



COLEGIO DE POSTGRADUADOS

INSTITUCIÓN DE ENSEÑANZA E INVESTIGACIÓN
EN CIENCIAS AGRÍCOLAS

CAMPUS MONTECILLO

POSTGRADO EN SOCIOECONOMÍA, ESTADÍSTICA E
INFORMÁTICA - ESTADÍSTICA

**PRUEBA DE BONDAD DE AJUSTE PARA LA
DISTRIBUCIÓN GUMBEL PARA DATOS
CENSURADOS TIPO II, BASADA EN LA
DIVERGENCIA DE KULLBACK-LEIBLER**

VICTOR SALINAS RUIZ

T E S I S

PRESENTADA COMO REQUISITO PARCIAL PARA
OBTENER EL GRADO DE:

MAESTRO EN CIENCIAS

MONTECILLO, TEXCOCO, EDO. DE MÉXICO
SEPTIEMBRE DEL 2011

La presente tesis titulada: **PRUEBA DE BONDAD DE AJUSTE PARA LA DISTRIBUCIÓN GUMBEL PARA DATOS CENSURADOS TIPO II, BASADA EN LA DIVERGENCIA DE KULLBACK-LEIBLER**, realizada por el alumno: **VICTOR SALINAS RUIZ**, bajo la dirección del Consejo Particular indicado, ha sido aprobada por el mismo y aceptada como requisito para obtener el grado de:

M A E S T R O E N C I E N C I A S

**SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA-
ESTADÍSTICA**

CONSEJO PARTICULAR

CONSEJERO:

Dr. Paulino Pérez Rodríguez

ASESOR:

Dra. Elizabeth González Estrada

ASESOR:

Dr. Leonardo Tijerina Chávez

SINODAL:

Dr. Humberto Vaquera Huerta

Montecillo, Texcoco, Estado de México, Septiembre del 2011

DEDICATORIA

A Dios, por acompañarme y apoyarme en toda esta etapa de mi vida.

A mis padres: Elíseo Salinas Manzano y Agustina Ruiz Guzmán, por darme la vida y apoyarme a continuar con mis estudios.

A mis hermanos: Cristina, Isaac, Filomena, Rosa María, Noé, Estela, Josafath, Elpidio, Margarita y María Inés ⁺, por todo el apoyo moral y económico brindado.

A todos mis sobrinos de las familias: Barrientos Salinas, Arenas Salinas, Pérez Salinas, Moreno Salinas, Salinas Ayala, Salinas Morales, Alaya Salinas y Salinas García; por su alegría que adorna a la familia.

A Liberia, por ser alguien muy especial para mi.

A mis compañeros de Maestría: María, David, Alberto, Alma, Emanuel, Sandra y Juan; por su amistad y compañía.

A G R A D E C I M I E N T O S

Al Consejo Nacional de Ciencia y Tecnología (CONACYT), por el apoyo económico brindado para poder concluir mis estudios de Maestría.

Al Colegio de Postgraduados, por abrirme las puertas y así continuar con mi formación académica.

Al Departamento de Estadística y en especial a todo el grupo de Académicos que me apoyaron durante mi estancia como estudiante.

Al Dr. Paulino Pérez Rodríguez, por todo el apoyo brindado como Consejero y en toda la realización de este trabajo de Tesis.

A la Dra. Elizabeth González Estrada, por todo el apoyo brindado en la realización de este trabajo de Tesis.

Al Dr. Leonardo Tijerina Chávez, por su colaboración en esta Tesis.

Al Dr. Humberto Vaquera Huerta, por todo el apoyo recibido en este presente trabajo de Tesis.

Al Dr. Javier Suárez Espinoza, por el apoyo que me otorgó en la etapa de estudiante.

Al Dr. Juan A. Burgueño Ferreira, por todos sus comentarios tan valiosos.

A Emma, Isabel, Griselda y a todo el personal de la Sala de Cómputo y Administrativo de Estadística, por todo el apoyo brindado.

Índice General

1. Introducción	1
2. Objetivos	2
2.1. Objetivo general	2
2.2. Objetivos particulares	2
3. Revisión de Literatura	3
3.1. Orígenes de la Distribución Gumbel	3
3.1.1. La Función de Distribución Gumbel	3
3.2. Pruebas para seleccionar un modelo	5
3.2.1. La prueba de bondad de ajuste de Pearson (Ji-Cuadrada)	5
3.2.2. La prueba de bondad de ajuste de Kolmogorov-Smirnov	6
3.3. La prueba del Coeficiente de Correlación	7
3.3.1. El estimador del Coeficiente de Correlación	8
3.4. Conceptos sobre Análisis de Supervivencia	8
3.4.1. Función de Riesgo	10
3.4.2. Función de Riesgo Acumulado	10
3.4.3. Estimador de Kaplan-Meier de la función de Supervi- vencia	11
3.4.4. Estimador de Nelson-Aalen	12
3.4.5. Censura por la derecha	13
4. Prueba de Kullback-Leibler para Datos Censurados Tipo II	15
4.1. Construcción de la Prueba	15
4.2. Distribución del estadístico de prueba	18
5. Estudio de Simulación Monte Carlo para estimar la potencia y tama- ño de la prueba, contra algunas pruebas alternativas	23
5.1. Potencia de la prueba	23
5.2. Prueba utilizando el estimador de Kaplan-Meier	24
5.2.1. Construcción de la prueba	24
5.2.2. Distribución del estadístico de prueba	25
5.2.3. Implementación de la prueba	30

5.3. Tamaño de la prueba	30
5.4. Potencia de la prueba	30
5.5. Prueba usando el estimador de Nelson-Aalen	31
5.5.1. Distribución del estadístico de prueba	31
5.5.2. Implementación de la prueba	33
5.6. Tamaño de la prueba	34
5.7. Potencia de la prueba	34
6. Estudios de comparación, tamaño y potencia de las pruebas	35
6.1. Comparación en el tamaño estimado de la prueba	35
6.2. Estudio comparativo de potencias	36
7. Ejemplo de Aplicación	49
8. Conclusiones y Recomendaciones	51

Índice de Figuras

3.1. Tiempos de vida de pacientes en observación.	14
4.1. Distribuciones empíricas de Kullback-Leibler ($KL^*(m, n, r)$), generadas con 10,000 réplicas de la distribución Gumbel, para varios valores de los parámetros de localidad y escala.	20
5.1. Distribución del estadístico de prueba de Kaplan-Meier para la distribución Gumbel, con diferentes valores de los parámetros de localidad y escala.	27
5.2. Distribución del estadístico de prueba de Nelson-Aalen para la distribución Gumbel, con diferentes valores de los parámetros de localidad y escala.	32
6.1. Comparación de potencias de las estadísticas de prueba, H_1 es <i>Weibull</i> (5, 10).	37
6.2. Comparación de potencias de las estadísticas de prueba, H_1 es <i>Normal</i> (0, 1).	39
6.3. Comparación de potencias de las estadísticas de prueba, H_1 es <i>Gamma</i> (5, 1).	41
6.4. Comparación de potencias de las estadísticas de prueba, H_1 es <i>Log-normal</i> (0, 1).	43
6.5. Comparación de potencias de las estadísticas de prueba, H_1 es <i>Logística</i> (0, 1).	45
6.6. Comparación de potencias de las estadísticas de prueba, H_1 es <i>Cauchy</i> (0, 0.5).	47

Índice de Tablas

3.1.	Tabla de contingencia para prueba de bondad de ajuste de Pearson. . .	5
4.1.	Valores críticos $K_{m,n,r}$ de la estadística $KL^*(m, n, r)$, obtenida mediante simulación Monte Carlo, con $B = 10,000$	21
4.2.	Valores críticos $K_{m,n,r}$ de la estadística $KL^*(m, n, r)$, obtenida mediante simulación Monte Carlo, con $B = 10,000$ (<i>Continuación</i>).	22
5.1.	Valores críticos ($K_{K-M}(\alpha)$) del Coeficiente de Correlación R_{K-M} , obtenidos mediante Simulación Monte Carlo con $B = 10,000$ y muestras de tamaño n , con un nivel de significancia $\alpha = 0.05$	29
5.2.	Valores críticos ($K_{N-A}(\alpha)$) del Coeficiente de Correlación R_{N-A} , obtenidos mediante Simulación Monte Carlo con $B = 10,000$ y muestras de tamaño n , con un nivel de significancia $\alpha = 0.05$	33
6.1.	Tamaño estimado de las estadísticas de prueba: Kaplan-Meier, Nelson-Aalen y Kullback-Leibler, considerando a la distribución <i>Gumbel</i> (0, 1).	35
6.2.	Comparación de potencias de las estadísticas de prueba: Kaplan-Meier, Nelson-Aalen y Kullback-Leibler, considerando como alternativa la distribución <i>Weibull</i> (5, 10).	36
6.3.	Comparación de potencias de las estadísticas de prueba: Kaplan-Meier, Nelson-Aalen y Kullback-Leibler, considerando como alternativa la distribución <i>Weibull</i> (20, 50).	38
6.4.	Comparación de potencias de las estadísticas de prueba: Kaplan-Meier, Nelson-Aalen y Kullback-Leibler, considerando como alternativa la distribución <i>Normal</i> (0, 1).	38
6.5.	Comparación de potencias de las estadísticas de prueba: Kaplan-Meier, Nelson-Aalen y Kullback-Leibler, considerando como alternativa la distribución <i>Gamma</i> (5, 1).	40
6.6.	Comparación de potencias de las estadísticas de prueba: Kaplan-Meier, Nelson-Aalen y Kullback-Leibler, considerando como alternativa la distribución <i>Gamma</i> (3, 9).	42
6.7.	Comparación de potencias de las estadísticas de prueba: Kaplan-Meier, Nelson-Aalen y Kullback-Leibler, considerando como alternativa la distribución <i>Log - normal</i> (0, 1).	42

6.8. Comparación de potencias de las estadísticas de prueba: Kaplan-Meier, Nelson-Aalen y Kullback-Leibler, considerando como alternativa la distribución <i>Logística</i> (0, 1).	44
6.9. Comparación de potencias de las estadísticas de prueba: Kaplan-Meier, Nelson-Aalen y Kullback-Leibler, considerando como alternativa la distribución <i>Cauchy</i> (0, 0.5).	46
6.10. Comparación de potencias de las estadísticas de prueba: Kaplan-Meier, Nelson-Aalen y Kullback-Leibler, considerando como alternativa la distribución <i>Cauchy</i> (0, 1).	48
7.1. Log de Tiempos de falla de un Aislante Eléctrico sometido a Tensión.	49
7.2. Comparación entre Pruebas.	50
8.1. Tamaño estimado de la prueba de Kullback-Leibler ($KL^*(m, n, r)$), para la distribución <i>Gumbel</i> (0, 1), obtenidos mediante simulación Monte Carlo con $B = 10,000$, y muestras de tamaño n	60
8.2. Potencia de la prueba de Kullback-Leibler ($KL^*(m, n, r)$), cuando H_1 es <i>Weibull</i> (5, 10).	60
8.3. Potencia de la prueba de Kullback-Leibler ($KL^*(m, n, r)$), cuando H_1 es <i>Weibull</i> (20, 50).	61
8.4. Potencia de la prueba de Kullback-Leibler ($KL^*(m, n, r)$), cuando H_1 es <i>Normal</i> (0, 1).	61
8.5. Potencia de la prueba de Kullback-Leibler ($KL^*(m, n, r)$), cuando H_1 es <i>Gamma</i> (5, 1).	61
8.6. Potencia de la prueba de Kullback-Leibler ($KL^*(m, n, r)$), cuando H_1 es <i>Gamma</i> (3, 9).	62
8.7. Potencia de la prueba de Kullback-Leibler ($KL^*(m, n, r)$), cuando H_1 es <i>Log - normal</i> (0, 1).	62
8.8. Potencia de la prueba de Kullback-Leibler ($KL^*(m, n, r)$), cuando H_1 es <i>Logística</i> (0, 1).	62
8.9. Potencia de la prueba de Kullback-Leibler ($KL^*(m, n, r)$), cuando H_1 es <i>Cauchy</i> (0, 0.5).	63
8.10. Valores críticos del Coeficiente de Correlación del estadístico de Kaplan-Meier (R_{K-M}), obtenidos mediante simulación Monte Carlo con $B = 10,000$ y muestras de tamaño n , con un nivel de significancia $\alpha = 0.01$	64
8.11. Valores críticos del Coeficiente de Correlación del estadístico de Kaplan-Meier (R_{K-M}), obtenidos mediante simulación Monte Carlo con $B = 10,000$ y muestras de tamaño n , con un nivel de significancia $\alpha = 0.02$	65
8.12. Valores críticos del Coeficiente de Correlación del estadístico de Kaplan-Meier (R_{K-M}), obtenidos mediante simulación Monte Carlo con $B = 10,000$ y muestras de tamaño n , con un nivel de significancia $\alpha = 0.10$	66

8.13. Valores críticos del Coeficiente de Correlación del estadístico de Kaplan-Meier (R_{K-M}), obtenidos mediante simulación Monte Carlo con $B = 10,000$ y muestras de tamaño n , con un nivel de significancia $\alpha = 0.15$.	67
8.14. Tamaño estimado de la prueba del estadístico de Kaplan-Meier (R_{K-M}), para la distribución <i>Gumbel</i> (0, 1), obtenidos mediante simulación Monte Carlo con $B = 10,000$, y muestras de tamaño n .	68
8.15. Potencia de la prueba del estadístico de Kaplan-Meier (R_{K-M}), cuando H_1 es <i>Weibull</i> (5, 10).	68
8.16. Potencia de la prueba del estadístico de Kaplan-Meier (R_{K-M}), cuando H_1 es <i>Weibull</i> (20, 50).	68
8.17. Potencia de la prueba del estadístico de Kaplan-Meier (R_{K-M}), cuando H_1 es <i>Normal</i> (0, 1).	69
8.18. Potencia de la prueba del estadístico de Kaplan-Meier (R_{K-M}), cuando H_1 es <i>Gamma</i> (5, 1).	69
8.19. Potencia de la prueba del estadístico de Kaplan-Meier (R_{K-M}), cuando H_1 es <i>Gamma</i> (3, 9).	69
8.20. Potencia de la prueba del estadístico de Kaplan-Meier (R_{K-M}), cuando H_1 es <i>Log - normal</i> (0, 1).	70
8.21. Potencia de la prueba del estadístico de Kaplan-Meier (R_{K-M}), cuando H_1 es <i>Logística</i> (0, 1).	70
8.22. Potencia de la prueba del estadístico de Kaplan-Meier (R_{K-M}), cuando H_1 es <i>Cauchy</i> (0, 0.5).	70
8.23. Valores críticos del Coeficiente de Correlación del estadístico de Nelson-Aalen (R_{N-A}), obtenidos mediante simulación Monte Carlo con $B = 10,000$ y muestras de tamaño n , con un nivel de significancia $\alpha = 0.01$.	71
8.24. Valores críticos del Coeficiente de Correlación del estadístico de Nelson-Aalen (R_{N-A}), obtenidos mediante simulación Monte Carlo con $B = 10,000$ y muestras de tamaño n , con un nivel de significancia $\alpha = 0.02$.	72
8.25. Valores críticos del Coeficiente de Correlación del estadístico de Nelson-Aalen (R_{N-A}), obtenidos mediante simulación Monte Carlo con $B = 10,000$ y muestras de tamaño n , con un nivel de significancia $\alpha = 0.10$.	72
8.26. Valores críticos del Coeficiente de Correlación del estadístico de Nelson-Aalen (R_{N-A}), obtenidos mediante simulación Monte Carlo con $B = 10,000$ y muestras de tamaño n , con un nivel de significancia $\alpha = 0.15$.	73
8.27. Tamaño estimado de la prueba del estadístico de Nelson-Aalen (R_{N-A}), para la distribución <i>Gumbel</i> (0, 1), obtenidos mediante simulación Monte Carlo con $B = 10,000$ y muestras de tamaño n .	73
8.28. Potencia de la prueba del estadístico de Nelson-Aalen (R_{N-A}), cuando H_1 es <i>Normal</i> (0, 1).	73
8.29. Potencia de la prueba del estadístico de Nelson-Aalen (R_{N-A}), cuando H_1 es <i>Cauchy</i> (0, 0.5).	74
8.30. Potencia de la prueba del estadístico de Nelson-Aalen (R_{N-A}), cuando H_1 es <i>Logística</i> (0, 1).	74

8.31. Potencia de la prueba del estadístico de Nelson-Aalen (R_{N-A}), cuando H_1 es <i>Weibull</i> (5, 10).	74
8.32. Potencia de la prueba del estadístico de Nelson-Aalen (R_{N-A}), cuando H_1 es <i>Weibull</i> (20, 50).	75
8.33. Potencia de la prueba del estadístico de Nelson-Aalen (R_{N-A}), cuando H_1 es <i>Gamma</i> (5, 1).	75
8.34. Potencia de la prueba del estadístico de Nelson-Aalen (R_{N-A}), cuando H_1 es <i>Gamma</i> (3, 9).	75
8.35. Potencia de la prueba del estadístico de Nelson-Aalen (R_{N-A}), cuando H_1 es <i>Log - normal</i> (0, 1).	76

R E S U M E N

Se propone una prueba de bondad de ajuste para la distribución Gumbel para datos censurados Tipo II. Esta prueba se basa en la metodología de discriminación de Lim & Park (2007), que utilizan la Divergencia de Kullback & Leibler (1951) adaptada para datos censurados Tipo II. Los valores críticos se obtuvieron mediante simulación Monte Carlo para diferentes tamaños de muestra y porcentajes de censura. La potencia y tamaño de la prueba se comparó con la prueba del Coeficiente de Correlación basado en los estimadores de Kaplan & Meier (1958) y Nelson (1972) - Aalen (1978). Los resultados de la simulación, muestran que la prueba basada en la Divergencia de Kullback-Leibler es superior a las otras pruebas en términos de potencia.

Palabras Clave: Entropía, Simulación Monte Carlo, Coeficiente de Correlación, Potencia y Tamaño de la Prueba.

A B S T R A C T

In this work a Goodness of fit test for the Gumbel distribution with Type II censored data is proposed. The test is based in the method proposed by Lim & Park (2007), using the Kullback & Leibler (1951) divergence modified to work with censored data. The critical values of the test were obtained through Monte Carlo Simulation for different sample sizes and proportions of censored data. The power and size of the proposed test was compared with the test based on the Correlation Coefficient using estimators proposed by Kaplan & Meier (1958) and Nelson (1972) - Aalen (1978). The simulation results have shown that the test based on the Kullback-Leibler divergence is superior in terms of power.

Key words: Entropy, Monte Carlo Simulation, Correlation Coefficient, Power and size of the test.

Capítulo 1

Introducción

El modelo Gumbel, es uno de los más utilizados para realizar estudios enfocados principalmente a valores extremos, en pruebas de confiabilidad y en experimentos de tiempos de vida. Esta distribución es adecuada para modelar fenómenos naturales, tales como la precipitación, inundaciones, niveles de ozono, entre otros. Debido a su amplia gama de aplicaciones, este documento pretende realizar una prueba de bondad de ajuste para esta distribución, pero con variables censuradas por la derecha Tipo II. Smith y Bain (1976), propusieron una prueba de bondad de ajuste basada en correlación con muestras censuradas. Al-Aboud (2009) propuso un estimador Bayesiano usando la distribución de valores extremos con censura progresiva. Pérez et al., (2009) usó la divergencia de Kullback-Leibler (1951) para la prueba de bondad de ajuste de la Distribución Gumbel. Arizono y Ohta (1989), realizaron una prueba de normalidad para la entropía muestral usando la divergencia de Kullback-Leibler (1951).

El presente trabajo reúne tres metodologías muy utilizadas en el campo de la estadística, la propuesta desarrollada por Kaplan-Meier (1958), Nelson (1972)-Aalen (1978) y Kullback-Leibler (1951), principalmente enfocadas al estudio de variables aleatorias que presentan un determinado tipo de censura. En este escrito, se trabajó con la Distribución Gumbel, utilizando la metodología de Lim y Park (2007), Park y Park (2003) y Pérez et al., (2005), la cual se basa en estimar la divergencia de Kullback-Leibler (1951) y la Entropía muestral.

Se generaron Tablas de valores críticos, estudios de potencia y estimación de tamaño de la prueba, para diferentes tamaños de muestra, niveles de significancia, porcentajes de censura y para cada uno de los métodos.

Capítulo 2

Objetivos

2.1. Objetivo general

- Realizar una Prueba de Bondad de Ajuste para la Distribución Gumbel con Datos Censurados Tipo II, utilizando la Divergencia de Kullback-Leibler.

2.2. Objetivos particulares

- Obtener tablas de valores críticos para la prueba de bondad de ajuste, con diferentes niveles de significancia.
- Obtener tablas de valores críticos para las pruebas de comparación de Nelson-Aalen y Kaplan-Meier.
- Estudiar el comportamiento de la prueba respecto al tamaño.
- Comparar la potencia de la prueba, con las pruebas del coeficiente de correlación basadas en el estimador de Kaplan-Meier y Nelson-Aalen bajo las distribuciones alternativas siguientes: Weibull, Normal, Log-normal, Gamma, Logística y Cauchy estándar.

Capítulo 3

Revisión de Literatura

3.1. Orígenes de la Distribución Gumbel

La distribución Gumbel es llamada de esta manera en honor a *Emil Julius Gumbel* (1891-1966), esta distribución es muy utilizada para modelar la distribución del máximo (o del mínimo), de manera que se presta para realizar cálculos de valores extremos. Debido a estas características es muy utilizada para modelar y predecir eventos de áreas tales como, la hidrología, finanzas y actualmente ha tenido una aplicación a los desastres naturales como terremotos e inundaciones.

