



COLEGIO DE POSTGRADUADOS

INSTITUCIÓN DE ENSEÑANZA E INVESTIGACIÓN EN CIENCIAS AGRÍCOLAS

CAMPUS MONTECILLO

POSTGRADO DE SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA
ESTADÍSTICA

UNA PRUEBA DE BONDAD DE AJUSTE PARA LA DISTRIBUCIÓN PARETO EN PRESENCIA DE CENSURA TIPO II

DAYNA PRISCILA SALDAÑA ZEPEDA

T E S I S

PRESENTADA COMO REQUISITO PARCIAL PARA
OBTENER EL GRADO DE:

MAESTRA EN CIENCIAS

MONTECILLO, TEXCOCO, EDO. DE MÉXICO
2008

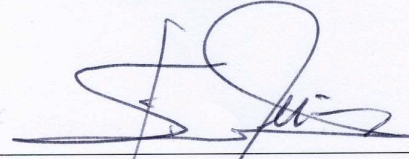
La presente tesis titulada: **UNA PRUEBA DE BONDAD DE AJUSTE PARA LA DISTRIBUCIÓN PARETO EN PRESENCIA DE CENSURA TIPO II**, realizada por la alumna: **DAYNA PRISCILA SALDAÑA ZEPEDA**, bajo la dirección del Consejo Particular indicado ha sido aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

MAESTRA EN CIENCIAS

**SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA
ESTADÍSTICA**

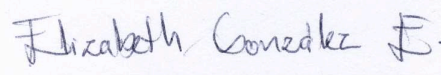
CONSEJO PARTICULAR

CONSEJERO



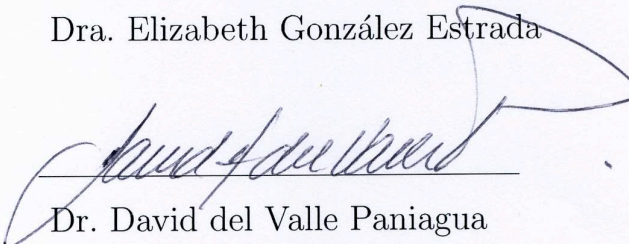
Dr. Gilberto Rendón Sánchez

ASESOR



Dra. Elizabeth González Estrada

ASESOR



Dr. David del Valle Paniagua

Montecillo, Texcoco, México, 29 de agosto de 2008

UNA PRUEBA DE BONDAD DE AJUSTE PARA LA DISTRIBUCIÓN PARETO EN PRESENCIA DE CENSURA TIPO II

Dayna Priscila Saldaña Zepeda, M.C.

Colegio de Postgraduados, 2008

Hay una escasez en la literatura de propuestas de bondad de ajuste para la distribución Pareto, menos aún con información incompleta (censurada). En este trabajo de investigación, una prueba para verificar la hipótesis H_0 : la distribución de los datos es Pareto, cuando las observaciones son sujetas a censura por la derecha Tipo II, es propuesta. La estadística de prueba involucra transformaciones de los datos originales y se basa en el estimador no paramétrico Nelson-Aalen de la Función de Riesgo Acumulada. Vía simulación Monte Carlo se obtuvo la distribución empírica y se investigó la potencia y el tamaño de la prueba. La potencia fue comparada contra adaptaciones para datos censurados Tipo II de las estadísticas Crámer-von Mises, Anderson-Darling y una prueba basada en la información de Kullback-Leibler. Para distribuciones alternativas con función de riesgo monótona decreciente la prueba propuesta tiene mayor potencia. La metodología es ilustrada con tres ejemplos de aplicación.

Palabras clave: Pruebas de bondad de ajuste, distribución Pareto, censura Tipo II, función de riesgo acumulada, estimador Nelson-Aalen.

A GOODNESS OF FIT TEST FOR THE PARETO DISTRIBUTION IN PRESENCE OF TYPE II CENSORING

Dayna Priscila Saldaña Zepeda, M.C.

Colegio de Postgraduados, 2008

There is a lack of literature regarding proposals of goodness-of-fit for the Pareto distribution, less even with incomplete (censored) information. In this research work, a test to check the hypothesis H_0 : the distribution of data is Pareto, when the observations are subjected to Type II right censoring, is proposed. The test statistic involves transformations of the original data and is based on the nonparametric Nelson-Aalen estimator of the cumulative hazard function. By Monte Carlo simulation, the empirical distribution was obtained and was investigated the power and the size of the test. The power is compared against adaptations for Type II censored data of the Cramer-von Mises, Anderson-Darling statistics, and a test based on the Kullback-Leibler information. For alternative distributions with monotone decreasing hazard function, the proposed test has higher power. The methodology is illustrated with three application examples.

Key words: Goodness of fit tests, Pareto distribution, Type II censoring, cumulative hazard function, Nelson-Aalen estimator.

*A mis padres, Catalina y Pedro
que me lo han dado todo...*

*A mis hermanos, Ely y Aarón
a quienes les doy todo...*

*A mi esposo, Luis
con quien comparto todo...*

AGRADECIMIENTOS

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo económico brindado durante la realización de mis estudios de maestría.

Al Colegio de Postgraduados, por la oportunidad de continuar mi formación académica en uno de sus postgrados.

Al Dr. Humberto Vaquera H., por su apoyo, dirección e infinita paciencia para conmigo durante toda mi estancia en el Colpos, y sobretodo durante la culminación del presente trabajo.

Al Dr. Barry C. Arnold por sus oportunos comentarios, sus observaciones y su tiempo dedicado a la revisión del trabajo.

Al Dr. Gilberto Rendón S., por su tiempo y dedicación brindados para continuar mi proceso de aprendizaje.

A la Dra. Elizabeth González E., por sus observaciones y tiempo dedicados de manera desinteresada a la revisión del presente trabajo.

Al Dr. David del Valle P., por su apoyo brindado para la culminación de mis estudios.

Al Dr. Paulio Pérez, R., por su gran ayuda, su paciencia y su valioso tiempo brindados durante toda mi estancia.

A todos mis compañeros, en especial a Blanca, Ángela, Luis y Alex por compartir conmigo dudas e inquietudes de clase, y sobretodo por su apoyo incondicional.

A todos mis profesores, por su guía y consejos brindados durante mi proceso de formación.

A Emma, Olga, Isabel, Carmen y todo el personal administrativo de Estadística por el gran apoyo que me brindaron.

A toda mi familia, por su comprensión y su infinito apoyo.

Índice

1. Introducción	1
2. Objetivos	10
3. Revisión de Literatura	11
3.1. Pruebas de hipótesis	11
3.1.1. Algunas pruebas de bondad de ajuste comunes	14
3.1.2. Coeficiente de correlación	16
3.2. La Distribución Pareto	18
3.2.1. Características	18
3.2.2. La distribución Pareto y su relación con otras distribuciones	20
3.2.3. Algunas aplicaciones del modelo Pareto	21
3.3. Algunos conceptos relacionados con el análisis de supervivencia	22
3.3.1. Censura por la derecha	23
3.3.2. Censura aleatoria	23
3.3.3. Censura por la izquierda	24

Índice

3.3.4. Censura por intervalo	24
3.3.5. Funciones básicas en el análisis de supervivencia	25
3.3.6. Estimación no paramétrica de la función de supervivencia y la función de riesgo acumulado	29
3.4. Otras distribuciones importantes en el análisis de supervivencia	32
3.4.1. Exponencial	32
3.4.2. Weibull	34
3.4.3. Log-normal	35
3.4.4. Gamma	35
4. Prueba de Bondad de Ajuste Propuesta	37
4.1. Derivación de la Prueba	38
4.2. Prueba de Bondad de Ajuste	39
4.3. Distribución del estadístico de prueba	40
4.4. Implementación de la prueba	42
4.5. Tamaño de la prueba	43
4.6. Potencia de la prueba ante algunas alternativas	44
5. Estudio Comparativo de Potencia	49
5.1. Cramér-von Mises y Anderson-Darling	49
5.2. Información de Kullback-Leibler	51
6. Ejemplos de aplicación	58

Índice

6.1. Datos de tiempos de falla	58
6.2. Datos de supervivencia de firmas	58
6.3. Datos de salarios anuales	59
7. Conclusiones	61
Referencias	61
Apéndice	65

Índice de tablas

5.1.	Valores del tamaño de ventana m	53
5.2.	Potencia para la prueba basada en la estadística $T(m, n, r)$	55
5.3.	Cuantiles de la estadística $T(m, n, r)$ para exponencialidad con Censura Tipo II	57
6.1.	Valores críticos del coeficiente de correlación R_{NA} para $n = 15, 30$ y Censura = 33.33%	59
1.	Valores críticos del coeficiente de correlación R_{NA}	65
2.	Tamaño de la prueba basada en R_{NA} con distribución $Par(3, 0.5)$	67
3.	Tamaño de la prueba basada en R_{NA} con distribución $Par(5, 1)$	69
4.	Tamaño de la prueba basada en R_{NA} con distribución $Par(0.5, 100)$	71
5.	Potencia de la prueba para alternativas $Weib(0.5, 1)$ y $Weib(3, 1)$	73
6.	Potencia de la prueba para alternativas $LNorm(5, 3)$ y $LNorm(1, 1)$	75
7.	Potencia de la prueba para alternativas $Ga(3, 1)$ y $Ga(0.8, 1)$	77
8.	Potencia para las pruebas basadas en las estadísticas $W_{r,n}^2$ y $A_{r,n}^2$	79
9.	Cuantiles de las estadísticas $W_{r,n}^2$ y $A_{r,n}^2$ para probar exponencialidad	82

Índice de figuras

3.1. Dos densidades Pareto con distinto parámetro de forma	19
4.1. Distribución de R_{N-A} con distintos valores de los parámetros y tres niveles de censura Tipo II	42
4.2. Tamaño de prueba para distribuciones Pareto	45
4.3. Potencia de la prueba basada en R_{N-A} para algunas alternativas y distintos tamaños de muestra	47
4.4. Potencia de la prueba basada en R_{N-A} para algunas alternativas y distintos niveles de censura	48
5.1. Potencia de las pruebas $W_{r,n}^2$ y $A_{r,n}^2$ para algunas alternativas y censura = 40 %	52
5.2. Potencia de la prueba basada en $T(m, n, r)$ para algunas alternativas y censura = 40 %	56

Capítulo 1

Introducción

En la modelación estadística, un problema importante es verificar la concordancia de los valores observados con el tipo de distribución sobre la cual se basará la inferencia. En este sentido, el investigador formula una aseveración o conjetura sobre el proceso que genera las observaciones, es decir, sobre la forma de la distribución, y la establece como una hipótesis estadística, conocida como la hipótesis de trabajo o hipótesis nula. El objetivo entonces será probar si dicha hipótesis es verdad o no.

Un tipo de prueba de hipótesis, llamada *prueba de bondad de ajuste*, es un procedimiento que nos permite decidir si el conjunto de datos observados es consistente con la forma de la distribución. Dicho de otro modo, es un procedimiento para decidir si rechazar o no la hipótesis nula. Esta decisión de rechazar o aceptar la hipótesis está basada en la información contenida en una muestra tomada de la población de interés, es decir, en los valores observados. Estos valores son usados para evaluar la estadística de prueba, que es función únicamente de las observaciones muestrales, y obtener un solo valor.

El conjunto de valores que la estadística de prueba puede asumir es particionado en dos subconjuntos o regiones, uno correspondiente a la región de rechazo y el otro a la región de aceptación. Si el valor de la estadística de prueba que ha sido obtenido con los datos observados cae en la región de rechazo, la hipótesis nula es rechazada. En caso contrario, si el valor cae en la región de aceptación, la hipótesis nula no se rechaza.

La importancia de las pruebas de bondad de ajuste radica en que el conocimiento de la distribución de los datos observados nos permite usar los métodos estadísticos apropiados

1. Introducción

para hacer inferencia sobre la población de la cual fue tomada la muestra.

Los tipos de pruebas de bondad de ajuste incluyen procedimientos formales y métodos menos formales. Entre los primeros se encuentran las pruebas basadas en la función de distribución empírica (FDE), como la prueba de Kolmogorov-Smirnov, la Cramér-von Mises y la Anderson-Darling; las pruebas basadas en regresión y correlación, entre otras. Los procedimientos informales son referidos a técnicas gráficas, tales como gráficas de probabilidad, gráficas cuantil-cuantil (gráficas Q-Q), y gráficas proporción-proporción (gráficas P-P). Un compendio más completo de las pruebas de bondad de ajuste se encuentra en [Stephens \(1986a\)](#):

Las técnicas gráficas son herramientas simples que pueden ser implementadas fácilmente, y pueden ser usadas para problemas de bondad de ajuste en al menos dos vías; [D'Agostino \(1986\)](#):

1. Como una técnica exploratoria. El objetivo es descubrir características de los datos que sugieran propiedades del fenómeno del cual provienen.
2. Conjuntamente con técnicas numéricas formales. El objetivo es probar formalmente una hipótesis preestablecida o una sugerida por las gráficas. Las gráficas pueden ayudar a revelar desviaciones del modelo asumido. Con frecuencia, descubren características de los datos que no eran anticipadas antes del análisis. Las técnicas numéricas cuantifican la información y evidencia en los datos o gráficas, y actúan como una verificación de inferencias sugeridas de estos. El uso de técnicas gráficas solas puede llevar a conclusiones espurias, por lo que acompañarlas de técnicas numéricas es con frecuencia esencial para evitar esto.

Una técnica gráfica comúnmente usada es la de gráficas de probabilidad. La idea básica de esta técnica es que un conjunto dado de observaciones ordenadas es graficado en contra de una función de los mismos datos, tal que la gráfica resultante debería ser una línea recta. Entonces, se dibuja una línea recta *a ojo* sobre la gráfica y se evalúa, subjetivamente, la linealidad del conjunto de datos. Sin embargo, distintos arreglos (posiciones) de la gráfica pueden llevar a distintas líneas subjetivas cuando estas son dibujadas independientemente. Asimismo, personas diferentes pueden también dibujar diferentes líneas subjetivas cuando se analiza de manera independiente la misma gráfica de probabilidad usando un arreglo dado en la gráfica.

1. Introducción

Kimball (1960) revisó el problema de elegir una posición gráfica para las observaciones ordenadas en una serie de n variables aleatorias independientes e idénticamente distribuidas. Kimball comparó algunas convenciones de la posición y sugirió que la elección de una fórmula a graficar depende en parte del propósito del análisis. En general, las fórmulas a graficar han sido descritas por Blom (1958) como:

$$P_i = (i - c)/(n + 1 - 2c) \quad (1.1)$$

donde P_i es la posición gráfica y c es una función de n y de la distribución muestreada, tal que $0 \leq c < 1$.

En la práctica, muchos autores simplifican el uso de (1.1) asumiendo que c es una constante. Una revisión de distintas elecciones de c que son encontradas en la literatura científica fue hecha por Mage (1982).

En general, en problemas de bondad de ajuste, es útil para las pruebas numéricas (o formales) ser precedidas y complementadas por análisis gráfico. En este sentido, muchos autores han especificado relaciones entre algunos procedimientos gráficos y pruebas numéricas formales que cuantifican la información revelada por las gráficas.

Beirlant *et al.* (2006) formularon dos modificaciones del estadístico de Jackson, originalmente propuesto como estadístico de bondad de ajuste para probar exponencialidad, de tal manera que miden la linealidad de las k observaciones mayores en una gráfica Q-Q Pareto. Para ambas estadísticas, derivaron la distribución límite. Gan *et al.* (1991) propuso el cuadrado del coeficiente de correlación como una medida cuantitativa para la evaluación de una gráfica P-P. Mediante simulación Monte Carlo, obtuvo los cuantiles empíricos de la distribución del estadístico para probar el ajuste de las distribuciones normal, exponencial o Gumbel con parámetros de localización y escala desconocidos. Por otra parte, Mage (1982) presentó un método basado en la minimización del estadístico de Kolmogorov-Smirnov que proporciona una base objetiva para rechazar una línea recta en una gráfica de probabilidad.

No obstante que las gráficas son, por su simplicidad, una técnica común usada para proporcionar una evaluación cualitativa de la bondad de ajuste de una distribución a un conjunto de datos, existen pruebas formales que proporcionan una medida objetiva para este problema, y por tanto, son en muchos casos preferidas. Entre estas se encuentran,

1. Introducción

como se ha mencionado, las estadísticas FDE.

La FDE es una función escalonada, calculada de la muestra, la cual estima a la función de distribución poblacional. Las estadísticas FDE son medidas de la discrepancia entre la FDE y una función de distribución hipotética o dada, y son usadas para probar el ajuste de la muestra a la distribución. Estas estadísticas están basadas en la diferencia vertical entre la FDE y la distribución hipotética, y son convenientemente divididas en dos clases, la clase suprema y la clase cuadrática.

La estadística FDE de la clase suprema más conocida es la estadística D de Kolmogorov. En la clase cuadrática se encuentran las estadísticas Anderson-Darling, A^2 y Cramér-von Mises, W^2 . Para una revisión más detallada de estas y otras estadísticas FDE, ver [Stephens \(1986b\)](#).

Si se desea hacer uso de estas estadísticas para probar el ajuste de un conjunto de valores observados a una distribución dada, cuando esta distribución contiene parámetros desconocidos, entonces se hace necesaria la estimación de dichos parámetros.

Cuando los parámetros desconocidos son de localización o escala, y si estos son estimados por métodos apropiados, las distribuciones de las estadísticas FDE no dependerán de los verdaderos valores de los parámetros desconocidos. Así, los puntos porcentuales de las distribuciones de las estadísticas FDE dependen solo de la familia probada y del tamaño de muestra n . No obstante, las distribuciones exactas de las estadísticas FDE son muy difíciles de encontrar. Sin embargo, para las estadísticas W^2 y A^2 existe teoría asintótica; además, los puntos porcentuales de esas estadísticas para finito n convergen rápidamente a los puntos asintóticos. Para la estadística D no hay una teoría asintótica general (excepto cuando los parámetros son todos conocidos) e incluso los puntos asintóticos deberían ser estimados. Por otra parte, cuando los parámetros desconocidos no son de localización o escala, por ejemplo cuando el parámetro desconocido es un parámetro de forma, la distribución, incluso asintótica, dependerá del verdadero valor del parámetro; [Stephens \(1986b\)](#).

Diversos autores han usado estas estadísticas, o formulado modificaciones de ellas, para probar el ajuste de varias distribuciones particulares. [Margolin y Maurer \(1976\)](#) derivaron analíticamente la distribución de la estadística D para el caso exponencial, y obtuvieron aproximaciones computacionales eficientes para la función de distribución. Para probar el ajuste de la distribución de valor extremo con parámetros desconocidos, [Stephens](#)

1. Introducción

(1977) dio los puntos porcentuales asintóticos de las estadísticas A^2 y W^2 . Chandra *et al.* (1981) obtuvieron los puntos porcentuales de la estadística D para la distribución Weibull y para la de valores extremos. Más recientemente, Porter *et al.* (1992) plantearon las estadísticas D , A^2 y W^2 como pruebas de bondad de ajuste de la distribución Pareto con parámetro de forma conocido. Usando simulación Monte Carlo, generaron los valores críticos para distintos valores del parámetro de forma. Choulakian y Stephens (2001) propusieron usar las estadísticas A^2 y W^2 para probar el ajuste de una distribución Pareto Generalizada (PG) con parámetros desconocidos, estimando tales parámetros con máxima verosimilitud. Ellos dieron tablas de los puntos porcentuales asintóticos de las estadísticas para distintos valores del parámetro de forma.

Los trabajos mencionados dan evidencia de la versatilidad de las estadísticas FDE en los problemas de bondad de ajuste, sin embargo, también queda claro que si la función de distribución que se desea probar es postulada con parámetros desconocidos, el inconveniente de usar estas estadísticas será la estimación de tales parámetros. Esta es una cuestión que debe tenerse en consideración, ya que, típicamente, pocos problemas realistas involucran una función de distribución completamente especificada (sin parámetros desconocidos).

Por otra parte, ya se ha mencionado que los métodos gráficos suelen acompañar a procedimientos cuantitativos que miden las características detectadas por las gráficas. La *prueba del coeficiente de correlación* es una técnica de bondad de ajuste estrechamente relacionada a gráficas de probabilidad.

