.5	1	3.7139	0.0008	1
9	35	1	0.5	1	2.6941	0.0282	1
10	35	1	0.5	1	2.1909	0.1139	0
11	35	1	0.5	1	2.6458	0.0326	1
12	35	1	0.5	1	2.3735	0.0705	0
13	35	1	0.5	1	2.1553	0.1244	0
14	35	1	0.5	1	1.5396	0.4782	0
15	35	1	0.5	1	0.9623	0.9541	0
16	35	1	0.5	1	1.7961	0.2870	0
17	35	1	0.5	1	2.4009	0.0655	0
18	35	1	0.5	1	1.8865	0.2356	0
19	35	1	0.5	1	1.8000	0.2846	0
20	35	1	0.5	1	2.5560	0.0423	1

*0= no se rechaza H_0 , 1= se rechaza H_0

Tabla 3-2.: Simulación de la prueba del supremo generalizado para $n = 35$

- Por último se promedian los valores de las variables indicadoras de los rechazos de la hipótesis nula (H_0) para encontrar el nivel de significancia o la potencia de la prueba alcanzados; en éste escenario se determina particularmente la potencia de la prueba del supremo generalizado, dado que el parámetro $\lambda_2=0.5$ indica que hay diferencia entre las CIF's, por tanto, la potencia estará dada por

$$\begin{aligned} 1 - \hat{\beta} &= \frac{1}{N} \sum I(\text{Rechazar } H_0 | H_0 \text{ es Falsa}) \\ &= \frac{6}{20} = 0.30 \end{aligned}$$

La potencia alcanzada en este estudio de simulación para la prueba del supremo generalizado es del 30%.

4. Análisis de resultados

En este capítulo se muestran los resultados y análisis del estudio de simulación realizado para cada escenario. La sección 4.1 presentan las tablas de los resultados de la simulación para cada tamaño de muestra. La sección 4.2 ilustra gráficamente los resultados de los errores tipo I obtenidos en los tres casos de censuras mostrando su comportamiento para los casos de dependencia en distintos tamaños de muestras. Por último, en la sección 4.3 se presentan los gráficos del comportamiento de las potencias empíricas para los diferentes valores de λ_2 en cada tamaño de muestra.

4.1. Tablas de resultados del estudio de simulación

		$\lambda_0 = 0.0$ y $\lambda_1 = 1.0$				$\lambda_0 = 1.0$ y $\lambda_1 = 1.0$				
		λ_2	Supremo	Bootstrap	RAS	Alternativa	Supremo	Bootstrap	RAS	Alternativa
SC		0.5	48.3	50.7	47.3	49.6	44.8	50.7	47.5	51.0
		1.0	5.0	7.1	7.1	7.6	3.2	4.3	3.9	5.0
		1.5	17.5	22.1	19.0	21.3	21.6	23.8	22.8	24.5
		2.0	43.4	49.7	50.0	49.4	42.4	55.3	49.1	52.1
		2.5	68.6	73.2	73.0	72.1	68.8	72.3	72.3	70.6
CM		0.5	33.6	39.1	43.1	44.7	34.7	38.3	40.1	43.8
		1.0	4.7	5.0	6.3	5.7	3.2	4.5	4.2	4.6
		1.5	12.6	16.6	17.5	16.9	14.6	21.5	11.7	20.4
		2.0	37.2	40.4	42.4	42.6	31.1	40.5	40.9	41.7
		2.5	59.6	57.2	64.2	66.1	55.3	62.2	64.7	64.5
CA		0.5	20.8	25.1	26.8	26.8	20.2	22.9	23.7	25.4
		1.0	2.9	5.0	5.6	3.9	1.7	3.9	3.5	2.7
		1.5	7.5	10.5	10.5	10.6	8.9	12.8	14.0	13.7
		2.0	19.6	23.5	27.3	24.8	20.4	21.5	25.0	27.1
		2.5	41.0	38.3	42.7	42.8	33.5	38.5	42.5	41.3

Tabla 4-1.: Tasas de error tipo I y potencias empíricas (en %) de la prueba $H_0 : \lambda_1(t) = \lambda_2(t)$ vs $H_1 : \lambda_1(t) \neq \lambda_2(t)$ con un nivel de significancia del 5% y $n = 50$.

La **Tabla 4-1** presenta la proporción de veces que se rechaza la hipótesis nula de manera que cuando λ_1 es igual a λ_2 (bajo H_0) estas proporciones corresponden a la tasa de error tipo I, mientras que en valores de λ_2 distintos a λ_1 (bajo H_1) tales porcentajes corresponden a las potencias empíricas; para los escenarios de independencia ($\lambda_1 = 0.0$) y dependencia ($\lambda_1 = 1.0$). Los valores de la tasa de error tipo I se presentan

		λ_2	$\lambda_0 = 0.0$ y $\lambda_1 = 1.0$				$\lambda_0 = 1.0$ y $\lambda_1 = 1.0$			
			Supremo	Bootstrap	RAS	Alternativa	Supremo	Bootstrap	RAS	Alternativa
Tipo de Censura	SC	0.5	80.2	77.8	77.5	79.2	79.2	79.6	78.4	79.0
		1.0	5.0	7.5	8.4	6.2	3.6	4.3	4.2	4.5
		1.5	35.0	34.3	37.0	35.7	39.7	38.5	38.6	38.4
		2.0	79.5	77.6	79.6	79.2	82.2	79.2	78.9	77.0
		2.5	95.4	95.4	94.3	95.4	95.1	94.8	89.0	93.3
	CM	0.5	68.8	69.7	69.4	71.7	69.2	70.2	66.2	75.8
		1.0	4.6	6.4	7.0	5.9	3.3	3.9	4.3	3.5
		1.5	30.7	30.4	27.7	33.1	31.8	32.2	29.2	34.7
		2.0	68.2	71.4	67.9	72.9	68.0	68.3	68.3	73.0
		2.5	90.7	89.2	90.6	92.8	90.4	89.5	88.9	92.6
	CA	0.5	45.5	44.7	48.5	45.9	43.6	45.1	48.9	46.9
		1.0	3.1	5.6	4.9	4.0	3.2	4.2	4.2	2.9
1.5		15.1	20.8	20.5	19.0	20.7	19.1	23.8	25.3	
2.0		44.8	44.9	49.5	48.5	45.7	44.8	44.0	48.4	
2.5		68.3	70.4	68.0	70.7	70.0	67.6	70.4	70.1	

Tabla 4-2.: Tasas de error tipo I y potencias empíricas (en %) de la prueba $H_0 : \lambda_1(t) = \lambda_2(t)$ vs $H_1 : \lambda_1(t) \neq \lambda_2(t)$ con un nivel de significancia del 5% y $n = 100$.

		λ_2	$\lambda_0 = 0.0$ y $\lambda_1 = 1.0$				$\lambda_0 = 1.0$ y $\lambda_1 = 1.0$			
			Supremo	Bootstrap	RAS	Alternativa	Supremo	Bootstrap	RAS	Alternativa
Tipo de Censura	SC	0.5	99.1	98.8	99.3	99.0	99.7	99.1	99.2	99.0
		1.0	7.5	6.6	6.5	7.0	3.3	4.1	4.6	3.7
		1.5	76.6	74.5	72.2	70.7	75.1	66.9	69.8	69.1
		2.0	99.3	99.0	99.5	99.2	99.7	99.4	98.6	99.5
		2.5	100	100	100	100	100	100	100	100
	CM	0.5	98.1	96.7	96.5	98.7	98.3	97.4	97.9	99.2
		1.0	6.0	6.1	7.4	7.3	3.8	4.1	5.3	4.3
		1.5	67.3	63.5	61.1	67.7	64.6	61.3	63.7	65.0
		2.0	98.6	97.7	97.2	98.6	98.9	97.7	97.3	98.7
		2.5	100	99.9	100	100	100	99.9	99.9	100
	CA	0.5	90.1	84.0	83.8	89.2	89.4	87.0	87.5	89.8
		1.0	4.9	6.1	6.6	4.1	2.9	3.8	4.9	3.8
1.5		44.7	44.7	43.4	42.1	45.5	42.0	47.1	46.8	
2.0		88.1	85.1	86.7	89.3	89.6	87.2	85.5	89.5	
2.5		99.1	98.4	97.9	98.6	98.6	98.5	97.7	97.9	

Tabla 4-3.: Tasas de error tipo II y potencias empíricas (en %) de la prueba $H_0 : \lambda_1(t) = \lambda_2(t)$ vs $H_1 : \lambda_1(t) \neq \lambda_2(t)$ con un nivel de significancia del 5% y $n = 250$.

en negrilla para facilitar la lectura de la tabla.

Las tablas 4-2 y 4-3 presentan los resultados del estudio de simulación cuando se tienen tamaños de muestras de $n = 100$ y $n = 250$ con sus respectivos resultados de las potencias empíricas y las tasa de error tipo I, tomando como referencia en los estudios de simulación un nivel de significancia de $\alpha = 0.05$.

Para analizar los resultados obtenidos en las tablas 4-1, 4-2 y 4-3 se contruyeron gráficos de líneas simultáneos del nivel de significancia alcanzado con relación a los diferentes tamaños de muestras. Las potencias empíricas se relacionan con los distintos valores asignados a λ_2 ; con el fin de evaluar la capacidad de las pruebas de detectar diferencias entre los dos riesgos a medida que se le asignan valores de λ_2 cercanos y alejados de $\lambda_2 = 1.0$, tanto en presencia y ausencia de censura para los escenarios de independencia y dependencia en cada tamaño de muestra.

A continuación se presenta un análisis gráfico del nivel de significancia y las potencias empíricas alcanzadas; las líneas de los gráficos se diferencia con distintos colores y formas para cada prueba como se muestra a continuación:

Supremo	—1—
Bootstrap	---2---
RAS3.....
Alternativa	-.-.-4-.-.-

4.2. Análisis gráfico del nivel de significancia alcanzado

En la **Figura 4-1** se observa el nivel de significancia bajo independencia (esto es, cuando $\lambda_0=0.0$ y $\lambda_2=1.0$), a medida que el tamaño de muestra crece, en ausencia de censura la prueba del supremo alcanza el nivel de significancia establecido en el estudio de simulación de $\alpha=0.05$ cuando tiene tamaños de muestra de $n=50$ y 100 , para muestras grandes el nivel de significancia supera el 5% pero es inferior al 9% mientras que las pruebas basadas en métodos de remuestreo Bootstrap, RAS y Bootstrap alternativo están por encima del nivel predeterminado. Se observa una pérdida en los niveles de significancia alcanzados debido a la presencia de censura. Las pruebas basadas en métodos de remuestreo en los tres casos de censura muestran un nivel de significancia por encima del nivel de significancia, mientras que la prueba basada en el supremo y la alternativa presentan mejores comportamientos bajo el escenario de independencia. En La **Figura 4-2** se presenta el comportamiento del nivel de significancia bajo el escenario de dependencia (esto es, cuando $\lambda_0=1.0$ y $\lambda_2=1.0$) observándose un buen comportamiento de las pruebas con relación al nivel de significancia dado tanto en ausencia como en presencia de censura.