3.1.1. La Función de Distribución Gumbel

Sea X una variable aleatoria con distribución Gumbel y con función de densidad dada por:

$$f_X(x) = \begin{cases} \frac{1}{\theta} \exp \left\{ -\frac{x-\xi}{\theta} - \exp \left\{ -\frac{x-\xi}{\theta} \right\} \right\} I_{(-\infty, \infty)}(x), & \xi \in \mathbb{R}, \theta > 0. \\ 0 & \text{de otra manera.} \end{cases} \quad (3.1)$$

La función de distribución de X , es representada por :

$$F_X(x) = \exp \left\{ -\exp \left(-\frac{x-\xi}{\theta} \right) \right\} I_{(-\infty, \infty)}(x). \quad (3.2)$$

Si se define la variable aleatoria $Y = \left(\frac{X-\xi}{\theta} \right)$, con $\xi=0$ y $\theta=1$, la función de densidad de y está dada por:

$$f_Y(y) = \exp\{-y - \exp\{-y\}\} I_{(-\infty, \infty)}(y). \quad (3.3)$$

Ahora si se define la variable aleatoria $Z = \exp\left\{-\left(\frac{X-\xi}{\theta}\right)\right\} = \exp\{-Y\}$, entonces Z tiene la distribución exponencial estándar dada por la siguiente expresión :

$$f_Z(z) = \exp\{-z\} I_{(0, \infty)}(z).$$

De manera que la función generatriz de momentos de Y , esta dada por la siguiente expresión.

$$m_Y(t) = E\{e^{ty}\} = E\left\{e^{t\left(\frac{X-\xi}{\theta}\right)}\right\} = E\{Z^{-1}\} = \Gamma(1-t) \quad \text{para } t < 1.$$

Al realizar la sustitución de t , por θt , la función generatriz de momentos de X esta dada por:

$$m_X(t) = \exp\{\xi t\} \Gamma(1-\theta t) \quad \text{para } \theta t < 1.$$

A partir de la función generatriz de momentos se puede obtener la media de la distribución:

$$\mu = E\{X\} = \xi + \theta\gamma.$$

Donde: $\gamma = 0.577216$, es la denominada constante de Euler.

De la misma manera para la varianza:

$$\sigma^2 = E\{(X - \mu)^2\} = \frac{\pi^2\theta^2}{6}.$$

3.2. Pruebas para seleccionar un modelo

3.2.1. La prueba de bondad de ajuste de Pearson (Ji-Cuadrada)

Esta prueba de hipótesis es una de las más conocidas en la actualidad, y fue propuesta por Pearson en el año de 1900.

Los datos consisten de N observaciones independientes, copias de una variable aleatoria X . Estas N observaciones están dentro de c clases posibles y el número de observaciones por clase se presenta en una tabla de contingencia de $1 \times c$ (Suárez, 2000).

Tabla 3.1: Tabla de contingencia para prueba de bondad de ajuste de Pearson.

Clases	1	2	...	c	Total
Frecuencias de Observaciones	O_1	O_2	...	O_c	N

Donde las O_j , denotan el número de observaciones en las clases j , con $j = 1, 2, \dots, c$, es decir:

$$N = \sum_{j=1}^c O_j.$$

Suposiciones:

- La muestra es aleatoria.
- La escala de medición es al menos nominal.

Hipótesis:

La función de distribución $F(x)$ verdadera de X , es desconocida, y sea $F^*(x)$ una función de distribución hipotética completamente especificada. Si $F^*(x)$ es completamente especificada excepto por algunos parámetros desconocidos que pueden ser estimados usando la muestra, esta prueba requiere de una pequeña modificación.

$$H_0 : F(x) = F^*(x); \text{ para toda } x.$$

$$H_1 : F(x) \neq F^*(x); \text{ para al menos una } x.$$

Estadística de prueba:

Sea p_j^* la probabilidad de que una observación caiga en una clase j , bajo la suposición de que $F^*(x)$ es la función de distribución de X .

Entonces defínase E_j como:

$$E_j = p_j^* N, \quad j = 1, 2, \dots, c.$$

Donde: E_j representa el número de observaciones esperadas en la clase j , cuando H_0 es verdadera.

La estadística de prueba T , esta dada por:

$$T = \sum_{j=1}^c \frac{(O_j - E_j)^2}{E_j} \sim \chi_{c-1}^2.$$

Regla de decisión:

La distribución exacta de T es difícil de obtener, pero puede obtenerse la distribución asintótica para muestras grandes. La distribución aproximada de T es válida para muestras grandes, y se aproxima a una distribución *ji - cuadrada* con $(c - 1)$ grados de libertad (si los parámetros son estimados con la muestra, los grados de libertad disminuyen uno por cada parámetro estimado).

3.2.2. La prueba de bondad de ajuste de Kolmogorov-Smirnov

La prueba de bondad de ajuste de *Kolmogorov-Smirnov*, fue introducida por *Kolmogorov* en 1933. Esta prueba es una de las más utilizadas, quizás muchas de las veces es más preferida sobre la prueba de bondad de ajuste *ji - cuadrada*, si la muestra de estudio es pequeña; en cambio esta prueba es generalmente más exacta cuando se trata con muestras pequeñas, mientras que la *ji - cuadrada* asume que la muestra es lo suficientemente grande para que la distribución de la estadística de prueba se aproxime a la distribución *ji - cuadrada*. Existen muchas controversias sobre su potencia, pero en general la prueba de *Kolmogorov* es más potente que la *ji - cuadrada* (Suárez, 2000).

Los datos para esta prueba, son provenientes de una muestra aleatoria n , asociada con alguna función de distribución desconocida, denotada por $F(x)$.

Suposiciones:

- Los datos son provenientes de una muestra aleatoria.
- La escala de medición de ese conjunto de datos es al menos nominal.

Hipótesis:

Sea $F^*(x)$ una función de distribución hipotética completamente especificada.

$$H_0 : F(x) = F^*(x) \text{ para toda } -\infty < x < \infty.$$

$$H_1 : F(x) \neq F^*(x) \text{ para al menos una } x.$$

Estadística de prueba:

Sea $S(x)$ la función de distribución empírica basada en la muestra aleatoria. Sea la estadística de prueba T el valor más grande (denotado por el “supremo” de la distancia vertical entre $S(x)$ y $F^*(x)$); en símbolos tenemos que:

$$T = \sup_x |F^*(x) - S(x)|.$$

Regla de decisión:

Rechazar H_0 con un nivel de significancia α si la estadística de prueba T , excede el cuantil $1 - \alpha$ de $W_{1-\alpha}$ (valor crítico) dado en tablas. Esta tabla es exacta solo si la $F^*(x)$ es continua.

3.3. La prueba del Coeficiente de Correlación

Es un método gráfico de bondad de ajuste, en el cual las estadísticas de orden x_j , de una muestra aleatoria X_1, \dots, X_n , son graficadas en el eje vertical, contra y_i (una función conveniente de i), en el eje horizontal; para ajustar una línea recta a los puntos y posteriormente, utilizar la estadística del coeficiente de correlación, para observar la dependencia entre estas dos variables; Saldaña (2008).

Por lo que si denotamos a X e Y , como dos variables aleatorias, el coeficiente de correlación denotado por $\rho_{X,Y}$, de las variables aleatorias X e Y , es definido de la siguiente manera.

$$\rho_{X,Y} = \frac{Cov[X, Y]}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

Donde:

$$\begin{aligned} Cov[X, Y] &= E\{[X - E(X)][Y - E(Y)]\} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [X - E(X)][Y - E(Y)] f_{X,Y}(x, y) dx dy. \end{aligned}$$

Es la covarianza entre las variables X e Y .

Y para la varianza de X .

$$Var(X) = E\{[X - E(X)]^2\} = \int_{-\infty}^{\infty} [X - E(X)]^2 f_X(x) dx$$

El coeficiente de correlación y la covarianza son estadísticas muy utilizadas para medir la relación lineal entre dos variables, sin embargo la $Cov[X, Y]$, carece de mucho significado ya que depende en gran medida de la variabilidad entre X e Y . En cierta forma, el coeficiente de correlación remueve esa variabilidad existente en cada una de las variables (Saldaña, 2008).

El coeficiente de correlación tiene las siguientes propiedades.

- Se puede demostrar que según la desigualdad de Cauchy-Schwarz, se satisface que $-1 \leq \rho_{X,Y} \leq 1$.
- El coeficiente de correlación es invariante ante cambios de localidad y escala en las variables.
- Si una variable X , es función lineal de otra variable Y , entonces el $\rho_{X,Y} = \pm 1$.
- Si X e Y , son dos variables aleatorias independientes, entonces $\rho_{X,Y} = 0$.

3.3.1. El estimador del Coeficiente de Correlación

Sean $(X_1, Y_1), \dots, (X_n, Y_n)$, una muestra aleatoria bivariada de tamaño n , por lo que un estimador del coeficiente de correlación $\rho_{X,Y}$, es denotado por:

$$R = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (3.4)$$

A R , se le conoce como el coeficiente de correlación muestral y cumple las siguiente propiedades:

- $-1 \leq R \leq 1$.
- Es invariante a cambios de localidad y escala en las variables.

3.4. Conceptos sobre Análisis de Supervivencia

Se le denomina función de supervivencia, a la función empleada para describir el tiempo necesario hasta observar un evento de interés, en términos estadísticos, es la probabilidad de que un individuo sobreviva más allá del tiempo x (experimentando el evento antes del tiempo x); (Klein y Moeschberger, 1997).

$$S(x) = 1 - F(x) \quad (3.5)$$

La función de supervivencia, es una función decreciente con valor 1 en el origen y 0 en el infinito. Si X es una variable aleatoria continua, la función de supervivencia es el complemento de la función de distribución acumulada.

$$S(x) = P(X \geq x) \quad (3.6)$$

Donde:

$$F(x) = P(X \leq x).$$

Por lo que la función de supervivencia, es la integral de la función de densidad de probabilidad $f_X(x)$, que es:

$$S(x) = \int_x^{\infty} f_X(x) dx.$$

Por lo tanto:

$$f_X(x) = -\frac{dS(x)}{d(x)}.$$

Note que, $f_X(x) \Delta x$, puede ser considerada como la probabilidad aproximada, de que el evento pueda ocurrir en el tiempo x y que la $f_X(x)$, es una función no-negativa y con área bajo la curva igual a uno.

3.4.1. Función de Riesgo

Se define como la tasa instantánea de muerte o falla a un tiempo x dado que la unidad de estudio ha sobrevivido hasta el tiempo x . En términos probabilísticos, es definida como la probabilidad de que una unidad experimente el suceso entre los tiempos x y Δx , teniendo en cuenta de antemano que ha llegado vivo al tiempo x . Esta función también es conocida como la tasa de falla condicional en estudios de confiabilidad, mortalidad (en estudios demográficos), en epidemiología (estudios de edades) y en la inversa de la razón de Mill en Economía. Esta función está definida como:

$$h(x) = \lim_{\Delta x \rightarrow 0} \frac{P[x \leq X < x + \Delta x \mid X \geq x]}{\Delta x}.$$

Si X es una variable aleatoria continua, entonces puede definirse de la siguiente manera:

$$h(x) = \frac{f(x)}{S(x)} = -\frac{d \ln[S(x)]}{dx}.$$

La función tiene la peculiaridad de que puede ser representada de formas diversas, la única restricción es que debe ser no negativa, es decir, $h(x) \geq 0$. Los modelos con tasa de riesgo creciente pueden originarse cuando existe un desgaste o envejecimiento natural de un individuo, los modelos de riesgo decrecientes se pueden encontrar cuando existe una prematura verosimilitud de falla, por ejemplo en fallas de aparatos electrónicos o en pacientes clínicos, en donde han experimentado algún tipo de trasplante.

La función de riesgo con forma de una bañera, es muy observada en estudios de poblaciones en donde existen nacimientos, seguidas de muertes; por ejemplo si seguimos a una población, la cual primero mostrará signos de crecimiento inicial y se estabilizará en la etapa de envejecimiento.

Una función de riesgo en forma de joroba, se presenta cuando una tasa de riesgo es creciente en un inicio y eventualmente comienza a declinar, en particular esta función de riesgo es muy utilizada para modelar la supervivencia después de una cirugía exitosa, donde existe un incremento inicial en el riesgo debido a infecciones, y decrece cuando el paciente empieza su recuperación.

3.4.2. Función de Riesgo Acumulado

La función de riesgo acumulado es definida de la siguiente manera:

$$H(x) = \int_0^x h(t) dt = -\ln S(x) \Big|_0^x = -\ln[S(x)] + \ln(1) = -\ln[S(x)]. \quad (3.7)$$

Esto indica que:

$$S(x) = \exp[-H(x)] = \exp\left[-\int_0^x h(t) dt\right].$$

3.4.3. Estimador de Kaplan-Meier de la función de Supervivencia

Este estimador de la función de supervivencia fue propuesto por Kaplan y Meier (1958), y está definido para todos los valores de x , en el rango donde existen datos y está planteado de la siguiente forma.

Sean X_1, \dots, X_n , los tiempos de vida (censurados o no) y sean $X_{(1)} \leq \dots \leq X_{(n)}$ las correspondientes estadísticas de orden, entonces:

$$\hat{S}(x) = \prod_{x_{(i)} \leq x} \left(\frac{n-i}{n-i+1} \right)^{\delta_i}, \quad (3.8)$$

con δ_i , la función indicadora :

$$\delta_i = \begin{cases} 0, & \text{si } x_i > c \\ 1, & \text{si } x_i \leq c \end{cases} \quad (3.9)$$

En donde:

- $\hat{S}(x)$, es la función de supervivencia estimada al tiempo x .
- n , es el tamaño de la muestra.
- i , es el número de supervivientes hasta el tiempo x .
- δ_i , es la función indicadora.
- $\hat{S}(x(0)) = 1$.
- Si dos o más $x_{(i)}$ son iguales, es decir que existen empates, se utiliza el valor más grande de i .
- Si dos o más $x_{(i)}$ son iguales, se colocan primero aquellos no censurados.
- $\hat{S}(x)$, es una función escalonada con brincos en cada falla.
- $\hat{S}(x)$, permanece constante sobre todos los intervalos en los que no hay fallas.

Usualmente el error estándar del estimador de Kaplan-Meier, se calcula utilizando la propuesta de Greenwood (1926), dada por:

$$V\left(\widehat{S}(x)\right) = \widehat{S}(x) \sqrt{\sum_{x_{(i)} \leq x} \frac{d_i}{n_i(n_i - d_i)}}. \quad (3.10)$$

Donde:

- d_i , es el número de fallas al tiempo x .
- n_i , es el número de supervivientes antes del tiempo x , incluyendo a casos censurados.

3.4.4. Estimador de Nelson-Aalen

Este estimador es un alternativo de la función de riesgo acumulado, el cual tiene como característica principal mejor desempeño para tamaños de muestra pequeños, fue un principio sugerido por Nelson (1972), en un enfoque de confiabilidad. Posteriormente Aalen (1978), extendió su uso a datos de supervivencia. Por lo que esta aportación es conocida como el estimador de Nelson-Aalen (N-A), de la función de riesgo acumulado, y esta definida por la siguiente ecuación.

$$\tilde{H}(x) = \begin{cases} 0, & \text{si } x < x_j \\ \sum_{x_j \leq x} \frac{d_j}{Y_j}, & \text{si } x_j \leq x \end{cases} \quad (3.11)$$

La forma gráfica de representar a $\tilde{H}(x)$, nos puede dar a conocer la forma de la función de riesgo, de manera que $\tilde{H}(x)$ es lineal si $h(x)$ es constante.

La varianza de $\tilde{H}(x)$, esta dada por :

$$V\left[\widehat{\tilde{H}}(x)\right] = \sum_{x_j \leq x} \frac{d_j}{Y_j^2}. \quad (3.12)$$

Si X , es una variable aleatoria continua, y suponiendo que no existen empates en los datos, entonces para una muestra aleatoria de tamaño n con censura Tipo II, en cual se dispondrán de las primeras r observaciones, $X_{(1)} < X_{(2)} < \dots < X_{(r)}$; el estimador de Nelson-Aalen evaluado en un tiempo específico x_j , puede ser calculado de la siguiente forma.

$$\tilde{H}(x_j) = \sum_{i=1}^j \frac{1}{n - i + 1} \quad (3.13)$$

3.4.5. Censura por la derecha

La censura por la derecha Tipo I, es definido como aquel evento que es observado solo si ocurre antes de algún tiempo pre-especificado (tiempo de censura o límite de observación), el cual puede ser fijo para las unidades de estudio o puede variar de unidad en unidad.

Sean X_1, \dots, X_n , variables aleatorias que representan los tiempos de vida de n unidades bajo estudio; estas variables aleatorias se asumen independientes e idénticamente distribuidas (i.i.d.) con función de densidad de probabilidad $f_X(x)$, y función de supervivencia $S(x)$. Ahora sean $L_i > 0$, con $i = 1, \dots, n$, los tiempos de vida especificados, ahora si la censura es considerada como uniforme para todas las unidades, esto puede significar que $L_1 = L_2 = \dots = L_n$; por lo que el tiempo de vida exacto podrá ser conocido si y solo si, $X_i \leq L_i$. Si X_i es mayor que L_i , se trata de una unidad sobreviviente, y su tiempo de censura corresponderá al tiempo L_i .

Ahora en vez de observar los tiempos de vida X_1, \dots, X_n , se pretende observar los datos por las variables T_1, \dots, T_n , con $T_i = \min(X_i, L_i)$; existe otra notación muy utilizada en el Análisis de supervivencia, dado por la forma (T_i, δ_i) , donde $\delta = 1$ si $T_i = X_i$ o $\delta = 0$ si $T_i = L_i$.

Existe otro tipo de censura por la derecha Tipo II, en la cual se basa este documento; lo cual indica que el estudio de análisis de supervivencia se efectúa hasta que las primeras r unidades experimentan el evento de interés (ver la Figura 3.1.), donde r es un número entero que se fija al inicio del estudio, cumpliendo la condición de que $r < n$.

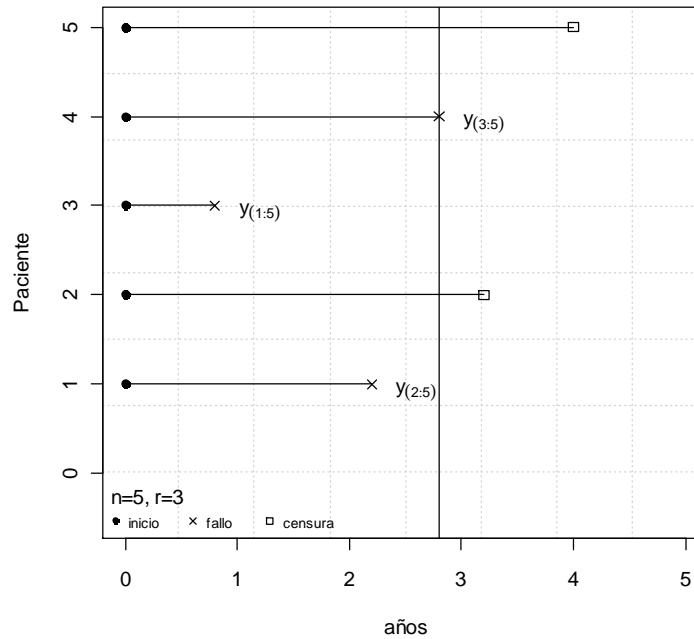


Figura 3.1: Tiempos de vida de pacientes en observación.

Las características de este tipo de censura, indican que solo se observan los r tiempos de vida menores; por lo que se puede utilizar nuevamente la notación (T_i, δ_i) , donde $T_i = X_{(r)}$, para aquellas unidades censuradas y $T_i = X_{(i)}$, con $\delta_i = 1$, donde las variables aleatorias X son dadas por estadísticas de orden.

Capítulo 4

Prueba de Kullback-Leibler para Datos Censurados Tipo II

4.1. Construcción de la Prueba

Sea $X_{(1)}, \dots, X_{(r)}$, una muestra aleatoria con censura por la derecha Tipo II, de una distribución conocida F , con función de densidad $f_X(x) \in \mathbb{R}$ y con media finita.

De tal manera que se requiere de probar el siguiente juego de hipótesis.

Para la hipótesis nula:

$$H_0 : f(x; \cdot) = f_0(x; \xi, \theta) = \frac{1}{\theta} \exp \left\{ -\frac{x - \xi}{\theta} - \exp \left\{ -\frac{x - \xi}{\theta} \right\} \right\} I_{(-\infty, \infty)}(x), \quad \xi \in \mathbb{R}, \theta > 0. \quad (4.1)$$

Para la hipótesis alterna:

$$H_1 : f(x; \cdot) \neq f_0(x; \xi, \theta). \quad (4.2)$$

Para tomar una decisión entre H_0 y H_1 , se pretende utilizar la divergencia de Kullback-Leibler para datos censurados Tipo II como proponen Lim y Park, (2007).

Supóngase que se está interesado en medir la distancia entre dos densidades conocidas, $f(x)$ y $g(x)$, con $x < c$; entonces se puede considerar la información incompleta de Kullback-Leibler de Lim y Park, (2007), como:

$$I(g, f : c) = \int_{-\infty}^c g(x) \log \frac{g(x)}{f(x)} dx. \quad (4.3)$$

Es muy fácil observar que $I(g, f) \geq 0$, si se cumple que $g(x) = f(x)$, sin embargo la información incompleta de Kullback-Leibler, no satisface la no negatividad y más si $F(c)$ no es igual $G(c)$. Es por eso que si $I(g, f : c) = 0$, no implica que $g(x)$ sea igual a $f(x)$, para toda x que se encuentre en el intervalo $(-\infty, c)$.

Lim y Park (2007), redefinen a la información de Kullback-Leibler para el caso censurado como:

$$I^*(g, f : c) = \int_{-\infty}^c g(x) \log \frac{g(x)}{f(x)} dx + F(c) - G(c), \quad (4.4)$$

la cual tiene las siguientes propiedades:

1. $I^*(g, f : c) \geq 0$.
2. $I^*(g, f : c) = 0$, si y solamente si $g(x) = f(x)$ en $(-\infty, c)$.
3. $I^*(g, f : c)$, es una función creciente de c .

Tomando como referencia esta propuesta, se planteó lo siguiente.

Sea KL^* , la información incompleta de Kullback-Leibler, para la distribución Gumbel y con datos censurados Tipo II y sean $f_1(x)$ y $f_0(x)$, dos funciones de densidad conocida, y por lo que se desea probar lo propuesto en las ecuaciones (4.1 y 4.2).