La prueba del coeficiente de correlación fue introducida por Filliben (1975) para probar el ajuste de una distribución normal. Looney y Gulledge (1985) usaron también el coeficiente de correlación como una técnica para construir estadísticas para varias posiciones gráficas comúnmente usadas para la distribución normal, y generaron la distribución de cada una usando métodos de muestreo empíricos. Kinnison (1989) usó el coeficiente de correlación entre los valores observados y sus valores esperados para probar el ajuste de una distribución de valor extremo; los valores esperados fueron calculados substituyendo el rango porcentual de los datos en la función de distribución inversa de la distribución de valor extremo estándar.

La idea básica de las pruebas del coeficiente de correlación es encontrar una función lineal de las observaciones conveniente, y medir tal linealidad computando el coeficiente de correlación muestral entre los datos y la función. El coeficiente de correlación muestral

1. Introducción

satisface las siguientes propiedades:

1. Su valor está limitado entre -1 y 1.
2. Es invariante ante cambios de escala y localidad en las variables.

Si bien las pruebas del coeficiente de correlación son procedimientos sencillos para evaluar el ajuste de los datos a una distribución dada, en muchas situaciones el estadístico de prueba resultante es intratable, de manera que no puede obtenerse su distribución exacta y es necesario estimarla mediante técnicas de simulación.

Como se ha visto, tratar con el problema de bondad de ajuste no es una tarea sencilla, y en varias situaciones puede complicarse aún más. Por ejemplo, hay muchos casos reales en los que es inconveniente o imposible hacer una completa medición de todos los miembros de una muestra aleatoria. En estudios médicos podría ser de interés determinar la distribución de los tiempos de supervivencia después de una operación, sin embargo, es común que el contacto con algunos individuos se pierda antes de su muerte, y que otros mueran por causas distintas a las de estudio. Similarmente, la observación de la vida (tiempo de funcionamiento) de un foco podría terminar por una rotura accidental del foco, o por una necesidad de usar el equipo o instalaciones de prueba para otros propósitos. En ambos ejemplos, las observaciones podrían también ser incompletas debido a la necesidad de terminar el estudio dentro de un tiempo razonable. Muestras de este tipo son llamadas muestras incompletas o censuradas.

Existen distintos tipos y mecanismos de censura. Los principales tipos de censura que se suelen considerar son la censura por la derecha, la censura por la izquierda y la censura por intervalo. La censura por la derecha se presenta cuando lo único que se sabe acerca de la variable de interés es que es mayor que algún valor. Simétricamente al caso anterior, se dice que la variable ha sido censurada por la izquierda si lo único que se sabe es que es menor que algún valor. Por último, la censura por intervalo combina los conceptos de censura por la izquierda y por la derecha, ya que sólo se sabe que la variable está entre dos valores.

Básicamente, se pueden distinguir los mecanismos de censura siguientes: censura fija tipo I, censura fija tipo II y censura aleatoria. La censura fija tipo I se presenta en la situación en donde se prefija, por parte del investigador, un límite superior, por ejemplo, el tiempo de duración del estudio o periodo de observación, por lo que una variable es observada

1. Introducción

solo si su valor es menor o igual al límite fijado. En el mecanismo de censura tipo II, el periodo de observación se determina después de haber alcanzado un número r prefijado de sucesos, es decir, el estudio continua solo hasta que se han observado un número r , determinado de antemano, de elementos. El tipo de censura aleatoria se produce cuando se supone que la censura viene dada por una variable aleatoria independiente de la variable de interés, tal que está última es censurada si su valor es mayor que el valor de la variable de censura, y es observada si su valor es menor o igual que el de la variable de censura.

En el problema clásico (no censurado) de bondad de ajuste uno observa una muestra aleatoria completa de una población con alguna función de distribución. La hipótesis nula afirma que la función de distribución es de un tipo dado. La necesidad de generalizar este problema para incluir datos censurados surge porque, como se señaló, en muchas situaciones prácticas no se dispone de una muestra completa.

Diversas técnicas de bondad de ajuste, como las ya mencionadas, han sido adaptadas a situaciones que incluyen muestras incompletas. Sin embargo, en la mayoría de los casos, los procedimientos se hacen más difíciles.

Como ejemplos de adaptaciones de pruebas de bondad de ajuste a distintos tipos y mecanismos de censura, se pueden citar los siguientes: [Michael y Schucany \(1986\)](#) presentaron extensiones del método de gráficas de probabilidad para incluir el caso de muestras incompletas para la mayoría de los tipos de censura. Para un conjunto de familias de distribuciones seleccionadas, dieron las fórmulas o valores a graficar en el eje de las abscisas y de las ordenadas. Adaptaciones de las estadísticas FDE a situaciones que incluyen algunos tipos de censura, se pueden encontrar, por ejemplo, en [Pettitt \(1977\)](#), [Koziol \(1980\)](#), [Sirvanci y Levent \(1982\)](#), [Stephens \(1986b\)](#) y [Guilbaud \(1988\)](#). [Chen \(1984\)](#) propuso una estadística de correlación como prueba de bondad de ajuste para datos con censura aleatoria.

En muchos de los trabajos de investigación sobre pruebas de bondad de ajuste, tanto para muestras completas como para muestras censuradas, se tiene interés en algún tipo de distribución en particular, ya sea por su relevancia en aplicaciones de la vida real, por la insuficiencia de trabajos publicados sobre ella, por su importancia para ayudar a explicar temas de actualidad, o por alguna otra razón de trascendencia para el investigador. El gran uso que se ha dado a la distribución Pareto para modelar diversos fenómenos en distintos campos de aplicación, motiva su estudio.

1. Introducción

La distribución Pareto fue introducida por el economista y sociólogo italiano Vilfredo Pareto (1848-1923) en el contexto de distribución de ingresos sobre una población cuando el ingreso excede un cierto límite. Sin embargo, su estudio ha extendido las áreas de aplicación más allá del contexto socioeconómico. Diversos autores han empleado esta distribución para modelar tamaños de ciudades, errores de clusters en circuitos de comunicación, tiempos de servicio y sistemas de espera, pruebas de vida, tiempos de falla, sustitución política, tamaños de empresas, stock de precios, entre otros.

La literatura sobre la distribución Pareto se ha desarrollado rápidamente a partir de la década de los 80's. [Arnold \(1983\)](#), quien ha dedicado varios trabajos de investigación al estudio de la distribución Pareto, aportó un excelente texto dedicado casi exclusivamente a esta distribución. Además de una minuciosa discusión de la caracterización y medidas de desigualdad que están cercanamente asociadas con la distribución Pareto, la monografía de Arnold provee una extensa revisión histórica.

No obstante del gran uso dado a la distribución Pareto, existe una escasez en la literatura de propuestas para probar su ajuste, menos aún si el conjunto de datos que se dispone está censurado. Esta es la principal motivación del presente trabajo, además de la convicción de que el uso de la distribución Pareto para aplicaciones prácticas puede ser incrementado por un método preciso de determinar si las observaciones muestrales vienen de una población gobernada por esta distribución.

Para el caso de muestras completas de una distribución Pareto Generalizada, que incluye como un caso especial a la distribución Pareto, [Porter et al. \(1992\)](#) desarrollaron modificaciones de las estadísticas D , A^2 y W^2 para probar el ajuste con parámetros de localización y escala desconocidos y conocido parámetro de forma; los parámetros desconocidos fueron estimados usando mejores estimadores linealmente insesgados. Se investigó la potencia de la prueba por simulación Monte Carlo bajo 8 distribuciones alternativas. En la mayoría de los casos, la potencia de las pruebas propuestas es considerablemente más alta que la prueba chi-cuadrada. [Choulakian y Stephens \(2001\)](#) dieron pruebas basadas en las estadísticas W^2 y A^2 , y mostraron cómo pueden ser usadas para ayudar a seleccionar el umbral en el modelo de *picos sobre un umbral*. La distribución asintótica de estas estadísticas de prueba depende del parámetro de forma. [Meintanis y Bassiakos \(2007\)](#) propusieron una prueba de bondad de ajuste del tipo Crámer-von Mises, basada en transformaciones de los datos originales. La prueba fue comparada contra los procedimientos desarrollados por [Choulakian y Stephens \(2001\)](#). Por los resultados del estudio Monte Carlo para investigar la potencia concluyeron que la prueba compite

1. Introducción

favorablemente con los procedimientos de las FDE.

Particularmente para la distribución Pareto, [Beirlant et al. \(2006\)](#) propusieron dos modificaciones de la estadística de Jackson, originalmente propuesta para probar exponencialidad. Para ambas estadísticas, la distribución límite fue derivada. El método se ilustró con dos estudios de caso prácticos. Al igual que las anteriores, estas estadísticas no se aplicaron a datos censurados.

Más recientemente, [Zeinab \(2007\)](#) propuso un procedimiento gráfico basado en la función de riesgo y dos procedimientos paramétricos, uno basado en la prueba clásica χ^2 de Pearson y otro en la estadística de razón de verosimilitud. Las pruebas se desarrollaron para *datos agrupados* provenientes de una distribución Pareto, considerando censura por la derecha Tipo I y Tipo II.

En esta investigación, se propone una metodología de bondad de ajuste para *datos no agrupados* de la distribución Pareto, en presencia de censura por la derecha Tipo II. La prueba involucra transformaciones de los datos originales, basadas en la relación que existe entre la distribución Pareto y la distribución exponencial, y utiliza el estimador Nelson-Aalen de la Función de Riesgo Acumulada.

El capítulo 3 constituye una revisión de los conceptos teóricos utilizados a lo largo de la investigación. En el capítulo 4 se presenta la metodología de bondad de ajuste propuesta; vía simulación Monte Carlo se obtiene la distribución empírica del estadístico de prueba y se investiga su tamaño y potencia. El capítulo 5 presenta un estudio comparativo de la potencia contra adaptaciones al caso de censura Tipo II de las estadísticas W^2 , A^2 y una basada en la información de Kullback-Leibler. Por último, en el capítulo 6 se ilustra con datos reales la aplicación de la prueba.

Capítulo 2

Objetivos

General

Proponer una metodología estadística que permita verificar si las observaciones de una muestra aleatoria censurada Tipo II se ajustan a una distribución de la forma Pareto para el caso univariado.

Particulares

1. Proponer una prueba de bondad de ajuste para la distribución Pareto cuando se tienen muestras censuradas Tipo II, usando el estimador de Nelson - Aalen.
2. Generar las tablas de valores críticos de la distribución del estadístico de prueba propuesto.
3. Realizar un análisis comparativo sobre el tamaño y la potencia de la prueba propuesta contra adaptaciones de las pruebas Anderson-Darling, Cramér-von Mises y Kullback-Leibler.
4. Ilustrar la metodología propuesta con aplicaciones de datos reales.

Capítulo 3

Revisión de Literatura

3.1. Pruebas de hipótesis

El objetivo de la estadística es hacer inferencia sobre ciertas características de una población, basada en la información contenida en una muestra. En el logro de este objetivo, las pruebas de bondad de ajuste desempeñan un papel primordial. Estas son usadas para determinar qué tan bien una muestra de datos concuerda, o es consistente, con la distribución sobre la cual será basada la inferencia.

Para entender la metodología concerniente a pruebas de bondad de ajuste, es necesario introducir algunas definiciones y establecer la notación. En principio, se asume que se tiene una muestra aleatoria X_1, \dots, X_n de alguna densidad $f(x; \theta)$ con su correspondiente función de distribución $F(x; \theta)$, donde $\theta \in \Omega \subset \mathbb{R}^m$, $m \geq 1$ y Ω el espacio de parámetros.

Definición 3.1. Hipótesis estadística Una hipótesis estadística es una aseveración o conjetura sobre la distribución de una o más variables aleatorias. Si la hipótesis estadística especifica completamente la distribución, entonces es llamada hipótesis simple, de otro modo es llamada hipótesis compuesta.

La hipótesis estadística se denota por la letra mayúscula H seguida de dos puntos y la aseveración que especifica la hipótesis.

Definición 3.2. Prueba de hipótesis

Una prueba de una hipótesis estadística es una regla o procedimiento para decidir cuando

3.1. Pruebas de hipótesis

rechazar H .

Denote por φ una prueba y por Ξ el espacio muestral de observaciones, o el conjunto de datos potencial; esto es,

$$\Xi = \{(x_1, \dots, x_n) : (x_1, \dots, x_n) \text{ es un posible valor de } (X_1, \dots, X_n)\}$$

Sea φ una prueba de una hipótesis estadística H definida como sigue: Rechazar H si y solo si $(x_1, \dots, x_n) \in \Xi_R$, donde Ξ_R es un subconjunto de Ξ . El espacio muestral es particionado en dos subconjuntos, Ξ_R y Ξ_A , tal que $\Xi = \Xi_R \cup \Xi_A$ y $\Xi_R \cap \Xi_A = \emptyset$. A Ξ_R se le denomina región de rechazo de la prueba y a Ξ_A región de aceptación.

En muchos problemas de pruebas de hipótesis se discuten dos hipótesis: la primera, llamada hipótesis *nula*, es aquella que está siendo probada y es denotada por H_0 ; la segunda es llamada hipótesis *alternativa*, denotada por H_1 . El razonamiento es que si la hipótesis nula es falsa, entonces la hipótesis alternativa es verdadera, y viceversa. Comúnmente se dice que H_0 es probada en contra de H_1 .

Entonces, usualmente se tiene interés en probar el siguiente juego de hipótesis:

$$H_0 : \theta \in \omega \subset \Omega \quad \text{vs} \quad H_1 : \theta \in \Omega - \omega$$

En el caso de pruebas de bondad de ajuste, el juego de hipótesis a ser contrastada se expresa como sigue:

$$H_0 : F(x; \theta) = F^*(x; \theta) \quad \text{vs} \quad H_1 : F(x; \theta) \neq F^*(x; \theta)$$

donde $F^*(x; \theta)$ es la función de distribución hipotética, y los demás elementos definidos como antes. Si la hipótesis nula no es rechazada, se dice que H_1 es aceptada. Con esto en mente, se habla de dos tipos de error que pueden ser cometidos:

1. Error Tipo I: Rechazar H_0 cuando H_0 es verdadera.
2. Error Tipo II: Aceptar H_0 cuando H_0 es falsa.

3.1. Pruebas de hipótesis

El *tamaño de un error Tipo I* es definido como la probabilidad de que sea cometido un error Tipo I:

$$P[\text{Rechazar } H_0 | H_0]$$

De manera similar, el *tamaño de un error Tipo II* es la probabilidad de que se cometa un error Tipo II:

$$P[\text{Aceptar } H_0 | H_1]$$

Como no es posible tener una prueba que minimice ambos tipos de error a la vez, lo que usualmente se hace es fijar la probabilidad de cometer el error tipo I y tratar de minimizar la probabilidad de cometer el error tipo II.

Definición 3.3. Función potencia

La función potencia de la prueba φ , denotada por $\beta_\varphi(\theta)$, es definida como la probabilidad de que H_0 sea rechazada cuando la distribución de la cual fue obtenida la muestra fue parametrizada por θ .

La expresión de esta función es:

$$\beta_\varphi(\theta) = P_\theta[\text{Rechazar } H_0]$$

donde P denota probabilidad. La función potencia es usualmente el criterio para evaluar la bondad de una prueba o para comparar dos pruebas en competencia. Una función potencia ideal es una función que es cero para aquellos valores θ correspondientes a la hipótesis nula. La idea es que uno no quiere rechazar H_0 si H_0 es verdad y se desea rechazar H_0 cuando H_0 es falsa.

Definición 3.4. Tamaño de prueba

El tamaño de la prueba φ de H_0 es definido como el $\sup_{\theta \in \omega} [\beta_\varphi(\theta)]$.

Se dice que una prueba φ es de *tamaño* $\alpha \in (0, 1)$ si el $\sup_{\theta \in \omega} [\beta_\varphi(\theta)] = \alpha$ y es de *nivel* $\alpha \in (0, 1)$ si $\sup_{\theta \in \omega} [\beta_\varphi(\theta)] \leq \alpha$.

3.1. Pruebas de hipótesis

El criterio para encontrar la prueba óptima es escoger la prueba φ^* tal que las probabilidades de ambos errores sean mínimas, es decir, tal que $\beta_{\varphi^*}(\theta_0) \leq \beta_{\varphi}(\theta_0)$ y $\beta_{\varphi^*}(\theta_1) \geq \beta_{\varphi}(\theta_1)$ para cualquier otra prueba φ . Note que es deseable encontrar una prueba de tamaño exactamente α , ya que es el máximo riesgo permitido y asegura la máxima potencia.

3.1.1. Algunas pruebas de bondad de ajuste comunes

Las pruebas formales de bondad de ajuste más usuales, las cuales pueden ser adaptadas a diversas formas de censura, están basadas en la función de distribución empírica (FDE).

La FDE es una función escalonada, calculada de la muestra, la cual estima la función de distribución de la población. Las estadísticas FDE son medidas de la discrepancia entre la FDE y una función de distribución dada, y son usadas para probar el ajuste de la muestra a la distribución. Valores suficientemente grandes de estas estadísticas son evidencia en contra del modelo hipotético.

La FDE se define como sigue: suponga que una muestra aleatoria de tamaño n es X_1, \dots, X_n , y sean $X_{(1)}, \dots, X_{(n)}$ las estadísticas de orden. Suponga también que la distribución de X es $F(x)$. Entonces, la FDE $F_n(x)$ es tal que

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I(-\infty, x](X_i)$$

Note que

$$F_n(x) = 0, \quad x < X_{(1)}$$

$$F_n(x) = \frac{i}{n}, \quad X_{(i)} \leq x < X_{(i+1)}$$

$$F_n(x) = 1, \quad X_{(n)} \leq x$$

Un tipo de estadísticas FDE, de la clase cuadrática, son de la familia

$$Q = n \int_{-\infty}^{\infty} \{F_n(x) - F(x)\}^2 \psi(x) dF(x)$$

3.1. Pruebas de hipótesis

donde $\psi(x)$ es una función conveniente, la cual da los pesos a la diferencia cuadrática $\{F_n(x) - F(x)\}^2$. Cuando $\psi(x) = 1$, se trata de la estadística de Cramér-von Mises, denotada por W^2 , y cuando $\psi(x) = \{[F(x)][1 - F(x)]\}^{-1}$, se trata de la estadística de Anderson-Darling, denotada por A^2 .

Si se desea hacer uso de estas estadísticas para probar la hipótesis nula H_0 : una muestra aleatoria $\{X_i\}_{i=1}^n$ proviene de $F(x; \theta)$, donde θ es un vector de parámetros desconocidos, entonces, se hace necesaria la estimación de tales parámetros.

Cuando los componentes de θ son parámetros desconocidos de localización o escala, y si estos son estimados por métodos apropiados, las distribuciones de las estadísticas FDE no dependerán de los verdaderos valores de los parámetros desconocidos. Así, los puntos porcentuales para las pruebas FDE de tales distribuciones dependen solo de la familia probada y del tamaño de muestra n . Para las estadísticas W^2 y A^2 existe teoría asintótica; además, los puntos porcentuales de esas estadísticas para finito n convergen rápidamente a los puntos asintóticos. Por otra parte, cuando los parámetros desconocidos no son de localización o escala, por ejemplo cuando el parámetro desconocido es un parámetro de forma, la distribución, incluso asintótica, dependerá del verdadero valor del parámetro; [Stephens \(1986b\)](#).

[Stephens \(1986b\)](#) presentaron fórmulas computacionales para las estadísticas W^2 y A^2 . Tales fórmulas son hechas usando la Transformación Integral de Probabilidad, $Z = F(x)$; cuando $F(x)$ es la verdadera distribución de X , la nueva variable aleatoria Z es uniformemente distribuida entre 0 y 1. Entonces, Z tiene función de distribución $F^*(z) = z$, $0 \leq z \leq 1$.

Las fórmulas para calcular las estadísticas FDE de los Z -valores son:

$$W^2 = \sum_{i=1}^n \{Z_{(i)} - (2i - 1)/(2n)\}^2 + 1/(12n)$$
$$A^2 = -n - (1/n) \sum_{i=1}^n (2i - 1) \{\log Z_{(i)} + \log[1 - Z_{(n+1-i)}]\}$$

donde $\log x$ significa $\log_e x$, y $Z_{(1)} < Z_{(2)} < \dots < Z_{(n)}$ son estadísticas de orden.