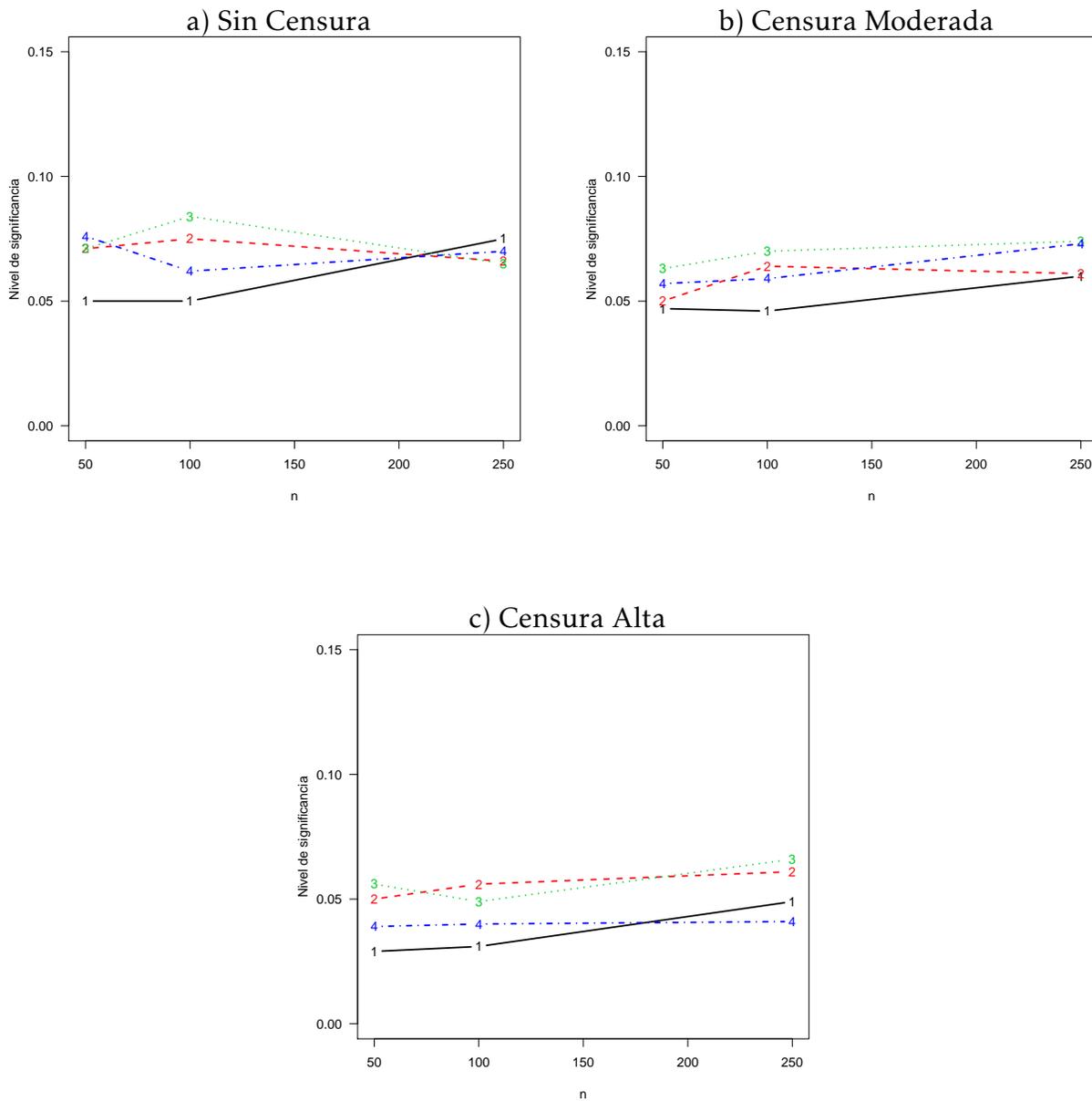


Figura 4-1.: Significancia alcanzada bajo independencia

La prueba del supremo en el estudio de simulación de ambos escenarios es más robusta que las pruebas basadas en los dos métodos de remuestreo propuestos por (Kam et al. 2002) debido a que sus errores tipo I no superan al nivel de significancia fijado en el escenario de dependencia, mientras que en el escenario de independencia los errores tipo I son más pequeños y cercanos a $\alpha = 0.05$ en comparación a las pruebas de remuestreo. Por último se puede observar que el método alternativo propuesto, presenta mejores potencias en el caso de independencia que la del método bootstrap y RAS; también presenta un buen comportamiento en el escenario de dependencia.

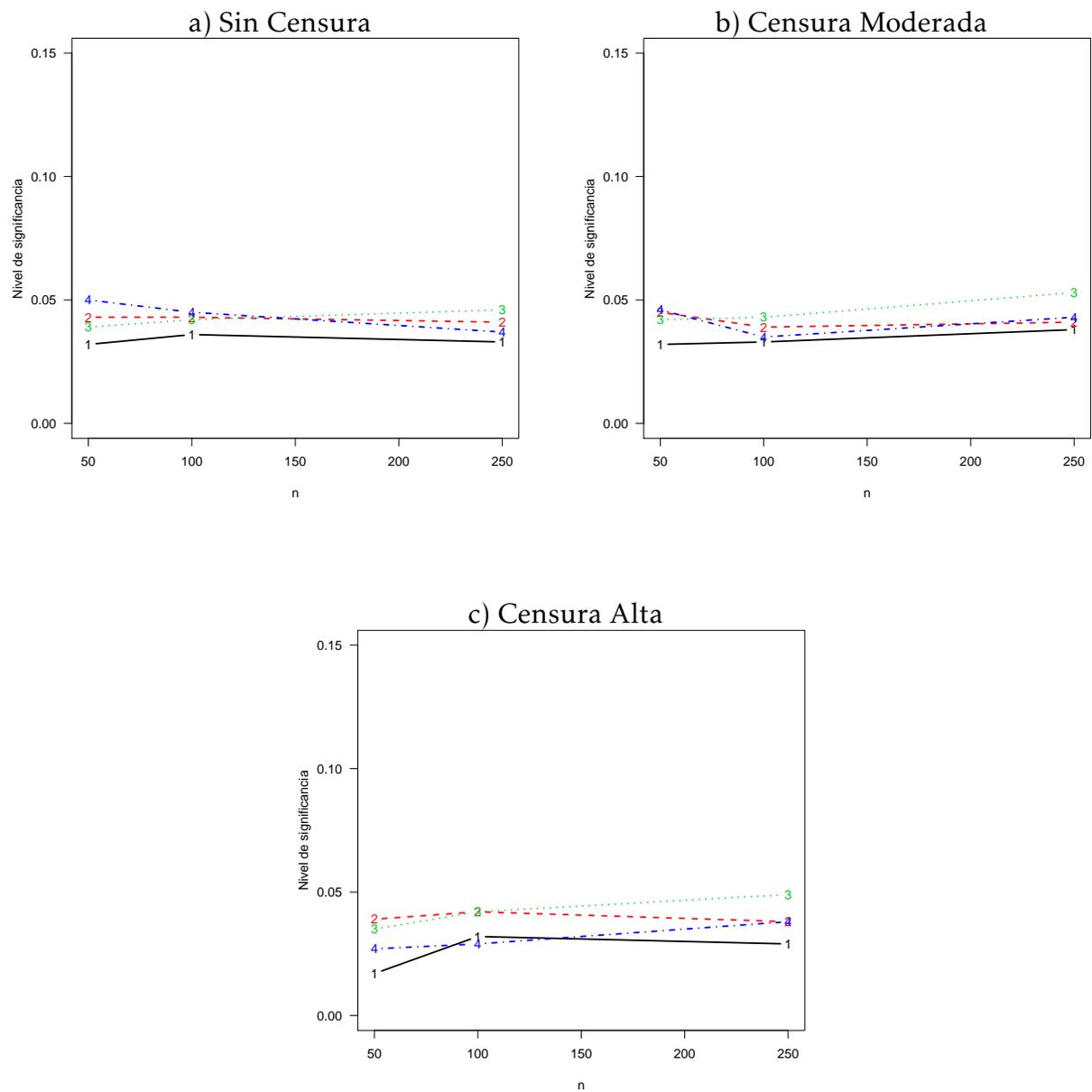


Figura 4-2.: Significancia alcanzada bajo dependencia

4.3. Análisis gráfico de potencia empírica

A continuación se presentan el análisis de las potencias empíricas para los diferentes tamaño de muestra tomados en el estudio de simulación a saber, $n = 50$, $n = 100$ y $n = 250$, a través de gráficos de líneas simultáneas que muestra el comportamiento de las potencias alcanzadas bajo dependencia e independencia y diferentes valores de λ_2 por las pruebas del supremo generalizado (1), método bootstrap (2) y **RAS** (3) , y el bootstrap alternativo (4).

Potencias empíricas para tamaños de muestra $n = 50$

En la **Figura 4-3** se visualizan los resultados presentados cuando hay independencia entre los tiempos donde se presenta el evento, se puede observar que en el caso de muestras pequeñas en presencia de censura hay una disminución considerable en las potencias empíricas alcanzando potencias por debajo del 15 %.

Cuando λ_2 toma el valor de 1.5 hay una pérdida de potencia muy marcada en las cuatro pruebas; las pruebas basadas en los métodos de remuestreo: bootstrap (2) y **RAS** (3) y el método alternativo presentan el mismo comportamiento a medida que el valor de λ_2 va cambiando mostrando una potencia máxima alrededor del 70 % pero inferior al 80 % en ausencia de censura. Se evidencia un mejor comportamiento en las pruebas basadas en métodos de remuestreo y la alternativa cuando los datos están en presencia de censura y en ausencia de ésta; mientras que la prueba basada en el supremo generalizado en los tres casos presenta potencias por debajo de las antes mencionadas.

En la **Figura 4-4** se presentan los resultados cuando hay dependencia entre los tiempos de falla potenciales. En ellos se revela que sigue el patrón de comportamiento de perdida de potencia en presencia de censura, tanto en el escenario de independencia como dependencia; hay una disminución considerable en las potencias empíricas alcanzando potencias por debajo del 15 % cuando $\lambda_2=1.5$. Además se puede observar en la **Figura 4-4** que las pruebas basadas en los métodos de remuestreo y la alternativa tienen comportamientos mas variables en cuanto a sus potencias cuando hay ausencia y presencia de censura, aunque la prueba basada en el método bootstrap (2) y la alternativa (4) presentan un comportamiento muy similar con mejores potencias cuando hay censura moderada y sin censura; en presencia de censura alta la alternativa sigue presentado un buen comportamiento, seguida de la prueba basada en el método **RAS** (3) y bootstrap (2). La prueba basada en el supremo generalizado en el escenario de dependencia sigue presentando potencias por debajo de las otras pruebas en