De la ecuación (4.4), y aplicado a nuestro caso, tenemos:

$$KL^*(f_1, f_0 : \nu_{r+1}) = -H(m, n, r) + F_0(\nu_{r+1}) - \frac{r}{n} - \int_{-\infty}^{\nu_{r+1}} f_1(x) \log f_0(x) dx. \quad (4.5)$$

Donde:

1. $KL^*(f_1, f_0 : \nu_{r+1})$, es la información incompleta de Kullback-Leibler, para la distribución Gumbel.
2. $-H(m, n, r)$, es la Entropía para el caso censurado.
3. $F(c) = F_0(\nu_{r+1})$, es la función de distribución Gumbel evaluado en ν_{r+1} , siendo $\nu_{r+1} \leq c$.
4. $G(c) = -\frac{r}{n}$.

De tal manera que la ecuación (4.5), puede escribirse como:

$$KL^*(f_1, f_0 : \nu_{r+1}) = -H(m, n, r) + \exp \left\{ -\exp \left(-\frac{\nu_{r+1} - \xi}{\theta} \right) \right\} - \frac{r}{n} - \int_{-\infty}^{\nu_{r+1}} f_1(x) \log f_0(x) dx. \quad (4.6)$$

De la ecuación (4.6), es necesario evaluar:

$$\int_{-\infty}^{\nu_{r+1}} f_1(x) \log f_0(x) dx = \int_{\nu_1}^{\nu_2} f_1(x) \log f_0(x) dx + \cdots + \int_{\nu_r}^{\nu_{r+1}} f_1(x) \log f_0(x) dx.$$

Por lo tanto:

$$\int_{-\infty}^{\nu_{r+1}} f_1(x) \log f_0(x) dx = \sum_{i=1}^r \int_{\nu_i}^{\nu_{i+1}} f_1(x) \log f_0(x) dx. \quad (4.7)$$

Park y Park (2003), proponen la siguiente expresión para estimar a $f_1(x)$.

$$f_1(x) = \begin{cases} 0 & \text{si } x < \nu_1 \\ n^{-1} \frac{2m}{x_{(i+m)} - x_{(i-m)}} & \text{si } \nu_i < x \leq \nu_{i+1}, i = 1, \dots, r \end{cases} \quad (4.8)$$

Donde:

1. $\nu_i = \left(\frac{x_{(i-m)} + \cdots + x_{(i+m-1)}}{2m} \right), i = 1, \dots, r.$
2. $f_1(x)$, es una función uniforme por tramos tal que $\int_{-\infty}^{\nu_{i+1}} f_1(x) dx = \frac{i}{n}, i = 1, \dots, r.$

Sustituyendo la ecuación (4.8), en cada uno de los términos de la ecuación (4.7), se tiene:

$$\begin{aligned} \int_{\nu_i}^{\nu_{i+1}} f_1(x) \log f_0(x) dx &= \frac{2m}{n} \times \frac{1}{x_{(i+m)} - x_{(i-m)}} \int_{\nu_i}^{\nu_{i+1}} \log f_0(x) dx \\ &= \frac{2m}{n} \times \frac{1}{x_{(i+m)} - x_{(i-m)}} \int_{-\nu_i}^{\nu_{i+1}} \left\{ -\log \theta - \frac{x - \xi}{\theta} - \exp \left(-\frac{x - \xi}{\theta} \right) \right\} dx \end{aligned}$$

$$\int_{-\infty}^{\nu_{r+1}} f_1(x) \log f_0(x) dx = \frac{2m}{n} \sum_{i=1}^r \times \frac{1}{x_{(i+m)} - x_{(i-m)}} \left\{ \left[-\log \theta x - \frac{1}{\theta} \left(\frac{x^2}{2} - \xi x \right) + \theta \exp \left(-\frac{x - \xi}{\theta} \right) \right] \right\}_{\nu_i}^{\nu_{i+1}}. \quad (4.9)$$

Finalmente el estimador de Kullback-Leibler para datos censurados Tipo II, presenta la siguiente forma:

$$KL^*(m, n, r) = - \left\{ \frac{1}{n} \sum_{i=1}^r \left[\log \frac{n}{2m} (x_{(i+m)} - x_{(i-m)}) \right] \right\} + \exp \left\{ - \exp \left(- \frac{\nu_{r+1} - \hat{\xi}}{\hat{\theta}} \right) \right\} + \frac{r}{n} - \frac{2m}{n} \times \sum_{i=1}^r \frac{1}{x_{(i+m)} - x_{(i-m)}} \left\{ \left[-\log \hat{\theta} x - \frac{1}{\hat{\theta}} \left(\frac{x^2}{2} - x \right) + \hat{\theta} \exp \left(- \frac{x - \hat{\xi}}{\hat{\theta}} \right) \right] \right\}_{\nu_i}^{\nu_{i+1}}. \quad (4.10)$$

El siguiente problema, es encontrar los Estimadores de Máxima Verosimilitud (**EMV**) de los parámetros de localidad y escala de la distribución *Gumbel*(ξ, θ); para obtenerlos se plantea lo siguiente.

Sea $\underline{\Theta} = (\xi, \theta)'$, el vector de parámetros de la distribución Gumbel, para el contexto censurado se plantea la siguiente función de verosimilitud.

$$L(\Theta) = \prod_{i=1}^n \{f_{X_i}(x_i; \Theta)\}^{\delta_i} \{1 - F_{X_i}(x_i; \Theta)\}^{1-\delta_i}.$$

El logaritmo de la función de verosimilitud.

$$\log(L(\Theta)) = \sum_{i=1}^n \delta_i \log f_{X_i}(x_i; \Theta) + (1 - \delta_i) \log \{1 - F_{X_i}(x_i; \Theta)\}.$$

Para poder obtener los **EMV**, de los parámetros de localidad y escala, se maximiza numéricamente la función de verosimilitud.

4.2. Distribución del estadístico de prueba

La distribución de la estadística de prueba es difícil de obtener en forma analítica, razón por la cual se utiliza simulación Monte Carlo. La distribución de $KL^*(m, n, r)$ puede obtenerse usando el procedimiento siguiente:

1. Fijar $r, n, \xi, \theta, m = 1$.
2. Generar una muestra censurada Tipo II de la distribución Gumbel, $(x_{(1)}, \dots, x_{(n)})$, $(\delta_1, \dots, \delta_n)$.
3. Obtener los estimadores de máxima verosimilitud de ξ y θ .
4. Obtener $KL^*(m, n, r)$ con la ecuación (4.10).

5. Repetir los pasos de 2 a 3, B veces.

En la Figura 4.1 se muestra la distribución de la estadística de prueba $KL^*(m, n, r)$ para $m = 6$, $n = 50$, $r = 5$, $B = 10,000$ y para diferentes valores de los parámetros ξ y θ . Debido a que la distribución de $KL^*(m, n, r)$ es invariante a transformaciones de localidad y escala, los valores críticos $K_{m,n,r}(\alpha)$ fueron obtenidos mediante simulación Monte Carlo. Los niveles de significancia utilizados fueron $\alpha = 0.01, 0.02, 0.05, 0.10$ y 0.15 . Se generaron muestras aleatorias de la distribución Gumbel estándar para $n \leq 200$, $r/n = 0.5, 0.6, 0.7, 0.8$ y 0.9 , y $B = 10,000$. El valor de $KL^*(m, n, r)$ fue calculado para cada $m < n/2$. El valor crítico fue calculado para cada m, n y r con el cuantil $(1 - \alpha) \times 100\%$ de la distribución empírica $KL^*(m, n, r)$; para cada valor de n y r , se tomó el valor de m que hacía mínimo el valor de $K_{m,n,r}(\alpha)$. En las Tablas 4.1 y 4.2 se presentan los valores críticos obtenidos mediante el proceso de simulación ya descrito. El programa que implementa el algoritmo está en el Anexo I del trabajo.

Regla de decisión:

- Rechazar H_0 con un nivel de significancia α si y solo si, $KL^*(m, n, r) \geq K_{m,n,r}(\alpha)$.

Donde $K_{m,n,r}(\alpha)$, debe cumplir la siguiente condición:

$$\begin{aligned} \alpha &= P(\text{Rechazar } H_0 \mid H_0) \\ &= P(KL^*(m, n, r) \geq K_{m,n,r}(\alpha)). \end{aligned}$$

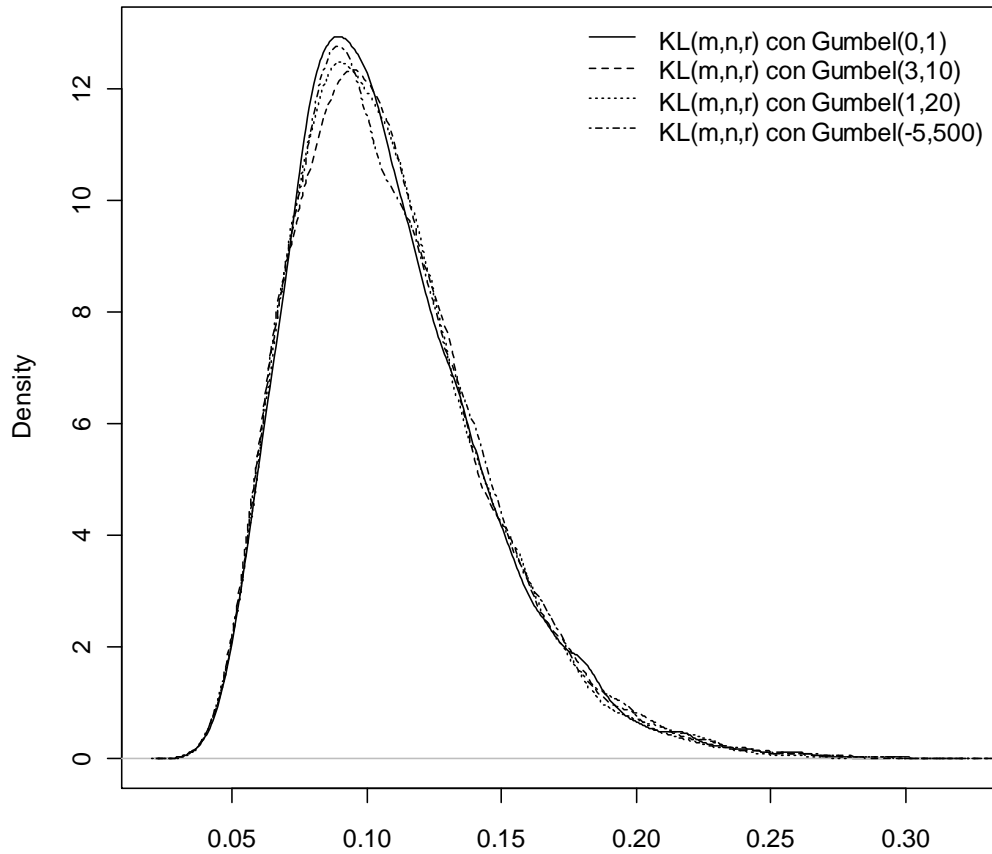


Figura 4.1: Distribuciones empíricas de Kullback-Leibler ($KL^*(m, n, r)$), generadas con 10,000 réplicas de la distribución Gumbel, para varios valores de los parámetros de localidad y escala.

Tabla 4.1: Valores críticos $K_{m,n,r}$ de la estadística $KL^*(m, n, r)$, obtenida mediante simulación Monte Carlo, con $B = 10,000$.

α	0.01		0.02		0.05		0.10		0.15		
n	r	m	$K_{m,n,r}$	m	$K_{m,n,r}$	m	$K_{m,n,r}$	m	$K_{m,n,r}$	m	$K_{m,n,r}$
10	5	4	0.1954	4	0.1908	3	0.1831	3	0.1589	3	0.1462
	6	4	0.2335	4	0.2180	4	0.1902	4	0.1774	4	0.1675
	7	5	0.2340	5	0.2260	5	0.2140	4	0.1997	4	0.1899
	8	6	0.2701	5	0.2634	5	0.2335	5	0.2201	5	0.2104
	9	6	0.3066	6	0.2912	6	0.2699	5	0.2569	5	0.2468
20	10	5	0.1662	6	0.1560	6	0.1361	5	0.1300	5	0.1237
	12	8	0.1741	8	0.1688	7	0.1560	5	0.1460	5	0.1403
	14	9	0.1912	8	0.1835	6	0.1724	6	0.1604	6	0.1566
	16	7	0.2119	10	0.2061	9	0.1919	7	0.1845	7	0.1747
	18	11	0.2470	11	0.2400	6	0.2225	5	0.2114	4	0.1990
30	15	10	0.1435	6	0.1379	6	0.1238	7	0.1122	6	0.1084
	18	8	0.1592	7	0.1477	8	0.1381	7	0.1290	7	0.1220
	21	10	0.1688	9	0.1609	8	0.1543	8	0.1424	5	0.1495
	24	11	0.1865	10	0.1779	9	0.1708	8	0.1591	4	0.1342
	27	14	0.2230	11	0.2075	6	0.1864	7	0.1732	4	0.1586
40	20	10	0.1302	10	0.1216	8	0.1105	8	0.1005	5	0.0981
	24	10	0.1405	10	0.1337	12	0.1248	9	0.1152	6	0.1092
	28	14	0.1540	11	0.1461	6	0.1385	8	0.1289	5	0.1155
	32	13	0.1704	12	0.1640	10	0.1540	4	0.1371	6	0.1247
	36	6	0.1989	7	0.1817	7	0.1604	7	0.1445	6	0.1318
50	25	11	0.1180	9	0.1107	10	0.1015	9	0.0954	7	0.0887
	30	11	0.1273	12	0.1201	8	0.1148	7	0.1040	6	0.0956
	35	12	0.1432	12	0.1342	6	0.1248	8	0.1103	5	0.1031
	40	7	0.1597	7	0.1464	6	0.1301	6	0.1166	6	0.1102
	45	6	0.1697	8	0.1559	7	0.1361	7	0.1274	6	0.1153
60	30	12	0.1084	12	0.1043	9	0.0949	5	0.0852	6	0.0784
	36	12	0.1201	13	0.1144	6	0.1040	7	0.0920	6	0.0861
	42	14	0.1342	11	0.1269	9	0.1116	7	0.0981	7	0.0900
	48	6	0.1422	9	0.1325	7	0.1177	7	0.1069	8	0.0987
	54	8	0.1499	8	0.1410	6	0.1248	7	0.1128	8	0.1043
70	35	12	0.1000	12	0.0953	5	0.0878	8	0.0787	5	0.0698
	42	11	0.1129	9	0.1052	10	0.0974	6	0.0847	6	0.0797
	49	8	0.1226	9	0.1116	6	0.1008	9	0.0922	7	0.0839
	56	8	0.1289	9	0.1181	7	0.1084	7	0.0968	8	0.0893
	63	7	0.1322	6	0.1282	9	0.1168	9	0.1027	7	0.0960
80	40	14	0.0949	7	0.0909	8	0.0827	6	0.0732	7	0.0670
	48	11	0.1080	8	0.0988	7	0.0891	8	0.0793	7	0.0736
	56	9	0.1051	9	0.1044	9	0.0940	8	0.0842	6	0.0779

Tabla 4.2: Valores críticos $K_{m,n,r}$ de la estadística $KL^*(m, n, r)$, obtenida mediante simulación Monte Carlo, con $B = 10,000$ (*Continuación*).

α	0.01		0.02		0.05		0.10		0.15		
n	r	m	$K_{m,n,r}$	m	$K_{m,n,r}$	m	$K_{m,n,r}$	m	$K_{m,n,r}$	m	$K_{m,n,r}$
90	45	10	0.0899	9	0.0854	7	0.0762	8	0.0680	9	0.0641
	54	10	0.0949	10	0.0930	8	0.0804	8	0.0748	8	0.0700
	63	10	0.1023	7	0.0954	9	0.0860	7	0.0785	10	0.0733
	72	9	0.1115	10	0.1028	8	0.0933	10	0.0831	8	0.0767
	81	9	0.1149	10	0.1085	8	0.0965	9	0.0866	8	0.0824
100	50	7	0.0877	8	0.0801	7	0.0709	7	0.0650	8	0.0596
	60	7	0.0907	8	0.0849	9	0.0770	7	0.0691	9	0.0648
	70	8	0.0981	9	0.0901	7	0.0817	7	0.0725	7	0.0694
	80	8	0.0981	12	0.0948	8	0.0857	10	0.0773	9	0.0723
	90	9	0.1077	11	0.1000	9	0.0899	8	0.0834	9	0.0772
120	60	8	0.0754	11	0.0711	8	0.0644	8	0.0573	8	0.0536
	72	10	0.0791	9	0.0744	10	0.0696	8	0.0630	8	0.0586
	84	7	0.0860	10	0.0809	8	0.0745	9	0.0659	8	0.0628
	96	9	0.0869	10	0.0841	9	0.0771	10	0.0714	10	0.0659
	108	12	0.0943	10	0.0903	9	0.0810	10	0.0742	8	0.0682
140	70	8	0.0692	11	0.0652	9	0.0572	8	0.0534	9	0.0491
	84	10	0.0746	11	0.0695	8	0.0632	8	0.0574	8	0.0524
	98	10	0.0767	10	0.0737	8	0.0660	9	0.0599	9	0.0566
	112	10	0.0812	14	0.0791	11	0.0701	11	0.0636	11	0.0602
	126	12	0.0871	11	0.0807	10	0.0734	10	0.0659	9	0.0643
160	80	11	0.0638	11	0.0606	12	0.0542	10	0.0481	9	0.0452
	96	8	0.0675	11	0.0622	10	0.0577	9	0.0530	9	0.0488
	112	13	0.0731	9	0.0674	11	0.0601	10	0.0564	10	0.0529
	128	11	0.0750	11	0.0704	10	0.0635	9	0.0594	12	0.0561
	144	12	0.0762	11	0.0755	11	0.0675	11	0.0615	11	0.0575
180	90	8	0.0592	11	0.0561	9	0.0504	8	0.0448	9	0.0432
	108	14	0.0628	10	0.0578	10	0.0536	12	0.0483	10	0.0454
	126	10	0.0652	13	0.0623	9	0.0565	9	0.0530	12	0.0486
	144	12	0.0709	14	0.0671	12	0.0600	11	0.0550	10	0.0523
	162	12	0.0723	13	0.0688	11	0.0628	10	0.0576	12	0.0555
200	100	12	0.0548	10	0.0522	12	0.0466	9	0.0423	10	0.0403
	120	14	0.0582	12	0.0564	12	0.0506	11	0.0459	9	0.0436
	140	13	0.0620	10	0.0591	12	0.0531	11	0.0488	13	0.0462
	160	10	0.0665	13	0.0623	12	0.0564	12	0.0523	13	0.0491
	180	13	0.0680	13	0.0631	14	0.0594	11	0.0541	13	0.0516

Capítulo 5

Estudio de Simulación Monte Carlo para estimar la potencia y tamaño de la prueba, contra algunas pruebas alternativas

5.1. Potencia de la prueba

El estudio de comparación de potencias se realizó tomando en consideración algunas distribuciones alternativas para diversos tamaños de muestras y niveles de significancia. Para realizar esta comparación se utilizaron las pruebas del coeficiente de correlación basadas en el estimador de *Kaplan-Meier* y *Nelson-Aalen*.

Se tomaron como distribuciones alternativas las siguientes:

- Distribución *Normal*

$$f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(x - \mu)^2\right\} I_{(-\infty, \infty)}(x), \mu \in \mathbb{R}, \sigma > 0.$$

- Distribución *Weibull*

$$f(x; \lambda, \beta) = \beta\lambda^\beta x^{\beta-1} \exp\left\{-(\lambda x)^\beta\right\} I_{(0, \infty)}(x), \beta > 0, \lambda > 1.$$

- Distribución *Gamma*

$$f(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp\left\{-\frac{x}{\beta}\right\} I_{(-\infty, \infty)}(x), \alpha, \beta > 0.$$

- Distribución *Log - normal*

$$f(x; \mu, \sigma^2) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(\log x - \mu)^2\right\} I_{(0,\infty)}(x), \mu \in \mathbb{R}, \sigma > 0.$$

- Distribución *Logística*

$$f(x; \mu, \sigma) = \frac{1}{\beta} \frac{\exp\left\{-\left(\frac{x-\mu}{\sigma}\right)\right\}}{\left\{1 + \exp\left\{-\left(\frac{x-\mu}{\sigma}\right)\right\}\right\}^2} I_{(-\infty,\infty)}(x), \mu \in \mathbb{R}, \beta > 0.$$

- Distribución *Cauchy estándar*

$$f(x) = \frac{1}{\pi(1+x^2)} I_{(-\infty,\infty)}(x).$$

5.2. Prueba utilizando el estimador de Kaplan-Meier

Sea $X_{(1)}, \dots, X_{(r)}$ una muestra aleatoria con censura por la derecha Tipo II, de una distribución conocida F , con función de densidad $f_X(x) \in \mathbb{R}$ y con media finita, por lo que se desea plantear el juego de hipótesis de la sección 4.1, ecuaciones (4.1 y 4.2).

5.2.1. Construcción de la prueba

Tomando la Función de Supervivencia dada por la ecuación (3.5) y sustituyendo la ecuación (3.2) de la función de distribución Gumbel, tenemos:

$$S(x) = 1 - F_X(x) = 1 - \exp\left\{-\exp\left\{-\frac{x-\xi}{\theta}\right\}\right\}. \quad (5.1)$$

Lo que es equivalente a:

$$1 - S(x) = \exp\left\{-\exp\left\{-\frac{x-\xi}{\theta}\right\}\right\}. \quad (5.2)$$

Aplicando logaritmos a ambos miembros de la igualdad de la ecuación (5.2).

$$\log\{1 - S(x)\} = \log\left\{\exp\left\{-\exp\left\{-\frac{x-\xi}{\theta}\right\}\right\}\right\}$$

$$-\log \{1 - S(x)\} = \exp \left\{ -\frac{x - \xi}{\theta} \right\}$$

Aplicado nuevamente logaritmos:

$$\log \{-\log \{1 - S(x)\}\} = \frac{x - \xi}{\theta}. \quad (5.3)$$

Finalmente podemos decir que bajo H_0 , existe una relación lineal entre $\log \{-\log \{1 - S(x)\}\}$ y x , por lo que tomando como referencia al modelo lineal tenemos:

$$y = \beta_0 + \beta_1 x. \quad (5.4)$$

Si observamos la ecuación (5.3) y la comparamos con la ecuación (5.4), se puede observar lo siguiente:

$$y = \log \{-\log \{1 - S(x)\}\}.$$

En donde y , es la variable respuesta.