[Stephens \(1986b\)](#) proporcionó una serie de adaptaciones de las estadísticas FDE a diversos tipos de censura. Cuando la censura es por la derecha Tipo II hay r valores Z_i , con

3.1. Pruebas de hipótesis

$Z_{(r)}$ el más grande y r fijo. Entonces, las fórmulas están dadas por:

$$W_{r,n}^2 = \sum_{i=1}^r \left\{ Z_{(i)} - \frac{2i-1}{2n} \right\}^2 + \frac{r}{12n^2} + \frac{n}{3} \left(Z_{(r)} - \frac{r}{n} \right)^3 \quad (3.1)$$

$$A_{r,n}^2 = -\frac{1}{n} \sum_{i=1}^r (2i-1) \{ \log Z_{(i)} - \log[1 - Z_{(i)}] \} - 2 \sum_{i=1}^r \log\{1 - Z_{(i)}\} \\ - \frac{1}{n} \{ (r-n)^2 \log[1 - Z_{(r)}] - r^2 \log Z_{(r)} + n^2 Z_{(r)} \} \quad (3.2)$$

3.1.2. Coeficiente de correlación

Un método gráfico de bondad de ajuste, en el cual las estadísticas de orden $U_{(j)}$ de una muestra aleatoria U_1, \dots, U_n son graficadas en el eje vertical, en contra de V_i , una función conveniente de i , en el eje horizontal, para ajustar una línea recta a los puntos y, entonces, basar la prueba en el coeficiente de correlación entre U y V , es llamada prueba del coeficiente correlación.

Permita que U y V sean dos variables aleatorias. El coeficiente de correlación, denotado por $\rho_{U,V}$, de las variables U y V es definido como

$$\rho_{U,V} = \frac{\text{Cov}[U, V]}{\sqrt{\text{Var}(U)} \sqrt{\text{Var}(V)}}, \quad (3.3)$$

donde

$$\text{Cov}[U, V] = \text{E}\{[U - \text{E}(U)][V - \text{E}(V)]\} \\ = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [U - \text{E}(U)][V - \text{E}(V)] f_{U,V}(U, V) dudv$$

es la covarianza entre las variables y

3.1. Pruebas de hipótesis

$$\text{Var}(U) = E\{[U - E(U)]^2\} = \int_{-\infty}^{\infty} [U - E(U)]^2 f_u(u) du$$

es la varianza de U , definida de manera similar para V . E denota esperanza y las definiciones tienen sentido solo si las esperanzas existen.

El coeficiente de correlación y la covarianza son medidas de una relación lineal entre las variables U y V , sin embargo, la magnitud de $\text{Cov}[U, V]$ no tiene mucho significado ya que depende de la variabilidad de U y V . El coeficiente de correlación remueve, en un sentido, la variabilidad individual de cada variable dividiendo la covarianza por el producto de las desviaciones estándar, por lo que $\rho_{U,V}$ es una mejor medida de la relación lineal de U y V que la covarianza.

El coeficiente de correlación tiene las siguientes propiedades:

1. Se puede demostrar con la desigualdad de Cauchy-Schwarz que se satisface $-1 \leq \rho_{U,V} \leq 1$.
2. El coeficiente de correlación entre dos variables aleatorias es invariante ante cambios de escala y localidad en las variables.
3. Si una variable aleatoria V es función lineal de otra variable aleatoria U , entonces $\rho_{U,V} = \pm 1$.
4. Si U y V son independientes entonces $\rho_{U,V} = 0$.

Estimador del coeficiente de correlación

Sea $(U_1, V_1), \dots, (U_n, V_n)$ una muestra aleatoria bivariada de tamaño n . Un estimador del coeficiente de correlación $\rho_{U,V}$, denotado por R , es

$$R = \frac{\sum_{i=1}^n (U_i - \bar{U})(V_i - \bar{V})}{\sqrt{\sum_{i=1}^n (U_i - \bar{U})^2} \sqrt{\sum_{i=1}^n (V_i - \bar{V})^2}} \quad (3.4)$$

3.2. La Distribución Pareto

R es conocido como el coeficiente de correlación muestral y se puede demostrar que satisface las siguientes propiedades:

1. $-1 \leq R \leq 1$.
2. El coeficiente de correlación muestral es invariante ante cambios de escala y localización en las variables.

Coeficiente de correlación para datos censurados

Suponga que los datos han sido censurados y solo se tiene disponible un subconjunto de las U_i . Siempre que los rangos i de las U_i conocidas sean también conocidos, la correspondiente V_i puede ser pareada con la U_i , y el coeficiente de correlación puede ser calculado como se indicó en (3.4), con las sumas solo sobre los valores i disponibles. Así, el cálculo de R es fácilmente adaptado a todos los tipos de censura.

3.2. La Distribución Pareto

3.2.1. Características

En estadística, la distribución Pareto es una distribución de probabilidad continua con dos parámetros, $\beta > 0$ (parámetro de forma) y $\theta > 0$ (parámetro de escala), cuya función de densidad está dada por

$$f_X(x; \beta, \theta) = \frac{\beta\theta^\beta}{x^{(\beta+1)}} I_{(\theta, \infty)}(x) \quad (3.5)$$

Su función de distribución es

$$F_X(x; \beta, \theta) = P(X \leq x) = \int_\theta^x \frac{\beta\theta^\beta}{x^{(\beta+1)}} = 1 - \left(\frac{\theta}{x}\right)^\beta \quad (3.6)$$

El parámetro θ marca el límite inferior de los valores posibles que una variable aleatoria Pareto puede tomar. El parámetro β tiene que ver con la dispersión; a mayores valores

3.2. La Distribución Pareto

de β , se obtienen densidades Pareto más concentradas cerca del mínimo (θ), es decir, menos dispersas; ver figura 3.1.

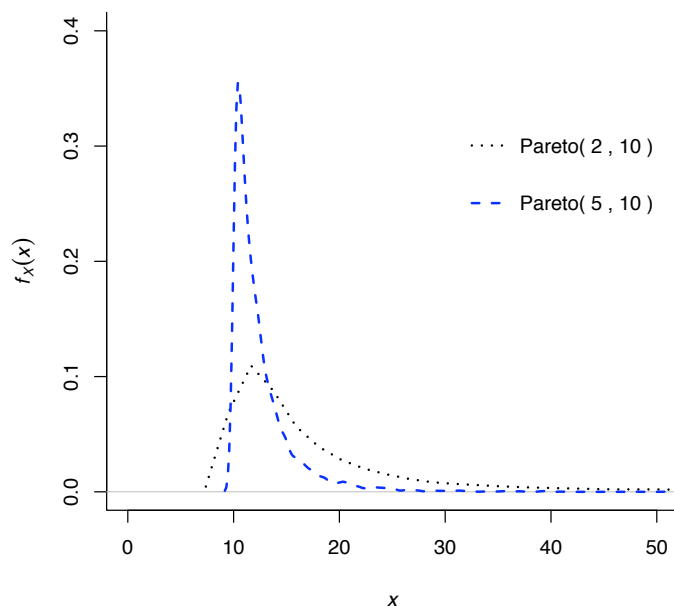


Figura 3.1: Dos densidades Pareto con distinto parámetro de forma

La distribución Pareto es una distribución de cola pesada. En consecuencia, la media y la varianza son finitas solo si el parámetro de forma es suficientemente grande. En efecto, sean

$$E(X) = \frac{\beta}{\beta - 1}\theta, \quad \beta > 1,$$
$$\text{Var}(X) = \frac{\beta}{(\beta - 1)^2(\beta - 2)}\theta^2, \quad \beta > 2$$

la media y la varianza, respectivamente. Entonces, se tiene que si $\beta \leq 1$ el valor esperado es infinito, al igual que la varianza si $\beta \leq 2$. Además, note que a medida que β crece, la media tiende a aproximarse a θ (por la derecha), lo que confirma que la distribución tiende a concentrarse cerca de θ para valores grandes de β .

3.2. La Distribución Pareto

En general, se tiene que

$$E(X^n) = \begin{cases} \frac{\beta}{\beta-n}\theta^n & \text{si } n < \beta \\ \infty & \text{si } n \geq \beta \end{cases}$$

La notación $Par(\beta, \theta)$ será usada para referir a una variable aleatoria con función de distribución (3.6).

3.2.2. La distribución Pareto y su relación con otras distribuciones

Permita que X sea una variable aleatoria (v.a.) con función de distribución $Par(\beta, 1)$. Entonces,

1. La v.a. $Y = \frac{1}{X}$ se distribuye $Beta(\beta, 1)$.

$$\begin{aligned} F_Y(y) &= P[Y \leq y] = P\left[\frac{1}{X} \leq y\right] = P\left[X > \frac{1}{y}\right] = 1 - F_X\left(\frac{1}{y}\right) \\ f_Y(y) &= F'_Y(y) = f_X\left(\frac{1}{y}\right) \cdot \frac{1}{y^2} = \frac{\beta}{\left(\frac{1}{y}\right)^{\beta+1}} \cdot \frac{1}{y^2} = \beta y^{\beta-1}, \quad 0 < y < 1 \end{aligned} \quad (3.7)$$

La función (3.7) corresponde a una densidad $Beta(\beta, 1)$.

2. La v.a. $Y = \ln(X)$ se distribuye $Exp(\beta)$.

$$\begin{aligned} F_Y(y) &= P[Y \leq y] = P[\ln(X) \leq y] = P[X \leq e^y] = F_X(e^y) \\ f_Y(y) &= F'_Y(y) = f_X(e^y) \cdot e^y = \frac{\beta}{(e^y)^{\beta+1}} \cdot e^y = \beta e^{-\beta y}, \quad y > 0 \end{aligned} \quad (3.8)$$

Es decir, se trata de una densidad exponencial con parámetro β .

Ahora, permita que X sea una v.a. con función de distribución $Par(\beta, \theta)$. Entonces,

3.2. La Distribución Pareto

3. La v.a. $Y = cX$, donde $c > 0$ es una constante, se distribuye $Par(\beta, c\theta)$

$$\begin{aligned} F_Y(y) &= P[Y \leq y] = P[cX \leq y] = P[X \leq \frac{y}{c}] = F_X(\frac{y}{c}) \\ f_Y(y) &= F'_Y(y) = f_X(\frac{y}{c}) \cdot \frac{1}{c} = \frac{\beta\theta^\beta}{(\frac{y}{c})^{\beta+1}} \cdot \frac{1}{c} = \frac{\beta(c\theta)^\beta}{y^{\beta+1}}, \quad y > c\theta \end{aligned} \quad (3.9)$$

4. La v.a. $Y = \ln(X)$ tiene función de distribución $Exp(\lambda, \beta)$ con $\lambda = \ln \theta$

$$\begin{aligned} F_Y(y) &= P[Y \leq y] = P[\ln(X) \leq y] = P[X \leq e^y] = F_X(e^y) \\ f_Y(y) &= F'_Y(y) = f_X(e^y) \cdot e^y = \frac{\beta\theta^\beta}{(e^y)^{\beta+1}} \cdot e^y = \beta\theta^\beta e^{-\beta y} = \beta e^{\ln\theta^\beta} e^{-\beta y} \\ &= \beta e^{\beta \ln\theta} e^{-\beta y} = \beta e^{-\beta(y-\lambda)}, \quad y > \lambda, \quad \lambda = \ln \theta \end{aligned} \quad (3.10)$$

3.2.3. Algunas aplicaciones del modelo Pareto

La distribución Pareto fue introducida por el economista y sociólogo italiano Vilfredo Pareto (1848-1923) para modelar la distribución de ingresos sobre una población, cuando el ingreso excede cierto límite. En este caso, el parámetro β , llamado índice de Pareto, es una medida de la amplitud de la distribución de los ingresos, e incorpora el principio de Pareto, que consistía en la observación de que el 20% de los miembros de la sociedad italiana poseían el 80% de la riqueza.

Una de las caracterizaciones más simples de la distribución de Pareto, cuando se usa para modelar la distribución del ingreso, dice que la proporción de la población cuyo ingreso excede cualquier número positivo $x > \theta$ es:

$$P(X > x) = 1 - F_X(x) = \left(\frac{\theta}{x}\right)^\beta$$

donde θ es el ingreso de la gente más pobre. Cuanto más grande es el índice de Pareto, más pequeña es la proporción de gente muy rica (con ingreso elevado).

Además del contexto socioeconómico, la distribución Pareto ha sido empleada para modelar diversos fenómenos en una gran variedad de aplicaciones. Varios autores la han usado como distribución de ingresos, de tiempos de vida o falla de componentes de equipos, stock de precios, tamaño de firmas, entre otros. Por ejemplo, [Ismail \(2004\)](#), [Soliman](#)

3.3. Algunos conceptos relacionados con el análisis de supervivencia

(2000) y [Dyer \(1981\)](#) presentaron aplicaciones en un contexto socioeconómico; [Beirlant et al. \(2006\)](#) la usaron en un contexto geoestadístico, concerniente a la valuación de diamantes; [Zeinab \(2007\)](#) ajustó la distribución a los tiempos de supervivencia de pacientes que fueron sujetos a diálisis, [Ouyang y Wu \(1994\)](#) a los tiempos de falla de componentes, [Nigm y Hamdy \(1987\)](#) a los tiempos de vida de negocios y [Harris \(1968\)](#) encontró útil la distribución Pareto en modelar tiempos de servicio y sistemas de espera.

3.3. Algunos conceptos relacionados con el análisis de supervivencia

El objetivo principal del Análisis de Supervivencia es usualmente el estudio del tiempo que transcurre hasta que ocurre un evento de interés (*tiempo hasta* o *tiempo transcurrido hasta*). Típicamente, el evento de interés es la muerte de una unidad biológica (paciente, animal, célula, etc.) o la falla de un componente físico (mecánico o eléctrico). Sin embargo, podría tratarse también de la recurrencia de una enfermedad, el aprendizaje de una tarea específica, etc. En todos los casos, el *tiempo transcurrido hasta* será la medida del tiempo desde algún punto particular de inicio hasta que se presenta el evento de interés. En un sentido figurativo, esta medida de tiempo es llamada *tiempo de vida* o *tiempo de supervivencia*.

Para generalizar, las personas, animales, componentes, etc. que son de interés en observar en el estudio, serán llamados *unidad de estudio* o simplemente *unidad*. Por otro lado, note que el hecho de estudiar la variable tiempo hace que los datos sean siempre no negativos.

Una característica inherente a estos estudios de supervivencia es conocida como **censura**. Existen varios tipos (y mecanismos) de censura, tales como censura por la derecha (Tipo I y Tipo II), censura por la izquierda, censura por intervalo y censura aleatoria, las cuales serán discutidas en este capítulo. Para una revisión más completa y ejemplos de cada una ver el Capítulo 3 de [Klein y Moeschberger \(1997\)](#).

3.3. Algunos conceptos relacionados con el análisis de supervivencia

3.3.1. Censura por la derecha

En la *censura por la derecha Tipo I* el evento es observado solo si ocurre antes de algún tiempo pre-especificado (tiempo de censura o límite de observación), el cual puede ser fijo para todas las unidades de estudio o puede variar de unidad a unidad.

Permita que X_1, \dots, X_n sean variables aleatorias que representen los tiempos de vida de n unidades bajo estudio. Las X 's se asumen independientes e idénticamente distribuidas (i.i.d.) con función de densidad de probabilidad $f(x)$ y función de supervivencia $S(x) = 1 - F(x)$. Sean $L_i > 0$, $i = 1, \dots, n$, los tiempos de censura especificados al inicio del estudio. Si la censura es fija para todas las unidades, significa que $L_1 = \dots = L_n$.

El tiempo de vida exacto de una unidad de estudio será conocido si, y sólo si, X_i es menor que o igual a L_i . Si X_i es mayor que L_i , se trata de una unidad sobreviviente y su tiempo de evento es censurado al tiempo L_i . Entonces, en lugar de observar los tiempos de vida X_1, \dots, X_n , se observan los datos por las variables T_1, \dots, T_n , con $T_i = \min(X_i, L_i)$. Los datos de tiempos de vida observados pueden ser convenientemente representados por pares de variables aleatorias de la forma (T_i, δ_i) , donde $\delta = 1$ si $T_i = X_i$ y $\delta = 0$ si $T_i = L_i$.

Un segundo tipo de censura por la derecha es la *censura Tipo II*, en la cual el estudio continúa solo hasta que las primeras r unidades experimentan el evento de interés, donde r es algún entero determinado al inicio del estudio, tal que $r < n$. En esta situación, solo se observan los r tiempos de vida menores. Entonces, los datos observados consistirán de (T_i, δ_i) , donde $T_i = X_{(r)}$ para aquellas unidades censuradas y $T_i = X_{(i)}$ con $\delta = 1$, siendo las $X_{(i)}$ estadísticas de orden.

3.3.2. Censura aleatoria

En algunas ocasiones ocurre que un evento independiente al evento de interés es causa de que una unidad bajo estudio sea censurada de manera aleatoria. Por ejemplo, en estudios médicos puede ocurrir que muertes accidentales, migraciones de las personas, pacientes que abandonen el experimento clínico, muerte por alguna causa distinta a la de interés, etc., sean la razón por la cual un individuo es removido del estudio (censurado). El mecanismo de censura que sigue a este ejemplo es denominado *censura aleatoria*.

3.3. Algunos conceptos relacionados con el análisis de supervivencia

Sea X_1, \dots, X_n una muestra aleatoria de una función de distribución continua F , y permita que L_1, \dots, L_n sea una muestra aleatoria independiente con función de distribución de censura G . Como antes, las X 's denotan tiempos de vida y las L 's tiempos de censura. Entonces, para datos con censura aleatoria por la derecha, las observaciones consisten de las parejas (T_i, δ_i) para $i = 1, \dots, n$, donde $T_i = \min(X_i, L_i)$ y $\delta_i = I(X_i \leq L_i)$.

Note que, a diferencia de la censura por la derecha Tipo I, las L 's son valores de otras variables aleatorias no conocidas de antemano, las cuales se asumen independientes de las X 's.

3.3.3. Censura por la izquierda

Se considera que un tiempo de vida X_i asociado con una unidad específica bajo estudio es censurado por la izquierda si es menor que un tiempo de censura C , es decir, el evento de interés ya ha ocurrido antes de que la unidad sea observada en el estudio al tiempo C . Para esa unidad, se sabe que ya ha experimentado el evento antes del tiempo C , pero el tiempo exacto al cual ocurrió es desconocido. El tiempo de vida exacto X_i será conocido si, y solo si, X_i es mayor que o igual a C . Los datos censurados por la izquierda pueden también ser representados por pares de variables aleatorias (T_i, δ_i) , donde como antes $T_i = X_i$ si se observa el tiempo de vida exacto y δ_i indica si se trata de un tiempo censurado o no.

3.3.4. Censura por intervalo

Un tipo más general de censura ocurre cuando solo se conoce que el tiempo de vida ocurre dentro de un intervalo, y se denomina censura por intervalo. Por ejemplo, esta censura ocurre cuando los pacientes en un estudio clínico o longitudinal se revisan periódicamente y solo se sabe que el tiempo de evento de los pacientes ocurre en un intervalo $(U_i, V_i]$. Este tipo de censura también puede ocurrir en experimentos industriales donde hay una inspección periódica del adecuado funcionamiento de algún componente físico.

Note que la censura por intervalo es una generalización de las censuras por la derecha y por la izquierda. Cuando el punto izquierdo del intervalo es cero y el punto derecho es C definido como en el apartado 3.3.3, entonces se tiene censura por la izquierda; cuando el

3.3. Algunos conceptos relacionados con el análisis de supervivencia

punto izquierdo es L_i definido en el apartado 3.3.1 y el punto derecho es infinito, se tiene censura por la derecha.

En todos los casos una censura, por definición, siempre impide el conocimiento de X_i . Por otro lado, el evento de interés, tal como una muerte o una falla, no siempre evita el conocimiento del correspondiente L_i o C , en caso de que los límites de las observaciones sean no aleatorios y previsible.

Para los distintos mecanismos de censura (cuando sea aplicable), es deseable suponer que la muerte o falla (o cualquier otro que sea el evento de interés) de una unidad y la pérdida (censura) de la misma, o de cualquiera otra, nunca ocurren al mismo tiempo. De otro modo, cuando el evento de interés para una unidad bajo estudio ocurre a un tiempo $X_i = L_i$, el tiempo de vida exacto de esa unidad es efectivamente conocido, por lo que es tratado como tal y no será censurado. En notación, esta unidad será registrada como (T_i, δ_i) , donde $T_i = X_i$ y $\delta_i = 1$.

3.3.5. Funciones básicas en el análisis de supervivencia

Como antes, permita que X denote el tiempo de algún evento especificado. Este evento podría ser una muerte, la aparición de un tumor, el desarrollo de alguna enfermedad, la falla de un equipo o maquinaria, el dejar de fumar, el aprendizaje de una tarea específica, etc. De manera más precisa, X es una variable aleatoria no negativa de una población homogénea con función de distribución $F(x)$.