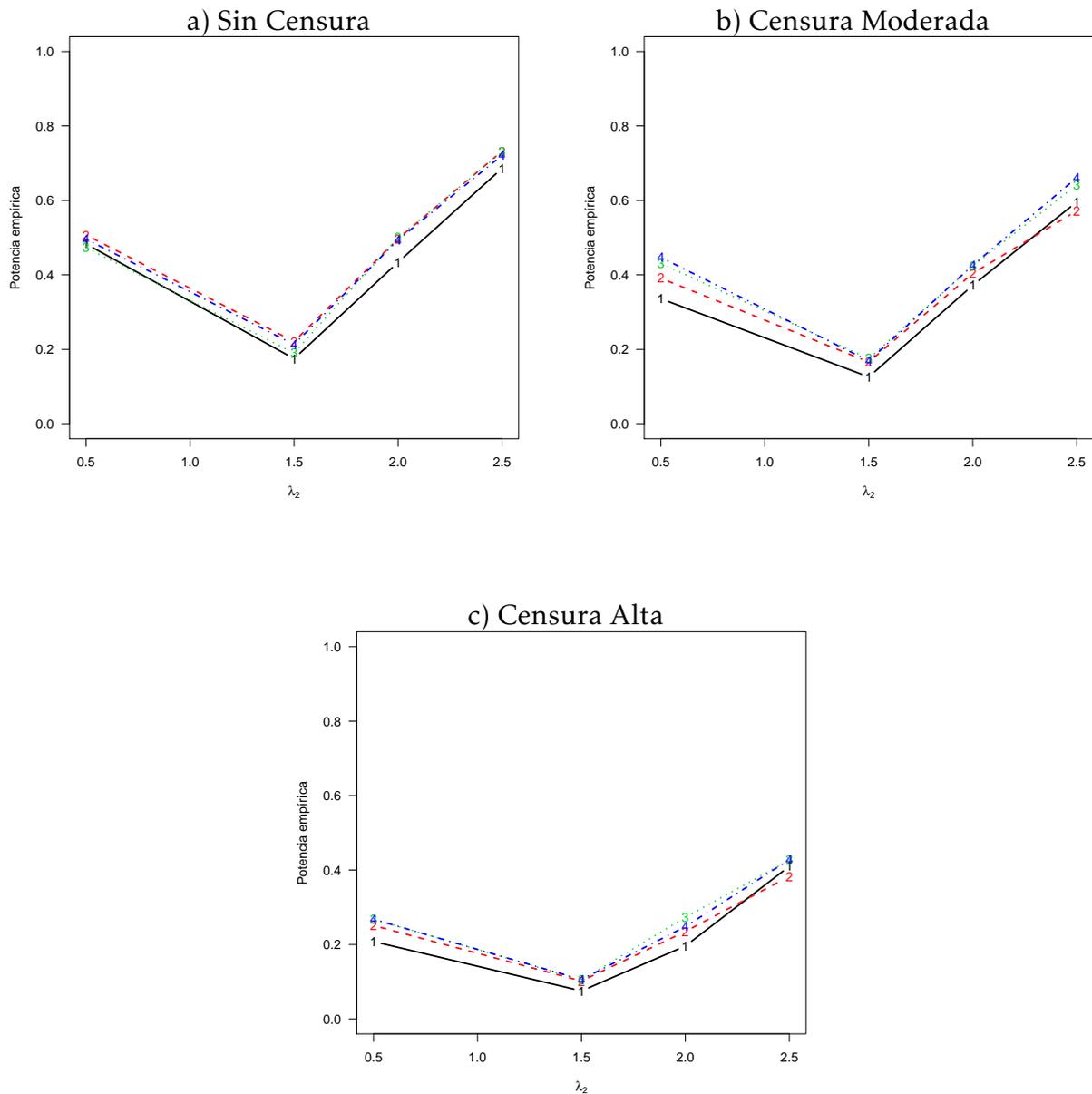


Figura 4-3.: Potencia empírica bajo independencia y tamaño de muestra $n = 50$

los tres casos de censura. Se puede evidenciar que bajo el escenario de independencia y dependencia en muestras pequeñas $n = 50$ las pruebas alcanzan potencias muy pequeñas cuando $\lambda=1.5$, en ambos casos las pruebas alcanzan mejores potencias en ausencia de censura, alcanzando su máxima potencia alrededor del 70 %, mientras que cuando están sometidas a censura alta sus potencias están por debajo del 50 %.

Por último en ambos escenarios las potencias basadas en las alternativa(4) propuesta, presenta un buen comportamiento en sus valores, en ocasiones superando a las pruebas de Kam et al. (2002) o igualándolas y superando la de Kochar et al. (2002).

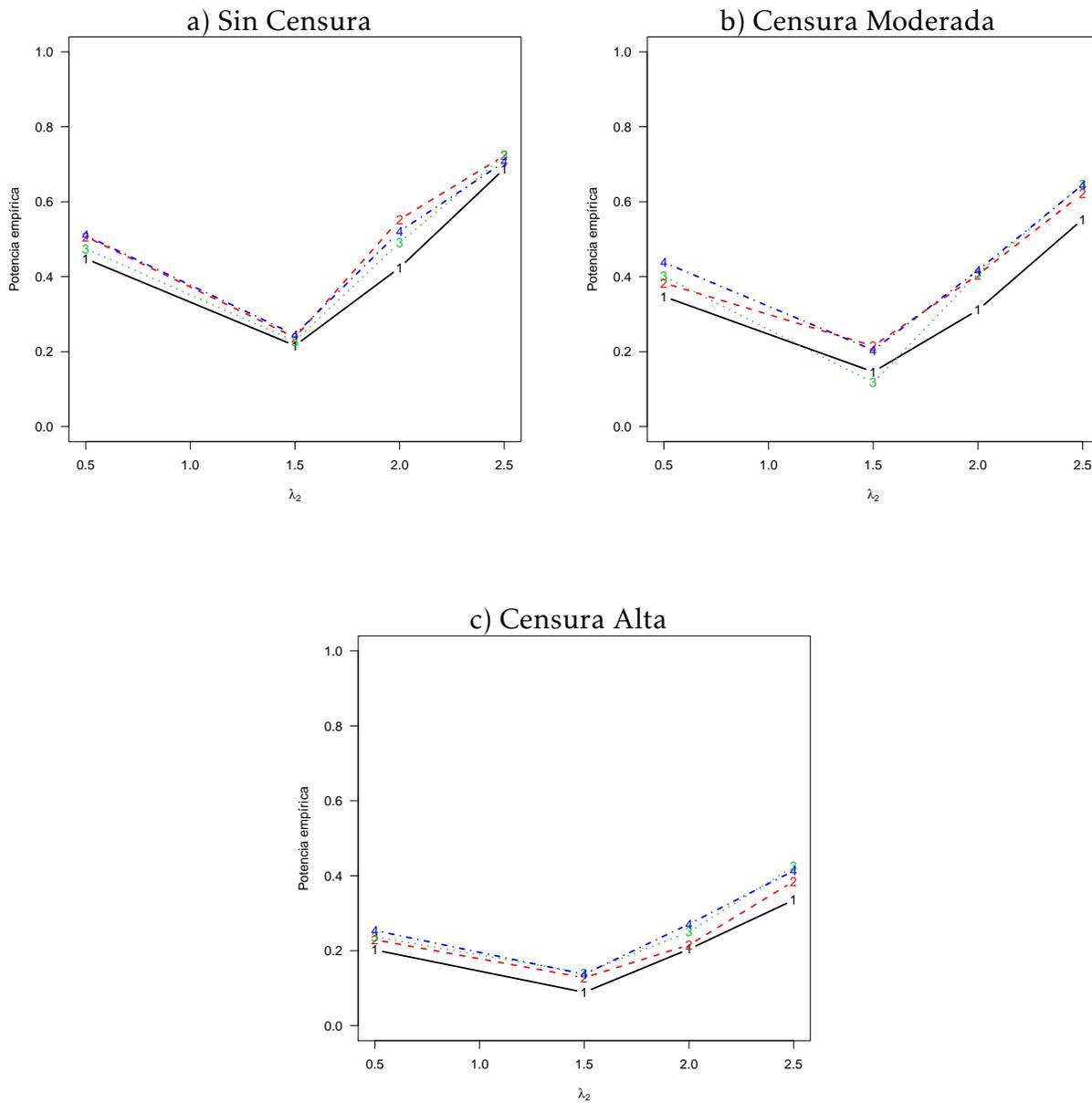


Figura 4-4.: Potencia empírica bajo dependencia y tamaño de muestra $n = 50$

Potencias empíricas para tamaños de muestra $n = 100$

Las Figuras 4-5 y Las 4-6 presentan los resultados para los casos de independencia y dependencia entre los tiempos de falla potenciales, para un tamaño de muestra $n = 100$; en ambos casos se observan comportamientos muy similares en las potencias empíricas para las diferentes pruebas, tanto en presencia de censura como en ausencia de ésta; se siguen observando mejores potencias en ausencia de censura, alcanzado potencias superiores al 90 % para las cuatro pruebas.

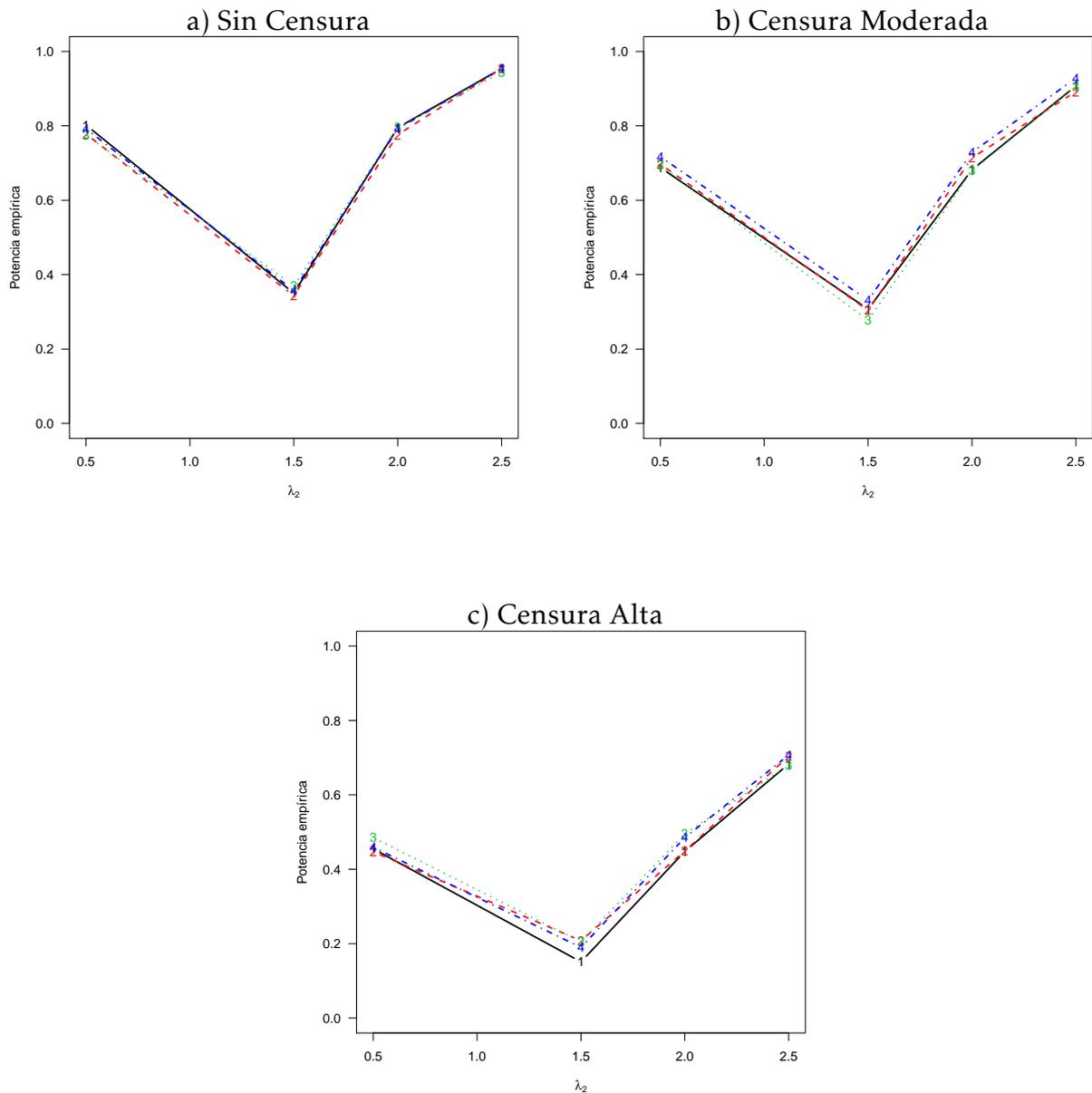


Figura 4-5.: Potencia empírica bajo independencia y tamaño de muestra $n = 100$

Se observa un poco de variabilidad cuando hay presencia de censura alta en las pruebas pero éstas se van estabilizando cuando λ_2 es igual o superior a 2.0; al igual que en el caso de muestras pequeñas cuando $\lambda_2 = 1.5$ se evidencia una caída en las potencias inferiores al 20% pero superiores al 10%. Teniendo como referencia este tamaño de muestra las potencias pueden alcanzar valores muy altos que superan el 90%. Nuevamente se ve un buen comportamiento en la alternativa (4).