Y los parámetros del modelo de regresión:

$$\beta_0 = -\frac{\xi}{\theta}, \quad \beta_1 = \frac{1}{\theta}.$$

Utilizando la Función de Riesgo Acumulado, se puede demostrar que se cumple la misma relación lineal.

$$\log \{-\log \{1 - \exp(-H(x))\}\} = \frac{x - \xi}{\theta} \quad (5.5)$$

De manera que una forma para medir la asociación lineal entre y e x , es mediante el Coeficiente de Correlación denotado por su estimador de la ecuación (3.4).

5.2.2. Distribución del estadístico de prueba

Los valores críticos de $K_{K-M}(\alpha)$ fueron determinados a partir de una aproximación, tomando como base a los percentiles muestrales de la distribución empírica de R_{K-M} (Coeficiente de Correlación del estadístico *Kaplan-Meier*), bajo la hipótesis nula. La estimación de los percentiles muestrales se realizó mediante simulación por el método de Monte Carlo en el lenguaje de programación **R** (Development Core Team, 2011).

El procedimiento consistió en generar muestras aleatorias de tamaño 10 (10), 100 y 120(20), 500, de la distribución *Gumbel* $(0, 1)$ con varios porcentajes de censura y se calculó el Coeficiente de Correlación R_{K-M} . El algoritmo utilizado, se muestra a continuación.

Algoritmo 5.2.2. Procedimiento para obtener los percentiles de la distribución del Coeficiente de Correlación del estadístico de Kaplan-Meier.

1. Fijar $n, r, \xi = 0$ y $\theta = 1$.
2. Generar una muestra aleatoria censurada Tipo II de la distribución *Gumbel* con los parámetros indicados en 1, que denotaremos por: $(x_{(1)}, \dots, x_{(n)}), (\delta_1, \dots, \delta_n)$.
3. Estimar la función de supervivencia con *K-M*, mediante la ecuación (3.8).
4. Calcular $\log(-\log(1 - \hat{S}(x_i))) = y_i$.
5. Calcular R_{K-M} , entre y_i e x_i , con las x 's ya ordenadas.
6. Repetir los pasos 2 a 5 B veces.

Al finalizar la simulación, se tendrán B muestras de R_{K-M} , con los cuales se estiman los percentiles.

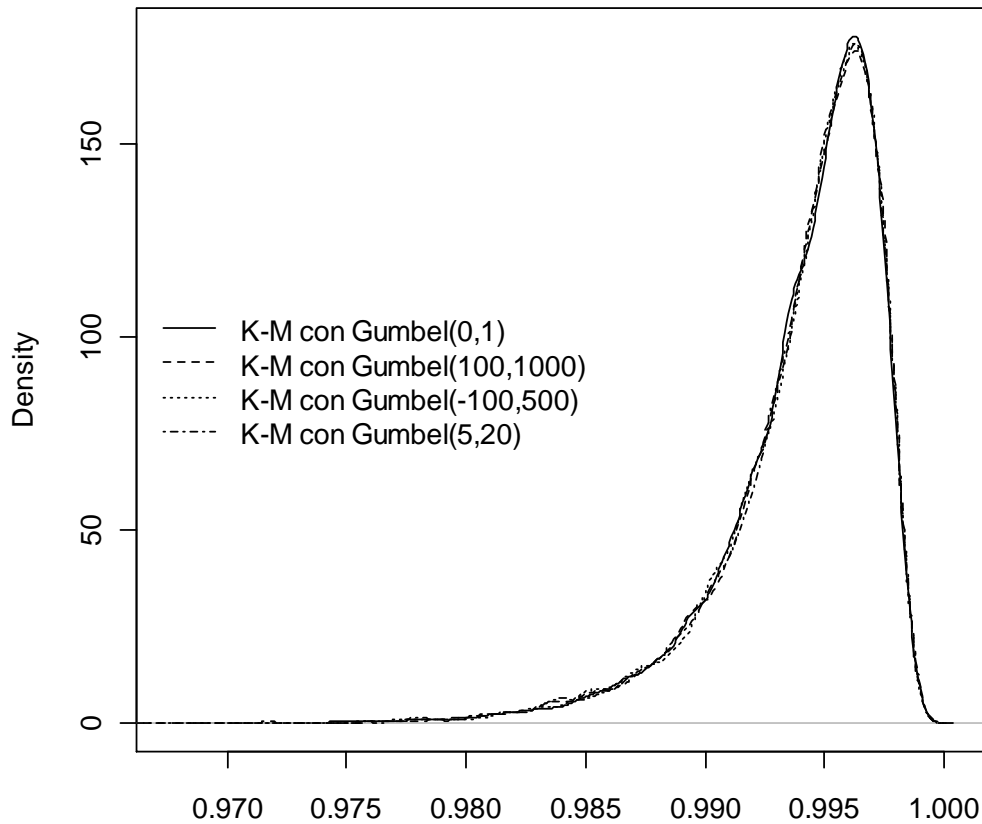


Figura 5.1: Distribución del estadístico de prueba de Kaplan-Meier para la distribución Gumbel, con diferentes valores de los parámetros de localidad y escala.

Si observamos la Figura 5.1, la forma de la distribución es invariante a transformaciones de localidad y escala; es por eso que con esta justificación, se generaron tablas de constantes críticas $K_{K-M}(\alpha)$ para la *Gumbel* $(0, 1)$, con sus respectivos porcentajes de censura y niveles de significancia $\alpha = 0.01, 0.02, 0.05, 0.1$ y 0.15 .

A continuación se presenta la demostración, del por que la forma de la distribución es invariante ante transformaciones de localidad y escala.

Demostración:

De la ecuación (5.3), tenemos:

$$\log \{-\log \{1 - S(x)\}\} = \frac{x - \xi}{\theta}.$$

Note que, si $X \sim \text{Gumbel}(\xi, \theta)$, y $Y \sim \text{Gumbel}(0, 1)$, entonces $X \stackrel{d}{=} \theta Y + \xi$, por lo tanto:

$$\log \{-\log \{1 - S(x)\}\} = \frac{\theta Y + \xi - \xi}{\theta} = Y.$$

Por lo que la relación lineal existente toma la siguiente forma:

$$\log \{-\log \{1 - S(x)\}\} = Y. \quad (5.6)$$

Lo que finalmente nos indica, es que las y_i 's, pueden ser representadas como estadísticas de orden $Y_{(1)} < Y_{(2)} < Y_{(r)}$, distribuidas como Gumbel $(0, 1)$ y dado que la función de supervivencia $\hat{S}(x)$ de Kaplan-Meier, planteado en la sección 3.4.3, ecuación (3.8), no depende de ningún parámetro de localidad y escala. Se puede observar que en ambas partes de la igualdad de la ecuación (5.6), no se depende de ningún parámetro de localidad y escala; por lo tanto la distribución del Coeficiente de Correlación, es invariante ante transformaciones de localidad y escala.

Tabla 5.1: Valores críticos ($K_{K-M}(\alpha)$) del Coeficiente de Correlación R_{K-M} , obtenidos mediante Simulación Monte Carlo con $B = 10,000$ y muestras de tamaño n , con un nivel de significancia $\alpha = 0.05$.

n	Censura (%)							
	10	20	30	40	50	60	70	80
10	0.9013	0.9017	0.8948	0.8871	0.8754	0.8629	0.8686	1.0000
20	0.9459	0.9424	0.9385	0.9296	0.9169	0.9048	0.8852	0.8686
30	0.9626	0.9619	0.9564	0.9483	0.9386	0.9271	0.9071	0.8859
40	0.9715	0.9707	0.9672	0.9608	0.9521	0.9414	0.9261	0.9006
50	0.9771	0.9757	0.9725	0.9685	0.9600	0.9507	0.9375	0.9135
60	0.9811	0.9799	0.9766	0.9722	0.9664	0.9576	0.9444	0.9238
70	0.9838	0.9824	0.9795	0.9763	0.9708	0.9632	0.9504	0.9337
80	0.9857	0.9846	0.9824	0.9789	0.9740	0.9670	0.9561	0.9398
90	0.9871	0.9863	0.9842	0.9806	0.9768	0.9703	0.9605	0.9428
100	0.9887	0.9878	0.9861	0.9830	0.9793	0.9729	0.9628	0.9460
120	0.9903	0.9897	0.9881	0.9855	0.9820	0.9764	0.9682	0.9530
140	0.9919	0.9908	0.9898	0.9870	0.9846	0.9796	0.9722	0.9587
160	0.9929	0.9923	0.9910	0.9890	0.9863	0.9821	0.9758	0.9627
180	0.9935	0.9929	0.9919	0.9903	0.9879	0.9840	0.9779	0.9673
200	0.9943	0.9937	0.9928	0.9910	0.9888	0.9855	0.9792	0.9686
220	0.9948	0.9943	0.9933	0.9920	0.9898	0.9865	0.9815	0.9708
240	0.9952	0.9947	0.9938	0.9926	0.9903	0.9876	0.9824	0.9732
260	0.9956	0.9953	0.9944	0.9932	0.9912	0.9884	0.9838	0.9743
280	0.9960	0.9954	0.9947	0.9936	0.9917	0.9893	0.9851	0.9770
300	0.9962	0.9959	0.9951	0.9940	0.9923	0.9901	0.9861	0.9779
320	0.9965	0.9960	0.9954	0.9943	0.9927	0.9906	0.9869	0.9793
340	0.9966	0.9963	0.9957	0.9948	0.9932	0.9911	0.9873	0.9801
360	0.9968	0.9965	0.9959	0.9949	0.9934	0.9916	0.9879	0.9809
380	0.9969	0.9967	0.9961	0.9952	0.9940	0.9921	0.9886	0.9818
400	0.9972	0.9968	0.9962	0.9955	0.9943	0.9925	0.9892	0.9831
420	0.9972	0.9970	0.9965	0.9957	0.9945	0.9928	0.9898	0.9836
440	0.9974	0.9971	0.9967	0.9959	0.9948	0.9930	0.9899	0.9839
460	0.9975	0.9973	0.9968	0.9961	0.9948	0.9933	0.9906	0.9847
480	0.9976	0.9974	0.9969	0.9962	0.9951	0.9935	0.9908	0.9854
500	0.9977	0.9975	0.9970	0.9964	0.9954	0.9938	0.9910	0.9856

En la Tabla 5.1, se muestran los valores críticos $K_{K-M}(\alpha)$, para un nivel de significancia $\alpha = 0.05$, para los otros niveles de significancia, se pueden revisar en las Tablas del Anexo III-K-M, de este documento.

5.2.3. Implementación de la prueba

Según las condiciones dadas por las variables aleatorias generadas, y sometidas a censura Tipo II; esta prueba consistirá de n unidades, la cual solo se dispondrán de las primeras r observaciones, para que finalmente se puedan probar los juegos de hipótesis planteados en la sección 4.1, ecuaciones (4.1 y 4.2).

Regla de decisión:

- Rechazar H_0 con un nivel de significancia α si y solo sí, $R_{K-M} \leq K_{K-M}(\alpha)$.

Donde $K_{K-M}(\alpha)$, debe cumplir la siguiente condición:

$$\begin{aligned}\alpha &= P(\text{Rechazar } H_0 \mid H_0) \\ &= P(R_{K-M} \leq K_{K-M}(\alpha)).\end{aligned}$$

5.3. Tamaño de la prueba

Para determinar el tamaño de la prueba, se calculó el coeficiente de correlación R_{K-M} , implementado este procedimiento en **R** (Development Core Team, 2011), con $B = 10,000$ repeticiones, para muestras aleatorias *Gumbel* (ξ, θ) , para tamaños de muestras $n = 20, 50$ y porcentajes de censura de 20% y 50%, el nivel de significancia utilizado fue para $\alpha = 0.01, 0.02, 0.05, 0.10$ y 0.15 .

El estudio del tamaño de la prueba fue consistente con el nivel de significancia utilizado, las Tablas generadas se encuentran en el Anexo III-K-M.

5.4. Potencia de la prueba

El estudio de potencia para la prueba estadística es de gran importancia, de esta manera podemos observar el comportamiento del Error Tipo I y II. Es por eso, que es necesario ver el comportamiento de la prueba del R_{K-M} , ante otras distribuciones alternativas. Para cada una de las distribuciones alternativas ya mencionadas, se generaron muestras de tamaño 20 y 50 mediante simulación Monte Carlo, y permitiendo niveles de censura de 50 y 20%, para cada una de las mismas. Para estimar la potencia, se calculó R_{K-M} y se estimó el número de veces que las $B = 10,000$ replicas rechazaban la hipótesis nula, siendo cierta la hipótesis alternativa. El Algoritmo utilizado, se presenta a continuación.

Algoritmo 5.4.1. Procedimiento para estimar la potencia de la prueba.

1. *Seleccionar el tamaño de la muestra n , r y α .*
2. *Seleccionar la distribución alternativa.*
3. *Hacer $Suma = 0$.*
4. *Generar una muestra censurada de tamaño n de la distribución del paso 2.*
5. *Calcular R_{K-M} y comparar con $K_{K-M}(\alpha)$, si existe rechazo hacer $Suma = Suma + 1$.*
6. *Repetir los pasos 4 y 5 B veces.*
7. *Para estimar la potencia, evaluar $Suma/B$.*

Los estudios de potencia para la distribuciones alternativas ya mencionadas, se encuentran en las Tablas del Anexo III-K-M.

5.5. Prueba usando el estimador de Nelson-Aalen

5.5.1. Distribución del estadístico de prueba

Tomando como base a los percentiles muestrales de la distribución empírica de R_{N-A} (Coeficiente de Correlación del estadístico Nelson-Aalen), bajo la hipótesis nula, se obtuvieron los valores críticos de $K_{N-A}(\alpha)$. La estimación de los percentiles muestrales se realizó mediante la Simulación por el método de Monte Carlo, en el lenguaje de programación **R** (Development Core Team, 2011).

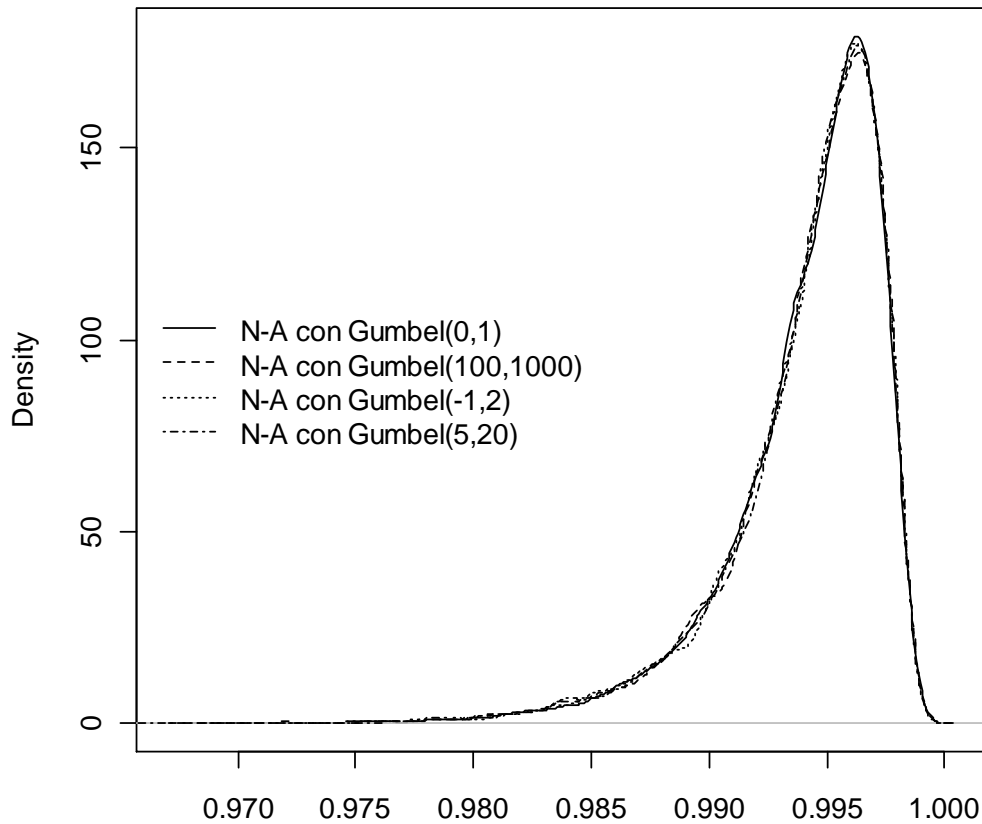


Figura 5.2: Distribución del estadístico de prueba de Nelson-Aalen para la distribución Gumbel, con diferentes valores de los parámetros de localidad y escala.

Se generaron muestras aleatorias con censura de 10 (10), 80%, por la derecha Tipo II y de la distribución *Gumbel* (0, 1), de tamaños 10 (10), 100. Se programó una rutina en **R** (Development Core Team, 2011), para describir el comportamiento de la distribución del Coeficiente de Correlación, con 10,000 repeticiones, manteniendo el mismo tamaño de la muestra y porcentaje de censura, pero con diferentes valores de los parámetros de localidad y escala. El procedimiento es semejante que el utilizado para Kaplan-Meier (Algoritmo 5.2.2), la única diferencia es utilizar la Función de Riesgo Acumulado de la ecuación (3.11).

Tabla 5.2: Valores críticos ($K_{N-A}(\alpha)$) del Coeficiente de Correlación R_{N-A} , obtenidos mediante Simulación Monte Carlo con $B = 10,000$ y muestras de tamaño n , con un nivel de significancia $\alpha = 0.05$.

n	Censura (%)							
	10	20	30	40	50	60	70	80
10	0.9097	0.9030	0.8960	0.8839	0.8779	0.8658	0.8671	1.0000
20	0.9484	0.9441	0.9383	0.9302	0.9188	0.9036	0.8866	0.8679
30	0.9642	0.9618	0.9568	0.9492	0.9408	0.9260	0.9084	0.8851
40	0.9724	0.9703	0.9666	0.9612	0.9539	0.9416	0.9246	0.8997
50	0.9778	0.9762	0.9727	0.9681	0.9608	0.9508	0.9351	0.9148
60	0.9818	0.9796	0.9765	0.9726	0.9664	0.9572	0.9441	0.9239
70	0.9839	0.9831	0.9806	0.9761	0.9712	0.9631	0.9514	0.9314
80	0.9862	0.9851	0.9826	0.9786	0.9740	0.9676	0.9557	0.9380
90	0.9875	0.9864	0.9841	0.9810	0.9762	0.9698	0.9608	0.9423
100	0.9887	0.9877	0.9857	0.9832	0.9790	0.9727	0.9630	0.9471

En la Figura 5.2, se observa que la distribución del estadístico de prueba es invariante a cambios en los parámetros de localidad y escala. La Tabla 5.2, muestra los valores críticos $K_{N-A}(\alpha)$, para un nivel de significancia $\alpha = 0.05$; también se generaron Tablas de $K_{N-A}(\alpha)$ para los niveles de significancia $\alpha=0.01, 0.02, 0.10$ y 0.15 , y se encuentran en el Anexo IV-N-A, de este documento.

5.5.2. Implementación de la prueba

Se dispondrán de variables aleatorias $X_{(1)}, \dots, X_{(r)}$, de la distribución de Gumbel bajo censura por la derecha Tipo II, y se probará el juego de hipótesis de las ecuaciones (4.1 y 4.2).

Regla de decisión:

- Rechazar H_0 con un nivel de significancia α si y solo sí, $R_{N-A} \leq K_{N-A}(\alpha)$.

Donde $K_{N-A}(\alpha)$, debe cumplir la siguiente condición:

$$\alpha = P(\text{Rechazar } H_0 \mid H_0)$$

$$= P(R_{N-A} \leq K_{N-A}(\alpha)).$$

5.6. Tamaño de la prueba

Para determinar el tamaño de la prueba, se calculó el Coeficiente de Correlación R_{N-A} , el procedimiento Monte Carlo fue simulado en **R** (Development Core Team, 2011), y es parecido al realizado para Kaplan-Meier; el cual se generaron $B = 10,000$ repeticiones, para muestras aleatorias *Gumbel* (ξ, θ) estándar, para tamaños de muestras $n = 20, 50$ y porcentajes de censura de 20 y 50% para cada una de las mismas, los niveles de significancia utilizados fueron, $\alpha = 0.01, 0.02, 0.05, 0.10$ y 0.15 .

El estudio del tamaño de la prueba indica que es consistente con los niveles de significancia utilizados, para mayor detalle revise el Anexo IV-N-A de este documento, en donde se encuentran los estudios de tamaño la prueba para diferentes valores de los parámetros de localidad y escala.

5.7. Potencia de la prueba

Para cada una de las distribuciones alternativas se generaron muestras de tamaños 20 y 50, mediante simulación Monte Carlo con $B = 10,000$ replicas y censuradas por la derecha al 20 y 50%, para cada una de las mismas. Para estimar la potencia, se calculó R_{N-A} y se estimó el número de veces que las B replicas rechazaban la hipótesis nula, siendo cierta la hipótesis alternativa. La rutina que se utilizó es muy semejante a la del Algoritmo 5.4.1.

Los estudios de potencia para la distribuciones alternativas ya mencionadas, se encuentran en las Tablas del Anexo IV-N-A.

Capítulo 6

Estudios de comparación, tamaño y potencia de las pruebas

En este capítulo se presenta el estudio comparativo entre las tres propuestas, comparando el tamaño y potencia de cada una; se utilizaron las distribuciones alternativas de la sección 5.1, con los niveles de significancia $\alpha = 0.01, 0.02, 0.05, 0.10$ y 0.15 , para tamaños de muestra $n = 20$ con censura del 50 % y $n = 50$ con censura del 20 %.