Existen varias funciones, equivalentes entre sí, que caracterizan la distribución de X : la función de supervivencia $S(x)$, la función de riesgo o función tasa de falla $h(x)$, la función de riesgo acumulado o función tasa de falla acumulada $H(x)$, y la función tiempo de vida media residual $m(x)$. Estas funciones son usadas para ilustrar diferentes aspectos de la distribución de X , y como se mostrará a continuación, están estrechamente relacionadas entre sí.

Función de supervivencia

La *función de supervivencia*, esto es, la probabilidad de que una unidad experimente el evento de interés después del tiempo x o, dicho de otro modo, sobreviva más allá del

3.3. Algunos conceptos relacionados con el análisis de supervivencia

tiempo x , es definida como:

$$S(x) = P(X > x)$$

En el contexto de falla de un equipo o artículo manufacturado, $S(x)$ se refiere a la función de confiabilidad. Note que la función de supervivencia es una función no creciente con valor 1 en el origen y 0 en el infinito. Si X es una variable aleatoria continua, entonces $S(x)$ es una función continua estrictamente decreciente.

Cuando X una variable aleatoria continua, la función de supervivencia es el complemento de la función de distribución acumulada, es decir, $S(x) = 1 - F(x)$, donde $F(x) = P(X \leq x)$. También, la función de supervivencia es la integral de la función de densidad de probabilidad, $f(x)$, esto es:

$$S(x) = P(X > x) = \int_x^{\infty} f(t) dt$$

Es fácil verificar que:

$$f(x) = -\frac{dS(x)}{dx}$$

Note que $f(x)$ es una función no negativa con área bajo $f(x)$ igual a 1.

Función de riesgo

Otra función fundamental en el análisis de supervivencia es la *función de riesgo*, la cual especifica la tasa instantánea de muerte o falla al tiempo x dado que la unidad en estudio ha sobrevivido hasta el tiempo x . Dicho de otra manera, indica la probabilidad de que una unidad experimente el suceso entre los tiempos x y $x + \Delta x$, sabiendo que ha llegado vivo al tiempo x .

Esta función se conoce también como la tasa de falla condicional en estudios de confiabilidad, la fuerza de mortalidad en demografía, la función de intensidad en procesos estocásticos, la tasa de falla con la edad en epidemiología, la inversa de la razón de Mill en economía, o simplemente como la tasa de riesgo. La función es definida por

3.3. Algunos conceptos relacionados con el análisis de supervivencia

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

Si X es una variable aleatoria continua, entonces

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

En efecto, note que

$$\begin{aligned} h(x) &= \lim_{\Delta x \rightarrow 0} \frac{P[x \leq X < x + \Delta x | X \geq x]}{\Delta x} = \frac{P[x \leq X < x + \Delta x \cap X \geq x]}{\Delta x \cdot P[X \geq x]} \\ &= \frac{P[x \leq X < x + \Delta x]}{\Delta x \cdot P[X \geq x]} = \frac{F(x + \Delta x) - F(x)}{\Delta x \cdot S(x)} = \frac{F'(x)}{S(x)} = \frac{f(x)}{S(x)} \end{aligned}$$

y como $f(x) = -\frac{dS(x)}{dx}$, entonces

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

La función de riesgo puede tomar diversas formas, la única restricción es que sea no negativa, es decir, $h(x) \geq 0$. Por ejemplo, uno podría creer que la tasa de riesgo para la ocurrencia de un evento particular es creciente, decreciente, constante, en forma de bañera, en forma de joroba, o que posea alguna otra característica que describa el mecanismo de falla.

Modelos con tasa de riesgo creciente podrían originarse cuando hay un desgaste o envejecimiento naturales. Usos de funciones de riesgo decrecientes se encuentran cuando hay una prematura verosimilitud de falla, tal como en ciertos tipos de aparatos electrónicos o en pacientes que hayan experimentado ciertos tipos de trasplantes. Una función de riesgo en forma de bañera es, con frecuencia, apropiada en poblaciones seguidas de su nacimiento. La mayoría de los registros de mortalidad de poblaciones siguen este tipo de función de riesgo donde, durante un periodo temprano, las muertes resultan, principalmente, de enfermedades infantiles, después del cual la tasa de muerte se estabiliza, seguida por un incremento debido al proceso de envejecimiento natural. Similarmente, algunos equipos manufacturados podrían experimentar una falla temprana debido a partes defectuosas,

3.3. Algunos conceptos relacionados con el análisis de supervivencia

seguida por una tasa de riesgo constante, la cual, en la última fase de la vida del equipo, se incrementa. Finalmente, si la tasa de riesgo es creciente al principio y eventualmente comienza a declinar, entonces, la función de riesgo es denominada en forma de joroba. Este tipo de tasa de riesgo es con frecuencia usada en modelar supervivencia después de una cirugía exitosa donde hay un incremento inicial en el riesgo debido a infección, seguida por una decaída estable en el riesgo conforme el paciente se recupera. Distribuciones específicas que originan estos diferentes tipos de tasa de riesgo son encontradas en [Klein y Moeschberger \(1997\)](#) y [Lawless \(2003\)](#).

Función de riesgo acumulada

La función de riesgo acumulada, $H(x)$, se define de la siguiente manera:

$$\begin{aligned} H(x) &= \int_0^x h(t)dt = -\ln S(t) \Big|_0^x = -\ln[S(x)] + \ln[S(0)] \\ &= -\ln[S(x)] + \ln[1] = -\ln[S(x)] \end{aligned}$$

Esto implica que:

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

Función de vida media residual

La función tiempo de vida media residual $m(x)$ o $mrl(x)$ se define como:

$$m(x) = E[X - x | X > x],$$

y representa la esperanza de vida para una unidad que ya haya sobrevivido x unidades de tiempo. Note que la vida media $\mu = m(0)$ es el área total bajo la curva de supervivencia.

Para una variable aleatoria continua,

$$m(x) = \frac{\int_x^\infty (t-x)f(t)dt}{S(x)} = \frac{\int_x^\infty S(t)dt}{S(x)}$$
$$\mu = E(X) = \int_0^\infty tf(t)dt = \int_0^\infty S(t)dt$$

3.3.6. Estimación no paramétrica de la función de supervivencia y la función de riesgo acumulado

Como se ha señalado, en el análisis de supervivencia una observación consiste típicamente de un *tiempo de estudio* y un *indicador* que expresa si tal tiempo es un *tiempo exacto de evento* o un *tiempo censurado*. Se asume aquí que el tiempo de censura no está relacionado al tiempo de evento. Los métodos desarrollados a continuación son apropiados para censura por la derecha (Tipo I y Tipo II) y aleatoria discutidos en la sección 3.3.

Para permitir empates en los datos (tiempos iguales de supervivencia), suponga que los eventos ocurren en k ($k \leq n$) distintos tiempos $t_1 < t_2 < \dots < t_k$, y que al tiempo t_j hay d_j , $j = 1, \dots, k$, eventos de interés (tales como muertes o fallas). Sea Y_j el número de unidades en riesgo (es decir, vivos y no censurados) justo antes del tiempo t_j . Note que Y_j es un conteo del número de unidades con un tiempo de estudio de t_j o más, es decir, el número de unidades que están vivas en t_j o experimentan el evento de interés en t_j . Si un tiempo de censura y un tiempo exacto de evento son registrados como iguales, se adopta la convención que el tiempo de censura es infinitesimalmente mayor. Entonces, cualquier unidad con tiempo de censura registrada como igual a t_j es incluida en el conjunto de Y_j unidades en riesgo al tiempo t_j , así como las unidades que experimentan el evento en t_j .

La cantidad $\frac{d_j}{Y_j}$ provee un estimador de la probabilidad condicional de que un individuo que sobrevive justo antes del tiempo t_j , experimente el evento al tiempo t_j . Como se verá, está es la cantidad básica de la cual se construyen los estimadores de la función de supervivencia y de la función de riesgo acumulada.

3.3. Algunos conceptos relacionados con el análisis de supervivencia

Estimador Kaplan-Meier

El estimador de la función de supervivencia propuesto por [Kaplan y Meier \(1958\)](#) es definido como sigue para todos los valores de t en el rango donde hay datos

$$\hat{S}(t) = \begin{cases} 1 & \text{si } t < t_j \\ \prod_{t_j < t} [1 - \frac{d_j}{Y_j}] & \text{si } t_j \leq t \end{cases} \quad (3.11)$$

Este estimador es llamado también estimador Producto-Límite (PL), y tiene las siguientes características:

1. Es continuo por la derecha.
2. $\hat{S}(0) = 1$ si y solo si ninguna unidad experimenta el evento de interés al inicio del experimento (al tiempo cero).
3. En la práctica $\hat{S}(\infty) = 0$.
4. Si el mayor tiempo de vida observado t^* corresponde a una censura, el estimador (3.11) no debería ser usado con $t > t^*$.
5. Si no hay observaciones censuradas, el estimador se reduce a uno menos la función de distribución empírica.

Note que (3.11) es una función escalonada con saltos en los tiempos a los cuales ocurre el evento de interés (tiempo exacto). El tamaño de los saltos depende no solo del número de eventos observados a cada tiempo exacto t_j , sino también de las observaciones censuradas antes de t_j .

Además de proporcionar una medida eficiente de la estimación de la función de supervivencia para datos censurados por la derecha, el estimador PL también puede ser usado para estimar la función de riesgo acumulada. Así,

$$\hat{H}(t) = -\ln[\hat{S}(t)] \quad (3.12)$$

3.3. Algunos conceptos relacionados con el análisis de supervivencia

Kaplan y Meier (1958) presentan una estimación de la varianza del estimador PL dada por:

$$\hat{V}[\hat{S}(t)] = \hat{S}^2(t) \sum_{t_j \leq t} \frac{d_j}{Y_j(Y_j - d_j)} \quad (3.13)$$

Estimador Nelson-Aalen

Un estimador alternativo para la función de riesgo acumulada, el cual tiene mejor desempeño para tamaños de muestra pequeños que aquel basado en el estimador PL, fue en principio sugerido por Nelson (1972) en un contexto de confiabilidad. Después, adoptando una formulación de procesos de conteo, Aalen (1978) extendió su uso más allá de datos de supervivencia. Este estimador, el cual es referido como el estimador Nelson-Aalen (N-A) de la función de riesgo acumulada, es definido como sigue:

$$\tilde{H}(t) = \begin{cases} 0 & \text{si } t < t_j \\ \sum_{t_j \leq t} \frac{d_j}{Y_j} & \text{si } t_j \leq t \end{cases} \quad (3.14)$$

La gráfica de $\tilde{H}(t)$ da información útil sobre la forma de la función de riesgo; note, por ejemplo, que $H(t)$ es lineal si $h(t)$ es constante.

El estimador Nelson-Aalen tiene varios usos en el análisis de datos. Uno de ellos es en la selección entre modelos paramétricos que involucren tiempos de evento. Aquí, se grafica el estimador N-A tal que, si un modelo paramétrico dado ajusta a los datos, el resultado de la gráfica sea aproximadamente lineal. La herramienta clave es encontrar una función lineal de la función de riesgo acumulada que sea lineal en alguna función del tiempo t . Entonces, se estima $H(t)$ con el estimador de la ecuación (3.14) y se grafica en contra de t .

La varianza para $\tilde{H}(t)$ puede ser estimada por

$$\hat{V}[\tilde{H}(t)] = \sum_{t_j \leq t} \frac{d_j}{Y_j^2} \quad (3.15)$$

3.4. Otras distribuciones importantes en el análisis de supervivencia

Note que si T es una variable aleatoria continua entonces, suponiendo que no hay empates en los datos, para una muestra aleatoria de tamaño n censurada Tipo II donde se dispone solo de las primeras r observaciones, $T_1 < T_2 < \dots < T_r$, el estimador Nelson-Aalen evaluado en un tiempo específico x_j puede ser calculado como

$$\tilde{H}(x_j) = \sum_{i=1}^j \frac{1}{n - i + 1} \quad (3.16)$$

3.4. Otras distribuciones importantes en el análisis de supervivencia

3.4.1. Exponencial

La distribución exponencial es probablemente la más usada en trabajos estadísticos después de la distribución normal. Esta distribución se relaciona de manera importante con pruebas de vida, teoría de confiabilidad y teoría de procesos estocásticos, y está estrechamente relacionada con algunas otras distribuciones con aplicaciones estadísticas, por ejemplo, las distribuciones Gamma y Weibull.

La forma general de la distribución exponencial es

$$F(x) = 1 - e^{-\beta(x-\lambda)}, \quad x > \lambda \quad (3.17)$$

donde λ y β son parámetros, con β positivo. En adelante, se usará la notación $Exp(\lambda, \beta)$ para referir a esta distribución o a una muestra de ella.

La media de $Exp(\lambda, \beta)$ es $\lambda + \frac{1}{\beta}$ y la varianza es $\frac{1}{\beta^2}$. Se puede verificar mediante integración que la función de densidad correspondiente a (3.17) toma la siguiente forma:

$$f(x) = \beta e^{-\beta(x-\lambda)}, \quad x > \lambda \quad (3.18)$$

Las función de supervivencia y la función de riesgo acumulada para una variable aleatoria

3.4. Otras distribuciones importantes en el análisis de supervivencia

con distribución $Exp(\lambda, \beta)$ son, respectivamente,

$$S(x) = e^{-\beta(x-\lambda)}, \quad x > \lambda \quad (3.19)$$

$$H(x) = -\ln[S(x)] = -\beta\lambda + \beta x, \quad x > \lambda \quad (3.20)$$

Algunas propiedades de la distribución $Exp(\lambda, \beta)$ son las siguientes:

1. Si X_1, \dots, X_n es una muestra aleatoria de tamaño n de la distribución $Exp(\lambda, \beta)$, entonces Y_1, \dots, Y_n , donde $Y_i = X_i - \lambda$ para $i = 1, \dots, n$ es una muestra aleatoria de $Exp(0, \beta) = Exp(\beta)$.
2. Si $X_{(1)}, \dots, X_{(n)}$ es una muestra ordenada de $Exp(\lambda, \beta)$, entonces la muestra obtenida de $Y_{(i)} = X_{(i+1)} - X_{(1)}$, para $i = 1, \dots, n-1$, es una muestra ordenada de tamaño $n-1$ de una distribución $Exp(0, \beta)$. Este resultado puede ser aplicado para dar el siguiente.
3. Sea $X_{(1)}, \dots, X_{(n)}$ una muestra ordenada de $Exp(\lambda, \beta)$. Entonces, la muestra obtenida de $Y_{(i)} = X_{(i+r)} - X_{(r)}$, para $i = 1, \dots, n-r$, donde r es fijo, tal que $1 \leq r \leq n-1$, es una muestra ordenada de tamaño $n-r$ de $Exp(0, \beta)$.

Para la función de distribución (3.17), el caso con $\lambda = 0$ se denomina distribución exponencial de un parámetro, denotada por $Exp(\beta)$, y todas las demás funciones caracterizan a esta última distribución haciendo $\lambda = 0$. Es decir, para una distribución exponencial de un parámetro:

$$F(x) = 1 - e^{-\beta x}, \quad x > 0 \quad (3.21)$$

$$f(x) = \beta e^{-\beta x}, \quad x > 0 \quad (3.22)$$

$$S(x) = e^{-\beta x}, \quad x > 0 \quad (3.23)$$

$$H(x) = -\ln[S(x)] = \beta x, \quad x > 0 \quad (3.24)$$

3.4. Otras distribuciones importantes en el análisis de supervivencia

3.4.2. Weibull

La distribución Weibull es tal vez el modelo de distribución de tiempo de vida más usado. Es común la aplicación a tiempos de vida o durabilidad de unidades manufacturadas. También es usada en aplicaciones biológicas o médicas, por ejemplo en estudios del tiempo de ocurrencia de tumores en poblaciones humanas o en animales de laboratorio; [Lawless \(2003\)](#).

La función de densidad de la distribución Weibull tiene la siguiente forma:

$$f(x) = \frac{1}{\theta} \beta \left(\frac{x}{\theta}\right)^{\beta-1} \exp \left[- \left(\frac{x}{\theta}\right)^\beta \right], \quad x > 0; \quad \theta > 0, \quad \beta > 0 \quad (3.25)$$

Note que para el caso especial $\beta = 1$ la densidad (3.25) es $Exp(1/\theta)$. El r -ésimo momento, $E(X^r)$, de la distribución es $\theta^r \Gamma[1 + \frac{r}{\beta}]$, donde $\Gamma(k) = \int_0^\infty u^{k-1} e^{-u} du$ con $k > 0$.

Entonces su media y varianza, respectivamente, son $\theta \Gamma[1 + \frac{1}{\beta}]$ y $\theta^2 [\Gamma(1 + \frac{2}{\beta}) - \Gamma(1 + \frac{1}{\beta})^2]$.

Las siguientes son la función de distribución, función de supervivencia y función de riesgo acumulada que también caracterizan a la distribución Weibull:

$$F(x) = 1 - \exp \left[- \left(\frac{x}{\theta}\right)^\beta \right] \quad (3.26)$$

$$S(x) = \exp \left[- \left(\frac{x}{\theta}\right)^\beta \right] \quad (3.27)$$

$$H(x) = \left(\frac{x}{\theta}\right)^\beta \quad (3.28)$$

Se usará la notación $Weib(\beta, \theta)$ para indicar que una v.a. X tiene la función de distribución (3.26). La forma de la distribución Weibull depende solo de β , el parámetro de forma, mientras que el efecto de distintos valores de θ (parámetro de escala) es un cambio de escala en la gráfica, y no de su forma básica.

Valores típicos de β varían de aplicación a aplicación, sin embargo, en muchas situaciones son apropiadas las distribuciones con β en el rango 0.5 a 3.0; [Lawless \(2003\)](#).

3.4.3. Log-normal

La distribución log-normal ha sido usada como un modelo en diversas aplicaciones en ingeniería, medicina y otras áreas. Se dice que el tiempo de vida X es log-normalmente distribuido si $Y = \log X$ es normalmente distribuido con media μ , varianza σ^2 y función de densidad dada por

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{y - \mu}{\sigma}\right)^2\right], \quad -\infty < y < \infty \quad (3.29)$$

Se puede verificar de (3.29) que la función de densidad para $X = \exp Y$ es

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left[-\frac{1}{2}\left(\frac{\log x - \mu}{\sigma}\right)^2\right], \quad x > 0 \quad (3.30)$$

La media y varianza de la distribución log-normal son, respectivamente, $\exp[\mu + \frac{\sigma^2}{2}]$ y $[\exp \sigma^2 - 1][\exp(2\mu + \sigma^2)]$.

La función de supervivencia de la distribución log-normal involucra a la función de distribución normal estándar $\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp[-\frac{u^2}{2}] du$

Se puede verificar que la función de supervivencia es

$$S(x) = 1 - \Phi\left(\frac{\log x - \mu}{\sigma}\right) \quad (3.31)$$

3.4.4. Gamma

La distribución gamma tiene la función de densidad de la forma

$$f(x) = \frac{\theta^\beta x^{\beta-1} \exp(-\theta x)}{\Gamma(\beta)} \quad x > 0 \quad (3.32)$$

donde $\beta > 0$ y $\theta > 0$ son parámetros. Esta distribución, como la distribución Weibull,

3.4. Otras distribuciones importantes en el análisis de supervivencia

incluyen la exponencial como un caso especial ($\beta = 1$). La correspondiente función de distribución acumulada no tiene una forma cerrada, sin embargo se ha tabulado extensamente. La función de supervivencia involucra la función gamma incompleta dada por $I(\beta, x) = \frac{1}{\Gamma(\beta)} \int_0^x u^{\beta-1} e^{-u} du$. Integrando (3.32) se encuentra que la función de supervivencia es

$$S(x) = 1 - I(\beta, \theta x) \tag{3.33}$$

La media de esta distribución es β/θ y su varianza β/θ^2 . Para indicar que una variable aleatoria tiene función de densidad gamma de la forma (3.32) se usará la notación $Ga(\beta, \theta)$.

Capítulo 4

Prueba de Bondad de Ajuste Propuesta

No obstante del gran uso que se da a la distribución Pareto (ver apartado 3.2.3), hay una escasez en la literatura de propuestas de bondad de ajuste para esta distribución, y menos aún si se trata de información incompleta (censurada). Zeinab (2007) destacó la importancia de la relación que existe entre la distribución Pareto y la distribución exponencial para la construcción de una prueba y propuso un procedimiento gráfico basado en la función de riesgo, usando el estimador no paramétrico de Kimball. Asimismo, presentó dos procedimientos paramétricos, uno basado en la prueba clásica χ^2 de Pearson y otro en la estadística de razón de verosimilitud. Las pruebas fueron desarrolladas para datos agrupados provenientes de una distribución $Par(\beta, \theta)$ bajo censura Tipo I y Tipo II.