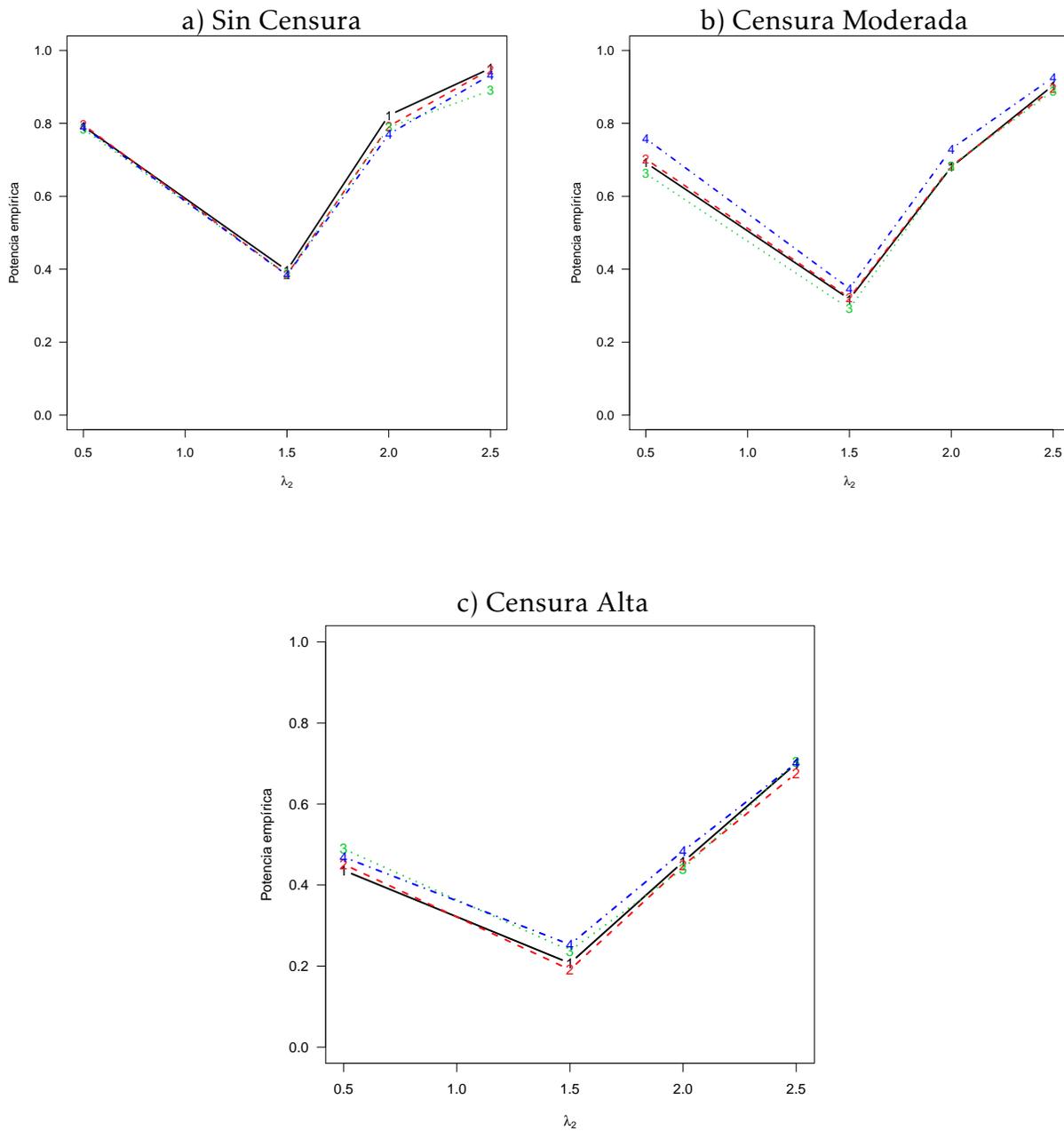


Figura 4-6.: Potencia empírica bajo dependencia y tamaño de muestra $n = 100$

Potencias empíricas para tamaños de muestra $n = 250$

En las Figuras 4-7 y 4-8 se visualizan los gráficos correspondientes a los casos de independencia y dependencia entre los tiempos de falla potenciales, para el tamaño de muestra $n = 250$; en el escenario de independencia se logra observar mejores potencias cuando se utiliza la prueba del supremo generalizado en los tres casos de censura, además cuando hay censura alta y moderada las cuatro pruebas tienen el mismo

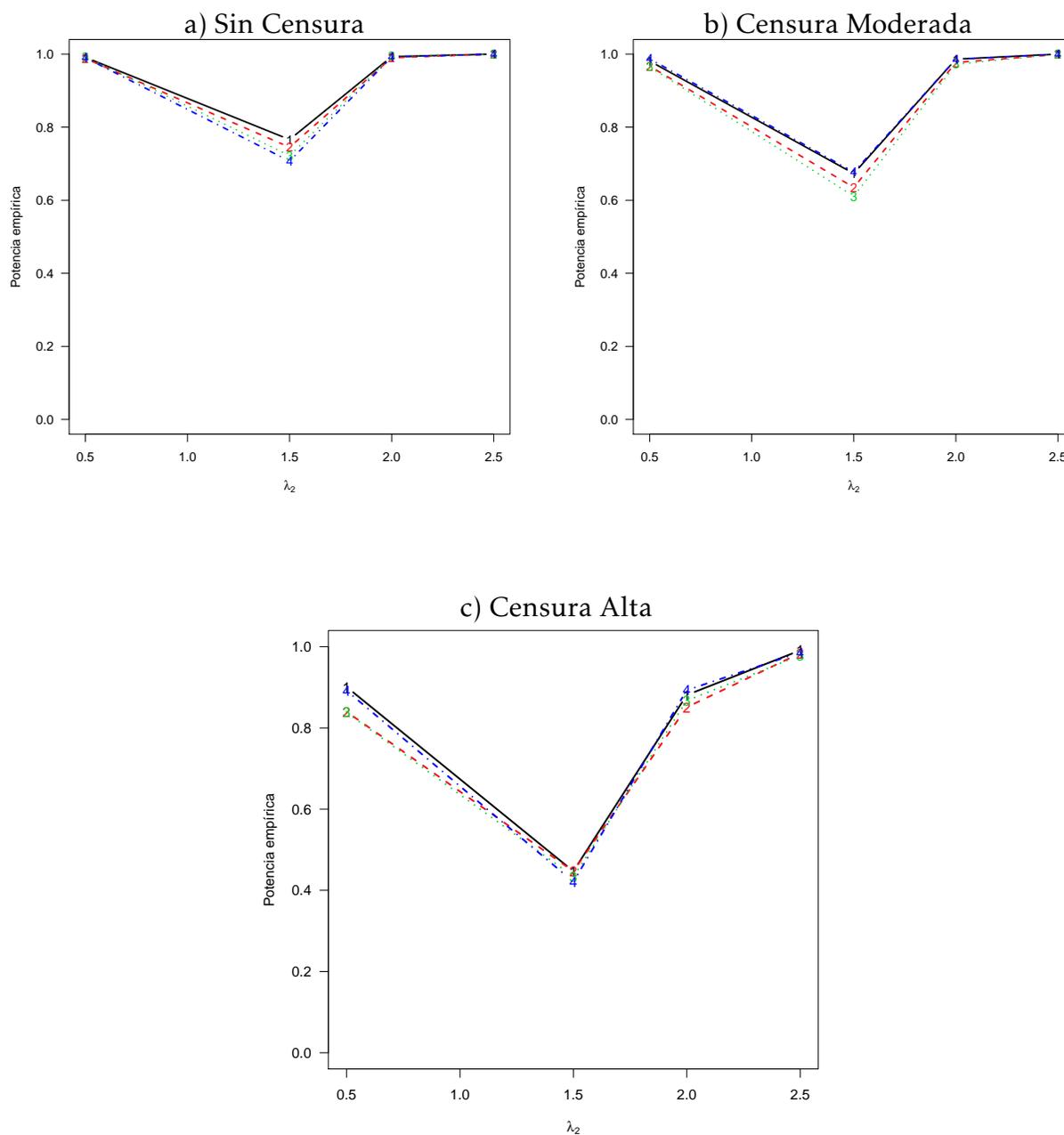


Figura 4-7.: Potencia empírica bajo independencia y tamaño de muestra $n = 250$

comportamiento. Observe que con este tamaño de muestra todas las pruebas alcanzan potencias cercanas o iguales a uno.

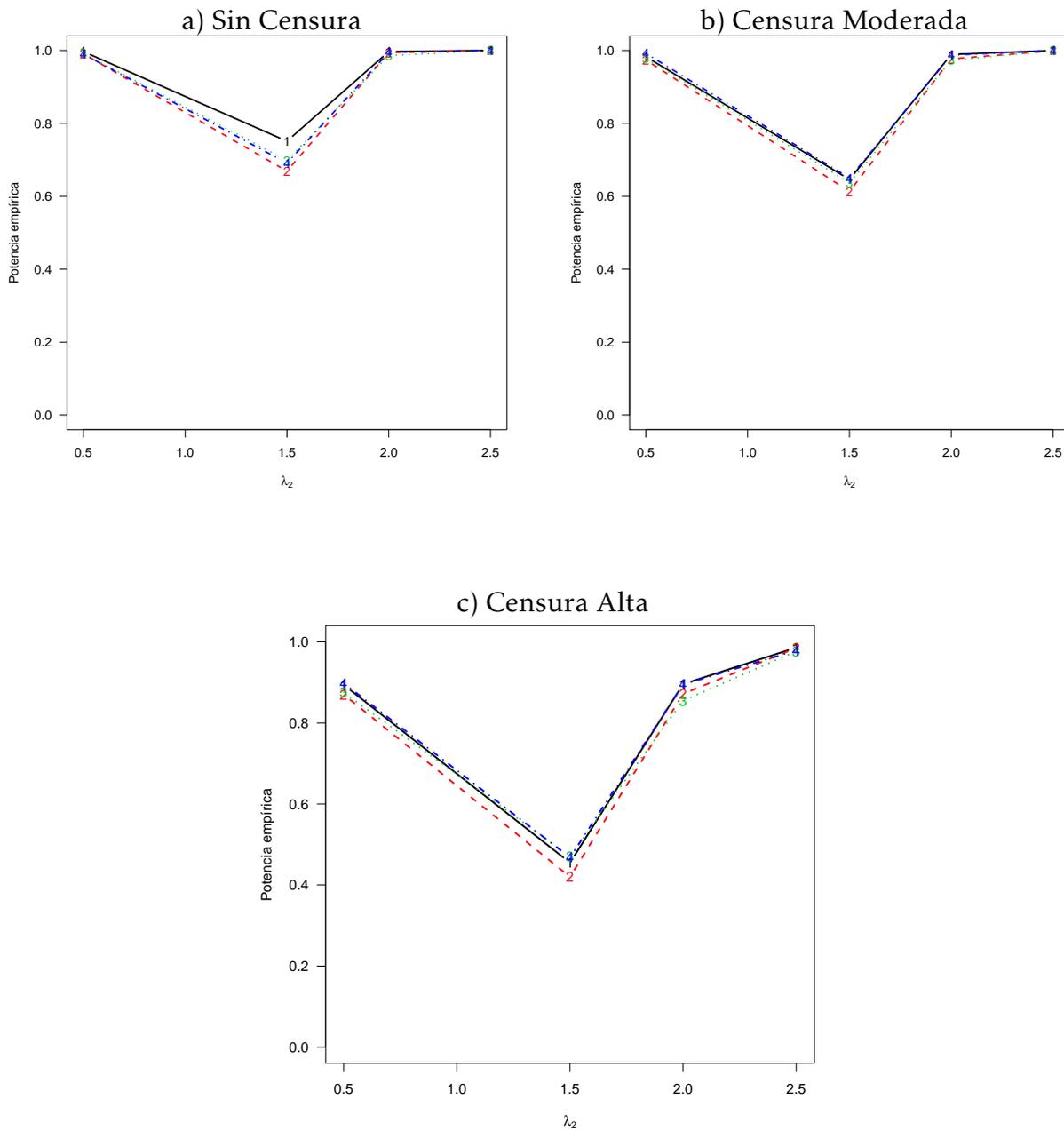


Figura 4-8.: Potencia empírica bajo dependencia y tamaño de muestra $n = 250$

Se sigue observando una caída en las potencias cuando $\lambda_2=1.5$ pero ya no tan bajas como en los tamaños de muestras anteriores, en este caso las potencias empíricas superan el 40 % .