6.1. Comparación en el tamaño estimado de la prueba

Tabla 6.1: Tamaño estimado de las estadísticas de prueba: Kaplan-Meier, Nelson-Aalen y Kullback-Leibler, considerando a la distribución *Gumbel* (0, 1).

n	Censura (%)	α	Kaplan-Meier	Nelson-Aalen	Kullback-Leibler
20	50	0.01	0.011	0.007	0.011
		0.02	0.019	0.016	0.019
		0.05	0.055	0.050	0.058
		0.10	0.099	0.103	0.113
		0.15	0.150	0.146	0.148
50	20	0.01	0.017	0.012	0.014
		0.02	0.018	0.019	0.025
		0.05	0.047	0.050	0.053
		0.10	0.097	0.101	0.107
		0.15	0.150	0.146	0.145

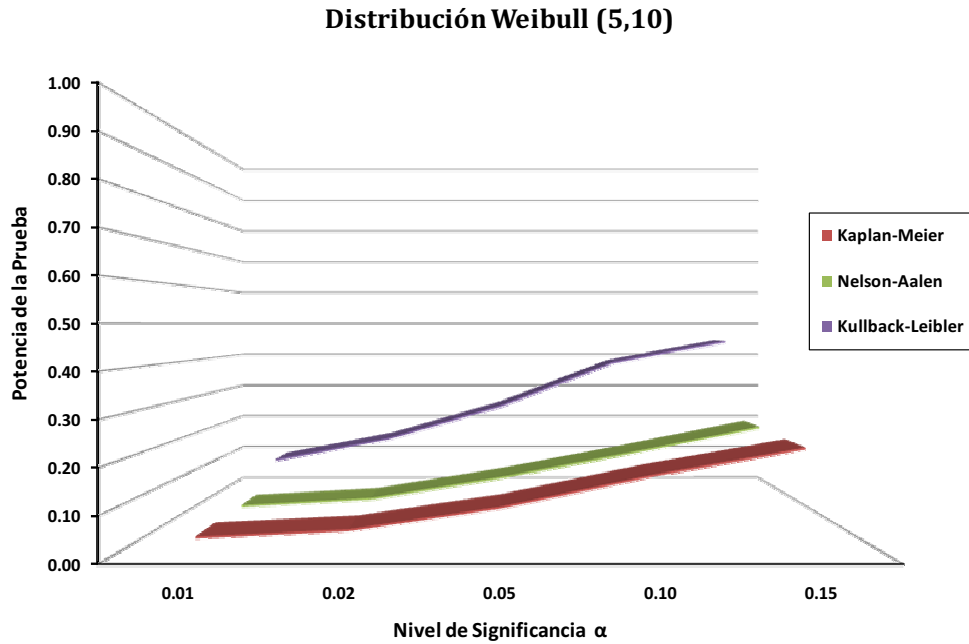
6.2. Estudio comparativo de potencias

Tabla 6.2: Comparación de potencias de las estadísticas de prueba: Kaplan-Meier, Nelson-Aalen y Kullback-Leibler, considerando como alternativa la distribución *Weibull* (5, 10).

n	Censura (%)	α	Kaplan-Meier	Nelson-Aalen	Kullback-Leibler
20	50	0.01	0.0297	0.0271	0.0922
		0.02	0.0460	0.0460	0.1559
		0.05	0.0956	0.1001	0.2526
		0.10	0.1655	0.1638	0.3819
		0.15	0.2252	0.2303	0.4454
50	20	0.01	0.3133	0.2962	0.6140
		0.02	0.3784	0.3722	0.7993
		0.05	0.5104	0.5063	0.9464
		0.10	0.6324	0.6165	0.9385
		0.15	0.7060	0.6846	0.9647

Figura 6.1: Comparación de potencias de las estadísticas de prueba, H_1 es *Weibull* (5, 10).

a) $n = 20$ y Censura = 50 %



b) $n = 50$ y Censura = 20 %

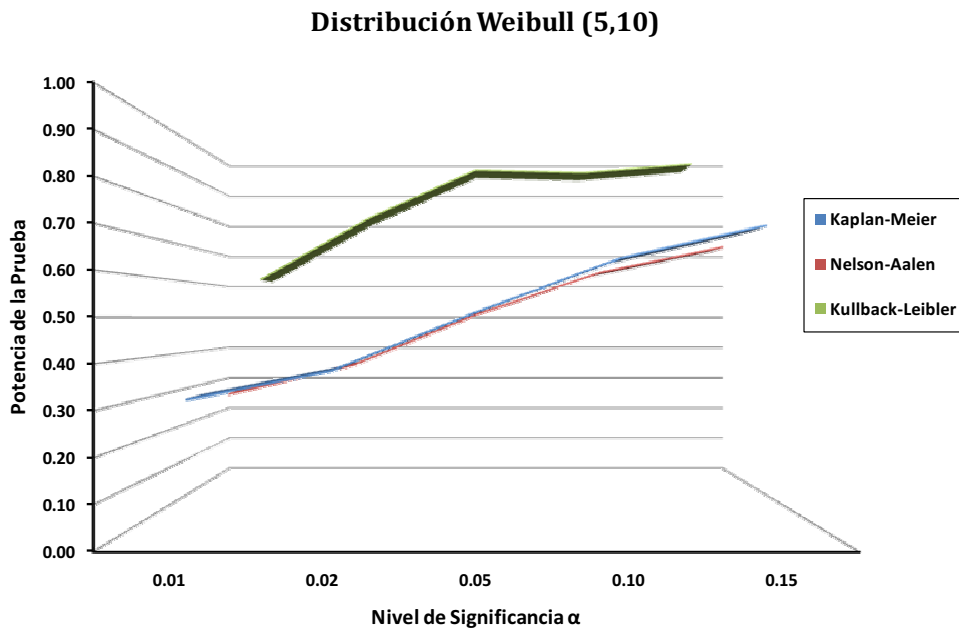


Tabla 6.3: Comparación de potencias de las estadísticas de prueba: Kaplan-Meier, Nelson-Aalen y Kullback-Leibler, considerando como alternativa la distribución *Weibull* (20, 50).

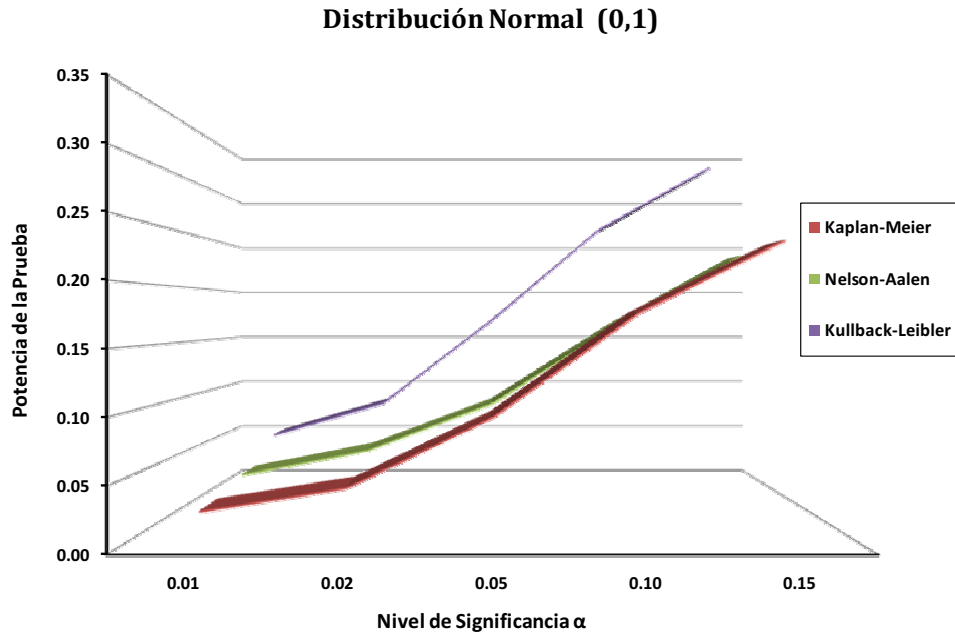
n	Censura (%)	α	Kaplan-Meier	Nelson-Aalen	Kullback-Leibler
20	50	0.01	0.0711	0.0713	0.3747
		0.02	0.1096	0.1048	0.4792
		0.05	0.1823	0.1817	0.6081
		0.10	0.2682	0.2693	0.7220
		0.15	0.3296	0.3324	0.7767
50	20	0.01	0.6950	0.6822	0.4375
		0.02	0.7536	0.7438	0.6049
		0.05	0.8323	0.8285	0.8582
		0.10	0.8943	0.8852	0.8526
		0.15	0.9186	0.9186	0.9106

Tabla 6.4: Comparación de potencias de las estadísticas de prueba: Kaplan-Meier, Nelson-Aalen y Kullback-Leibler, considerando como alternativa la distribución *Normal* (0, 1).

n	Censura (%)	α	Kaplan-Meier	Nelson-Aalen	Kullback-Leibler
20	50	0.01	0.0232	0.0302	0.0502
		0.02	0.0412	0.0522	0.0833
		0.05	0.0955	0.0945	0.1692
		0.10	0.1758	0.1656	0.2639
		0.15	0.2317	0.2275	0.3268
50	20	0.01	0.4199	0.2095	0.7324
		0.02	0.5021	0.2701	0.8780
		0.05	0.6269	0.3990	0.9646
		0.10	0.7438	0.5043	0.9713
		0.15	0.8091	0.5781	0.9794

Figura 6.2: Comparación de potencias de las estadísticas de prueba, H_1 es *Normal* (0, 1).

a) $n = 20$ y Censura = 50 %



b) $n = 50$ y Censura = 20 %

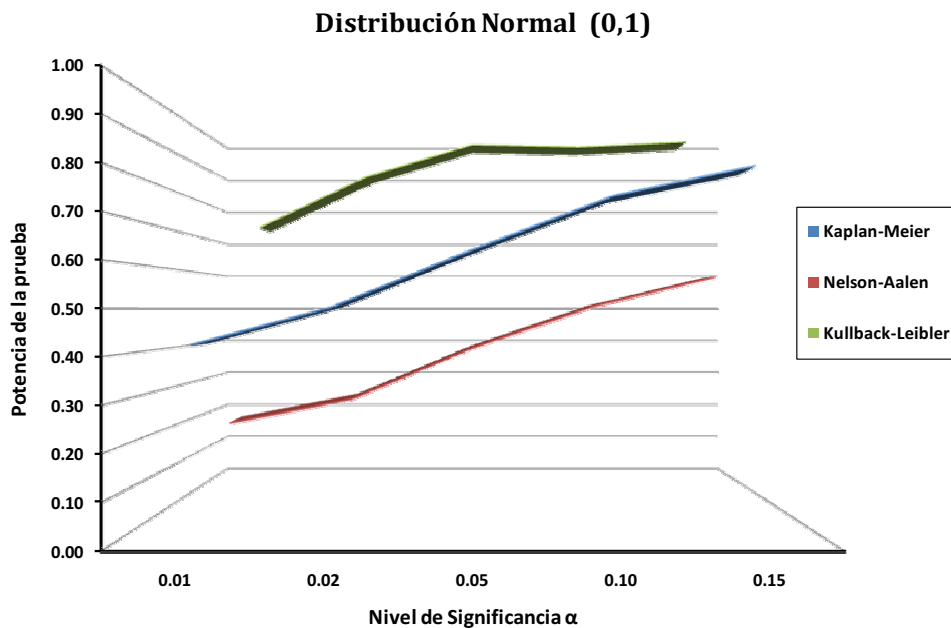
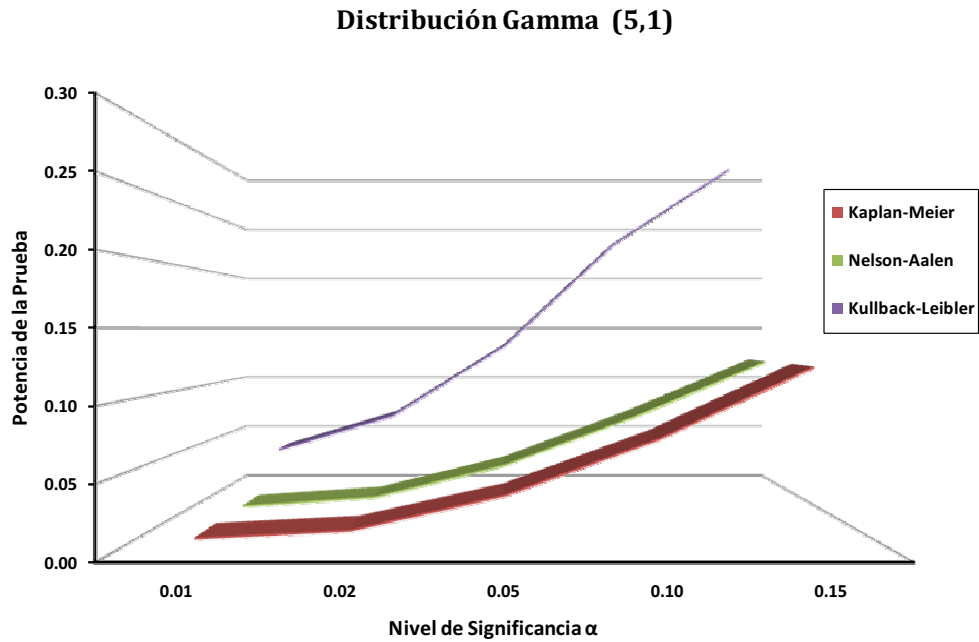


Tabla 6.5: Comparación de potencias de las estadísticas de prueba: Kaplan-Meier, Nelson-Aalen y Kullback-Leibler, considerando como alternativa la distribución *Gamma* (5, 1).

n	Censura (%)	α	Kaplan-Meier	Nelson-Aalen	Kullback-Leibler
20	50	0.01	0.0070	0.0063	0.0363
		0.02	0.0124	0.0135	0.0677
		0.05	0.0360	0.0392	0.1328
		0.10	0.0742	0.0781	0.2299
		0.15	0.1229	0.1221	0.2973
50	20	0.01	0.0084	0.0067	0.4570
		0.02	0.0165	0.0145	0.6893
		0.05	0.0417	0.0462	0.9484
		0.10	0.0927	0.0928	0.9779
		0.15	0.1358	0.1378	0.9992

Figura 6.3: Comparación de potencias de las estadísticas de prueba, H_1 es $Gamma(5, 1)$.

a) $n = 20$ y Censura = 50 %



b) $n = 50$ y Censura = 20 %

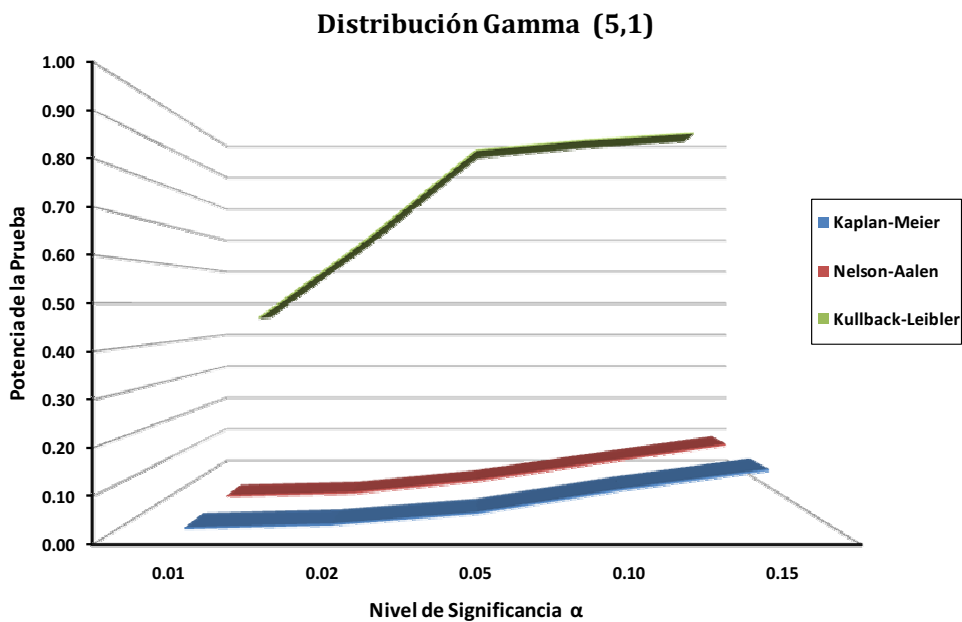


Tabla 6.6: Comparación de potencias de las estadísticas de prueba: Kaplan-Meier, Nelson-Aalen y Kullback-Leibler, considerando como alternativa la distribución *Gamma* (3, 9).

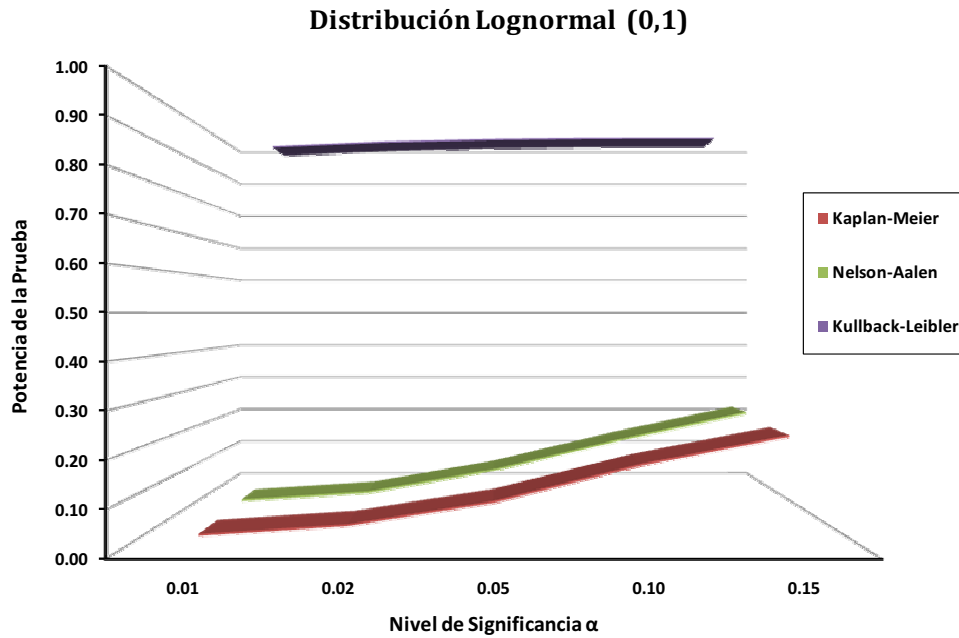
n	Censura (%)	α	Kaplan-Meier	Nelson-Aalen	Kullback-Leibler
20	50	0.01	0.0060	0.0071	0.0863
		0.02	0.0135	0.0151	0.1525
		0.05	0.0410	0.0458	0.2690
		0.10	0.0866	0.0826	0.3215
		0.15	0.1294	0.1322	0.4973
50	20	0.01	0.0083	0.0103	0.6799
		0.02	0.0129	0.0192	0.7965
		0.05	0.0393	0.0471	0.8993
		0.10	0.0842	0.0931	0.9995
		0.15	0.1320	0.1391	0.9998

Tabla 6.7: Comparación de potencias de las estadísticas de prueba: Kaplan-Meier, Nelson-Aalen y Kullback-Leibler, considerando como alternativa la distribución *Log-normal* (0, 1).

n	Censura (%)	α	Kaplan-Meier	Nelson-Aalen	Kullback-Leibler
20	50	0.01	0.0240	0.0298	0.9737
		0.02	0.0449	0.0502	0.9877
		0.05	0.0948	0.1057	0.9952
		0.10	0.1741	0.1809	0.9987
		0.15	0.2357	0.2486	0.9995
50	20	0.01	0.4148	0.4365	0.9996
		0.02	0.4889	0.5197	0.9999
		0.05	0.6289	0.6693	1.0000
		0.10	0.7549	0.7725	1.0000
		0.15	0.8148	0.8288	1.0000

Figura 6.4: Comparación de potencias de las estadísticas de prueba, H_1 es *Log-normal* (0, 1).

a) $n = 20$ y Censura = 50 %



b) $n = 50$ y Censura = 20 %

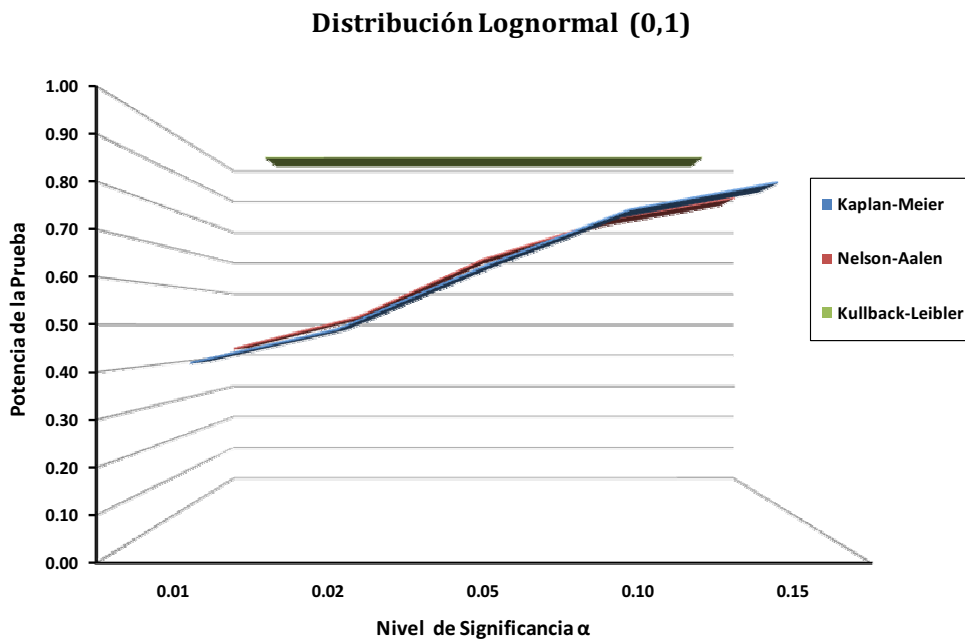
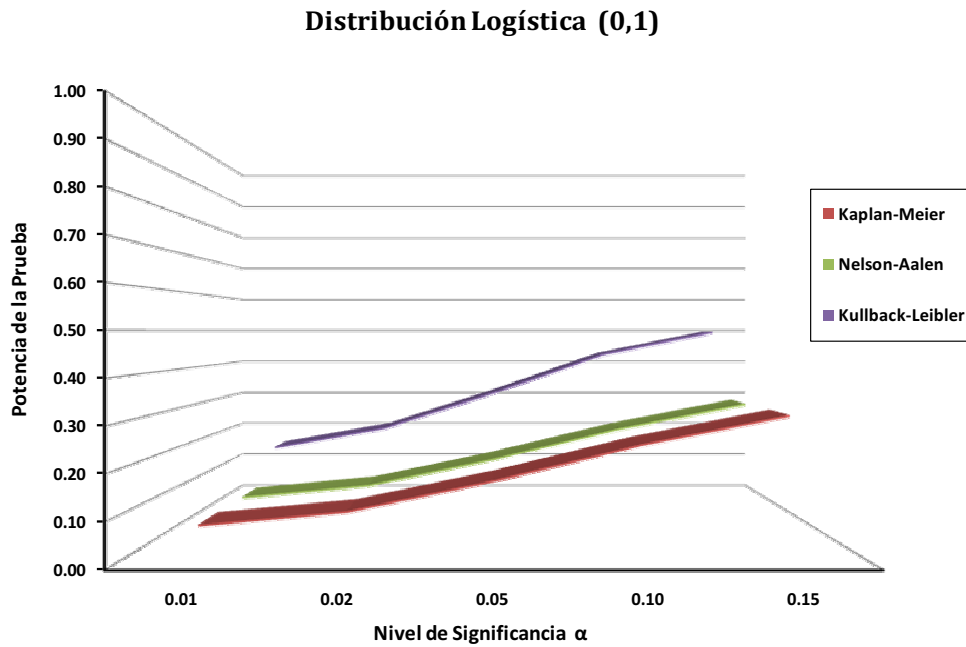