En este capítulo se presenta una prueba de bondad de ajuste para verificar la validez de que una muestra aleatoria de observaciones se ajusta a una distribución de la forma Pareto, definida como en (3.6). La prueba involucra ciertas transformaciones de los datos originales y está basada en el coeficiente de correlación de una muestra ordenada de $Exp(\beta)$ y su correspondiente función de riesgo acumulada (FRA).

Los datos originales son transformados primero a $Exp(\lambda, \beta)$ y después a $Exp(\beta)$ (ver secciones 3.2.2 y 3.4.1). La razón por la cual se realiza la doble transformación es debido a que la variable aleatoria involucrada con la FRA es no negativa, mientras que la función de densidad exponencial de dos parámetros puede tomar valores negativos, según sea el

4.1. Derivación de la Prueba

valor del parámetro de localidad; en cambio, una densidad exponencial de un parámetro toma siempre valores positivos, definida como en (3.22).

4.1. Derivación de la Prueba

Sea $T_{(1)}, \dots, T_{(r)}$ una muestra ordenada censurada Tipo II, tal como se discutió en la sección 3.3, proveniente de una población con distribución $Par(\beta, \theta)$. Para construir el estadístico de prueba realice las siguientes transformaciones:

1. $W_{(i)} = \ln(T_{(i)})$ con $i = 1, \dots, r$.
2. $Z_{(i)} = W_{(i+1)} - W_{(1)}$ con $i = 1, \dots, r - 1$

Note que las $W_{(i)}$ s se distribuyen como las estadísticas de orden de una muestra aleatoria de tamaño r de una distribución $Exp(\delta, \beta)$, donde $\delta = \ln \theta$, (ver sección 3.2.2); y que $Z_{(i)}$ son un conjunto de $(r - 1)$ variables distribuidas como las estadísticas de orden de una muestra aleatoria de tamaño $r - 1$ de una distribución $Exp(\beta)$, ver sección 3.4.1.

La FRA de una variable aleatoria Z con función de distribución $Exp(\beta)$ es

$$H(z) = -\ln[S(z)] = -\ln[\exp(-\beta z)] = \beta z, \quad (4.1)$$

y un buen estimador de ésta es el estimador Nelson-Aalen; ver expresión (3.14). De la función (4.1) es evidente que existe una relación lineal entre $H(z)$ y z .

Por tratarse de variables aleatorias continuas, se supondrá que no hay empates en los datos. Entonces, permita que $Z_{(1)} < Z_{(2)} < \dots < Z_{(r-1)}$ sean los valores de las $(r - 1)$ observaciones transformadas. Sea R_{N-A} el coeficiente de correlación muestral entre el estimador Nelson-Aalen de la FRA dada en (4.1) y Z . Calcule R_{N-A} como en (3.4) permitiendo que $Z_{(i)} = U_i$ y $\tilde{H}(Z_{(i)}) = V_i$, donde $\tilde{H}(Z_{(i)})$ es el estimador Nelson-Aalen de (4.1), esto es:

4.2. Prueba de Bondad de Ajuste

$$R_{N-A} = \frac{\sum_{i=1}^{r-1} \{Z_{(i)} - \bar{Z}\} \{\tilde{H}(Z_{(i)}) - \tilde{\bar{H}}\}}{\sqrt{\sum_{i=1}^{r-1} \{Z_{(i)} - \bar{Z}\}^2} \sqrt{\sum_{i=1}^{r-1} \{\tilde{H}(Z_{(i)}) - \tilde{\bar{H}}\}^2}} \quad (4.2)$$

donde $\bar{Z} = (r-1)^{-1} \sum_{i=1}^{r-1} Z_{(i)}$ y $\tilde{\bar{H}} = (r-1)^{-1} \sum_{i=1}^{r-1} \tilde{H}(Z_{(i)})$. Note que las sumas son solo sobre los valores i disponibles, es decir, los no censurados. Si la muestra aleatoria proviene de la distribución Pareto, se espera que R_{N-A} sea cercano a 1.

4.2. Prueba de Bondad de Ajuste

Sea T_1, T_2, \dots, T_n una muestra aleatoria i.i.d. de tamaño n con función de distribución común F_T . Bajo censura tipo II se dispone solo de las primeras r observaciones, $T_{(1)} < T_{(2)} < \dots < T_{(r)}$. Sea \mathfrak{S}_P la familia de distribuciones Pareto con parámetros de forma y escala desconocidos. Se desea probar la siguiente hipótesis nula:

$$H_0 : F_T \in \mathfrak{S}_P \quad (4.3)$$

Realice las transformaciones $W_{(i)} = \ln(T_{(i)})$, $i = 1, \dots, r$, y $Z_{(i)} = W_{(i+1)} - W_{(1)}$, $i = 1, \dots, r-1$.

Prueba: Calcule el coeficiente de correlación R_{N-A} entre el estimador Nelson-Aalen de la función (4.1) y Z , expresado en (4.2). Rechazar H_0 de (4.3) si $R_{N-A} < K_\alpha$, y no rechazar H_0 si $R_{N-A} \geq K_\alpha$, donde K_α es la constante crítica que se obtiene de tal forma que

$$P[\text{Rechazar } H_0 | H_0] = P[R_{N-A} < K_\alpha | H_0] \leq \alpha$$

4.3. Distribución del estadístico de prueba

Bajo la hipótesis nula, la distribución del estadístico R_{N-A} no depende de β , ni de ningún otro parámetro desconocido. Note que bajo H_0 las $Z_{(i)}$, $i = 1, \dots, r-1$, son distribuidas como las estadísticas de orden de una $Exp(\beta)$. Sea $V_{(i)} = \beta Z_{(i)}$, $i = 1, \dots, r-1$. Entonces, las $V_{(i)}$'s son distribuidas como las estadísticas de orden de una distribución esponencial estándar ($Exp(1)$). Entonces, R_{N-A} puede ser escrito en términos de V como:

$$\begin{aligned}
 R_{N-A} &= \frac{\sum_{i=1}^{r-1} \left(\frac{1}{\beta} V_{(i)} - \frac{1}{\beta} \bar{V} \right) \left(\tilde{H} \left(\frac{V_{(i)}}{\beta} \right) - \tilde{H} \right)}{\sqrt{\sum_{i=1}^{r-1} \left(\frac{1}{\beta} V_{(i)} - \frac{1}{\beta} \bar{V} \right)^2} \sqrt{\sum_{i=1}^{r-1} \left(\tilde{H} \left(\frac{V_{(i)}}{\beta} \right) - \tilde{H} \right)^2}} \\
 &= \frac{\frac{1}{\beta} \sum_{i=1}^{r-1} (V_{(i)} - \bar{V}) \left(\tilde{H} \left(\frac{V_{(i)}}{\beta} \right) - \tilde{H} \right)}{\frac{1}{\beta} \sqrt{\sum_{i=1}^{r-1} (V_{(i)} - \bar{V})^2} \sqrt{\sum_{i=1}^{r-1} \left(\tilde{H} \left(\frac{V_{(i)}}{\beta} \right) - \tilde{H} \right)^2}} \\
 &= \frac{\sum_{i=1}^{r-1} (V_{(i)} - \bar{V}) \left(\tilde{H} \left(\frac{V_{(i)}}{\beta} \right) - \tilde{H} \right)}{\sqrt{\sum_{i=1}^{r-1} (V_{(i)} - \bar{V})^2} \sqrt{\sum_{i=1}^{r-1} \left(\tilde{H} \left(\frac{V_{(i)}}{\beta} \right) - \tilde{H} \right)^2}}
 \end{aligned}$$

De la ecuación (3.16) se tiene que $\tilde{H} \left(\frac{V_{(i)}}{\beta} \right)$ no depende de β . Por tanto, el estadístico R_{N-A} no depende de β ni de cualquier otro parámetro desconocido.

Los valores críticos K_α fueron aproximados con base en los percentiles muestrales de la distribución empírica de R_{N-A} bajo la hipótesis nula. La estimación de los percentiles muestrales se realizó mediante simulación por el Método Monte Carlo en el lenguaje R. Para ello, se generaron muestras aleatorias de tamaño 10(10)100 y 500 de una distribución $Par(3, 0.5)$, censurando el 10, 20(20)80 % de las muestras¹. Para cada muestra censurada se calculó el valor del coeficiente R_{N-A} con los datos transformados en $M = 10,000$ ocasiones y con esta información se obtuvieron los percentiles empíricos 0.005, 0.01, 0.02, 0.05, 0.10 y 0.15. Las tablas y el código en R se muestran en la Tabla 1 del Apéndice, . El algoritmo empleado es el siguiente:

¹No se realiza censura de 80 % para $n = 10$

4.3. Distribución del estadístico de prueba

Algoritmo 4.1. Para obtener los percentiles de la distribución empírica del coeficiente de correlación:

1. Fijar los valores n , $r < n$, β , θ .
2. Generar una muestra aleatoria, X_1, \dots, X_n , de tamaño n de $Par(\beta, \theta)$.
3. Ordenar los datos de menor a mayor, $X_{(1)}, \dots, X_{(n)}$.
4. Extraer de la muestra ordenada las primeras r observaciones. El porcentaje de censura será $C = [1 - (r/n)] * 100$.
5. Hacer $W_{(i)} = \ln X_{(i)}$, $i = 1, \dots, r$.
6. Hacer $Z_{(i)} = W_{(i+1)} - W_{(1)}$, $i = 1, \dots, r - 1$.
7. Con los valores $Z_{(i)}$, calcular el estimador Nelson - Aalen de la FRA con la ecuación (3.14).
8. Obtener el coeficiente de correlación R_{N-A} entre el estimador del paso 7 y los tiempos $Z_{(i)}$ mediante la ecuación (4.2).
9. Repetir los pasos 2 a 8 M veces.
10. Con la información del paso 9, obtener los percentiles de la distribución empírica del coeficiente de correlación muestral.

De acuerdo con los resultados obtenidos mediante la simulación, se verifica que la distribución del estadístico R_{N-A} es invariante ante distintos valores de los parámetros, incluso para tamaños de muestra pequeños. Sin embargo, el porcentaje de censura influye en la forma de la distribución. Como ejemplo, la figura 4.1 muestra la distribución empírica de la estadística de prueba con cinco distintas combinaciones de valores de los parámetros a tres niveles de censura Tipo II. Para esta figura se simularon $M = 10,000$ repeticiones de muestras tamaño 50. Como se puede observar, con niveles de censura pequeños la distribución está más concentrada cerca del uno; con alto porcentaje de censura la distribución es más dispersa.

Por lo anterior, fue necesario obtener tablas de los cuantiles empíricos del coeficiente de correlación para distintos porcentajes de censura.

4.4. Implementación de la prueba

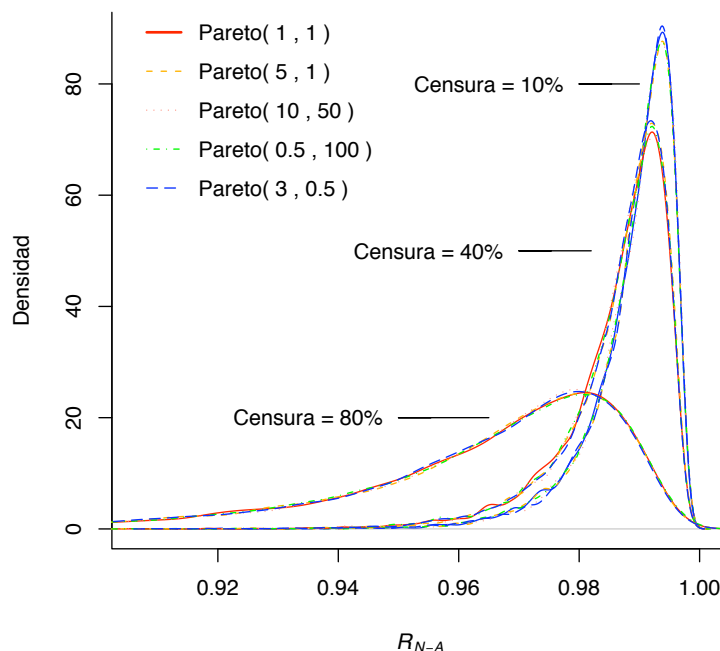


Figura 4.1: Distribución de la estadística de prueba con distintos valores de parámetros y tres niveles de censura Tipo II

4.4. Implementación de la prueba

Bajo censura Tipo II, en un experimento con n unidades puestas a prueba se dispondrá solo de las primeras r observaciones. La hipótesis es que la muestra aleatoria viene de una distribución $Par(\beta, \theta)$, y el objetivo será tener un criterio que permita rechazar o no rechazar tal hipótesis.

Sea $T_{(1)} < \dots < T_{(r)}$ la muestra observada (es decir, los datos no censurados) ordenada, con censura Tipo II. Denote por \mathfrak{S}_P a la familia de distribución Pareto con parámetros de forma y escala desconocidos.

Para probar $H_0 : F_T \in \mathfrak{S}_P$ realice las transformaciones $W_{(i)} = \ln(T_{(i)})$ y $Z_{(i)} = W_{(i+1)} - W_{(1)}$. Calcule el valor del coeficiente de correlación muestral R_{N-A} entre el estimador Nelson-Aalen de la función (4.1) y los valores observados transformados $Z_{(i)}$, para $i = 1, 2, \dots, r - 1$, con la ecuación (4.2).

Regla de decisión: Rechazar H_0 con un nivel de significancia α si la estadística de prueba R_{N-A} es menor que el cuantil $K_{\alpha, C}$ de la tabla 1, donde $C = [1 - (r/n)] * 100$ es

4.5. Tamaño de la prueba

el porcentaje de censura.

Note que la prueba tiene las siguientes características:

1. No requiere de la estimación de parámetros, es decir, es una prueba no paramétrica.
2. El nivel de significancia depende del tamaño de muestra, como se puede observar de las tablas de cuantiles empíricos.

4.5. Tamaño de la prueba

Para determinar el tamaño de la prueba, se calculó el coeficiente de correlación R_{NA} en $M = 10,000$ ocasiones para muestras aleatorias $Par(\beta, \theta)$ con distintas combinaciones de valores de los parámetros, $(3,0.5)$, $(5,1)$ y $(0.5, 100)$, permitiendo censura de 10, 20(20)80%¹. El valor del coeficiente de correlación de cada repetición se comparó con el cuantil correspondiente a los niveles de significancia $\alpha = 0.005, 0.01, 0.02, 0.05, 0.10$ y 0.15 , presentados en la tabla 1 del Apéndice. La proporción del número de veces que las 10,000 réplicas rechazaban la hipótesis nula es una estimación del tamaño de error tipo I, es decir, de la probabilidad de rechazar la hipótesis nula cuando esta es cierta. El algoritmo usado para la simulación es el siguiente:

Algoritmo 4.2. Para estimar el tamaño de la prueba.

1. Fijar los valores $n, r < n, \beta, \theta$.
2. Hacer $Suma = 0$.
3. Realizar los pasos 2 a 8 del Algoritmo 4.1
4. Comparar el valor obtenido de R_{NA} con el cuantil $K_{\alpha,C}$ de la Tabla 1. Si $R_{NA} < K_{\alpha,C}$, $Suma = Suma + 1$
5. Repetir los pasos 3 y 4 M veces.
6. Calcular la proporción de veces que las M réplicas rechazan H_0 , es decir, $Suma/M$. Esta es la estimación del tamaño de la prueba.

4.6. Potencia de la prueba ante algunas alternativas

Los resultados obtenidos de la simulación se presentan en las tablas 2, 3 y 4 del Apéndice. La prueba resulta ser consistente de tamaño $= \alpha$, incluso para niveles de censura muy severa. Las siguientes figuras ejemplifican este resultado; se ha graficado el tamaño de la prueba con censura ligera (20 %) y censura muy alta (80 %) para diversos tamaños de muestra (n) en las tres distintas distribuciones Pareto simuladas; ver figura 4.2. Note que el tamaño de la prueba está muy cercano al nivel de significancia para cualquier n . Las variaciones que se presentan son tan pequeñas que se pueden suponer debidas a la simulación, por lo que se considera que se trata de una prueba de tamaño $= \alpha$ que se mantiene igual ante distintos n .

4.6. Potencia de la prueba ante algunas alternativas

Con el fin de estudiar el comportamiento de la potencia de la prueba del coeficiente de correlación R_{NA} ante distintas distribuciones alternativas a la Pareto, se consideraron las siguientes:

1. $Weib(3, 1)$ y $(0.5, 1)$
2. $LNorm(1, 1)$ y $(5, 3)$
3. $Ga(3, 1)$ y $(0.8, 1)$

Definidas como en la sección 3.4. Para cada una de ellas se realizó, como antes, simulación Monte Carlo de muestras de tamaño $10(10)100,500$ con $M = 10,000$ réplicas, permitiendo los niveles de censura ya usados. Para estimar la potencia, se calculó en cada repetición el coeficiente R_{NA} y se obtuvo la proporción del número de veces que las 10,000 simulaciones rechazaban la hipótesis nula de forma correcta, siendo cierta la hipótesis alternativa. El algoritmo empleado es el siguiente:

Algoritmo 4.3. Para estimar el potencia de la prueba.

1. Definir los valores n y r .
2. Hacer $Suma = 0$
3. Generar una muestra aleatoria de tamaño n de la distribución alternativa.

4.6. Potencia de la prueba ante algunas alternativas

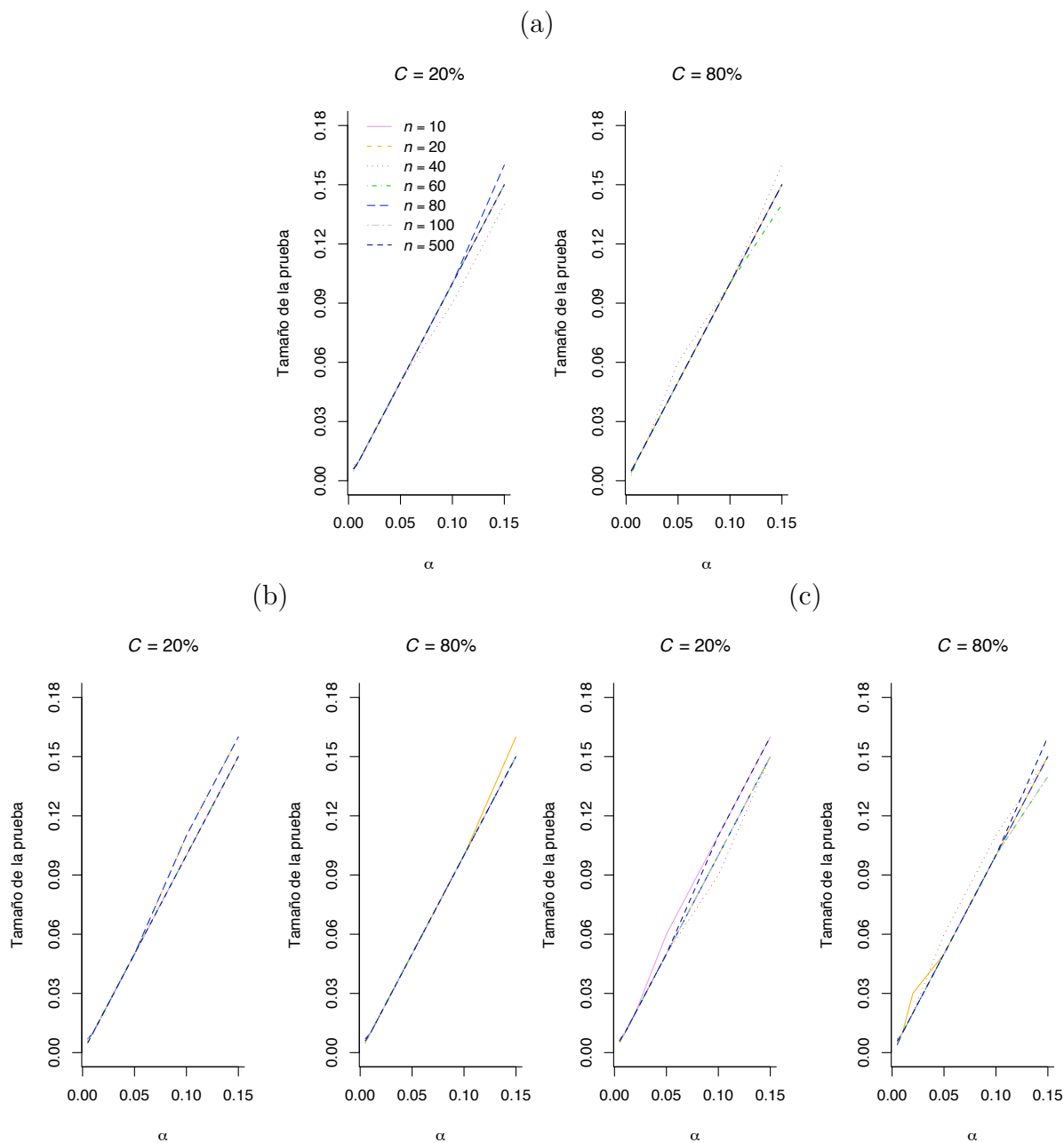


Figura 4.2: Tamaño de prueba para distribuciones Pareto con niveles de censura del 20 y 80% a distintos n , (a)= $Par(3, 0.5)$, (b)= $Par(5, 1)$, (c)= $Par(0.5, 100)$, α =nivel de significancia, C = Censura.