En general de las secciones 4.2 y 4.3 se puede ver que:

- En el escenario de dependencia, la prueba del supremo sigue funcionando muy

bien, tiene un mejor comportamiento cuando no hay presencia de censura, comparada con las otras pruebas, pero en presencia de censura las cuatro pruebas tienen comportamientos parecidos, con potencias empíricas altas cercanas o iguales al 100 %.

- Se puede observar que en el caso de muestras pequeñas cuando no hay presencia de censura la prueba del supremo generalizado presenta errores Tipo I más cercanos al nivel de significancia fijado, tanto para el escenario de independencia como el de dependencia. En el caso de las pruebas **RAS**, bootstrap y en la prueba alternativa, se observa que aunque el error tipo I es pequeño en los tres casos, en el escenario de independencia los errores tipo I se alejan un poco del nivel de significancia establecido.
- Se observa que en tamaños de muestra grandes hay mayor estabilidad de resultados en las potencias de las pruebas para los dos escenarios, ya sea en presencia de censura o en ausencia de esta. Cuando $\lambda_2 = 1.5$ en el caso de muestra grandes y en presencia de censura hay una disminución en las potencias empíricas pero no tan marcados como en el caso de muestras pequeñas, en el caso de la muestra $n = 250$ se puede ver que ya las potencias empíricas tienen mayor estabilidad tanto en presencia de censura como en ausencia de ésta, en las cuatro pruebas.
- En general las pruebas basadas en métodos de remuestreo bootstrap y **RAS** tienen un buen funcionamiento en ambos escenarios, pero especialmente en los casos de dependencia entre los tiempos de fallas, sus errores tipos I son más cercanos al nivel de significancia.
- En todos los casos, las potencias empíricas convergen de manera constante al número uno cuando aumenta el tamaño de la muestra. También se observa errores tipo I más cercanos al nivel de significancia en la pruebas del supremo generalizado y alternativa cuando la censura es alta para ambos escenarios, mientras que los métodos bootstrap y **RAS** siguen teniendo mejor comportamiento en en los escenarios de dependencia.
- En los diferentes tamaños de muestra se observa que a medida que el valor de la tasa del segundo evento aumenta las potencias empíricas de las pruebas son muy altas, mientras que, cuando los valores de dicha tasa son cercanos a uno, las pruebas son un poco inestables alcanzando potencias muy bajas que se van estabilizando a medida que el tamaño de muestra va creciendo.

- Por último en la diferentes gráficos, los resultados de la simulación para la propuesta bootstrap usando los estimadores de las CIF's y Kaplan-Meier, presentan comportamientos similares a las pruebas basadas en los métodos de remuestreo: bootstrap y **RAS**, donde se ve evidenciado nuevamente que en presencia de censura las potencias empíricas disminuyen pero a medida que el tamaño de la muestra crece su potencias empíricas también; en cuanto a los errores tipo I, la prueba alternativa muestra un mejor comportamiento que los propuestos por Kam et al. (2002).

En el Capítulo 5 se presenta un ejemplo con datos reales de pacientes con linfoma para evaluar los resultados de las pruebas diferentes pruebas estudiadas, de acuerdo al escenario de censura de los datos a estudiar.

5. Ejemplo: Pacientes con linfoma

Pintilie (2006) presenta una base de datos de los pacientes con linfoma creada en el hospital Princess Margaret, Toronto, con registros que datan desde 1967. En la actualidad, la base de datos es de tipo prospectivo, pacientes que están siendo introducidos y registrados para recibir tratamiento en el hospital. Esta base de datos contiene un subgrupo de 541 pacientes de todos los pacientes identificados con linfoma tipo folicular, registrados para tratamiento en el hospital entre 1967 y 1996, con enfermedad en la etapa temprana y tratados con sólo radioterapia o con radiación y quimioterapia. El objetivo de este estudio era mostrar los resultados a largo plazo en este grupo de pacientes. El resultado registrado incluye la respuesta al tratamiento, la primera recaída y la muerte. El tiempo hasta la primera falla se calcula en años a partir de la fecha de diagnóstico. Un informe sobre una parte de este conjunto de datos se puede encontrar en Petersen et al. (2004). De la anterior base de datos se tomó una muestra aleatoria de 74 pacientes, para determinar si hay diferencias entre los riesgos de recaída y muerte sin recaída, después del tratamiento. A continuación se presenta una parte de la base de datos de la muestra analizada

ID PACIENTE	TIEMPO (DÍAS)	TIPO DE EVENTO*
246	9.919233402	1
440	4.010951403	1
428	7.797399042	0
97	3.058179329	1
136	6.844626968	2
41	0.002737851	1
348	0.826830938	2
522	3.455167693	0
469	7.449691992	0
516	1.385352498	1
115	11.824777550	2

*0 = no se observó el evento, 1 = recaída y 2 = muerte sin recaída

Tabla 5-1.: Muestra de recaída y muerte sin recaída de pacientes con linfoma folicular

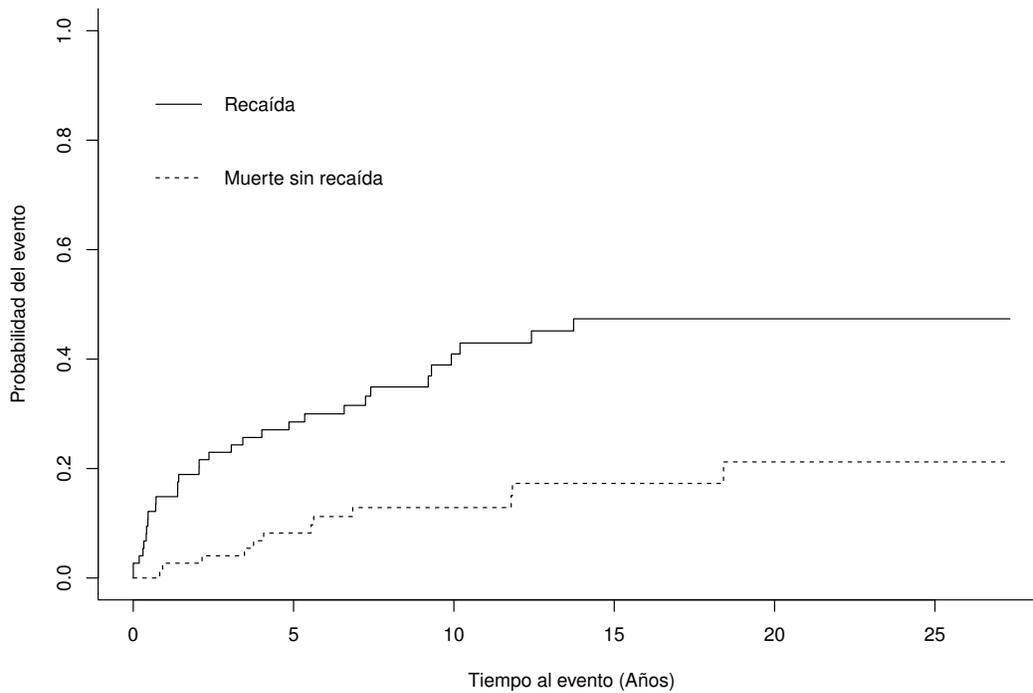


Figura 5-1.: Comparación del riesgo por recaída con el riesgo por muerte sin recaída de los datos de pacientes con linfoma

Para comparar la incidencia de recaída con la incidencia de morir sin recaída después del tratamiento en el conjunto de datos antes mencionado; la recaída se codificó como 1, la muerte sin recaída como 2 y los datos censurados con 0. Los valores p correspondientes a los estadísticos del supremo generalizado, los métodos bootstrap, RAS y la propuesta se muestran a continuación

Estadístico de prueba	Valor del estadístico	Valor - p
Supremo	2.7449	0.0242
RAS	1.4779	0.0010
Bootstrap	1.4779	0.0004
Propuesta	4.5625	0.0030

Tabla 5-2.: Estadísticos de prueba (Valor- p) para los datos de pacientes con linfoma

Tomando un nivel de significancia de $\alpha = 0.05$ los valores p de las pruebas son significativos, estos valores sugieren que hay una diferencia entre la incidencia de recaídas y

la incidencia de morir sin recaída. En la **Figura 5-1** la CIF de recaída está muy por encima de la CIF para la muerte sin recaída, apoyando la conclusión encontrada usando los estadísticos de prueba estudiados.

Aunque las tres pruebas arrojan la misma decisión cabe mencionar que de acuerdo a la muestra que se tomó de 74 pacientes, ésta posee un 41.89% de datos censurados, por tanto de acuerdo al estudio de simulación se está en un caso de censura moderada con una muestra relativamente pequeña; si existe alguna sospecha de que hay diferencias entre los dos riesgos de acuerdo al estudio de simulación en el caso de muestra pequeña es recomendable usar la pruebas basadas en los métodos de remuestreo, dado que durante el estudio presentaron mejores comportamientos en comparación a la prueba del supremo generalizado en el momento en que hay presencia de censura.

Si existe alguna sospecha de que hay diferencias entre los dos riesgos de acuerdo al estudio de simulación en el caso de muestra pequeña es recomendable usar la prueba basada en el método bootstrap alternativo, dado que durante el estudio arrojó mejores potencias en comparación a las otras tres y en presencia de censura el nivel de significancia alcanzado es cercanos al valor de referencia $\alpha = 0.05$.

6. Conclusiones y trabajo futuro

En el presente trabajo se estudió vía simulación cuatro pruebas existentes para probar la igualdad de dos riesgos competitivos a través de las funciones de incidencia acumulada. Se ilustró la utilidad práctica de los métodos y la propuesta usando datos reales.

Se pudo evidenciar bajo los escenarios estudiados que las pruebas tienen un buen funcionamiento, cuando los tamaños de muestras son grandes. La prueba del supremo presenta errores tipo I más cercanos al nivel nominal fijado en comparación a las otras pruebas en todos los escenarios. En relación a la potencia cuando $n = 50$ esta prueba tiene un desempeño inferior a las pruebas basadas en remuestreo y para muestras grandes su desempeño depende de si se tiene o no censura.

Las pruebas basadas en métodos de remuestreo alcanzaron errores tipo I pequeños pero mostraron un mejor comportamiento cuando se encuentra en un escenario de dependencia; computacionalmente el método **RAS** es más eficiente que el bootstrap. En relación a la potencia cuando $n=50$ los métodos de remuestreo funcionan mejor en comparación a la prueba del supremo, destacándose el método alternativo que está a la cabeza de sus pares y para muestras grandes cuando no hay censura el método del supremo alcanzó mayores potencias pero en escenarios con censura se observó que el método alternativo es muy buen competidor.

Una buena alternativa cuando hay presencia de censura es usar la propuesta del método de bootstrap alternativo usando las estimaciones de las CIF's y Kaplan-Meier, dado que estas metodologías están diseñadas especialmente para tratar datos en presencia de riesgos competitivos y eventos censurados, respectivamente. En los estudios de simulación realizados este método presentó un mejor comportamiento que los obtenidos por el método bootstrap y la aproximación **RAS**, e incluso supera en algunos escenarios al método del supremo.