Tabla 6.8: Comparación de potencias de las estadísticas de prueba: Kaplan-Meier, Nelson-Aalen y Kullback-Leibler, considerando como alternativa la distribución *Logística* (0, 1).

n	Censura (%)	α	Kaplan-Meier	Nelson-Aalen	Kullback-Leibler
20	50	0.01	0.0694	0.0667	0.1509
		0.02	0.0996	0.0977	0.2084
		0.05	0.1677	0.1661	0.3137
		0.10	0.2469	0.2441	0.4274
		0.15	0.3109	0.3076	0.4931
50	20	0.01	0.4751	0.4613	0.6981
		0.02	0.5375	0.5305	0.8470
		0.05	0.6503	0.6293	0.9548
		0.10	0.7349	0.7272	0.9625
		0.15	0.7859	0.7730	0.9770

Figura 6.5: Comparación de potencias de las estadísticas de prueba, H_1 es Logística (0, 1).

a) $n = 20$ y Censura = 50 %



b) $n = 50$ y Censura = 20 %

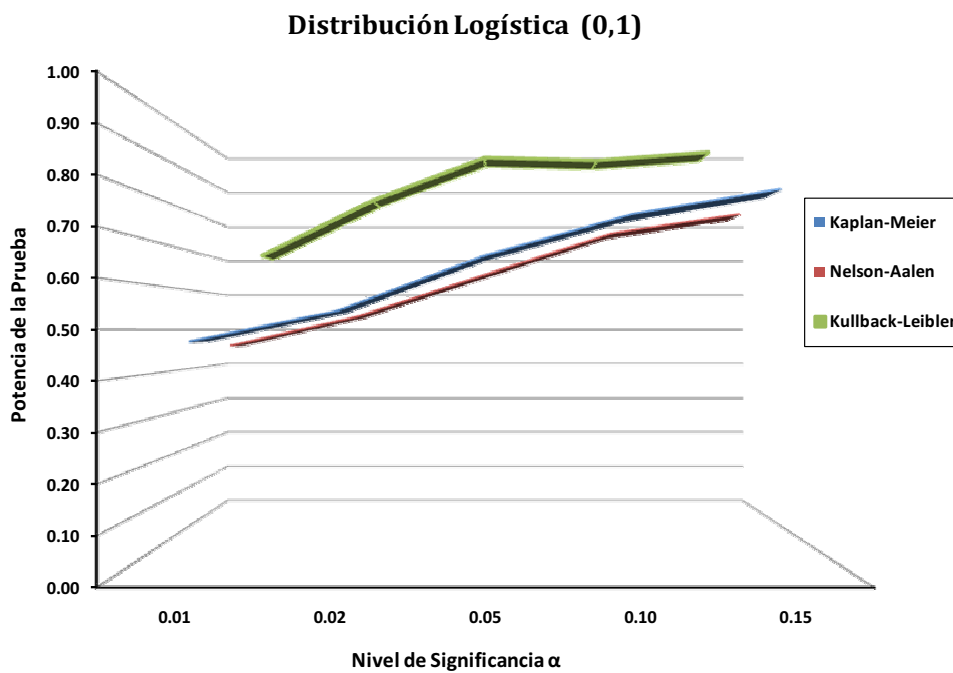
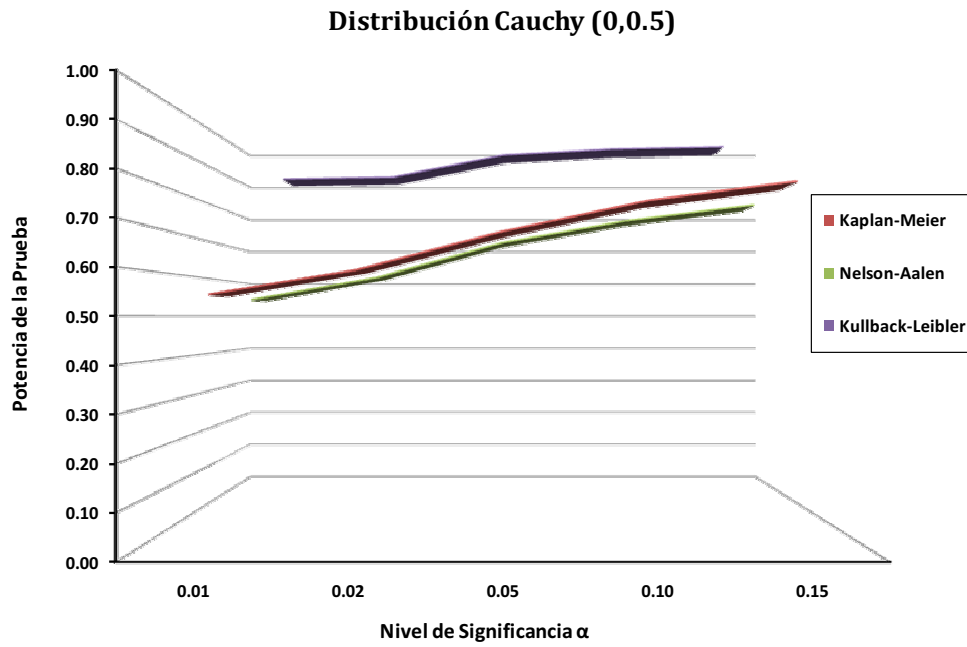


Tabla 6.9: Comparación de potencias de las estadísticas de prueba: Kaplan-Meier, Nelson-Aalen y Kullback-Leibler, considerando como alternativa la distribución *Cauchy* (0, 0.5).

n	Censura (%)	α	Kaplan-Meier	Nelson-Aalen	Kullback-Leibler
20	50	0.01	0.5462	0.5408	0.8923
		0.02	0.5984	0.5962	0.8977
		0.05	0.6812	0.6837	0.8981
		0.10	0.7478	0.7393	0.9007
		0.15	0.7888	0.7783	0.9081
50	20	0.01	0.9867	0.9859	0.9100
		0.02	0.9899	0.9889	0.9351
		0.05	0.9941	0.9927	0.9602
		0.10	0.9966	0.9951	0.9781
		0.15	0.9975	0.9963	0.9908

Figura 6.6: Comparación de potencias de las estadísticas de prueba, H_1 es *Cauchy* (0, 0.5).

a) $n = 20$ y Censura = 50 %



b) $n = 50$ y Censura = 20 %

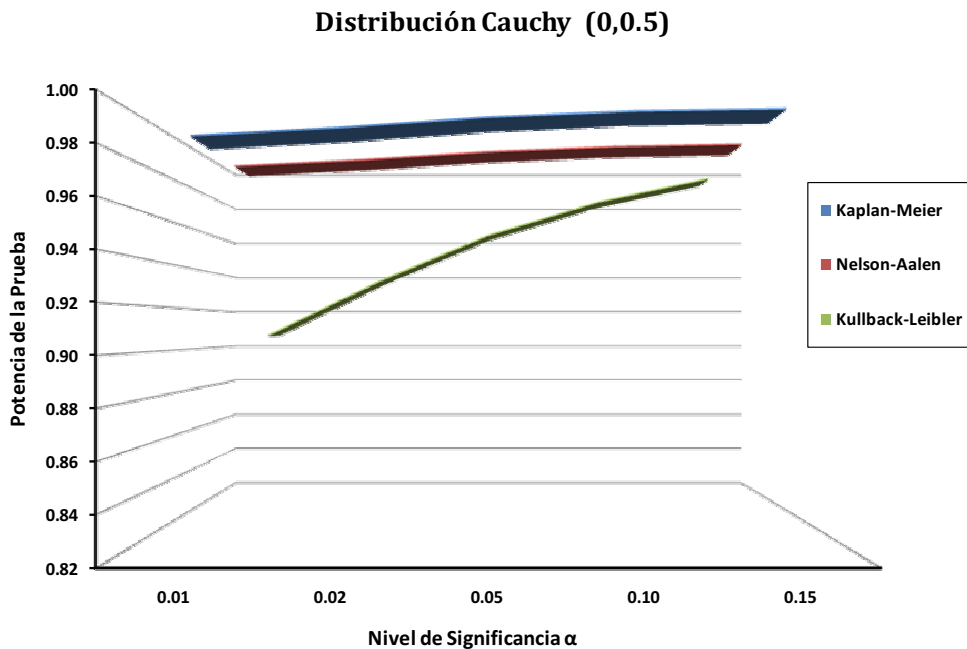


Tabla 6.10: Comparación de potencias de las estadísticas de prueba: Kaplan-Meier, Nelson-Aalen y Kullback-Leibler, considerando como alternativa la distribución *Cauchy* $(0, 1)$.

n	Censura (%)	α	Kaplan-Meier	Nelson-Aalen	Kullback-Leibler
20	50	0.01	0.5453	0.5414	0.8836
		0.02	0.5942	0.5927	0.9153
		0.05	0.6829	0.6889	0.9599
		0.10	0.7482	0.7463	0.9572
		0.15	0.7882	0.7853	0.9693
50	20	0.01	0.9859	0.9855	0.8929
		0.02	0.9881	0.9875	0.9011
		0.05	0.9932	0.9934	0.9612
		0.10	0.9947	0.9956	0.9798
		0.15	0.9972	0.9969	0.9850

Capítulo 7

Ejemplo de Aplicación

En este capítulo, se presenta un ejemplo de aplicación, en el cual se probará el juego de hipótesis de la sección 4.1, ecuaciones (4.1 y 4.2), y que nos permitirá realizar la prueba de bondad de ajuste de la Distribución Gumbel, usando los estadísticos de prueba de Kullback-Leibler, Kaplan-Meier y Nelson-Aalen.

Tabla 7.1: Log de Tiempos de falla de un Aislante Eléctrico sometido a Tensión.

x_i	δ_i	x_i	δ_i	x_i	δ_i	x_i	δ_i	x_i	δ_i
-1.66073	1	1.02245	1	1.54116	0	1.54116	0	1.54116	0
-0.24846	1	1.15057	1	1.54116	0	1.54116	0	1.54116	0
-0.04082	1	1.42311	1	1.54116	0	1.54116	0	1.54116	0
0.27002	1	1.54116	1	1.54116	0	1.54116	0		

La censura progresiva, es de gran importancia durante la planeación de experimentos de confiabilidad, en las industrias se utiliza para estudiar el tiempo de vida de maquinas y unidades. Los datos de este ejemplo, son de un experimento de tiempos de vida, reportados por Nelson (1982), la cual se generaron unas muestras de pruebas de esfuerzos a tensión (34 KV/ minuto) de un tipo de aislante eléctrico, al analizar los datos completos, Nelson asumió una distribución Weibull para los 19 tiempos de rotura (en escala logarítmica), y se muestran en la Tabla 7.1.

Tabla 7.2: Comparación entre Pruebas.

Estimador	n	r	m	Valores Críticos	Estadístico de Prueba	Decisión
Kullback-Leibler	19	8	6	$KL_{6,20,10}(0.05) = 0.1361$	$KL^*(m, n, r) = 0.1477$	Se rechaza H_0
Kaplan-Meier	19	8	-	$K_{K-M}(0.05) = 0.9169$	$R_{K-M} = 0.9526$	No se rechaza H_0
Nelson-Allen	19	8	-	$K_{N-A}(0.05) = 0.9188$	$R_{N-A} = 0.9341$	No se rechaza H_0

De los datos de la Tabla 7.1, para $n = 19$, $r = 8$, $m = 6$, $\alpha = 0.05$ y con los **EMV** $\hat{\xi} = 1.3756$ y $\hat{\theta} = 2.0165$; tomando los valores críticos (mediante interpolación) de las Tablas 4.1, 5.1 y 5.2, para los parámetros ya especificados y realizando la comparación con los estadísticos de prueba, se obtienen resultados contradictorios, ya que para Kaplan-Meier y Nelson Aalen no se rechaza H_0 y para Kullback-Leibler se rechaza H_0 , tal y como lo muestra la Tabla 7.2.

Capítulo 8

Conclusiones y Recomendaciones

- Para las distribuciones alternativas que se consideraron, se observó que la potencia de la prueba utilizando la divergencia de Kullback-Leibler fue más potente, en comparación con las estadísticas de prueba.
- Para las estadísticas de comparación de Nelson-Aalen y Kaplan-Meier, y utilizando las distribuciones alternativas; mostraron valores semejantes en los estudios de potencia y tamaño de la prueba.
- En lo correspondiente al tamaño de la prueba, la divergencia de Kullback-Leibler fue consistente con los niveles de significancia α utilizados; por lo que nos indica que la prueba trabaja adecuadamente para cualquier n y niveles de censura, rechazando o no H_0 cuando sea necesario.
- El tamaño de las pruebas de comparación de Nelson-Aalen y Kaplan-Meier, fueron consistentes con los niveles de significancia α utilizados, lo que indica que las pruebas trabajan bien para cualquier valor de n y valores de censura, rechazando o no H_0 cuando sea necesario.

Referencias Bibliográficas

- Al-Aboud, Fahad M. (2009). Bayesian Estimations for the Extreme Value Distribution Using Progressive Censored Data and Asymmetric Loss. Department of Statistics, Faculty of Science King Abdul-Aziz University, Saudi Arabia. International Mathematical Forum, 4, 2009, no. 33, 1603 - 1622.
- Aalen, O. (1978). Nonparametric Inference for a Family of Counting Processes. The Analysis of Statistics, 6, 701–726.
- Arizono, I. y Ohta, H. (1989). A test for normality based on Kullback-Leibler information, American Statistician, 43, pp. 20–22.
- David W. Hosmer & Stanley Lemeshow, (1999). Applied Survival Analysis: Regression Modeling of Time to Event Data. Wiley Series in Probability and Statistics.
- Greenwood, M. (1926). The natural duration of cancer. Reports on Public Health and Medical Subject, 33, 1–26. Her Majesty's Stationery Office, London.
- Kaplan, E. L. y Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. Journal of the American Statistical Association, 53, 282, 457–481.
- Klein, J. P. y Moeschberger, M. L. (1997). Survival Analysis: Techniques for Censored and Truncated Data. Statistics for Biology and Health.
- Kullback, S. and Leibler, R. A. (1951). On information and sufficiency, Analysis of Mathematical Statistics, 4, 49-70.
- Lim, Jonggun and Park, Sangun (2007). Censored Kullback-Leibler Information and Goodness-of-Fit Test with Type II Censored Data, Journal of Applied Statistics, 34: 9, 1051-1064.
- Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. Technometrics, 14, 945–965.
- Nelson, W. (1982). Applied Life Data Analysis. Wiley, New York.
- Park, S. & Park, D. (2003) Correcting moments for goodness of fit tests based on two entropy estimates, Journal of Statistical Computation and Simulation, 73, pp. 685–69.

- Pérez-Rodríguez, Paulino, Vaquera-Huerta, Humberto and Villaseñor-Alva, José A (2009). A Goodness-of-Fit Test for the Gumbel Distribution Based on Kullback-Leibler Information, Communications in Statistics - Theory and Methods, 38: 6, 842-855.
- R Development Core Team (2011). R: A lenguaje and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Saldaña Zepeda D.P.(2008). Prueba de bondad de ajuste para la Distribución Pareto en presencia de Censura Tipo II. Tesis de Maestría. Colegio de Postgraduados, Campus Montecillo, Texcoco Edo. de México.
- Smith, R. y Bain, L. (1976) Correlation type goodness-of-fit statistic with censored sampling, Communication in Statistics, Part A – Theory and Methods, 5, pp. 119–132
- Suárez Espinoza J.(2000). Una prueba de bondad de ajuste para la Distribución Pareto Generalizada. Tesis de Maestría. Colegio de Postgraduados, Campus Montecillo, Texcoco Edo. México.
- Viveros, R. and Balakrishnan, N.(1994). Interval estimation of parameters of life from progressively censored data. Tech-nometrics, 36, 84-91.

Anexo I

Programa de Simulación en R, para la prueba de Kullback-Leibler

```
#PRUEBA DE BONDAD DE AJUSTE PARA LA DISTRIBUCIÓN GUMBEL, PARA DATOS CENSURADOS TIPO II,
BASADA EN LA DIVERGENCIA DE KULLBACK-LEIBLER
#Eliminar todos los objetos en el área de trabajo
rm(list=ls())
#Genera números aleatorios de la distribución gumbel
#n es tamaño de muestra
#alfa es parámetro de localidad
#beta es parámetro de escala
rgumbel=function(n,alfa,beta)
{
  z<-runif(n)
  return(alfa-beta*log(-log(z)))
}
#Genera muestra censurada tipo II de la distribución Gumbel
#p=proporción de observaciones con información completa, numero menor que uno y mayor que 0
rgumbel_censura_2=function(n,alfa,beta,p=0.90)
{
  r=as.integer(n*p)
  muestra=rgumbel(n,alfa,beta)
  delta=rep(0,n)
  muestra=sort(muestra)
  delta[1:r]=1
  muestra[(r+1):n]=muestra[r]
  out=list(y=muestra,delta=delta,r=r)
}
#Función de densidad Gumbel
dgumbel=function(x,alfa,beta)
{
  z=(x-alfa)/beta
  1/beta*exp(-z-exp(-z))
}
#Función de distribución Gumbel
pgumbel=function(x,alfa,beta)
{
  z=(x-alfa)/beta
  exp(-exp(-z))
}
#El negativo de la Log-Verosimilitud para datos censurados tipo II de la Distribución Gumbel, Omega=(alfa,beta)
vero_gumbel=function(theta,x,delta)
{
  n=length(x)
  alfa=theta[1]
  beta=theta[2]
  suma=0;
  for(i in 1:n)
```

```

{
suma=suma+delta[i]*log(dgumbel(x[i],alfa,beta))+(1-delta[i])*log(1-pgumbel(x[i],alfa,beta))
}
return(-suma)
}
#Hmnr, es una función general para calcular la entropía muestral##
Hmnr=function(x,m,n,r)
{
suma=0
ximasm=0
ximenosm=0
for(i in 1:r)
{
if((i+m)>r) ximasm=x[r] else ximasm=x[i+m]
if((i-m)<1) ximenosm=x[1] else ximenosm=x[i-m]
suma=suma+log(n/(2*m))*(ximasm-ximenosm)
}
return(suma/n)
}
#xi, esta función general, definida en la ec. 7, de Lim y Park
#Límites de integración
xi=function(x,r,m,i)
{
index=seq(i-m,i+m-1,by=1)
index[which(index<1)]=1
index[which(index>r)]=r
sum(x[index])/(2*m)
}
#Esta función evalúa  $\int f_{nu}(x) \log f_0(x) dx$ , con
#f_0(x) la función de densidad de la Gumbel(alfa,beta)
Integral=function(x,alfa,beta)
{
-x*log(beta)-1/beta*(0.5*x^2-alfa*x)+beta*exp(-(x-alfa)/beta)
}
#Escribir la función para la Gumbel
#Esta es la estadística de prueba (Ec. 4.10). Valores Grandes de I favorecen la hipótesis alterna,
#es decir, los datos no son una m.a. de Gumbel.
I_gumbel=function(x,delta,m)
{
r=sum(delta==1)
n=length(x)
index=order(x)
x=x[index]
delta=delta[index]
H=Hmnr(x,m,n,r)
#Estimadores de máxima verosimilitud
Sy=sd(x)
Eulergamma=0.5772;
betaini=Sy*sqrt(6)/pi
alfaini=mean(x)-Eulergamma*betaini
out_optim=optim(c(alfaini,betaini),vero_gumbel,x=x,delta=delta)
if(out_optim$convergence!=0) stop ("Error: No puedo encontrar emv\n");
theta=out_optim$par
alfa=theta[1]
beta=theta[2]
#j1[1],...,j1[r+1]
jis=rep(NA,r+1)
for(i in 1:(r+1)) jis[i]=xi(x,r,m,i)
#La integral
suma=0
for(i in 1:r)
{
if((i+m)>r) ximasm=x[r] else ximasm=x[i+m]
if((i-m)<1) ximenosm=x[1] else ximenosm=x[i-m]
suma=suma+2*m/n*1/(ximasm-ximenosm)*(Integral(jis[i+1],alfa,beta)-Integral(jis[i],alfa,beta))
}