4.6. Potencia de la prueba ante algunas alternativas

4. Realizar los pasos 3 a 8 del Algoritmo 4.1
5. Comparar el valor obtenido de R_{NA} con el cuantil $K_{\alpha,C}$ de la tabla 1. Si $R_{NA} < K_{\alpha,C}$, $Suma = Suma + 1$
6. Repetir los pasos 3 a 5 M veces.
7. Calcular $Suma/M$. Esta es la estimación de la potencia de la prueba.

Los resultados obtenidos se encuentran en las tablas 5, 6 y 7 del Apéndice. En general, para todas las alternativas se presenta una tendencia de incremento en la potencia conforme aumenta el tamaño de muestra, así como un descenso de la misma al elevar el nivel de censura; ver figuras 4.3 y 4.4.

Los resultados para las dos distribuciones Weibull son muy similares. Con un nivel de significancia $\alpha = 0.05$, si se censura solo el 10 % de una muestra de tamaño 30 la potencia de la prueba es superior a 0.95. Permitiendo mayor censura, como por ejemplo del 40 %, con tamaños de muestra $n = 50$ y más se obtienen potencias superiores a 0.94. Sin embargo, con censura extrema, en este caso del 80 %, una muestra de tamaño 100 da una potencia alrededor de 0.68.

En el caso de las distribuciones gamma, con $n = 50$ o más se tiene una potencia superior a 0.9 permitiendo censura de hasta 40 %. Para censura = 60 %, con $n = 80$ la potencia es ligeramente mayor a 0.90; y si la censura es del 80 %, un tamaño de muestra 100 dará una potencia a penas superior a 0.61, todos a un nivel de significancia $\alpha = 0.05$.

Para las distribuciones log-normal el comportamiento de la potencia es distinto. Considerando el mismo α que en los casos anteriores, con censura del 10 % y 20 % se obtienen potencias mayores a 0.91 a partir de $n = 50$. Al censurar el 60 %, es necesario tamaños de muestra de 100 o más para asegurar que la potencia sea superior a 0.90; y para la censura más severa considerada (80 %), aun con $n = 100$ la potencia es solo por encima de 0.51.

4.6. Potencia de la prueba ante algunas alternativas

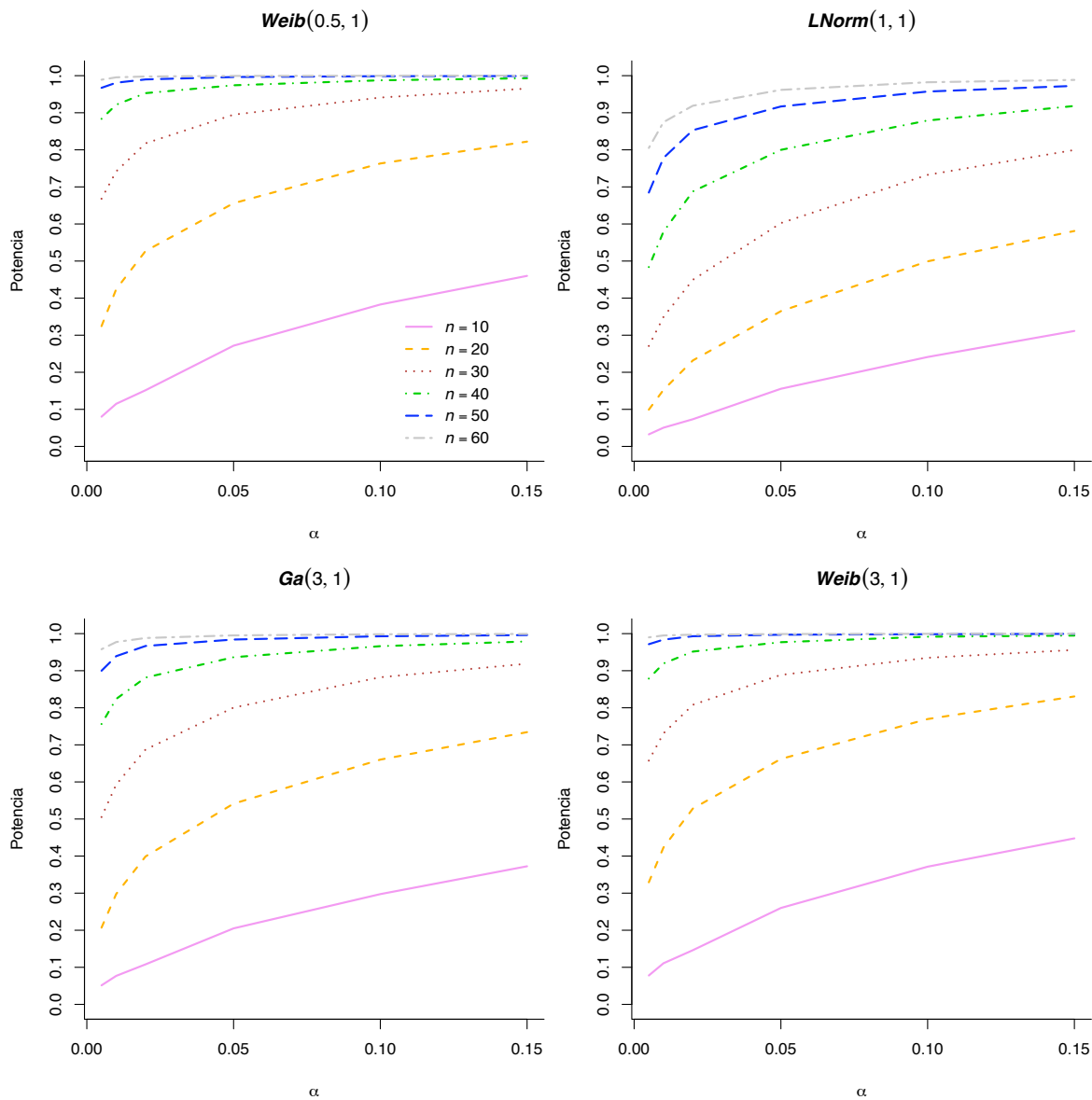


Figura 4.3: Potencia de la prueba basada en R_{N-A} para algunas alternativas, censura=20%, $M = 10,000$ réplicas de tamaño n , $\alpha =$ nivel de significancia.

4.6. Potencia de la prueba ante algunas alternativas

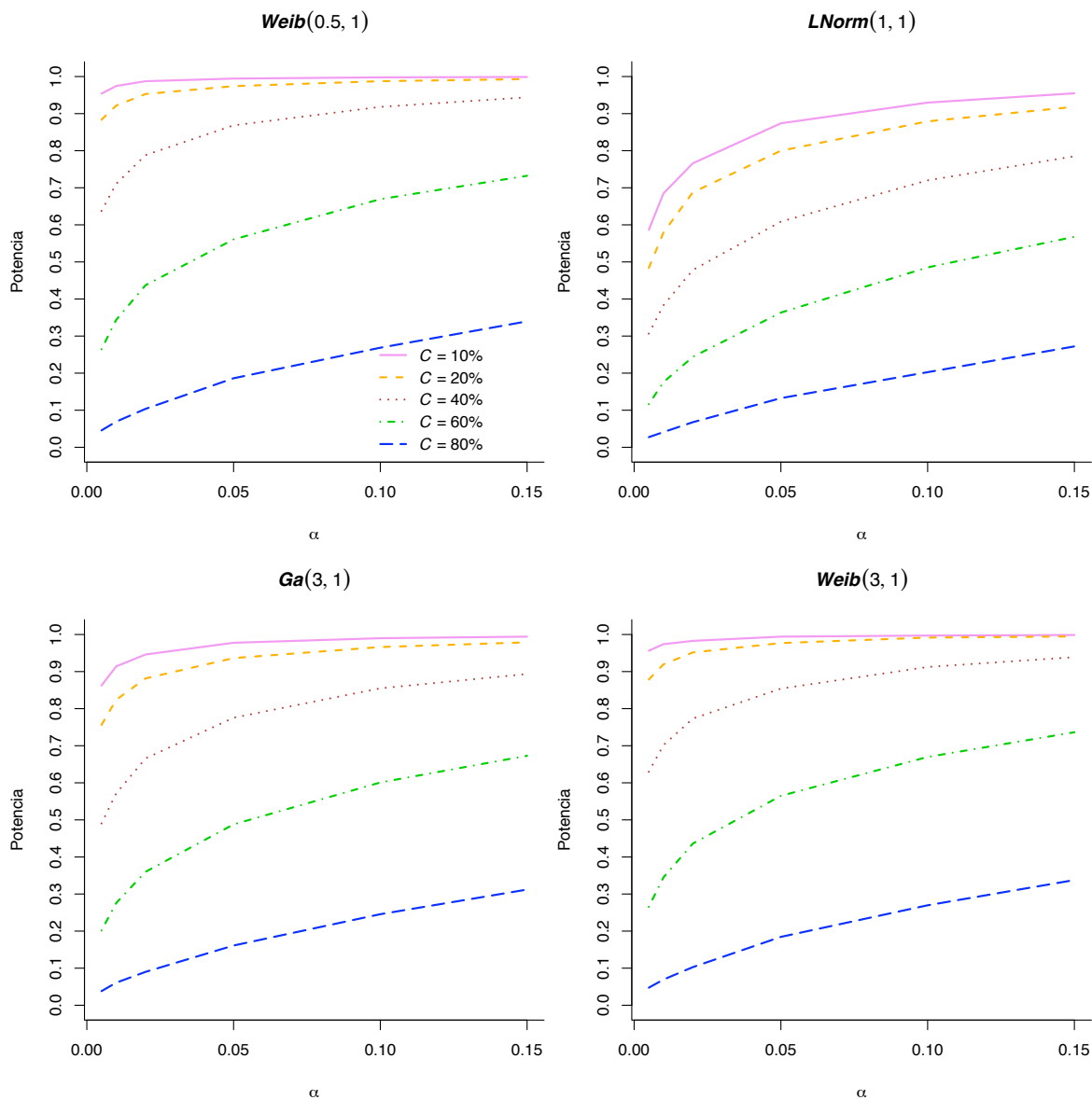


Figura 4.4: Potencia de la prueba para algunas alternativas con distintos niveles de censura, $n = 40$, $C =$ Censura $M = 10,000$ réplicas, $\alpha =$ nivel de significancia.

Capítulo 5

Estudio Comparativo de Potencia

El estudio comparativo de potencia se realiza contra tres pruebas de exponencialidad aplicables a datos no agrupados con censura Tipo II. La comparación resulta pertinente ya que la prueba basada en el coeficiente de correlación R_{NA} involucra transformaciones tales que la muestra resultante proviene de una población con función de distribución exponencial de un parámetro.

5.1. Cramér-von Mises y Anderson-Darling

[Stephens \(1986b\)](#) propuso adaptaciones de las ya conocidas estadísticas Crámer-von Mises y Anderson-Darling para probar exponencialidad con datos censurados Tipo II, y dio también las tablas de valores críticos. Las pruebas consisten en lo siguiente:

Para una variable aleatoria X con función de distribución exponencial dada en [\(3.21\)](#), bajo cenura Tipo II el parámetro β es estimado por:

$$\hat{\beta} = r / \left\{ \sum_{i=1}^r X_{(i)} + (n - r) X_{(r)} \right\} \quad (5.1)$$

donde $X_{(1)}, \dots, X_{(r)}$ son estadísticas de orden, con $X_{(r)}$ la estadística de orden mayor.

Ahora, la prueba se realiza como sigue:

5.1. Cramér-von Mises y Anderson-Darling

1. Calcule $Z_{(i)} = 1 - \exp(-X_{(i)}\hat{\beta})$, $i = 1, \dots, r$.
2. Use $Z_{(i)}$ en las fórmulas 3.1 y 3.2 del apartado 3.1.1 para calcular las estadísticas $W_{r,n}^2$ y $A_{r,n}^2$.
3. Refiera $W_{r,n}^2$ o $A_{r,n}^2$ a los puntos críticos dados en la tabla 9 del Apéndice. La prueba rechaza para valores mayores que el cuantil.

La potencia de las pruebas a comparar se estimó para las distribuciones alternativas $Weib(0.5, 1)$ y $Ga(0.8, 1)$, con función de riesgo monótona decreciente; $Weib(3, 1)$ y $Ga(3, 1)$, con función de riesgo monótona creciente; $LNorm(5, 3)$ y $LNorm(1, 1)$ con función de riesgo no monótona, parametrizadas como se hizo en la sección 3.4. Para cada una se generaron muestras de tamaño 10, 20(20)100 y se sometieron a la prueba con $M = 10,000$ réplicas a niveles de significancia $\alpha = 0.01, 0.05$ y 0.10 . La proporción del número de veces que las 10,000 repeticiones rechazaron la hipótesis nula es una estimación de la potencia. El algoritmo empleado es el siguiente:

Algoritmo 5.1. Para estimar la potencia de las pruebas $W_{r,n}^2$ y $A_{r,n}^2$ con censura Tipo II.

1. Definir $n, r, C = [1 - (r/n)] * 100$
2. $Suma = 0$
3. Generar una muestra aleatoria de tamaño n de la distribución alternativa.
4. Ordenar los datos de menor a mayor.
5. Extraer de la muestra las primeras r observaciones no censuradas.
6. Estimar el parámetro de escala de una distribución exponencial con la fórmula (5.1).
7. Calcular $Z_{(i)} = 1 - \exp\{-X_{(i)}\hat{\beta}\}$, $i = 1, \dots, r$.
8. Usar las $Z_{(i)}$ en las fórmulas 3.1 y 3.2 del apartado 3.1.1 para calcular las estadísticas $W_{r,n}^2$ y $A_{r,n}^2$.
9. Comparar el valor de $W_{r,n}^2$ y $A_{r,n}^2$ con el cuantil $K_{\alpha,C}$ de la Tabla 9. Si el valor de la estadística es mayor que el cuantil, entonces $Suma = Suma + 1$

5.2. Información de Kullback-Leibler

10. Repetir los pasos anteriores M veces y calcular $Suma/M$. La proporción de rechazos es una estimación de la potencia de la prueba.

Los resultados de la comparación se resumen en la figura 5.1. Para mayores detalles, las potencias estimadas de las estadísticas $W_{r,n}^2$ y $A_{r,n}^2$ se encuentran en la tabla 8 del Apéndice.

De acuerdo con los resultados, es difícil tomar una decisión sobre qué prueba resulta ser la mejor. Cuando la alternativa tiene función de riesgo monótona decreciente, la estadística de prueba propuesta tiene mejor comportamiento. Para la distribución $LNorm(1, 1)$, las pruebas basadas en las estadísticas $W_{r,n}^2$ y $A_{r,n}^2$ tienen potencias bajas, más aun con poca censura. Esto puede deberse a que la distribución log-normal tiene la propiedad de equivalencia aproximada a una distribución exponencial negativa conforme σ se acerca a la unidad.

Para las distribuciones con función de riesgo monótona creciente consideradas, las estadísticas $W_{r,n}^2$ y $A_{r,n}^2$ tienen potencias muy cercanas y superiores a la estadística R_{N-A} . Para la distribución alternativa $LNorm(5, 3)$ con función de riesgo no monótona, la estadística $A_{r,n}^2$ tuvo potencias más altas que el resto.

5.2. Información de Kullback-Leibler

[Sangun \(2005\)](#) propuso una prueba de exponencialidad con censura Tipo II, basada en la información de Kullback-Leibler. El estadístico de prueba es:

$$T(m, n, r) = -H(m, n, r) + \frac{r}{n} \left\{ \log \left[\frac{1}{r} \left(\sum_{i=1}^r X_{(i:n)} + (n-r)X_{(r:n)} \right) \right] + 1 \right\} \quad (5.2)$$

donde

n = tamaño de muestra.

m = tamaño de ventana, es un entero positivo menor que $n/2$.

5.2. Información de Kullback-Leibler

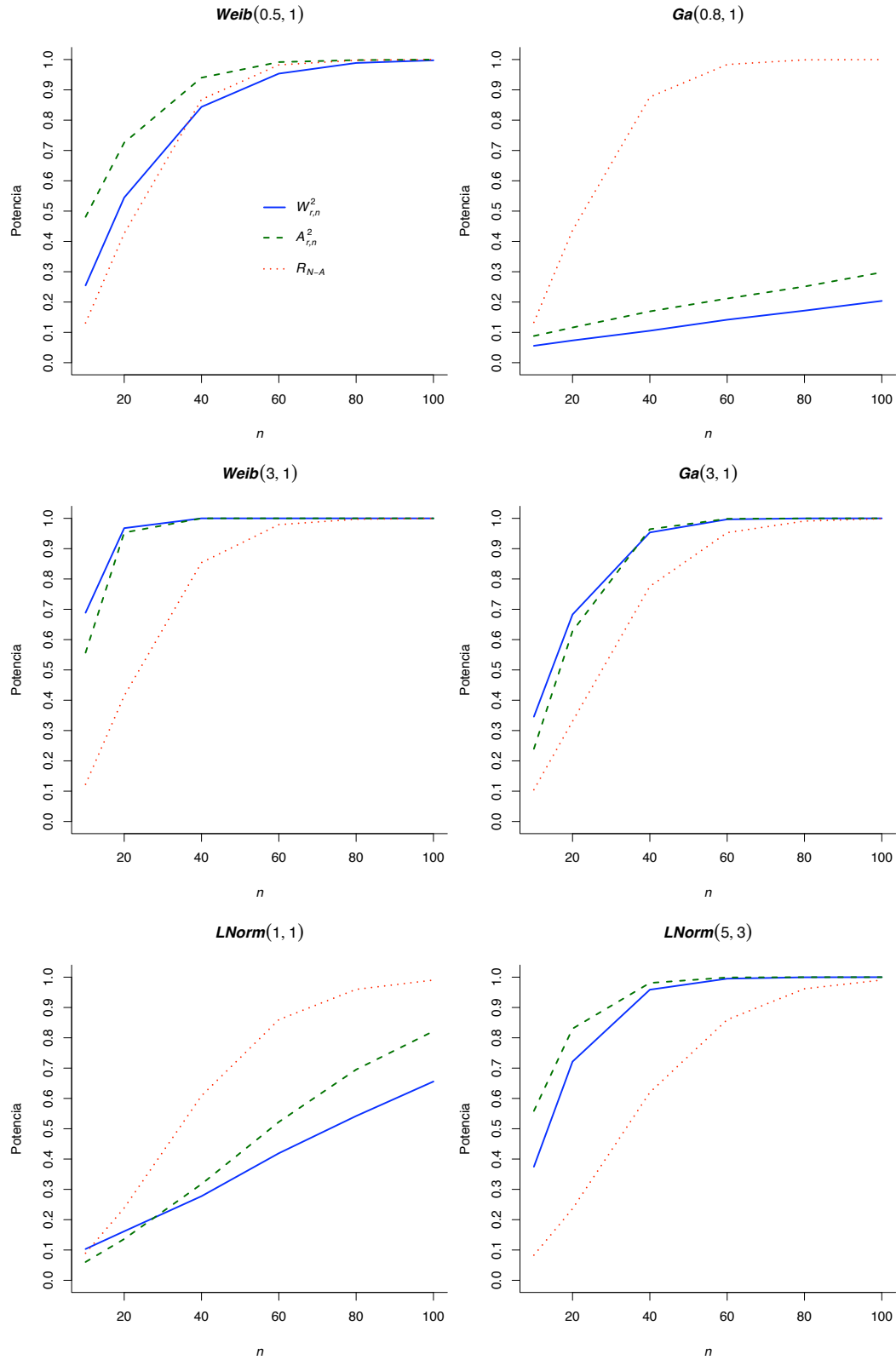


Figura 5.1: Potencia de las pruebas $W^2_{r,n}$ y $A^2_{r,n}$ para algunas alternativas, censura = 40%, $\alpha = 0.05$, n = tamaño de muestra, $M = 10,000$ réplicas.