Para trabajos futuros sería importante evaluar el desempeño de las pruebas cuando hay más de dos riesgos competitivos presentes, usando como alternativa la distribución geométrica bivariada discreta para modelar los datos de tiempo de vida, según

Davarzani et al. (2015) esta podría traer algunas ventajas computacionales en comparación a la distribución de Block & Basu (1974). Además, queda como trabajo futuro demostrar la propiedades de convergencia del método de bootstrap alternativo propuesto basado en las estimaciones de las CIF's y Kaplan-Meier.

Apéndice

A. Anexos: Códigos en R de la simulación

En el siguiente código se presenta un ejemplo de la de los parámetros de entrada de la simulación con sus respectivas probabilidades, estas probabilidades cambian de acuerdo a los valores que tomen los parámetros de la distribución de (Block & Basu 1974), usando la técnica descrita en (Leandro & Achcar 1996) para cada uno de los casos.

```
setwd("C:\\Users\\supremo")

# Preparar el archivo de resultados
library(snowfall)
sfInit(parallel=TRUE, cpus=4, type="SOCK")
sfLibrary(survival)
sfLibrary(cmprsk)
sfSource("funcioneSupremo.R")
sfSource("parametroSupremoCaso1i1.R")

# Correr un caso de la simulación 1000 veces (paralelo=F)
system.time(resultado <- sfLapply(1:Ns, compara.riesgos, n, lambda,
censura, celdas, probs))
sfStop()

res <- c(Ns, n, lambda1, lambda2, lambda3, censura,
colMeans(do.call(rbind, resultado)))
write(res, file=paste("Resultado_Ns", Ns, "n", n, "l1_", lambda1,
"l2_", lambda2, "l3_", lambda3, "cens_", censura, ".txt", sep=""),
append=T, ncolumns=7)
```

```
# INICIO Parámetros de la simulación
n <- 250 # Valores de simulación 50,100,250
N <- 1000*(n/100)
Nr <- 5000
Ns <- 1000
lambda1 <- 1
lambda2 <- 0.5 # Valores de simulación 0.5,1.0,1.5,2.0,2.5
lambda3 <- 1 # Valores de simulación 0.0,1.0
lambda12 <- lambda1+lambda2
lambda13 <- lambda1+lambda3
lambda23 <- lambda2+lambda3
lambda <- lambda12+lambda3
p <- lambda1/lambda12
alfa <- 0.05

# Probabilidades de celda de acuerdo a cada escenario
probs <- c(0.5941,0.1826,0.0407,0.0091,0.0767,0.0488,0.0150,0.0033,
           0.0104,0.0063,0.0040,0.0012,0.0014,0.0009,0.0005,0.0003)

censura <- 0 # Valores de censura 0:sin cens, 0.2:18-35, 0.5:45-60

# Determinar las clases de celda en una matriz
#(dependen de cada escenario)
lmaxx <- 4 #ceiling(maxi[1]/1)*1
lmaxy <- 4 #ceiling(maxi[2]/1)*1
limx <- seq(0,lmaxx,by=1)
limy <- seq(0,lmaxy,by=1)
xl <- limx[-length(limx)]
yl <- limy[-length(limy)]
xu <- limx[-1]
yu <- limy[-1]
xlc <- unlist(lapply(xl,rep,length(limy)-1))
xuc <- unlist(lapply(xu,rep,length(limy)-1))
ylc <- rep(yl,length(limx)-1)
yuc <- rep(yu,length(limx)-1)
celdas <- cbind(xlc,xuc,ylc,yuc)
# FIN Parámetros de la simulación

# Preparar el archivo de resultados
namesres <- c("n","lambda1","lambda2","lambda3","censura","replica",
             "epkly","vpkly","Decisionkly")
```

```

write(namesres,file=paste("n_",n,"lambda1_",lambda1,"lambda2_",lambda2,
"lambda3_",lambda3,"censura_",censura,".txt",sep=""),
append=F,ncolumns=9)

nameres <- c("Ns","n","lambda1","lambda2","lambda3","censura",
"errortipoI_kly")
write(nameres,file=paste("Resultado_Ns",Ns,"n",n,"l1_",lambda1,"l2_",
lambda2,"l3_",lambda3,"cens_",censura,".txt",sep=""),
append=F,ncolumns=7)

```

Código de funciones en R para generar los datos de la distribución de Block & Basu en R:

```

# Generación de valores de T_1, T_2 a partir de
#un mínimo y un delta, ambos en "x"
GeneraPareja <- function(x){
  if(x[2]==1){
    X <- x[1]
    py <- (lambda/lambda12)*exp(-lambda23*X)-
    lambda3/lambda12*exp(-lambda*X)
    Y <- (lambda/(lambda12*py))*(X*exp(-lambda23*X)+
    (1/lambda23)*exp(-lambda23*X)-(lambda3/lambda)*X*exp(-lambda*X)
    -(lambda3/lambda^2)*exp(-lambda*X)) #hay un error en la fórmula
    px <- 0
  } else {
    Y <- x[1]
    px <- (lambda/lambda12)*exp(-lambda13*Y)-
    (lambda3/lambda12)*exp(-lambda*Y)
    X <- (lambda/(lambda12*px))*(Y*exp(-lambda13*Y)+
    (1/lambda13)*exp(-lambda13*Y)-(lambda3/lambda)*Y*exp(-lambda*Y)
    -(lambda3/lambda^2)*exp(-lambda*Y)) #hay un error en la fórmula
    py <- 0
  }
  data <- c(X,Y)
  data
}

```

```
# Sacar probabilidades de "celda" de la distribución ACBVE
#basada en celdas cuadradas
probs.j1 <- function(celda){
x1 <- celda[1]
xu <- celda[2]
y1 <- celda[3]
yu <- celda[4]
if(xu<=y1){
phi1 <- (lambda/lambda12)*(exp(-lambda1*xu)-
        exp(-lambda1*x1))*(exp(-lambda23*yu)-exp(-lambda23*y1))
phi1} else if(yu<=x1){
phi2 <- (lambda/lambda12)*(exp(-lambda2*y1)-
        exp(-lambda2*yu))*(exp(-lambda13*x1)-exp(-lambda13*xu))
phi2} else {
phi3 <- (lambda/lambda12)*exp(-lambda23*yu)*(exp(-lambda1*xu)-
        exp(-lambda1*x1))+(lambda/lambda12)*exp(-lambda2*y1)
        *(exp(-lambda13*x1)-exp(-lambda13*xu))+
        (lambda3/lambda12)*(exp(-lambda*xu)-exp(-lambda*x1))
phi3}
}
# Sacar frecuencias de "celda" de una matrix de "datos"
frecs.j1 <- function(celda,datos=datos){
x1 <- celda[1]
xu <- celda[2]
y1 <- celda[3]
yu <- celda[4]
tmp <- subset(datos,datos[,1]<=xu & datos[,1]>x1)
tmp1 <- subset(tmp,tmp[,2]<=yu & tmp[,2]>y1)
frec <- nrow(tmp1)
}
# Obtener muestra de tamaño n_j1*p_j1 de cada celda
muestra.j1 <- function(muestra,datos=datos){
x1 <- muestra[1]
xu <- muestra[2]
y1 <- muestra[3]
yu <- muestra[4]
tmp <- subset(datos,datos[,1]<=xu & datos[,1]>x1)
```

```
tmp1 <- subset(tmp,tmp[,2]<=yu & tmp[,2]>y1)
if(nrow(tmp1)>=muestra[5] & muestra[5]>0){
tmp2 <- sample(nrow(tmp1),muestra[5],replace=F)
tmp3 <- datos[tmp2,]} else tmp3 <- NULL
tmp3
}

# Incorporar censura a derecha en datos generados
datoscens <- function(fulldata){
datamin <- min(fulldata)
delta <- which(fulldata==datamin)%% 3
cbind(datamin,delta)
}

# Remuestras de una base de datos
remuestras <- function(index,datos,n){
datosrem <- datos[index,]
while(table(factor(datosrem[,2],levels=c(0,1,2)))[[2]]
==0|table(factor(datosrem[,2],levels=c(0,1,2)))[[3]]==0){
datosrem <- datos[sample(n,replace=T),]
}
datosrem
}

#ordenar los datos de la muestra
ordena <- function(df){
orden <- order(df$min)
dford <- df[orden,]
dford
}

#Remuestreos de los datos y organización de tiempos de la remuestra
#para el método RAS
remuest <- replicate(Nr,sample(n,replace=T))
datosrem <- apply(remuest,2,remuestras,datoscr,n) #Remuestra
datoscr.ord.rem <- datosrem #Organizan los tiempos
for(i in 1:Nr)
{
datoscr.ord.rem[[i]] <- datosrem[[i]][order(datosrem[[i]][,1]),]
}
```

El siguiente código en R de la prueba del supremo generalizado C_n^* , es tomado de (Pintilie 2006):

```
kly=function(time,cens)
{
#####
# based on Kochar, Lam and Yip #
# Lifetime Data Analysis,2002 #
# time=time to first event    #
# cens=censor variable        #
#     1=one type of event     #
#     2=another type of event #
#     0=the rest of events and#
#     censored records        #
#####

dd0=table(time,cens)
if (sum(cens==0)==0) {dd=cbind(rep(0,nrow(dd)),dd0)}
else {dd=dd0}
if (sum(cens==1)==0 | sum(cens==2)==0 | sum(cens>2)>0)
{stop ("Codes for events should be 1 or 2")}
n=length(time)
sigma=sqrt(n*sum(cens>0))
nt=apply(dd,1,sum)
ntt=cumsum(nt)
atrisk=n-ntt
lnti=(dd[,3]-dd[,2])
lnt=cumsum(lnti)
k=nrow(dd)
maxlnt=vector("numeric",k)

for (i in 1:k)
{
maxlnt[i]=max(abs(lnt-lnt[i]))
}
cnw=sqrt(n)*max(maxlnt)/sigma
## for C, (without absolute value)
pcnw=function(cc)
```

```

{
k=c(0:15)
parti=(-1)^k*exp(-pi^2*(2*k+1)^2/(8*cc^2))/(2*k+1)
p=sum(parti)*4/pi
return(p)
}

## for Cstar (with absolute value)
ph=function(z)
{
f1=function(x1,x2,k,xi)
{
ki=c(1:k)
x=c((x1*xi):(x2*xi))/xi
n=length(x)
h=vector("numeric",n)

for (i in 1:n)
{
h[i]=sum(8*(-1)^(ki-1)*ki^2*dnorm(ki*x[i]))
}

parth=sum(h)/xi
return(parth)}

if (z<=0.1)
ph=f1(0.01,z,1000,100)
else if (z<=0.5)
ph=f1(0.01,0.1,1000,100)+f1(0.11,z,100,100)
else ph=f1(0.01,0.1,1000,100)+f1(0.11,0.5,100,100)+
f1(0.51,z,30,1000)
return(ph)
}
pvalue=1-ph(cnw)
if (pvalue>=0.00001) r=data.frame(cnw,pvalue)
if (pvalue<0.00001) r=data.frame(cnw,pvalue=0) #"p-value<0.00001"
row.names(r)=""
return(r)
}