```

```

}
-H-r/n+pgumbel(jis[r+1],alfa,beta)-suma
}
#DISTRIBUCIÓN EMPÍRICA DEL ESTADÍSTICO DE PRUEBA
Empirica_KL_c2=function(n,m,alfa,beta,p,B=1000)
{
distr_KL=rep(NA,B)
for(i in 1:B)
{
cat("i=",i,"\n")
a=rgumbel_censura_2(n,alfa,beta,p)
distr_KL[i]=l_gumbel(a$a,a$delta,m)
}
return(distr_KL)
}
#Distribución del estadístico de Kullback-Leibler para distintos valores de los parámetros de localidad y escala
d1=Empirica_KL_c2(n=50,m=3,alfa=0,beta=1,p=0.90,B=10000)
d2=Empirica_KL_c2(n=50,m=3,alfa=3,beta=10,p=0.90,B=10000)
d3=Empirica_KL_c2(n=50,m=3,alfa=1,beta=20,p=0.90,B=10000)
d4=Empirica_KL_c2(n=50,m=3,alfa=-5,beta=500,p=0.90,B=10000)
plot(density(d1),main="",xlab="",ylab="Density",lty=1)
lines(density(d2),lty=2)
lines(density(d3),lty=3)
lines(density(d4),lty=4)
legend("topright",c("KL(m,n,r) con Gumbel(0,1)", "KL(m,n,r) con Gumbel(3,10)", "KL(m,n,r) con Gumbel(1,20)", "KL(m,n,r)
con Gumbel(-5,500)"),bty="n",lty=c(1,2,3,4))
#Obtencion de los Cuantiles del min(C,n,r,alfa)
#Seleccionar m para n=(10-200), r=50%-90%
n=90
r=63
p=r/n
B=1000 #Cambiar a 1000
nivel_sig=0.15
cuantiles=rep(NA,15)
for(m in 1:15)
{
cuantiles[m]=quantile(Empirica_KL_c2(n=n,m=m,alfa=0,beta=1,p=p,B=B),1-nivel_sig)
}
m=which.min(cuantiles) #da m
K=min(cuantiles) #
#Tamano
rechazos=rep(0,1000)
for(i in 1:1000)
{
# cat("i=",i,"\n")
a=rgumbel_censura_2(n=n,alfa=200,beta=500,p=p)
if(l_gumbel(a$a,a$delta,m=m)>K) rechazos[i]=1
}
sum(rechazos)/1000
#FUNCIÓN PARA GENERAR NÚMEROS ALEATORIOS DE DISTRIBUCIONES ALTERNATIVAS
rgumbel=function(n,ji=0,theta=1)
{
z=runif(n)
ji-theta*log(-log(z))
}
rgumbelc2=function(n,r,ji=0,theta=1)
{
x=rgumbel(n,ji,theta)
x=sort(x)
delta=rep(0,n)
delta[1:r]=1
return(list(x=x,delta=delta))
}
#FUNCIÓN PARA GENERAR NÚMEROS ALEATORIOS DE LA DISTRIBUCIÓN LOG-NORMAL ESTANDAR, PARA
DATOS CENSURADOS TIPO II

```

```

rlnormc2=function(n,r,meanlog=0,sdlog=1)
{
x=rlnorm(n,meanlog,sdlog)
x=sort(x)
delta=rep(0,n)
delta[1:r]=1
return(list(x=x,delta=delta))
}
ccorrelacion.rlnorm=function(n,r,meanlog,sdlog)
{
out=rlnormc2(n,r,meanlog,sdlog)
mi.corr(out$x,out$delta)
}
#FUNCIÓN PARA GENERAR NÚMEROS ALEATORIOS DE LA DISTRIBUCIÓN GAMMA, PARA DATOS CENSU-
RADOS TIPO II
rgammac2=function(n,r,shape=0,scale=1)
{
x=rgamma(n,shape,scale)
x=sort(x)
delta=rep(0,n)
delta[1:r]=1
return(list(x=x,delta=delta))
}
ccorrelacion.rgamma=function(n,r,shape,scale)
{
out=rgammac2(n,r,shape,scale)
mi.corr(out$x,out$delta)
}
#FUNCIÓN PARA GENERAR NÚMEROS ALEATORIOS DE LA DISTRIBUCIÓN NORMAL, PARA DATOS CENSU-
RADOS TIPO II
rnormc2= function(n,r, mean = 0, sd = 1)
{
x=rnorm(n,mean,sd)
x=sort(x)
delta=rep(0,n)
delta[1:r]=1
return(list(x=x,delta=delta))
}
ccorrelacion.rnorm=function(n,r,mean,sd)
{
out=rnormc2(n,r,mean,sd)
mi.corr(out$x,out$delta)
}
#FUNCIÓN PARA GENERAR NÚMEROS ALEATORIOS DE LA DISTRIBUCIÓN LOGISTICA, PARA DATOS CEN-
SURADOS TIPO II
rlogisc2=function(n,r,location = 0, scale = 1)
{
x=rlogis(n,location,scale)
x=sort(x)
delta=rep(0,n)
delta[1:r]=1
return(list(x=x,delta=delta))
}
ccorrelacion.rlogis=function(n,r,location,scale)
{
out=rlogisc2(n,r,location,scale)
mi.corr(out$x,out$delta)
}
#FUNCIÓN PARA GENERAR NÚMEROS ALEATORIOS DE LA DISTRIBUCIÓN CAUCHY, PARA DATOS CENSU-
RADOS TIPO II
rcauchyc2=function(n,r, location = 0, scale = 1)
{
x=rcauchy(n,location,scale)
x=sort(x)
delta=rep(0,n)

```

```

delta[1:r]=1
return(list(x=x,delta=delta))
}
ccorrelacion.rcauchy=function(n,r,location,scale)
{
out=rcauchyc2(n,r,location,scale)
mi.corr(out$x,out$delta)
}
#FUNCIÓN PARA GENERAR NÚMEROS ALEATORIOS DE LA DISTRIBUCIÓN WEIBULL, PARA DATOS CENSU-
RADOS TIPO II
rweibullc2=function(n,r, shape=0, scale = 1)
{
x=rweibull(n,shape,scale)
x=sort(x)
delta=rep(0,n)
delta[1:r]=1
return(list(x=x,delta=delta))
}
ccorrelacion.rweibull=function(n,r,shape,scale)
{
out=rweibullc2(n,r,shape,scale)
mi.corr(out$x,out$delta)
}
#FUNCIÓN PARA GENERAR NÚMEROS ALEATORIOS DE LA DISTRIBUCIÓN T-STUDENT, PARA DATOS CEN-
SURADOS TIPO II
rstudentc2=function(n,r,params)
{
x=rt(n,params)
x=sort(x)
delta=rep(0,n)
delta[1:r]=1
return(list(x=x,delta=delta))
}
ccorrelacion.rstudent=function(n,r,params)
{
out=rstudentc2(n,r,params)
mi.corr(out$x,out$delta)
}
#FUNCIÓN PARA SELECCIONAR EL VALOR CRÍTICO DE LA PRUEBA DE KULLBACK-LEIBLER (NIVEL DE SIG.
ALFA=0.05)
#n es el tamaño de muestra, y el % de Censura es igual a r/n
#si n o alpha no estan en la tabla regresa NA NA
CriticoKL.05<-function(n,pcensura)
{
#Crea la tabla de valores críticos
Tabla<-matrix(nrow=15,ncol=10)
Tabla[1,]<-c(0.1831165,3,0.1901636,4,0.2139569,5,0.2334548,5,0.2699408,6)
Tabla[2,]<-c(0.1361004,6,0.1560067,7,0.1724471,6,0.1919255,9,0.2224663,6)
Tabla[3,]<-c(0.1237775,6,0.1381355,8,0.1543177,8,0.1708361,9,0.1864447,6)
Tabla[4,]<-c(0.1104802,8,0.1248387,12,0.1385143,6,0.15397,10,0.1603581,7)
Tabla[5,]<-c(0.1015404,10,0.1148255,8,0.1248174,6,0.1301082,6,0.1360578,7)
Tabla[6,]<-c(0.09486525,9,0.1040339,6,0.1116112,9,0.1176781,7,0.1247775,6)
Tabla[7,]<-c(0.08775952,5,0.09742928,10,0.1008187,6,0.1084396,7,0.1167954,9)
Tabla[8,]<-c(0.08268052,8,0.08911468,7,0.0940463,9,0.09908681,7,0.1028219,10)
Tabla[9,]<-c(0.0762286,7,0.08038718,8,0.08603441,9,0.0932973,8,0.09654952,8)
Tabla[10,]<-c(0.07090386,7,0.07700086,9,0.08174014,7,0.08569215,8,0.08986158,9)
Tabla[11,]<-c(0.06441966,8,0.06960382,10,0.07454375,8,0.07713713,9,0.08101058,9)
Tabla[12,]<-c(0.05720314,9,0.06320506,8,0.06602238,8,0.0700754,11,0.07337088,10)
Tabla[13,]<-c(0.05420215,12,0.05773384,10,0.06011011,11,0.06348211,10,0.06745487,11)
Tabla[14,]<-c(0.05037521,9,0.05361361,10,0.05645131,9,0.06001988,12,0.06275206,11)
Tabla[15,]<-c(0.04660686,12,0.05067103,12,0.05312082,12,0.056468,12,0.05949097,14)
i<-c(10,20,30,40,50,60,70,80,90,100,120,140,160,180,200)# n's
j<-c(0.5,NA,0.6,NA,0.7,NA,0.8,NA,0.8,NA,0.9,NA)#censuras, los NA'S son los valores de m en la tabla (pcensura,m)
critico<-Tabla[match(n,i),match(pcensura,j)]
m<-Tabla[match(n,i),match(pcensura,j)+1]

```

```

return(c(critico,m))#Regresa un vector con la constante critica y el valor de m
}
#FUNCIÓN PARA COMPARAR LA CONSTANTE CRÍTICA vs EL ESTADÍSTICO DE PRUEBA
Phix.KL=function(x,delta)###duda en esta funcion
{
n=length(x)
r=sum(delta)
pcensura=r/n
critico_m= CriticoKL.05(n,pcensura)#### Tiene que regresar m y Ccritica
#Una vez que tienes m
m=critico_m[2]
critico=critico_m[1]
valor_ep=l_gumbel(x,delta,m)
if(valor_ep>critico) return(1) else return(0)
}
#FUNCIÓN PARA CALCULAR LA POTENCIA DE LA PRUEBA
Pot.KL=function(n=100,r=80,B=10000,alternativa="cauchy",params=c(0,1))
{
rechazos=rep(NA,B)
for(i in 1:B)
{
muestra=switch(alternativa,weibull=rweibullc2(n,r,params[1],params[2]),gumbel=rgumbelc2(n,r,params[1],params[2]),lognormal=rlnormc2(n,r,p
rechazos[i]=Phix.KL(muestra$x,muestra$delta) ### que muestras se necesitan cargar
}
}
sum(rechazos)/B
}
Pot.KL(n=20,r=16,B=10000,alternativa="cauchy",params=c(0,0.5))#c=20
Pot.KL(n=20,r=10,B=10000,alternativa="cauchy",params=c(0,0.5))#c=50
Pot.KL(n=50,r=40,B=10000,alternativa="cauchy",params=c(0,0.5))#c=20
Pot.KL(n=50,r=25,B=10000,alternativa="cauchy",params=c(0,0.5))#c=50
Pot.KL(n=20,r=16,B=10000,alternativa="cauchy",params=c(-2,1))#c=20
Pot.KL(n=20,r=10,B=10000,alternativa="cauchy",params=c(-2,1))#c=50
Pot.KL(n=50,r=40,B=10000,alternativa="cauchy",params=c(-2,1))#c=20
Pot.KL(n=50,r=25,B=10000,alternativa="cauchy",params=c(-2,1))#c=50
rechazos=rep(NA,1000)
for(i in 1:1000)
{
#muestra=rweibullc2(n=100,r=80,100,200)
#muestra=rlnormc2(n=100,r=80,0,1)
#muestra=rgammac2(n=100,r=80,0,1)
#muestra=rlogisc2(n=100,r=80,0,1)
#muestra=rnormc2(n=100,r=80,0,1)
#muestra=rcauchyc2(n=100,r=80,0,1)
muestra=rgumbelc2(n=50,r=25,0,1)
rechazos[i]=Phix.KL(muestra$x,muestra$delta)
}
}
sum(rechazos)/1000

```


Anexo II-K-L

Tabla 8.1: Tamaño estimado de la prueba de Kullback-Leibler ($KL^*(m, n, r)$), para la distribución *Gumbel* (0, 1), obtenidos mediante simulación Monte Carlo con $B = 10,000$, y muestras de tamaño n .

n	20	50	20	50
α	C = 20 %	C = 20 %	C = 50 %	C = 50 %
0.01	0.017	0.014	0.011	0.013
0.02	0.027	0.025	0.019	0.024
0.05	0.063	0.053	0.058	0.046
0.10	0.121	0.107	0.113	0.108
0.15	0.144	0.145	0.148	0.177

Tabla 8.2: Potencia de la prueba de Kullback-Leibler ($KL^*(m, n, r)$), cuando H_1 es *Weibull* (5, 10).

n	20	50	20	50
α	C = 50 %	C = 50 %	C = 20 %	C = 20 %
0.01	0.097	0.092	0.614	0.842
0.02	0.172	0.156	0.799	0.798
0.05	0.308	0.253	0.946	0.959
0.10	0.294	0.382	0.939	0.959
0.15	0.412	0.445	0.965	0.916

Tabla 8.3: Potencia de la prueba de Kullback-Leibler ($KL^*(m, n, r)$), cuando H_1 es *Weibull* (20, 50).

n	20	50	20	50
α	C = 50 %	C = 50 %	C = 20 %	C = 20 %
0.01	0.227	0.375	0.438	0.590
0.02	0.238	0.479	0.605	0.664
0.05	0.408	0.608	0.858	0.886
0.10	0.440	0.722	0.853	0.906
0.15	0.554	0.777	0.911	0.859

Tabla 8.4: Potencia de la prueba de Kullback-Leibler ($KL^*(m, n, r)$), cuando H_1 es *Normal* (0, 1).

n	20	50	20	50
α	C = 50 %	C = 250 %	C = 20 %	C = 20 %
0.01	0.085	0.050	0.732	0.911
0.02	0.210	0.083	0.878	0.881
0.05	0.325	0.169	0.971	0.981
0.10	0.284	0.264	0.965	0.977
0.15	0.419	0.327	0.979	0.956

Tabla 8.5: Potencia de la prueba de Kullback-Leibler ($KL^*(m, n, r)$), cuando H_1 es *Gamma* (5, 1).

n	20	50	20	50
α	C = 50 %	C = 50 %	C = 20 %	C = 20 %
0.01	0.176	0.036	0.957	1.000
0.02	0.478	0.068	0.989	0.999
0.05	0.568	0.133	0.998	1.000
0.10	0.449	0.230	0.998	1.000
0.15	0.594	0.297	0.999	1.000

Tabla 8.6: Potencia de la prueba de Kullback-Leibler ($KL^*(m, n, r)$), cuando H_1 es *Gamma* (3, 9).

n	20	50	20	50
α	C = 50 %	C = 50 %	C = 20 %	C = 20 %
0.01	0.280	0.086	0.980	1.000
0.02	0.610	0.153	0.997	1.000
0.05	0.698	0.269	0.999	1.000
0.10	0.586	0.422	1.000	1.000
0.15	0.728	0.497	1.000	1.000

Tabla 8.7: Potencia de la prueba de Kullback-Leibler ($KL^*(m, n, r)$), cuando H_1 es *Log - normal* (0, 1).

n	20	50	20	50
α	C = 50 %	C = 50 %	C = 20 %	C = 20 %
0.01	0.922	0.974	1.000	1.000
0.02	0.980	0.988	1.000	1.000
0.05	0.988	0.995	1.000	1.000
0.10	0.981	0.999	1.000	1.000
0.15	0.991	1.000	1.000	1.000

Tabla 8.8: Potencia de la prueba de Kullback-Leibler ($KL^*(m, n, r)$), cuando H_1 es *Logística* (0, 1).

n	20	50	20	50
α	C = 50 %	C = 50 %	C = 20 %	C = 20 %
0.01	0.164	0.151	0.698	0.913
0.02	0.309	0.208	0.847	0.882
0.05	0.450	0.314	0.963	0.982
0.10	0.386	0.427	0.955	0.981
0.15	0.521	0.493	0.977	0.961

Tabla 8.9: Potencia de la prueba de Kullback-Leibler ($KL^*(m, n, r)$), cuando H_1 es *Cauchy* (0, 0.5).

n	20	50	20	50
α	C = 50 %	C = 50 %	C = 20 %	C = 20 %
0.01	0.881	0.886	0.892	0.889
0.02	0.933	0.915	0.898	0.906
0.05	0.959	0.960	0.961	0.961
0.10	0.937	0.958	0.979	0.968
0.15	0.956	0.971	0.985	0.978

Anexo III -K-M

Tabla 8.10: Valores críticos del Coeficiente de Correlación del estadístico de Kaplan-Meier (R_{K-M}), obtenidos mediante simulación Monte Carlo con $B = 10,000$ y muestras de tamaño n , con un nivel de significancia $\alpha = 0.01$.

n	Censura (%)							
	10	20	30	40	50	60	70	80
10	0.8554	0.8506	0.8428	0.8339	0.8231	0.8153	0.8419	1.0000
20	0.9200	0.9158	0.9065	0.8981	0.8766	0.8541	0.8299	0.8052
30	0.9451	0.9424	0.9331	0.9213	0.9085	0.8903	0.8616	0.8371
40	0.9591	0.9576	0.9513	0.9399	0.9264	0.9073	0.8877	0.8536
50	0.9669	0.9657	0.9603	0.9538	0.9391	0.9245	0.9030	0.8712
60	0.9732	0.9715	0.9666	0.9580	0.9504	0.9355	0.9130	0.8860
70	0.9768	0.9756	0.9707	0.9650	0.9564	0.9430	0.9243	0.8985
80	0.9795	0.9780	0.9744	0.9694	0.9610	0.9502	0.9318	0.9045
90	0.9815	0.9805	0.9774	0.9718	0.9652	0.9555	0.9392	0.9105
100	0.9836	0.9821	0.9798	0.9757	0.9683	0.9580	0.9433	0.9156
120	0.9864	0.9851	0.9835	0.9783	0.9734	0.9649	0.9508	0.9285
140	0.9880	0.9874	0.9853	0.9808	0.9774	0.9696	0.9580	0.9355
160	0.9900	0.9892	0.9876	0.9843	0.9798	0.9739	0.9616	0.9416
180	0.9910	0.9899	0.9886	0.9860	0.9819	0.9759	0.9649	0.9488
200	0.9921	0.9912	0.9896	0.9871	0.9832	0.9790	0.9689	0.9488
220	0.9928	0.9920	0.9903	0.9885	0.9846	0.9795	0.9724	0.9545
240	0.9933	0.9926	0.9909	0.9894	0.9861	0.9817	0.9738	0.9584
260	0.9938	0.9933	0.9921	0.9901	0.9873	0.9836	0.9756	0.9606
280	0.9940	0.9936	0.9926	0.9906	0.9877	0.9838	0.9775	0.9635
300	0.9946	0.9941	0.9930	0.9914	0.9885	0.9852	0.9788	0.9657
320	0.9951	0.9943	0.9936	0.9919	0.9898	0.9859	0.9801	0.9692
340	0.9953	0.9948	0.9940	0.9926	0.9902	0.9867	0.9811	0.9695
360	0.9956	0.9949	0.9943	0.9928	0.9906	0.9877	0.9823	0.9707
380	0.9957	0.9954	0.9944	0.9933	0.9915	0.9886	0.9829	0.9718

Tabla 8.11: Valores críticos del Coeficiente de Correlación del estadístico de Kaplan-Meier (R_{K-M}), obtenidos mediante simulación Monte Carlo con $B = 10,000$ y muestras de tamaño n , con un nivel de significancia $\alpha = 0.02$.

n	Censura (%)							
	10	20	30	40	50	60	70	80
10	0.8726	0.8743	0.8637	0.8550	0.8411	0.8328	0.8491	1.0000
20	0.9307	0.9287	0.9202	0.9124	0.8928	0.8763	0.8520	0.8382
30	0.9529	0.9504	0.9442	0.9344	0.9208	0.9062	0.8805	0.8574
40	0.9643	0.9632	0.9584	0.9497	0.9379	0.9229	0.9050	0.8727
50	0.9711	0.9697	0.9655	0.9601	0.9478	0.9360	0.9193	0.8895
60	0.9766	0.9747	0.9710	0.9648	0.9572	0.9449	0.9262	0.9007
70	0.9797	0.9785	0.9742	0.9703	0.9627	0.9520	0.9358	0.9138
80	0.9825	0.9807	0.9780	0.9735	0.9664	0.9564	0.9421	0.9195
90	0.9839	0.9830	0.9801	0.9757	0.9704	0.9628	0.9488	0.9253
100	0.9858	0.9844	0.9825	0.9789	0.9726	0.9642	0.9509	0.9291
120	0.9881	0.9872	0.9854	0.9815	0.9772	0.9694	0.9584	0.9381
140	0.9897	0.9889	0.9873	0.9838	0.9808	0.9737	0.9643	0.9464
160	0.9912	0.9906	0.9890	0.9863	0.9823	0.9775	0.9684	0.9510
180	0.9921	0.9911	0.9902	0.9878	0.9842	0.9798	0.9706	0.9568
200	0.9930	0.9922	0.9910	0.9888	0.9858	0.9817	0.9738	0.9577
220	0.9937	0.9930	0.9916	0.9899	0.9870	0.9826	0.9763	0.9622
240	0.9942	0.9936	0.9922	0.9909	0.9880	0.9842	0.9775	0.9648
260	0.9946	0.9941	0.9931	0.9915	0.9891	0.9856	0.9794	0.9670
280	0.9949	0.9944	0.9935	0.9920	0.9894	0.9861	0.9806	0.9692
300	0.9954	0.9950	0.9939	0.9923	0.9901	0.9873	0.9821	0.9713
320	0.9957	0.9951	0.9944	0.9930	0.9910	0.9882	0.9829	0.9735
340	0.9958	0.9954	0.9947	0.9935	0.9914	0.9886	0.9840	0.9740
360	0.9961	0.9956	0.9950	0.9937	0.9918	0.9893	0.9848	0.9753
380	0.9963	0.9959	0.9951	0.9942	0.9925	0.9900	0.9857	0.9764
400	0.9965	0.9961	0.9954	0.9945	0.9928	0.9905	0.9861	0.9782
420	0.9966	0.9963	0.9957	0.9947	0.9932	0.9909	0.9870	0.9789
440	0.9968	0.9965	0.9959	0.9950	0.9936	0.9912	0.9873	0.9797
460	0.9970	0.9966	0.9961	0.9952	0.9936	0.9915	0.9880	0.9802
480	0.9970	0.9967	0.9961	0.9954	0.9940	0.9918	0.9883	0.9814
500	0.9972	0.9969	0.9963	0.9955	0.9943	0.9923	0.9888	0.9813