5.2. Información de Kullback-Leibler

r = número de observaciones no censuradas.

$X_{(i)}$ = la i -ésima estadística de orden de una muestra aleatoria i.i.d. de tamaño n de $F(x)$, with $F(x) \sim Exp(1/\beta)$

$$H(m, n, r) = \frac{1}{n} \sum_{i=1}^r \log \left\{ \frac{n}{2m} (x_{(i+m)} - x_{(i-m)}) \right\} - \left(1 - \frac{r}{n} \right) \log \left(1 - \frac{r}{n} \right) \quad (5.3)$$

donde $x_{(i:n)} = x_{(1:n)}$ si $i < 1$ y $x_{(i:n)} = x_{(r:n)}$ si $i > r$. Bajo la hipótesis nula $T(m, n, r)$ debería ser cercano a 0. Los valores críticos de $T(m, n, r)$ fueron estimados usando muestras Monte Carlo; ver tabla 5.3.

El tamaño de ventana m depende de n , r y el nivel de significancia α . Sin embargo, de los resultados de simulación, Sangun (2005) encontró que el tamaño de ventana óptimo m varía mucho con respecto a r en lugar de a n , y no varía mucho con respecto a α si $\alpha \leq 0.1$. En vista de esto, recomienda valores de m para diferentes r ; ver tabla 5.1.

Tabla 5.1: Valores del tamaño de ventana m , los cuales dan Valores Críticos Mínimos de $\alpha \leq 0.1$. Tomado de Sangun (2005)

r	m
5-7	2
8-15	3
16-30	4
31-40	5
41-50	6

Para fines de comparación, se estimó la potencia de la prueba basada en la estadística $T(m, n, r)$ para las distribuciones $Weib(0.5, 1)$, $Weib(3, 1)$, $Ga(0.8, 1)$, $Ga(3, 1)$, $LNorm(5, 3)$ y $LNorm(1, 1)$, considerando censura de 20 y 40 % y tamaños de muestra 10(10)50, usando los valores m de la tabla 5.1. Los resultados se encuentran en la tabla 5.2. Para estimar la potencia de la prueba se empleó el siguiente algoritmo:

Algoritmo 5.2. Para estimar la potencia de la prueba $T(m, n, r)$ con censura Tipo II.

1. Siga los pasos 1 a 5 del Algoritmo 5.1
2. Según el valor r , elija de la tabla 5.1 el correspondiente valor de m .

5.2. Información de Kullback-Leibler

3. Calcule $H(m, n, r)$ con la ecuación (5.3)
4. Use $H(m, n, r)$ en la ecuación (5.2) para calcular la estadística $T(m, n, r)$.
5. Compare el valor de $T(m, n, r)$ con el cuantil $K_{n,r}$ de la tabla 5.3 al nivel de significancia deseado. Si $T(m, n, r) > K_{n,r}$, $Suma = Suma + 1$
6. Repita los pasos 1 a 5 M veces. La proporción de rechazos, $Suma/M$, es una estimación de la potencia de la prueba.

En la figura 5.2 se observa que, al igual que con las estadísticas $W_{r,n}^2$ y $A_{r,n}^2$, la estadística de prueba basada en el coeficiente R_{N-A} tiene mayor potencia que $T(m, n, r)$ para alternativas con función de riesgo monótona decreciente. Asimismo, para distribuciones con función de riesgo monótona creciente la potencia de R_{N-A} es menor que la de $T(m, n, r)$. Con distribuciones con función de riesgo no monótona la estadística propuesta presenta un comportamiento favorable. Es importante notar que en el caso de la distribución $LNorm(1, 1)$, la diferencia a favor de la prueba del R_{N-A} puede deberse a que, como se mencionó en los resultados de comparación con las pruebas $W_{r,n}^2$ y $A_{r,n}^2$, esta distribución es muy aproximada a una exponencial.

5.2. Información de Kullback-Leibler

Tabla 5.2: Potencia estimada para la prueba basada en la estadística $T(m, n, r)$

n	Distribución	Censura							
		20 %				40 %			
		Nivel de significancia							
		0.1	0.05	0.025	0.01	0.1	0.05	0.025	0.01
10	<i>LNorm</i> (5, 3)	0.2368	0.1688	0.1221	0.0746	0.1202	0.0694	0.0412	0.0187
	<i>LNorm</i> (1, 1)	0.2312	0.1232	0.0646	0.0261	0.2499	0.1303	0.0688	0.0275
	<i>Weib</i> (0.5, 1)	0.0596	0.0319	0.0173	0.0082	0.0638	0.0297	0.0155	0.0063
	<i>Weib</i> (3, 1)	0.9713	0.9090	0.8183	0.6324	0.8413	0.6839	0.5248	0.3259
	<i>Ga</i> (0.8, 1)	0.0504	0.0205	0.0103	0.0036	0.0612	0.0270	0.0146	0.0063
	<i>Ga</i> (3, 1)	0.7277	0.5447	0.3792	0.2024	0.5715	0.3741	0.2340	0.1139
20	<i>LNorm</i> (5, 3)	0.7878	0.7108	0.6504	0.5657	0.4779	0.3729	0.2951	0.2088
	<i>LNorm</i> (1, 1)	0.3964	0.2247	0.1335	0.0622	0.4178	0.2511	0.1474	0.0725
	<i>Weib</i> (0.5, 1)	0.3974	0.2950	0.2285	0.1604	0.2747	0.1878	0.1321	0.0820
	<i>Weib</i> (3, 1)	0.9997	0.9983	0.9947	0.9794	0.9878	0.9611	0.9194	0.8269
	<i>Ga</i> (0.8, 1)	0.0489	0.0209	0.0112	0.0041	0.0570	0.0239	0.0100	0.0037
	<i>Ga</i> (3, 1)	0.9375	0.8452	0.7419	0.5853	0.8506	0.7199	0.5762	0.3999
30	<i>LNorm</i> (5, 3)	0.9729	0.9570	0.9363	0.9038	0.7289	0.6436	0.5671	0.4633
	<i>LNorm</i> (1, 1)	0.4901	0.3259	0.2062	0.1063	0.5628	0.3870	0.2642	0.1362
	<i>Weib</i> (0.5, 1)	0.7401	0.6558	0.5741	0.4814	0.4662	0.3665	0.2929	0.2112
	<i>Weib</i> (3, 1)	1.0000	1.0000	0.9999	0.9990	0.9997	0.9986	0.9944	0.9791
	<i>Ga</i> (0.8, 1)	0.0713	0.0353	0.0171	0.0076	0.0649	0.0295	0.0152	0.0055
	<i>Ga</i> (3, 1)	0.9839	0.9549	0.9037	0.8043	0.9648	0.9055	0.8236	0.6732
40	<i>LNorm</i> (5, 3)	0.9951	0.9905	0.9857	0.9749	0.8803	0.8333	0.7772	0.7118
	<i>LNorm</i> (1, 1)	0.6287	0.4552	0.3145	0.1796	0.6173	0.4496	0.3124	0.1974
	<i>Weib</i> (0.5, 1)	0.8797	0.8255	0.7638	0.6833	0.6901	0.6044	0.5178	0.4305
	<i>Weib</i> (3, 1)	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9991	0.9969
	<i>Ga</i> (0.8, 1)	0.0754	0.0378	0.0195	0.0069	0.0684	0.0362	0.0169	0.0080
	<i>Ga</i> (3, 1)	0.9977	0.9917	0.9763	0.9403	0.9823	0.9563	0.9103	0.8320
50	<i>LNorm</i> (5, 3)	0.9992	0.9984	0.9976	0.9960	0.9501	0.9247	0.8940	0.8509
	<i>LNorm</i> (1, 1)	0.6713	0.5059	0.3686	0.2299	0.6525	0.5055	0.3626	0.2288
	<i>Weib</i> (0.5, 1)	0.9494	0.9190	0.8896	0.8443	0.8071	0.7480	0.6804	0.5965
	<i>Ga</i> (3, 1)	0.9992	0.9971	0.9920	0.9773	0.9917	0.9766	0.9504	0.8951
	<i>Weib</i> (3, 1)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9996
	<i>Ga</i> (0.8, 1)	0.0852	0.0448	0.0266	0.0125	0.0682	0.0379	0.0185	0.0078

5.2. Información de Kullback-Leibler

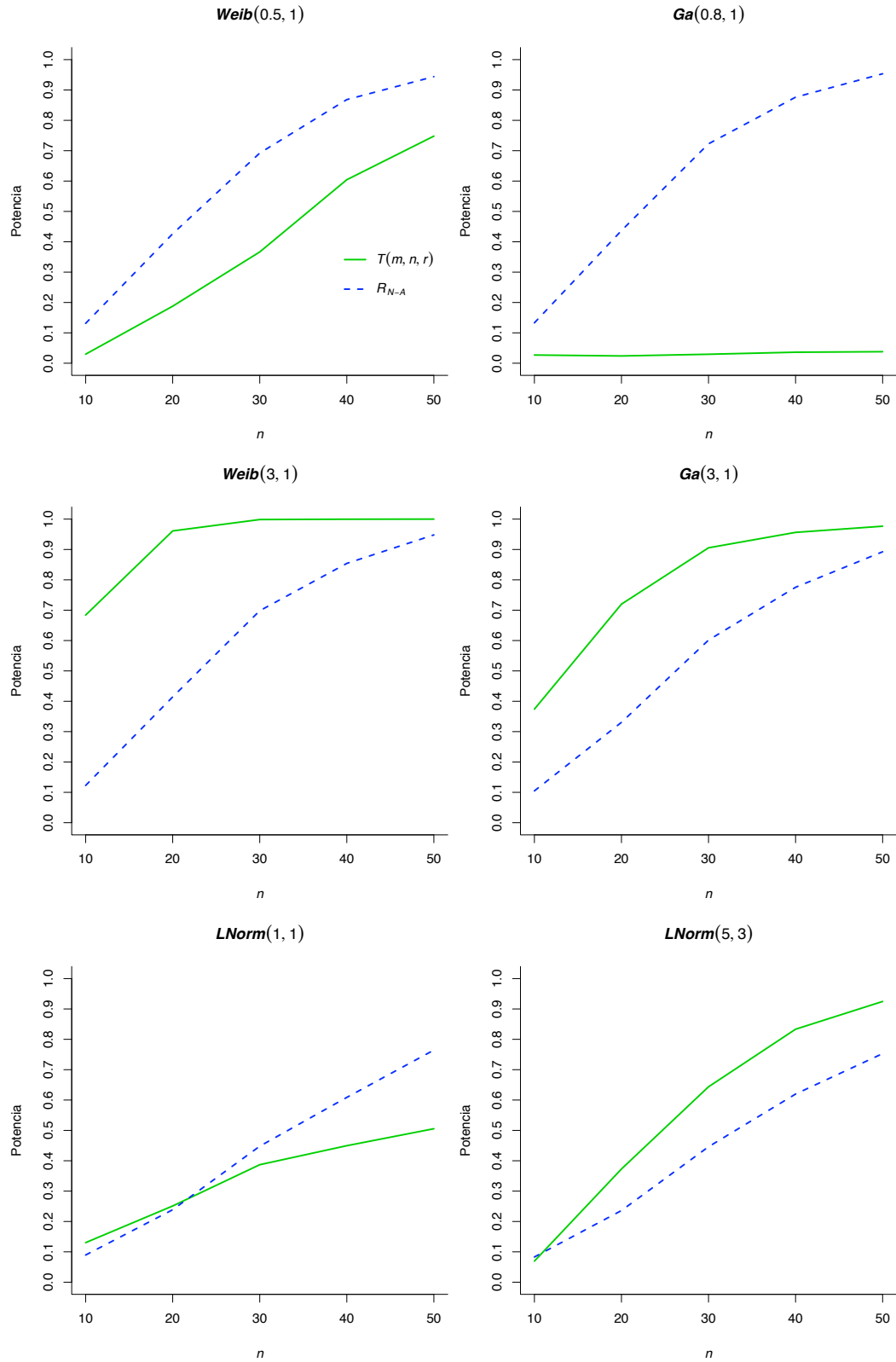


Figura 5.2: Potencia de la prueba basada en $T(m, n, r)$ para algunas alternativas, censura = 40 %, $\alpha = 0.05$, n = tamaño de muestra, $M = 10,000$ réplicas.

5.2. Información de Kullback-Leibler

Tabla 5.3: Cuantiles de la estadística $T(m, n, r)$ para probar exponencialidad cuando existe censura por la derecha Tipo II. Tomado de [Sangun \(2005\)](#)

n	r	Nivel de significancia			
		0.1	0.05	0.025	0.01
10	6	0.4490	0.5429	0.6289	0.7485
	8	0.4906	0.5769	0.6572	0.7733
20	12	0.2620	0.3040	0.3428	0.3956
	16	0.2901	0.3356	0.3739	0.4227
30	18	0.1973	0.2246	0.2495	0.2844
	24	0.2159	0.2434	0.2713	0.3052
40	24	0.1649	0.1846	0.2045	0.2267
	32	0.1782	0.1990	0.2191	0.2448
50	30	0.1452	0.1606	0.1769	0.1962
	40	0.1565	0.1742	0.1899	0.2097

Capítulo 6

Ejemplos de aplicación

6.1. Datos de tiempos de falla

Ouyang y Wu (1994) dieron las primeras $r = 20$ estadísticas de orden de un estudio en el que fueron puestas a prueba simultáneamente $n = 30$ unidades. La prueba terminó cuando fallaron 20 unidades. Ouyang y Wu (1994) usaron los datos para obtener intervalos de predicción de un conjunto de observaciones ordenadas provenientes de la distribución Pareto. Los tiempos de vida observados (en horas) fueron los siguientes:

30.101	30.150	30.374	30.581	30.871	31.086	31.398	31.752	31.792	31.960
32.260	32.517	32.636	33.002	33.552	33.721	34.002	34.023	34.150	35.274

Al calcular el valor de R_{N-A} de los datos usando la fórmula (3.4) con $r = 20$, se tiene $R_{N-A} = 0.9899$. Los cuantiles K_α para $n = 15, 30$ y censura de 33.33% se estimaron como en la sección 4.3, y se muestran en la tabla 6.1. De la tabla 6.1 con $n = 30$ se observa que $R_{N-A} = 0.9899 > K_\alpha$, por lo que *no se rechaza* H_0 : la muestra aleatoria proviene de una distribución Pareto.

6.2. Datos de supervivencia de firmas

Nigm y Hamdy (1987) asumieron que la medida de tiempo en años que un negocio opera

6.3. Datos de salarios anuales

Tabla 6.1: Valores críticos del coeficiente de correlación R_{NA} para $n = 15, 30$ y Censura = 33.33 %.

Censura	n	α					
		0.005	0.01	0.02	0.05	0.10	0.15
33.33 %	15	0.8408	0.8591	0.8779	0.9073	0.9285	0.9390
	30	0.9218	0.9304	0.9425	0.9539	0.9632	0.9688

hasta que *falla* tiene una distribución Pareto. Ellos consideraron disponibles solo los primeros 10 de una muestra aleatoria de 15 negocios. Soliman (2000) y Fernández (2008) usaron las mismas observaciones en ejemplos, asumiendo que tienen distribución Pareto. El conjunto de datos se muestra a continuación:

1.01 1.05 1.08 1.14 1.28 1.30 1.33 1.43 1.59 1.62

Al calcular el valor de R_{N-A} con la muestra censurada usando (3.4), donde $r = 10$, se obtiene $R_{N-A} = 0.9820$. Comparando este valor con los cuantiles correspondientes a $n = 15$ de la tabla 6.1, se concluye que no hay evidencia en contra de la hipótesis de que los datos provienen de una distribución Pareto.

6.3. Datos de salarios anuales

Dyer (1981) reportó datos de salarios anuales de una muestra aleatoria de 30 trabajadores de una línea de producción en una compañía industrial de Estados Unidos (en múltiplos de 100 dolares). Dyer (1981) mostró también el buen ajuste de la distribución Pareto en este caso. Fernández (2008) usó los mismos datos, censurando de manera artificial, para ilustrar la estimación de parámetros de la distribución Pareto en un contexto Bayesiano. La muestra completa es la siguiente:

112 154 119 108 112 156 123 103 115 107
 125 119 128 132 107 151 103 104 116 140
 108 105 158 104 119 111 101 157 112 115

Como ejemplo, suponga que solo se dispone de las primeras r observaciones de la muestra ordenada. Al aplicar la prueba del coeficiente de correlación R_{N-A} a los datos no

6.3. Datos de salarios anuales

censurados se obtienen los siguientes resultados:

r	Censura	R_{N-A}
27	10 %	0.9901
24	20 %	0.9966
18	40 %	0.9924
12	60 %	0.9848
6	80 %	0.9859

Al comparar los valores R_{N-A} con los cuantiles de la tabla 1 para $n = 30$, se tiene que $R_{N-A} > K_\alpha$ para $\alpha = 0.005, 0.01, 0.02, 0.05, 0.10, 0.15$ y $C = 10, 20(20)80\%$. Por tanto, *no se rechaza* H_0 : la muestra aleatoria proviene de una distribución Pareto.

Capítulo 7

Conclusiones

- La distribución del estadístico R_{N-A} es independiente de valores de los parámetros, pero depende del porcentaje de censura. Con censura ligera la distribución del estadístico tiende a concentrarse cerca del 1. A medida que la censura es más severa, la distribución se hace más dispersa.
- La prueba basada en el estadístico R_{N-A} es de tamaño α para cualquier tamaño de muestra y cualquier porcentaje de censura.
- La potencia de la prueba se incrementa conforme aumenta el tamaño de muestra. Esto sugiere que la prueba basada en el estadístico R_{N-A} es una prueba consistente.
- La potencia de la prueba tiende a disminuir conforme la censura es más severa.
- Para alternativas con función de riesgo monótona decreciente, la potencia de la prueba propuesta tiene mejor desempeño que las pruebas clásicas basadas en $W_{r,n}^2$ y $A_{r,n}^2$, y que la prueba basada en $T(m, n, r)$.
- Para alternativas con función de riesgo monótona creciente y no monótona, la potencia de la prueba basada en R_{N-A} es menor que las basadas en $W_{r,n}^2$, $A_{r,n}^2$ y $T(m, n, r)$. Sin embargo, conforme aumenta el tamaño de muestra la prueba propuesta se hace muy competitiva.
- En particular, para la distribución $LNorm(1, 1)$ la potencia de la prueba propuesta es mayor que las basadas en los estadísticos $W_{r,n}^2$, $A_{r,n}^2$ y $T(m, n, r)$.

Referencias

- Aalen, O. (1978). Nonparametric Inference for a Family of Counting Processes. *The Annals of Statistics*, 6, 701–726.
- Andersen, P., Borgan, Q., Gill, R. y Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer Verlag.
- Arnold, B. C. (1983). *Pareto Distributions*. International Cooperative Publishing House.
- Beirlant, J., de Wet, T. y Goegebeur, Y. (2006). A goodness-of-fit statistic for Pareto-type behaviour. *Journal of Computational and Applied Mathematics*, 186, 99–116.
- Blom, G. (1958). *Statistical Estimates and Transformed Beta Variables*. John Wiley.
- Chandra, M., Singpurwalla, N. D. y Stephens, M. A. (1981). Kolmogorov Statistic for Test of Fit for the Extreme Value and Weibull Distributions. *Journal of the American Statistical Association*, 76, 729–731.
- Chen, C.-H. (1984). A Correlation Goodness-of-fit Test for Randomly Censored Data. *Biometrika*, 71, 315–322.
- Choulakian, V. y Stephens, M. A. (2001). Goodness-of-fit Test for the Generalized Pareto Distribution. *Technometrics*, 43, 478–484.
- D’Agostino, R. B. (1986). *Goodness-of-fit Techniques*, cap. Graphical Analysis, 7–62. Marcel Dekker, Inc.
- Dyer, D. (1981). Structural Probability Bounds for the Strong Pareto Law. *The Canadian Journal of Statistics*, 9, 71–77.
- Fernández, A. J. (2008). Highest posterior density estimation from multiply censored Pareto data. *Statistical Papers*, 49, 333–341.
- Filliben, J. J. (1975). The Probability Plot Correlation Coefficient Test for Normality. *Technometrics*, 17, 111–117.