```

Código de las funciones en R para evaluar la prueba del supremo generalizado C_n^* :

```
# Evaluar los estadísticos de prueba en una simulación
compara.riesgos<-function(replica,n,lambd,censura,celdas,probs)

{
# Genera N valores de un mínimo y le asocia un modo
#de falla con probabilidad p
minimo <- rexp(N,lambd)
delta <- rbinom(N,1,p)
datmin <- as.data.frame(cbind(minimo,delta))
datos <- as.data.frame(t(apply(datmin,1,GeneraPareja)));
names(datos) <- c("X","Y")

# Frecuencias de celda
allfrecs <- cbind(celdas,apply(celdas,1,frecs.jl,datos))

# Probabilidades de celda desde derive y en la forma
#de la tabla de frecuencias
allprobs <- cbind(celdas,probs)

# Sacar una muestra de tamaño n con base en las probs de
#celda (sin censura)
nm <- 1
while(nm!=0){
muestra1d <- rmultinom(1,n*1.1,prob=probs)
muestra1d <- cbind(celdas,muestra1d)
muestra <- apply(muestra1d,1,muestra.jl,datos)
muestrafull <- do.call(rbind, muestra)
nm <- n-nrow(muestrafull)
}

# Modifica los datos incorporando censura a derecha
if(censura==0) ctime <- rep(N,n) else{ #censura=P(C<MIN(T1,T2))
gama <- censura*lambd/(1-censura)
ctime <- rexp(n,gama)
}
```

```

datafull <- cbind(muestrafull,ctime)
datoscr <- t(apply(datafull,1,datoscens));
colnames(datoscr) <- c("min","delta")
datoscr <- as.data.frame(datoscr)

# Estadístico de Kochar et al (2002)
estadkly <- kly(datoscr$min,datoscr$delta)
epkly <- estadkly$cnw
vpkly <- estadkly$pvalue
Decisionkly <- (vpkly<alfa)+0

result <- c(n,lambda1,lambda2,lambda3,censura,replica,epkly,
           vpkly,Decisionkly)

write(result,file=paste("n_",n,"lambda1_",lambda1,
"lambda2_",lambda2,"lambda3_",lambda3,
"censura_",censura,".txt",sep=""),append=T,ncolumns=9)
c(Ddecisionkly)
}

```

Código de las funciones en R de la prueba basada en el método bootstrap D_n^* :

```

# Calculos en las remuestras
estadrem <- function(Datos.ordenados.rem,c1,c2){
eventosrem <- ifelse(Datos.ordenados.rem$delta>'0',1,0)/n
evento1rem <- ifelse(Datos.ordenados.rem$delta=='1',1,0)/n
evento2rem <- ifelse(Datos.ordenados.rem$delta=='2',1,0)/n

#Acumuladas empíricas remuestradas
Acum.Emp.rem <- cumsum(eventosrem)
Acum.Emp1.rem <- cumsum(evento1rem)
Acum.Emp2.rem <- cumsum(evento2rem)
Datosfull.rem <- data.frame(Datos.ordenados.rem,eventosrem,
evento1rem,evento2rem,Acum.Emp.rem,
Acum.Emp1.rem,Acum.Emp2.rem)

```

```

#Funciones Empíricas
f.acum.emp1r <- stepfun(c(Datosfull.rem$min),
c(0,Datosfull.rem$Acum.Emp1.rem))
f.acum.emp2r <- stepfun(c(Datosfull.rem$min),
c(0,Datosfull.rem$Acum.Emp2.rem))
f.acum.empr <- stepfun(c(Datosfull.rem$min),
c(0,Datosfull.rem$Acum.Emp.rem))
#Tamaños de los saltos
saltosr <- diff(c(0,Datosfull.rem$Acum.Emp.rem))
c1r <- f.acum.emp1r(Datosfull.rem$min)
c2r <- f.acum.emp2r(Datosfull.rem$min)

Wcif1r <- n^(1/2)*(c1r-c1) #Construcción del W para cif1r
Wcif2r <- n^(1/2)*(c2r-c2) #Construcción del W para cif2r
Wcif12r <- Wcif2r-Wcif1r

Dnr <- sum((Wcif12r)^2*saltosr)
Dnr #Estadístico Remuestro Cramer Von Mises
}

```

Código de la funciones en R para evaluar la prueba basada en el método bootstrap D_n^* :

```

# Evaluar los estadísticos de prueba en una simulación
compara.riesgos <- function(replica,n,lambd,censura,celdas,probs){
# Genera N valores de un mínimo y le asocia un modo de falla
#con probabilidad p
minimo <- rexp(N,lambd)
delta <- rbinom(N,1,p)
datmin <- as.data.frame(cbind(minimo,delta))
datos <- as.data.frame(t(apply(datmin,1,GeneraPareja)));
names(datos) <- c("X","Y")
# Frecuencias de celda
allfrecs <- cbind(celdas,apply(celdas,1,frecs.jl,datos))

# Probabilidades de celda desde derive y en la forma de
#la tabla de frecuencias
allprobs <- cbind(celdas,probs)

```

```
# Sacar una muestra de tamaño n con base en las probs de
#celda (sin censura)
nm <- 1
while(nm!=0){
muestra1d <- rmultinom(1,n*1.1,prob=probs)
muestra1d <- cbind(celdas,muestra1d)
muestra <- apply(muestra1d,1,muestra.jl,datos)
muestra1full <- do.call(rbind, muestra)
nm <- n-nrow(muestra1full)
}
print(replica)

# Modifica los datos incorporando censura a derecha
if(censura==0) ctime<-rep(N,n) else{ #censura=P(C<MIN(T1,T2))
gama <- censura*lambda/(1-censura)
ctime <- rexp(n,gama)
}
datafull <- cbind(muestra1full,ctime)
datoscr <- t(apply(datafull,1,datoscens));
colnames(datoscr)<-c("min", "delta")
datoscr <- as.data.frame(datoscr)

#Remuestreos de los datos y organización de tiempos de la remuestra
remuest <- replicate(Nr,sample(n,replace=T))
datosrem <- apply(remuest,2,remuestras,datoscr,n) #Remuestra
Datos.ordenados.rem <- lapply(datosrem,ordena) #Organizan los tiempos

# Estadístico de Kam et al (2002)
# construcción de empíricas
datoscr.ord <- datoscr[order(datoscr$min),] #Organizan los tiempos
eventos <- ifelse(datoscr.ord$delta>'0',1,0)/n
evento1 <- ifelse(datoscr.ord$delta=='1',1,0)/n
evento2 <- ifelse(datoscr.ord$delta=='2',1,0)/n
#Datos2 <- data.frame(Datos.ordenados,eventos,evento1,evento2)
```

```

#Acumuladas empíricas
Acum.Emp <- cumsum(eventos)
Acum.Emp1 <- cumsum(evento1)
Acum.Emp2 <- cumsum(evento2)
Datosfull <- data.frame(datoscr.ord,eventos,evento1,evento2,
Acum.Emp,Acum.Emp1,Acum.Emp2)

#Funciones Empíricas
f.acum.emp1<-stepfun(c(Datosfull$min),c(0,Datosfull$Acum.Emp1))
f.acum.emp2<-stepfun(c(Datosfull$min),c(0,Datosfull$Acum.Emp2))
f.acum.emp<-stepfun(c(Datosfull$min),c(0,Datosfull$Acum.Emp))

#Tamaños de los saltos
saltos <- diff(c(0,Datosfull$Acum.Emp))
c1 <- f.acum.emp1(datoscr.ord$min)
c2 <- f.acum.emp2(datoscr.ord$min)
Dn<-n*sum((c2-c1)^2*saltos) #Estadístico Cramer Von Mises

#####
#Estadístico Cramer Von Mises Remuestreo#
#####

remuest <- replicate(Nr,sample(n,replace=T))
ep.rem <- unlist(lapply(Datos.ordenados.rem,estadrem,c1,c2))
ep.rem <- unlist(ep.rem)
vcRem <- as.vector(quantile(ep.rem,probs=1-alfa))
DecisionRem <- (Dn>vcRem)+0
result <- c(n,lambda1,lambda2,lambda3,censura,replica,
Dn,vcRem,DecisionRem)

write(result,file=paste("n_",n,"lambda1_",lambda1,"lambda2_",lambda2,
"lambda3_",lambda3,"censura_",censura,".txt",sep=""),
append=T,ncolumns=9)
c(DecisionRem)
}

```

Código de las funciones en R de la prueba basada en el método de aproximación por simetrización aleatoria D_n^Z :

```

estadras <- function(datos.ras,datosfull,saltos){
ras.event <- datos.ras*datosfull$probas
ras.event1 <- datos.ras*datosfull$probs1
ras.event2 <- datos.ras*datosfull$probs2

#Acumuladas ras
Acum.ras <- cumsum(ras.event)
Acum.ras1 <- cumsum(ras.event1)
Acum.ras2 <- cumsum(ras.event2)

#funcion empírica con proceso RAS versión RAS
f.ras.emp<-stepfun(datosfull$min,c(0,Acum.ras)) #Empririca RAS DE F
f.ras1.emp<-stepfun(datosfull$min,c(0,Acum.ras1)) #Empririca RAS DE F1
f.ras2.emp<-stepfun(datosfull$min,c(0,Acum.ras2)) #Empririca RAS DE F2

#creación de las funciones para estadístico de prueba
c2ras<-f.ras2.emp(datosfull$min)
c1ras<-f.ras1.emp(datosfull$min)

Wz_12<-sqrt(n)*(c2ras-c1ras)

Dn_p<-sum(((Wz_12)^2)*saltos) #Estadistico Dn_ras
Dn_p
}

```

Código de las funciones en R para evaluar la prueba basada en el método de aproximación por simetrización aleatoria D_n^Z

```

compara.riesgos.Ras <- function(replica,n,lambd,censura,celdas,probs){
# Genera N valores de un mínimo y le asocia un modo de falla
#con probabilidad p
minimo <- rexp(N,lambd)
delta <-rbinom(N,1,p)
datmin <- as.data.frame(cbind(minimo,delta))
datos <- as.data.frame(t(apply(datmin,1,GeneraPareja)));
names(datos)<-c("X","Y")

```

```
# Frecuencias de celda
allfrecs <- cbind(celdas, apply(celdas, 1, frecs.jl, datos))

# Probabilidades de celda desde derive y en la forma de la tabla
#de frecuencias
allprobs <- cbind(celdas, probs)

# Sacar una muestra de tamaño n con base en las probs de celda
#(sin censura)
nm <- 1
while(nm!=0){
muestraaid <- rmultinom(1, n*1.1, prob=probs)
muestraaid <- cbind(celdas, muestraaid)
muestra <- apply(muestraaid, 1, muestra.jl, datos)
muestrafull <- do.call(rbind, muestra)
nm <- n-nrow(muestrafull)
print(nm)
}
# Modifica los datos incorporando censura a derecha
if(censura==0) ctime <- rep(N, n) else{ #censura=P(C<MIN(T1,T2))
gama<-censura*lambda/(1-censura)
ctime <- rexp(n, gama)
}
datafull <- cbind(muestrafull, ctime)
datoscr <- t(apply(datafull, 1, datoscens));
colnames(datoscr) <- c("min", "delta")
datoscr <- as.data.frame(datoscr)

#construcción de empíricas
datoscr.ord <- datoscr[order(datoscr$min),] #Organizan los tiempos
probas <- ifelse(datoscr.ord$delta>'0', 1, 0)/n
probs1 <- ifelse(datoscr.ord$delta=='1', 1, 0)/n
probs2 <- ifelse(datoscr.ord$delta=='2', 1, 0)/n

#Acumuladas empíricas
Acum.Emp <- cumsum(probas)
Acum.Emp1 <- cumsum(probs1)
Acum.Emp2 <- cumsum(probs2)
datosfull <- data.frame(datoscr.ord, probas, probs1, probs2,
Acum.Emp, Acum.Emp1, Acum.Emp2)
```

```
#Funciones Empíricas
f.acum.emp1 <- stepfun(c(datosfull$min),
c(0,datosfull$Acum.Emp1)) #Empririca De F1
f.acum.emp2 <- stepfun(c(datosfull$min),
c(0,datosfull$Acum.Emp2)) #Empririca De F
f.acum.emp<-stepfun(c(datosfull$min),
c(0,datosfull$Acum.Emp)) #Empririca De F1

#Tamaños de los saltos
saltos <- diff(c(0,datosfull$Acum.Emp))
c1 <- f.acum.emp1(datoscr.ord$min)
c2 <- f.acum.emp2(datoscr.ord$min)

#Construcción de Dz_0
Dz_0 <- n*sum((c2-c1)^2*saltos) #Estadistico Cramer Von Mises

# Permutaciones método RAS
Np <- Nr
permutacion <- function(){
b1 <- rbinom(n,size=1,prob=0.5)
signos <- b1
signos[which(signos==0)]<--1
signos
}
datos.ras <- replicate(Np,permutacion())

#Creando el valor p
ep.ras <- apply(datos.ras,2,estadras,datosfull,saltos)
a <- sum((ep.ras>Dz_0)+0)
p <- a/(Np+1)
Decision.Ras <- (p<=alfa)+0

result<-c(n,lambda1,lambda2,lambda3,censura,replica,Dz_0,p,
Decision.Ras)
write(result,file=paste("n_",n,"lambda1_",lambda1,"lambda2_",lambda2,
"lambda3_",lambda3,"censura_",censura,".txt",sep=""),
append=T,ncolumns=9)
c(Decision.Ras)
}
```