Tabla 8.12: Valores críticos del Coeficiente de Correlación del estadístico de Kaplan-Meier (R_{K-M}), obtenidos mediante simulación Monte Carlo con $B = 10,000$ y muestras de tamaño n , con un nivel de significancia $\alpha = 0.10$.

n	Censura (%)							
	10	20	30	40	50	60	70	80
10	0.9217	0.9217	0.9162	0.9093	0.9006	0.8918	0.8960	1.0000
20	0.9572	0.9545	0.9508	0.9438	0.9346	0.9239	0.9098	0.8942
30	0.9700	0.9685	0.9647	0.9588	0.9517	0.9425	0.9269	0.9077
40	0.9770	0.9759	0.9732	0.9684	0.9619	0.9535	0.9415	0.9212
50	0.9815	0.9803	0.9778	0.9746	0.9685	0.9616	0.9503	0.9314
60	0.9847	0.9836	0.9810	0.9778	0.9734	0.9664	0.9563	0.9394
70	0.9867	0.9855	0.9836	0.9807	0.9765	0.9706	0.9615	0.9468
80	0.9885	0.9874	0.9855	0.9830	0.9792	0.9735	0.9655	0.9518
90	0.9896	0.9887	0.9871	0.9843	0.9813	0.9761	0.9682	0.9553
100	0.9907	0.9899	0.9886	0.9862	0.9832	0.9786	0.9703	0.9572
120	0.9921	0.9915	0.9903	0.9882	0.9854	0.9812	0.9749	0.9638
140	0.9932	0.9926	0.9915	0.9897	0.9874	0.9837	0.9782	0.9682
160	0.9941	0.9937	0.9925	0.9911	0.9888	0.9856	0.9806	0.9703
180	0.9947	0.9942	0.9934	0.9920	0.9901	0.9873	0.9824	0.9740
200	0.9953	0.9948	0.9941	0.9927	0.9908	0.9882	0.9836	0.9755
220	0.9957	0.9953	0.9945	0.9934	0.9918	0.9891	0.9851	0.9770
240	0.9960	0.9956	0.9949	0.9939	0.9923	0.9899	0.9861	0.9791
260	0.9963	0.9961	0.9954	0.9943	0.9929	0.9906	0.9871	0.9800
280	0.9966	0.9962	0.9957	0.9947	0.9933	0.9913	0.9880	0.9819
300	0.9969	0.9966	0.9959	0.9951	0.9938	0.9920	0.9887	0.9828
320	0.9971	0.9967	0.9962	0.9953	0.9941	0.9924	0.9893	0.9839
340	0.9972	0.9969	0.9964	0.9957	0.9945	0.9927	0.9898	0.9842
360	0.9974	0.9970	0.9966	0.9958	0.9947	0.9931	0.9904	0.9851
380	0.9975	0.9973	0.9967	0.9961	0.9951	0.9935	0.9909	0.9856
400	0.9977	0.9974	0.9969	0.9963	0.9953	0.9939	0.9912	0.9866
420	0.9977	0.9975	0.9971	0.9964	0.9955	0.9941	0.9918	0.9870
440	0.9978	0.9976	0.9972	0.9966	0.9958	0.9943	0.9919	0.9873
460	0.9979	0.9977	0.9973	0.9968	0.9958	0.9945	0.9924	0.9880
480	0.9980	0.9978	0.9974	0.9969	0.9960	0.9948	0.9925	0.9884
500	0.9981	0.9979	0.9975	0.9970	0.9961	0.9950	0.9928	0.9887

Tabla 8.13: Valores críticos del Coeficiente de Correlación del estadístico de Kaplan-Meier (R_{K-M}), obtenidos mediante simulación Monte Carlo con $B = 10,000$ y muestras de tamaño n , con un nivel de significancia $\alpha = 0.15$.

n	Censura (%)							
	10	20	30	40	50	60	70	80
10	0.9330	0.9333	0.9286	0.9242	0.9155	0.9093	0.9065	1.0000
20	0.9627	0.9607	0.9578	0.9520	0.9443	0.9348	0.9225	0.9085
30	0.9741	0.9729	0.9695	0.9651	0.9590	0.9511	0.9377	0.9206
40	0.9800	0.9791	0.9767	0.9729	0.9675	0.9605	0.9505	0.9336
50	0.9841	0.9830	0.9808	0.9778	0.9729	0.9669	0.9579	0.9419
60	0.9866	0.9857	0.9835	0.9808	0.9771	0.9712	0.9628	0.9493
70	0.9884	0.9875	0.9858	0.9833	0.9798	0.9752	0.9670	0.9550
80	0.9900	0.9891	0.9877	0.9852	0.9820	0.9774	0.9707	0.9587
90	0.9910	0.9902	0.9888	0.9865	0.9840	0.9794	0.9729	0.9618
100	0.9919	0.9912	0.9899	0.9881	0.9855	0.9816	0.9747	0.9638
120	0.9931	0.9925	0.9914	0.9897	0.9875	0.9838	0.9785	0.9694
140	0.9941	0.9936	0.9926	0.9911	0.9890	0.9860	0.9814	0.9730
160	0.9948	0.9944	0.9935	0.9922	0.9903	0.9876	0.9833	0.9750
180	0.9954	0.9950	0.9942	0.9930	0.9913	0.9890	0.9848	0.9777
200	0.9959	0.9955	0.9948	0.9936	0.9921	0.9898	0.9860	0.9793
220	0.9962	0.9959	0.9952	0.9943	0.9928	0.9906	0.9874	0.9804
240	0.9965	0.9962	0.9955	0.9946	0.9933	0.9914	0.9881	0.9824
260	0.9968	0.9965	0.9959	0.9950	0.9938	0.9920	0.9890	0.9833
280	0.9970	0.9967	0.9962	0.9954	0.9942	0.9926	0.9896	0.9845
300	0.9972	0.9970	0.9964	0.9957	0.9946	0.9930	0.9903	0.9853
320	0.9974	0.9971	0.9967	0.9959	0.9949	0.9934	0.9908	0.9863
340	0.9975	0.9973	0.9969	0.9962	0.9952	0.9937	0.9913	0.9866
360	0.9977	0.9974	0.9970	0.9963	0.9954	0.9940	0.9917	0.9872
380	0.9978	0.9976	0.9971	0.9966	0.9957	0.9944	0.9922	0.9879
400	0.9980	0.9977	0.9973	0.9967	0.9959	0.9947	0.9925	0.9885
420	0.9980	0.9978	0.9974	0.9968	0.9961	0.9948	0.9929	0.9889
440	0.9981	0.9979	0.9976	0.9970	0.9963	0.9950	0.9930	0.9892
460	0.9982	0.9980	0.9977	0.9972	0.9963	0.9952	0.9933	0.9897
480	0.9982	0.9981	0.9977	0.9972	0.9965	0.9954	0.9936	0.9902
500	0.9983	0.9981	0.9978	0.9974	0.9967	0.9956	0.9939	0.9900

Tabla 8.14: Tamaño estimado de la prueba del estadístico de Kaplan-Meier (R_{K-M}), para la distribución *Gumbel* (0, 1), obtenidos mediante simulación Monte Carlo con $B = 10,000$, y muestras de tamaño n .

n	20	50	20	50
α	C= 20 %	C= 20 %	C= 50 %	C= 50 %
0.01	0.005	0.017	0.011	0.009
0.02	0.024	0.018	0.019	0.021
0.05	0.042	0.047	0.055	0.045
0.10	0.085	0.097	0.099	0.100
0.15	0.173	0.150	0.150	0.132

Tabla 8.15: Potencia de la prueba del estadístico de Kaplan-Meier (R_{K-M}), cuando H_1 es *Weibull* (5, 10).

n	20	50	80	100	500
α	C = 50 %	C = 20 %	C = 10 %	C = 20 %	C = 30 %
0.01	0.030	0.313	0.712	0.600	0.994
0.02	0.046	0.378	0.807	0.686	0.997
0.05	0.096	0.510	0.883	0.811	0.999
0.10	0.166	0.632	0.938	0.871	1.000
0.15	0.225	0.706	0.959	0.915	1.000

Tabla 8.16: Potencia de la prueba del estadístico de Kaplan-Meier (R_{K-M}), cuando H_1 es *Weibull* (20, 50).

n	20	50	80	100	500
α	C = 50 %	C = 20 %	C = 10 %	C = 20 %	C = 30 %
0.01	0.071	0.695	0.976	0.955	1.000
0.02	0.110	0.754	0.985	0.969	1.000
0.05	0.182	0.832	0.993	0.993	1.000
0.10	0.268	0.894	0.997	0.992	1.000
0.15	0.330	0.919	0.999	0.996	1.000

Tabla 8.17: Potencia de la prueba del estadístico de Kaplan-Meier (R_{K-M}), cuando H_1 es *Normal* (0, 1).

n	20	50	80	100	500
α	C = 50 %	C = 20 %	C = 10 %	C = 20 %	C = 30 %
0.01	0.023	0.420	0.840	0.844	1.000
0.02	0.041	0.502	0.902	0.899	1.000
0.05	0.096	0.627	0.950	0.957	1.000
0.10	0.176	0.744	0.973	0.977	1.000
0.15	0.232	0.809	0.984	0.986	1.000

Tabla 8.18: Potencia de la prueba del estadístico de Kaplan-Meier (R_{K-M}), cuando H_1 es *Gamma* (5, 1).

n	20	50	80	100	500
α	C = 50 %	C = 20 %	C = 10 %	C = 20 %	C = 30 %
0.01	0.008	0.007	0.006	0.005	0.010
0.02	0.017	0.012	0.016	0.011	0.018
0.05	0.042	0.036	0.041	0.033	0.047
0.10	0.093	0.074	0.086	0.071	0.093
0.15	0.136	0.123	0.137	0.108	0.139

Tabla 8.19: Potencia de la prueba del estadístico de Kaplan-Meier (R_{K-M}), cuando H_1 es *Gamma* (3, 9).

n	20	50	80	100	500
α	C = 50 %	C = 20 %	C = 10 %	C = 20 %	C = 30 %
0.01	0.006	0.008	0.010	0.013	0.281
0.02	0.014	0.013	0.021	0.025	0.356
0.05	0.041	0.039	0.047	0.073	0.521
0.10	0.087	0.084	0.104	0.132	0.666
0.15	0.129	0.132	0.157	0.183	0.738

Tabla 8.20: Potencia de la prueba del estadístico de Kaplan-Meier (R_{K-M}), cuando H_1 es *Log - normal* (0, 1).

n	20	50	80	100	500
α	C = 50 %	C = 20 %	C = 10 %	C = 20 %	C = 30 %
0.01	0.024	0.415	0.841	0.845	1.000
0.02	0.045	0.489	0.894	0.901	1.000
0.05	0.095	0.629	0.944	0.957	1.000
0.10	0.174	0.755	0.975	0.980	1.000
0.15	0.236	0.815	0.985	0.989	1.000

Tabla 8.21: Potencia de la prueba del estadístico de Kaplan-Meier (R_{K-M}), cuando H_1 es *Logística* (0, 1).

n	20	50	80	100	500
α	C = 50 %	C = 20 %	C = 10 %	C = 20 %	C = 30 %
0.01	0.069	0.475	0.755	0.797	1.000
0.02	0.100	0.538	0.819	0.841	1.000
0.05	0.168	0.650	0.876	0.901	1.000
0.10	0.247	0.735	0.924	0.934	1.000
0.15	0.311	0.786	0.946	0.950	1.000

Tabla 8.22: Potencia de la prueba del estadístico de Kaplan-Meier (R_{K-M}), cuando H_1 es *Cauchy* (0, 0.5).

n	20	50	80	100	500
α	C = 50 %	C = 20 %	C = 10 %	C = 20 %	C = 30 %
0.01	0.546	0.987	0.999	1.000	1.000
0.02	0.598	0.990	0.999	1.000	1.000
0.05	0.681	0.994	1.000	1.000	1.000
0.10	0.748	0.997	1.000	1.000	1.000
0.15	0.789	0.998	1.000	1.000	1.000

Anexo IV-N-A

Tabla 8.23: Valores críticos del Coeficiente de Correlación del estadístico de Nelson-Aalen (R_{N-A}), obtenidos mediante simulación Monte Carlo con $B = 10,000$ y muestras de tamaño n , con un nivel de significancia $\alpha = 0.01$.

n	Censura (%)							
	10	20	30	40	50	60	70	80
10	0.8658	0.8524	0.8517	0.8266	0.8187	0.8221	0.8397	1.0000
20	0.9261	0.9162	0.9108	0.8930	0.8762	0.8547	0.8306	0.7984
30	0.9486	0.9444	0.9363	0.9261	0.9096	0.8884	0.8640	0.8303
40	0.9597	0.9548	0.9506	0.9416	0.9304	0.9111	0.8855	0.8500
50	0.9677	0.9658	0.9588	0.9526	0.9424	0.9241	0.9012	0.8713
60	0.9733	0.9711	0.9662	0.9598	0.9506	0.9336	0.9145	0.8826
70	0.9770	0.9752	0.9715	0.9653	0.9573	0.9425	0.9251	0.8942
80	0.9803	0.9787	0.9757	0.9685	0.9608	0.9510	0.9301	0.9045
90	0.9824	0.9800	0.9775	0.9723	0.9648	0.9555	0.9401	0.9094
100	0.9834	0.9824	0.9793	0.9751	0.9678	0.9591	0.9438	0.9171

Tabla 8.24: Valores críticos del Coeficiente de Correlación del estadístico de Nelson-Aalen (R_{N-A}), obtenidos mediante simulación Monte Carlo con $B = 10,000$ y muestras de tamaño n , con un nivel de significancia $\alpha = 0.02$.

n	Censura (%)							
	10	20	30	40	50	60	70	80
10	0.8836	0.8734	0.8713	0.8500	0.8443	0.8376	0.8464	1.0000
20	0.9355	0.9283	0.9218	0.9097	0.8935	0.8757	0.8522	0.8311
30	0.9547	0.9531	0.9446	0.9362	0.9237	0.9050	0.8837	0.8524
40	0.9654	0.9617	0.9573	0.9512	0.9405	0.9258	0.9035	0.8742
50	0.9727	0.9701	0.9653	0.9583	0.9500	0.9367	0.9158	0.8915
60	0.9774	0.9747	0.9706	0.9653	0.9577	0.9440	0.9278	0.9013
70	0.9799	0.9785	0.9750	0.9700	0.9633	0.9527	0.9368	0.9102
80	0.9832	0.9815	0.9787	0.9731	0.9662	0.9576	0.9421	0.9183
90	0.9845	0.9831	0.9805	0.9757	0.9700	0.9619	0.9488	0.9253
100	0.9859	0.9847	0.9821	0.9788	0.9732	0.9654	0.9522	0.9316

Tabla 8.25: Valores críticos del Coeficiente de Correlación del estadístico de Nelson-Aalen (R_{N-A}), obtenidos mediante simulación Monte Carlo con $B = 10,000$ y muestras de tamaño n , con un nivel de significancia $\alpha = 0.10$.

n	Censura (%)							
	10	20	30	40	50	60	70	80
10	0.9262	0.9235	0.9170	0.9090	0.9005	0.8922	0.8966	1.0000
20	0.9581	0.9550	0.9502	0.9436	0.9349	0.9229	0.9087	0.8935
30	0.9708	0.9693	0.9650	0.9590	0.9524	0.9418	0.9280	0.9081
40	0.9775	0.9757	0.9729	0.9691	0.9627	0.9538	0.9411	0.9206
50	0.9820	0.9806	0.9779	0.9743	0.9684	0.9616	0.9495	0.9323
60	0.9849	0.9834	0.9814	0.9776	0.9731	0.9663	0.9560	0.9401
70	0.9868	0.9860	0.9840	0.9809	0.9769	0.9704	0.9616	0.9459
80	0.9886	0.9878	0.9857	0.9828	0.9792	0.9740	0.9648	0.9504
90	0.9897	0.9888	0.9872	0.9846	0.9811	0.9762	0.9691	0.9541
100	0.9906	0.9899	0.9883	0.9864	0.9828	0.9782	0.9709	0.9581

Tabla 8.26: Valores críticos del Coeficiente de Correlación del estadístico de Nelson-Aalen (R_{N-A}), obtenidos mediante simulación Monte Carlo con $B = 10,000$ y muestras de tamaño n , con un nivel de significancia $\alpha = 0.15$.

n	Censura (%)							
	10	20	30	40	50	60	70	80
10	0.9367	0.9351	0.9294	0.9237	0.9147	0.9086	0.9064	1.0000
20	0.9639	0.9619	0.9575	0.9520	0.9445	0.9341	0.9220	0.9083
30	0.9747	0.9735	0.9698	0.9648	0.9594	0.9505	0.9385	0.9223
40	0.9804	0.9788	0.9767	0.9732	0.9680	0.9605	0.9498	0.9337
50	0.9843	0.9831	0.9808	0.9777	0.9731	0.9668	0.9573	0.9429
60	0.9868	0.9856	0.9840	0.9806	0.9769	0.9712	0.9634	0.9492
70	0.9885	0.9877	0.9861	0.9835	0.9799	0.9748	0.9667	0.9543
80	0.9899	0.9892	0.9876	0.9852	0.9821	0.9778	0.9702	0.9577
90	0.9911	0.9902	0.9888	0.9867	0.9836	0.9795	0.9733	0.9609
100	0.9918	0.9912	0.9899	0.9881	0.9852	0.9812	0.9753	0.9643

Tabla 8.27: Tamaño estimado de la prueba del estadístico de Nelson-Aalen (R_{N-A}), para la distribución *Gumbel* (0,1), obtenidos mediante simulación Monte Carlo con $B = 10,000$ y muestras de tamaño n .

n	20	50	20	50
α	C = 20 %	C = 20 %	C = 50 %	C = 50 %
0.01	0.007	0.012	0.007	0.012
0.02	0.023	0.019	0.016	0.023
0.05	0.054	0.050	0.050	0.049
0.10	0.069	0.101	0.103	0.092
0.15	0.152	0.146	0.146	0.143

Tabla 8.28: Potencia de la prueba del estadístico de Nelson-Aalen (R_{N-A}), cuando H_1 es *Normal* (0,1).

n	20	50	80	100
α	C = 50 %	C = 20 %	C = 10 %	C = 30 %
0.01	0.030	0.210	0.504	0.309
0.02	0.052	0.270	0.606	0.392
0.05	0.095	0.399	0.725	0.518
0.10	0.166	0.504	0.803	0.625
0.15	0.228	0.578	0.849	0.697

Tabla 8.29: Potencia de la prueba del estadístico de Nelson-Aalen (R_{N-A}), cuando H_1 es *Cauchy* (0, 0.5).

n	20	50	80	100
α	C = 50 %	C = 20 %	C = 10 %	C = 30 %
0.01	0.541	0.986	0.999	1.000
0.02	0.596	0.989	1.000	1.000
0.05	0.684	0.993	1.000	1.000
0.10	0.739	0.995	1.000	1.000
0.15	0.778	0.996	1.000	1.000

Tabla 8.30: Potencia de la prueba del estadístico de Nelson-Aalen (R_{N-A}), cuando H_1 es *Logística* (0, 1).

n	20	50	80	100
α	C = 50 %	C = 20 %	C = 10 %	C = 30 %
0.01	0.067	0.461	0.752	0.700
0.02	0.098	0.531	0.814	0.760
0.05	0.166	0.629	0.882	0.832
0.10	0.244	0.727	0.919	0.888
0.15	0.308	0.773	0.937	0.918

Tabla 8.31: Potencia de la prueba del estadístico de Nelson-Aalen (R_{N-A}), cuando H_1 es *Weibull* (5, 10).

n	20	50	80	100
α	C = 50 %	C = 20 %	C = 10 %	C = 30 %
0.01	0.027	0.296	0.715	0.388
0.02	0.046	0.372	0.799	0.482
0.05	0.100	0.506	0.876	0.614
0.10	0.164	0.617	0.928	0.727
0.15	0.230	0.685	0.947	0.788

Tabla 8.32: Potencia de la prueba del estadístico de Nelson-Aalen (R_{N-A}), cuando H_1 es *Weibull* (20, 50).

n	20	50	80	100
α	C = 50 %	C = 20 %	C = 10 %	C = 30 %
0.01	0.071	0.682	0.975	0.842
0.02	0.105	0.744	0.988	0.884
0.05	0.182	0.829	0.993	0.932
0.10	0.269	0.885	0.996	0.961
0.15	0.332	0.919	0.997	0.972

Tabla 8.33: Potencia de la prueba del estadístico de Nelson-Aalen (R_{N-A}), cuando H_1 es *Gamma* (5, 1).

n	20	50	80	100
α	C = 50 %	C = 20 %	C = 10 %	C = 30 %
0.01	0.007	0.006	0.007	0.004
0.02	0.015	0.014	0.016	0.011
0.05	0.046	0.039	0.043	0.032
0.10	0.093	0.078	0.082	0.068
0.15	0.138	0.122	0.118	0.114

Tabla 8.34: Potencia de la prueba del estadístico de Nelson-Aalen (R_{N-A}), cuando H_1 es *Gamma* (3, 9).

n	20	50	80	100
α	C = 50 %	C = 20 %	C = 10 %	C = 30 %
0.01	0.007	0.010	0.014	0.016
0.02	0.015	0.019	0.027	0.028
0.05	0.046	0.047	0.065	0.065
0.10	0.083	0.093	0.113	0.123
0.15	0.132	0.139	0.161	0.191

Tabla 8.35: Potencia de la prueba del estadístico de Nelson-Aalen (R_{N-A}), cuando H_1 es *Log-normal* (0, 1).

n	20	50	80	100
α	C = 50%	C = 20%	C = 10%	C = 30%
0.01	0.030	0.437	0.875	0.698
0.02	0.050	0.520	0.920	0.783
0.05	0.106	0.669	0.961	0.882
0.10	0.181	0.773	0.978	0.938
0.15	0.249	0.829	0.987	0.961