Código de la funciones en R de la prueba del método bootstrap alternativo usando los estimadores de las CIF's y Kaplan Meier

```
estadrem <- function(datosrem,c1,c2){
fit <- cuminc(datosrem$min,datosrem$delta)
delta2 <- (datosrem$delta>0)+0
KM1 <- survfit(Surv(datosrem$min,delta2)~1,type="kaplan-meier")
dis.emp <- stepfun(c(KM1$time,max(KM1$time)),c(0,1-KM1$surv,1))
Min.ordr <- sort(datosrem$min)
dis.emp.evalr <- dis.emp(Min.ordr)
saltosr <- diff(dis.emp.evalr)
fitr <- cuminc(datosrem$min,datosrem$delta)
cif1r <- fitr$"1 1"
cif2r <- fitr$"1 2"
cif1fr <- stepfun(cif1r$time[-1],cif1r$est)
c1r <- cif1fr(Min.ordr)
cif2fr <- stepfun(cif2r$time[-1],cif2r$est)
c2r <- cif2fr(Min.ordr)
Wcif1r <- n^(1/2)*(c1r-c1) #Construcción del W para cif1r
Wcif2r <- n^(1/2)*(c2r-c2) #Construcción del W para cif2r
Wcif12r <- Wcif2r-Wcif1r
Dnr <- sum((Wcif12r)^2*c(0,saltosr))
Dnr #Estadistico Cramer Von Mises Remuestro
}
```

Código de la funciones en R para evaluar la prueba del método bootstrap alternativo usando las CIF y Kaplan Meier.

```
compara.riesgos<-function(replica,n,lambd,censura,celdas,probs){
# Genera N valores de un mínimo y le asocia un modo de
#falla con probabilidad p
minimo <- rexp(N,lambd)
delta <- rbinom(N,1,p)
datmin <- as.data.frame(cbind(minimo,delta))
datos <- as.data.frame(t(apply(datmin,1,GeneraPareja)));
names(datos) <- c("X","Y")

# Frecuencias de celda
allfrecs <- cbind(celdas,apply(celdas,1,frecs.jl,datos))
```

```

# Probabilidades de celda desde derive y en la forma de
#la tabla de frecuencias
allprobs <- cbind(celdas,probs)

# Sacar una muestra de tamaño n con base en las probs de celda
#(sin censura)
nm <- 1
while(nm!=0){
muestraaid <- rmultinom(1,n*1.1,prob=probs)
muestraaid <- cbind(celdas,muestraaid)
muestra <- apply(muestraaid,1,muestra.jl,datos)
muestrafull <- do.call(rbind, muestra)
nm <- n-nrow(muestrafull)
}

# Modifica los datos incorporando censura a derecha
if(censura==0) ctime <- rep(N,n) else{ #censura=P(C<MIN(T1,T2))
gama <- censura*lambda/(1-censura)
ctime <- rexp(n,gama)
}
datafull <- cbind(muestrafull,ctime)
datoscr <- t(apply(datafull,1,datoscr));colnames(datoscr)
<-c("min","delta")
datoscr <- as.data.frame(datoscr)

# Estadístico de Kam et al (2002)
fit <- cuminc(datoscr$min,datoscr$delta)
delta2 <- (datoscr$delta>0)+0
KM1 <- survfit(Surv(datoscr$min,delta2)~1,type="kaplan-meier")
dis.emp <- stepfun(c(KM1$time,max(KM1$time)),c(0,1-KM1$surv,1))
Min.ord <- sort(datoscr$min)
dis.emp.eval <- dis.emp(Min.ord)
saltos <- diff(dis.emp.eval)
cif1 <- fit$"1 1"
cif2 <- fit$"1 2"
c1 <- cif1f(Min.ord)
cif2f <- stepfun(cif2$time[-1],cif2$est)
c2 <- cif2f(Min.ord)
Dn <- n*sum((c2-c1)^2*c(0,saltos)) #Estadistico Cramer Von Mises

```

```
#####  
#Estadistico Cramer Von Mises Remuestreo#  
#####  
remuest <- replicate(Nr,sample(n,replace=T))  
datosrem <- apply(remuest,2,remuestras,datoscr,n) #Remuestra  
ep.rem <- unlist(lapply(datosrem,estadrem,c1,c2))  
ep.rem <- unlist(ep.rem)  
vcRem <- as.vector(quantile(ep.rem,probs=1-alfa))  
DecisionRem <- (Dn>vcRem)+0  
result <- c(n,lambda1,lambda2,lambda3,censura,replica,Dn,  
           vcRem,DecisionRem)  
write(result,file=paste("n_",n,"lambda1_",lambda1,"lambda2_",lambda2,  
"lambda3_",lambda3,"censura_",censura,".txt",sep=""),  
append=T,ncolumns=9)  
c(DecisionRem)  
}
```

Bibliografía

- Aly, E. A. A., Kochar, S. C. & McKeague, I. W. (1994), 'Some tests for comparing cumulative incidence functions and cause-specific hazard rates', *Journal of the American Statistical Association* **89**, 994–999.
- Block, H. & Basu, A. (1974), 'A Continuous bivariate exponential extension', *Journal of the American Statistical Association* **69**, 1031– 1037.
- Burke, M. D. & Yuen, K. C. (1995), 'Goodness-of-fit tests for the Cox model via bootstrap method', *Journal of Statistical Planning Inference* **47**, 237–256.
- Carriere, K. C. & Kochar, S. C. (2000), 'Comparing sub-survival functions in a competing risks model', *Lifetime Data Analysis* **6**, 85– 97.
- Davarzani, N., Achcar, J.A., a. S. E. & Peeters, R. (2015), 'Bivariate lifetime geometric distribution in presence of cure fractions', *Journal of Data Science* **13**, 755–770.
- Feller, W. (1951), 'The asymptotic distribution of the range of sums of independent random variables', *Annals of Mathematical Statistics* **22**, 427– 432.
- Fleming, T. R. & Harrington, D. P. (1991), *Counting Processes and Survival Analysis*, John Wiley & Sons, New York.
- Friday, D. S. & Patil, G. P. (1977), 'A Bivariate Exponential Model With Applications to Reliability and Computer Generation of Random Variables', *The Theory and Applications of Reliability With Emphasis on Bayesian and Nonparametric Methods* **1**, 527–549. eds. C. P. Tsokos and I. N. Shimi, New York: Academic Pres.
- Gine, E. & Zinn, J. (1984), 'On the central limit theorems for empirical processes (with discussion)', *Annals of Probability* **12**, 929– 989.
- Gooley, T. A., Leisenring, W., Crowley, J. & Storer, B. E. (1999), 'Estimation of failure probabilities in the presence of competing risks: new representations of old estimators', *Statistics in Medicine* **18**, 695–706.
- Gray, R. J. (1988), 'A class of k-sample tests for comparing the cumulative incidence of a competing risk', *Ann. Statist.* **16**, 1141– 1154.

- Kalbfleisch, J. D. & Prentice, R. L. (1980), *The Statistical Analysis of Failure Time Data*, John Wiley & Sons Ltd, New York.
- Kam, C. Y., Lixing, Z. & Dixin, Z. (2002), 'Comparing k Cumulative Incidence Functions Through Resampling Methods', *Lifetime Data Analysis* **8**, 401–412.
- Kaplan, E. L. & Meier, P. (1958), 'Nonparametric estimation from incomplete observations', *Journal of the American Statistical Association* **53**, 457–481.
- Kochar, S. C., Law, K. F. & Yip, P. (2002), 'Generalized Supremum Tests for the Equality of Cause Specific Hazard Rates', *Lifetime Data Analysis* **8**, 277–288.
- Leandro, R. & Achcar, J. (1996), 'Generation of bivariate lifetime data assuming the Block & Basu exponential distribution', *Revista de matemática e estatística, Sao Paulo* **14**, 43– 52.
- Luo, X. & Turnbull, B. W. (1999), 'Comparing two treatments with multiple competing risks endpoints', *Statistica Sinica* **9**, 985– 997.
- Meeker, W. & Escobar, L. (1998), *Statistical Methods for Reliability Data*, John Wiley & Sons Ltd, New York.
- Moeschberger, M. & David, H. (1978), *The Theory of Competing Risks*, Griffin's Statistical, London.
- Neuhaus, G. & Zhu, L. X. (1998), 'Permutation tests for reflected symmetry', *Journal of Multivariate Analysis* **67**, 129– 153.
- Petersen, P. M., Gospodarowicz, M., Tsang, R., Pintilie, M., Wells, W., Hodgson, D., Sun, A., Crump, M., Patterson, B. & Bailey, D. (2004), 'Long-term outcome in stage I and II follicular lymphoma following treatment with involved field radiation therapy alone', *Journal of Clinical Oncology* **22**, 563S.
- Pintilie, M. (2006), *Competing risks: A practical perspective*, John Wiley & Sons Ltd, Canadá.
- Pintilie, M. (2011), 'Análisis de riesgos competitivos', *Revista Española de Cardiología* **64**, 599–605.
- Pollard, D. (1984), *Convergence of Stochastic Processes*, Springer-Verlag, New York.
- Sankaran, P. G., Unnikrishnan, N. & Sreedevi, E. (2010), 'A quantile based test for comparing cumulative incidence functions of competing risks models', *Statistics and Probability Letters* **80**, 886–891.

- Sun, Y. & Tiwari, R. C. (1995), 'Comparing Cause-Specific Hazard Rates of a Competing Risks Model with Censored Data', *Institute of Mathematical Statistics* **27**, 225–270.
- Yáñez, S., Brango, H., Jaramillo, M. J. & Lopera, C. M. (2011), 'Comparación entre riesgos competitivos vía el estimador cópula-gráfico', *Revista Colombiana de Estadística* **34**(2), 231 – 248.
- Yuen, K. C. & Burke, M. D. (1997), 'A test of fit for a semiparametric additive risk model', *Biometrika* **84**, 631– 639.
- Yuen, K. C., Zhu, L. X. & Tang, N. Y. (2003), 'On the mean residual life regression model', *Journal of Statistical Planning Inference* **113**, 685–698.
- Zhu, L. X. & Neuhaus, G. (2000), 'Nonparametric Monte Carlo tests for multivariate distributions', *Biometrika* **87**, 919– 928.
- Zhu, L. X., Yuen, K. C. & Tang, N. Y. (2002), 'Resampling methods for testing a semi-parametric random censorship model', *Scandinavian Journal of Statistics* **29**, 111 – 123.