Referencias

- Gan, F., Koehler, K. J. y Thompson, J. C. (1991). Probability Plots and Distribution Curves for Assessing the Fit of Probability Models. *The American Statistician*, 45, 14–21.
- Guilbaud, O. (1988). Exact Kolmogorov-Type Test for Left-Truncated and/or Right-Censored Data. *Journal of the American Statistical Association*, 83, 213–221.
- Harris, C. M. (1968). The Pareto Distribution as a Queue Service Discipline. *Operations Research*, 16, 307–313.
- Ismail, S. (2004). A Simple Estimator for the Shape Parameter of the Pareto Distribution with Economics and Medical Applications. *Journal of Applied Statistics*, 31, 3–13.
- Kaplan, E. L. y Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, 53, 282, 457–481.
- Kimball, B. F. (1960). On the Choice of Plotting Positions on Probability Paper. *Journal of the American Statistical Association*, 55, 546–560.
- Kinnison, R. (1989). Correlation Coefficient Goodness-of-fit Test for Extreme Value Distribution. *The American Statistician*, 43, 98–100.
- Klein, J. P. y Moeschberger, M. L. (1997). *Survival Analysis: Techniques for Censored and Truncated Data*. Statistics for Biology and Health.
- Koziol, J. A. (1980). Goodness-of-fit Test for Randomly Censored Data. *Biometrika*, 67, 693–696.
- Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*. John Wiley and Sons, Inc.
- Looney, S. W. y Gullledge, T. R. (1985). Use of Correlation Coefficient with Normal Probability Plots. *The American Statistician*, 39, 75–79.
- Mage, D. T. (1982). An Objective Graphical Method for Testing Normal Distributional Assumptions Using Probability Plots. *The American Statistician*, 36, 116–120.
- Margolin, B. H. y Maurer, W. (1976). Test of the Kolmogorov-Smirnov Type for Exponential Data with Unknown Scale, and Related Problems. *Biometrika*, 63, 149–160.
- Meintanis, S. G. y Bassiakos, Y. (2007). Data-Transformation and Test of Fit for the Generalized Pareto Hypothesis. *Communications in Statistics-Theory and Methods*, 36, 833–849.
- Michael, J. R. y Schucany, W. R. (1986). *Goodness-of-fit Techniques*, cap. Analysis of Data Censored Samples, 461–496. Marcel Dekker, Inc.

Referencias

- Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics*, 14, 945–965.
- Nigm, A. M. y Hamdy, H. I. (1987). Bayesian prediction bounds for the Pareto lifetime model. *Communications in Statistics*, A16, 1761–1772.
- Ouyang, L.-Y. y Wu, S.-J. (1994). Prediction Intervals for an Ordered Observation from a Pareto Distribution. *IEEE Transactions on Reliability*, 43, 264–269.
- Pettitt, A.Ñ. (1977). Test for the Exponential Distribution with Censored Data Using Cramer-von Mises Statistics. *Biometrika*, 64, 629–632.
- Porter, J. E., Coleman, J. W. y Moore, A. H. (1992). Modified KS, AD, and C-v M Test for the Pareto Distribution with Unknown Location & Scale Parameters. *IEEE Transactions on Reliability*, 41, 112–117.
- R Development Core Team (2007). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Sangun, P. (2005). Testing Exponentiality Based on the Kullback-Leibler Information With the Type II Censored Data. *IEEE Transactions on Reliability*, 54, 22–26.
- Sirvanci, M. y Levent, I. (1982). Cramer-von Mises Statistic for Testing Exponentiality with Censored Samples. *Biometrika*, 69, 641–646.
- Soliman, A. A. (2000). Bayes Prediction in a Pareto Lifetime Model with Random Sample Size. *The Statistician*, 49, 51–62.
- Stephens, M. A. (1977). Goodness-of-fit for the Extreme Value Distribution. *Biometrika*, 64, 583–588.
- Stephens, M. A. (1986a). *Goodness-of-fit Techniques*. Marcel Dekker, Inc.
- Stephens, M. A. (1986b). *Goodness-of-fit Techniques*, cap. Test Based on EDF Statistics, 97–193. Marcel Dekker, Inc.
- Zeinab, A. H. (2007). Tests for the Validity of the Assumption that the Underlying Distribution of Life is Pareto. *Journal of Applied Statistics*, 34, 195–201.

Apéndice

Tabla 1: Valores críticos del coeficiente de correlación R_{NA} , obtenidos mediante Simulación Monte Carlo con $M = 10,000$ muestras de tamaño n .

C	n	α					
		0.005	0.01	0.02	0.05	0.10	0.15
10 %	10	0.8248	0.8469	0.8685	0.8942	0.9178	0.9296
	20	0.8932	0.9080	0.9216	0.9394	0.9517	0.9588
	30	0.9289	0.9383	0.9470	0.9582	0.9664	0.9715
	40	0.9479	0.9548	0.9608	0.9692	0.9753	0.9789
	50	0.9565	0.9623	0.9674	0.9744	0.9796	0.9826
	60	0.9634	0.9684	0.9731	0.9783	0.9827	0.9851
	70	0.9681	0.9721	0.9764	0.9817	0.9853	0.9876
	80	0.9712	0.9759	0.9791	0.9840	0.9875	0.9891
	90	0.9751	0.9781	0.9814	0.9855	0.9886	0.9901
	100	0.9785	0.9810	0.9835	0.9872	0.9899	0.9913
500	0.9957	0.9962	0.9968	0.9975	0.9979	0.9982	
20 %	10	0.8203	0.8409	0.8578	0.8923	0.9142	0.9266
	20	0.8925	0.9076	0.9219	0.9385	0.9513	0.9586
	30	0.9278	0.9363	0.9453	0.9568	0.9659	0.9710
	40	0.9461	0.9528	0.9600	0.9680	0.9745	0.9783
	50	0.9578	0.9639	0.9694	0.9752	0.9800	0.9828
	60	0.9632	0.9687	0.9729	0.9785	0.9831	0.9855
	70	0.9680	0.9723	0.9767	0.9821	0.9857	0.9877
	80	0.9735	0.9765	0.9796	0.9841	0.9873	0.9892
	90	0.9758	0.9793	0.9823	0.9860	0.9887	0.9902
	100	0.9796	0.9816	0.9838	0.9875	0.9898	0.9912
500	0.9959	0.9963	0.9969	0.9975	0.9980	0.9983	

continúa...

C	n	α					
		0.005	0.01	0.02	0.05	0.10	0.15
40 %	10	0.7828	0.8032	0.8289	0.8654	0.8926	0.9105
	20	0.8637	0.8828	0.9003	0.9247	0.9404	0.9491
	30	0.9140	0.9250	0.9351	0.9499	0.9596	0.9652
	40	0.9351	0.9429	0.9511	0.9610	0.9688	0.9735
	50	0.9465	0.9539	0.9599	0.9687	0.9752	0.9788
	60	0.9577	0.9626	0.9679	0.9740	0.9789	0.9821
	70	0.9626	0.9670	0.9710	0.9776	0.9821	0.9846
	80	0.9692	0.9722	0.9753	0.9804	0.9845	0.9866
	90	0.9720	0.9751	0.9785	0.9828	0.9859	0.9880
	100	0.9723	0.9759	0.9799	0.9844	0.9875	0.9892
	500	0.9947	0.9955	0.9962	0.9970	0.9976	0.9979
60 %	10	0.8478	0.8508	0.8551	0.8716	0.8900	0.8999
	20	0.8037	0.8326	0.8596	0.8917	0.9159	0.9286
	30	0.8665	0.8877	0.9054	0.9257	0.9411	0.9495
	40	0.9015	0.9157	0.9283	0.9435	0.9551	0.9618
	50	0.9218	0.9315	0.9406	0.9538	0.9630	0.9684
	60	0.9347	0.9431	0.9512	0.9618	0.9697	0.9740
	70	0.9434	0.9515	0.9587	0.9677	0.9741	0.9777
	80	0.9509	0.9572	0.9632	0.9716	0.9778	0.9808
	90	0.9575	0.9626	0.9684	0.9746	0.9798	0.9827
	100	0.9618	0.9673	0.9711	0.9775	0.9821	0.9846
	500	0.9924	0.9934	0.9943	0.9955	0.9964	0.9968
80 %	20	0.8602	0.8634	0.8682	0.8775	0.8877	0.8985
	30	0.7894	0.8100	0.8297	0.8685	0.8972	0.9115
	40	0.8233	0.8429	0.8640	0.8961	0.9167	0.9297
	50	0.8449	0.8652	0.8862	0.9121	0.9292	0.9399
	60	0.8683	0.8843	0.9045	0.9253	0.9411	0.9488
	70	0.8921	0.9048	0.9176	0.9355	0.9488	0.9562
	80	0.9050	0.9154	0.9280	0.9437	0.9556	0.9617
	90	0.9173	0.9273	0.9380	0.9511	0.9606	0.9661
	100	0.9239	0.9350	0.9448	0.9553	0.9642	0.9691
		500	0.9854	0.9873	0.9889	0.9911	0.9928

Tabla 2: Tamaño estimado de la prueba R_{NA} con distribución $Par(3, 0.5)$, obtenidos mediante Simulación Monte Carlo con $M = 10,000$ muestras de tamaño n .

C	n	α					
		0.005	0.01	0.02	0.05	0.10	0.15
10 %	10	0.0058	0.0117	0.0221	0.0504	0.1057	0.1499
	20	0.0073	0.0140	0.0256	0.0541	0.1007	0.1494
	30	0.0074	0.0131	0.0252	0.0526	0.0990	0.1534
	40	0.0092	0.0163	0.0270	0.0581	0.1057	0.1555
	50	0.0086	0.0145	0.0252	0.0565	0.1009	0.1552
	60	0.0061	0.0116	0.0228	0.0485	0.0952	0.1382
	70	0.0062	0.0121	0.0221	0.0547	0.1042	0.1532
	80	0.0052	0.0115	0.0199	0.0536	0.1121	0.1582
	90	0.0050	0.0087	0.0184	0.0481	0.0970	0.1458
	100	0.0050	0.0111	0.0197	0.0524	0.1081	0.1584
	500	0.0046	0.0095	0.0206	0.0513	0.0989	0.1486
20 %	10	0.0064	0.0107	0.0188	0.0514	0.1011	0.1474
	20	0.0055	0.0109	0.0217	0.0518	0.0991	0.1481
	30	0.0068	0.0111	0.0218	0.0490	0.1001	0.1479
	40	0.0059	0.0112	0.0214	0.0477	0.0903	0.1403
	50	0.0060	0.0120	0.0242	0.0511	0.0998	0.1537
	60	0.0052	0.0116	0.0215	0.0493	0.1019	0.1538
	70	0.0055	0.0098	0.0201	0.0524	0.1062	0.1565
	80	0.0061	0.0101	0.0204	0.0506	0.1039	0.1588
	90	0.0066	0.0115	0.0229	0.0507	0.1012	0.1509
	100	0.0081	0.0115	0.0209	0.0511	0.0977	0.1463
	500	0.0059	0.0106	0.0220	0.0501	0.1042	0.1522

continúa...

C	n	α					
		0.005	0.01	0.02	0.05	0.10	0.15
40 %	10	0.0043	0.0083	0.0173	0.0453	0.0957	0.1493
	20	0.0042	0.0080	0.0172	0.0491	0.1008	0.1507
	30	0.0074	0.0135	0.0260	0.0578	0.1050	0.1539
	40	0.0061	0.0124	0.0234	0.0518	0.1037	0.1553
	50	0.0047	0.0102	0.0193	0.0495	0.1018	0.1525
	60	0.0071	0.0116	0.0198	0.0476	0.0946	0.1461
	70	0.0055	0.0102	0.0184	0.0530	0.1019	0.1512
	80	0.0069	0.0112	0.0203	0.0484	0.1027	0.1505
	90	0.0056	0.0109	0.0213	0.0492	0.0948	0.1445
	100	0.0041	0.0082	0.0191	0.0526	0.1019	0.1503
	500	0.0043	0.0095	0.0199	0.0548	0.1075	0.1584
60 %	10	0.0000	0.0034	0.0125	0.0475	0.1089	0.1592
	20	0.0054	0.0094	0.0213	0.0508	0.1059	0.1552
	30	0.0045	0.0099	0.0199	0.0519	0.1016	0.1511
	40	0.0048	0.0117	0.0226	0.0528	0.1041	0.1526
	50	0.0046	0.0096	0.0172	0.0476	0.0946	0.1444
	60	0.0053	0.0109	0.0207	0.0494	0.1006	0.1461
	70	0.0048	0.0105	0.0205	0.0495	0.0929	0.1450
	80	0.0029	0.0081	0.0168	0.0479	0.1064	0.1540
	90	0.0070	0.0114	0.0228	0.0493	0.1042	0.1522
	100	0.0050	0.0104	0.0190	0.0486	0.1038	0.1559
	500	0.0050	0.0105	0.0205	0.0518	0.1017	0.1508
80 %	20	0.0038	0.0116	0.0223	0.0537	0.1015	0.1519
	30	0.0049	0.0094	0.0181	0.0465	0.0957	0.1406
	40	0.0053	0.0111	0.0205	0.0555	0.1047	0.1624
	50	0.0046	0.0097	0.0211	0.0537	0.0990	0.1474
	60	0.0031	0.0068	0.0176	0.0462	0.0961	0.1406
	70	0.0054	0.0097	0.0185	0.0467	0.0956	0.1460
	80	0.0052	0.0086	0.0181	0.0475	0.1047	0.1513
	90	0.0069	0.0122	0.0230	0.0535	0.1054	0.1578
	100	0.0058	0.0124	0.0241	0.0508	0.1004	0.1481
	500	0.0052	0.0095	0.0207	0.0499	0.0995	0.1533

Tabla 3: Tamaño estimado de la prueba R_{NA} con distribución $Par(5, 1)$, obtenidos mediante Simulación Monte Carlo con $M = 10,000$ muestras de tamaño n .

C	n	α					
		0.005	0.01	0.02	0.05	0.10	0.15
10 %	10	0.0058	0.0115	0.0240	0.0515	0.1054	0.1515
	20	0.0072	0.0138	0.0248	0.0543	0.1015	0.1499
	30	0.0072	0.0124	0.0232	0.0504	0.0960	0.1464
	40	0.0090	0.0155	0.0247	0.0552	0.1036	0.1544
	50	0.0082	0.0139	0.0246	0.0561	0.1012	0.1568
	60	0.0064	0.0114	0.0222	0.0483	0.1000	0.1481
	70	0.0072	0.0140	0.0258	0.0550	0.1044	0.1548
	80	0.0057	0.0127	0.0214	0.0535	0.1074	0.1508
	90	0.0055	0.0109	0.0198	0.0462	0.0989	0.1481
	100	0.0064	0.0108	0.0205	0.0497	0.1023	0.1530
	500	0.0049	0.0106	0.0209	0.0549	0.0987	0.1485
20 %	10	0.0065	0.0123	0.0200	0.0547	0.1022	0.1462
	20	0.0050	0.0119	0.0240	0.0534	0.1102	0.1618
	30	0.0062	0.0113	0.0219	0.0470	0.0959	0.1462
	40	0.0066	0.0106	0.0233	0.0502	0.0964	0.1473
	50	0.0069	0.0156	0.0272	0.0539	0.0991	0.1504
	60	0.0056	0.0119	0.0220	0.0500	0.1020	0.1492
	70	0.0044	0.0098	0.0217	0.0569	0.1064	0.1587
	80	0.0072	0.0122	0.0211	0.0539	0.1063	0.1618
	90	0.0046	0.0115	0.0229	0.0544	0.1034	0.1511
	100	0.0072	0.0117	0.0205	0.0503	0.1017	0.1489
	500	0.0054	0.0097	0.0230	0.0523	0.1047	0.1530

continúa...

C	n	α					
		0.005	0.01	0.02	0.05	0.10	0.15
40 %	10	0.0043	0.0081	0.0174	0.0439	0.0870	0.1394
	20	0.0050	0.0106	0.0205	0.0546	0.1046	0.1549
	30	0.0064	0.0137	0.0241	0.0544	0.1053	0.1561
	40	0.0056	0.0096	0.0177	0.0444	0.0952	0.1494
	50	0.0041	0.0093	0.0181	0.0443	0.0971	0.1517
	60	0.0057	0.0100	0.0211	0.0459	0.0918	0.1475
	70	0.0055	0.0114	0.0189	0.0502	0.1019	0.1512
	80	0.0059	0.0108	0.0202	0.0482	0.1002	0.1474
	90	0.0057	0.0109	0.0205	0.0466	0.0900	0.1355
	100	0.0039	0.0088	0.0190	0.0488	0.0997	0.1497
	500	0.0046	0.0089	0.0218	0.0566	0.1062	0.1544
60 %	10	0.0000	0.0046	0.0128	0.0442	0.1016	0.1485
	20	0.0030	0.0088	0.0171	0.0467	0.1048	0.1510
	30	0.0042	0.0112	0.0204	0.0498	0.0986	0.1465
	40	0.0047	0.0101	0.0201	0.0496	0.1004	0.1537
	50	0.0049	0.0094	0.0172	0.0488	0.0996	0.1480
	60	0.0042	0.0091	0.0179	0.0464	0.0990	0.1462
	70	0.0041	0.0102	0.0198	0.0480	0.0936	0.1461
	80	0.0037	0.0089	0.0186	0.0503	0.1069	0.1549
	90	0.0061	0.0112	0.0210	0.0514	0.1010	0.1562
	100	0.0048	0.0106	0.0195	0.0465	0.0992	0.1550
	500	0.0044	0.0097	0.0191	0.0520	0.1016	0.1521
80 %	20	0.0051	0.0120	0.0219	0.0529	0.1043	0.1564
	30	0.0072	0.0147	0.0225	0.0546	0.1013	0.1476
	40	0.0061	0.0103	0.0210	0.0541	0.0985	0.1542
	50	0.0060	0.0112	0.0222	0.0546	0.0989	0.1484
	60	0.0054	0.0089	0.0203	0.0520	0.1031	0.1484
	70	0.0053	0.0090	0.0179	0.0461	0.0949	0.1449
	80	0.0065	0.0105	0.0214	0.0480	0.1005	0.1481
	90	0.0065	0.0123	0.0226	0.0550	0.1048	0.1542
	100	0.0053	0.0118	0.0217	0.0511	0.1018	0.1468
	500	0.0055	0.0097	0.0203	0.0486	0.0999	0.1513

Tabla 4: Tamaño estimado de la prueba R_{NA} con distribución $Par(0.5, 100)$, obtenidos mediante Simulación Monte Carlo, $M = 10,000$ muestras de tamaño n

C	n	α					
		0.005	0.01	0.02	0.05	0.10	0.15
10 %	10	0.0061	0.0129	0.0244	0.0518	0.1068	0.1509
	20	0.0080	0.0132	0.0244	0.0555	0.1034	0.1485
	30	0.0092	0.0165	0.0269	0.0576	0.1054	0.1547
	40	0.0106	0.0174	0.0281	0.0596	0.1084	0.1589
	50	0.0091	0.0143	0.0246	0.0531	0.1026	0.1523
	60	0.0054	0.0129	0.0240	0.0526	0.0988	0.1436
	70	0.0057	0.0102	0.0204	0.0511	0.1006	0.1537
	80	0.0044	0.0107	0.0204	0.0570	0.1142	0.1598
	90	0.0068	0.0115	0.0207	0.0499	0.0996	0.1467
	100	0.0076	0.0134	0.0230	0.0570	0.1131	0.1604
	500	0.0054	0.0105	0.0212	0.0530	0.0987	0.1520
20 %	10	0.0070	0.0145	0.0242	0.0577	0.1094	0.1580
	20	0.0048	0.0109	0.0192	0.0528	0.1008	0.1532
	30	0.0054	0.0103	0.0187	0.0460	0.0958	0.1467
	40	0.0061	0.0116	0.0227	0.0499	0.0947	0.1452
	50	0.0067	0.0137	0.0237	0.0530	0.1007	0.1475
	60	0.0055	0.0115	0.0204	0.0490	0.0995	0.1512
	70	0.0052	0.0107	0.0231	0.0574	0.1093	0.1621
	80	0.0061	0.0103	0.0202	0.0496	0.1031	0.1546
	90	0.0047	0.0107	0.0208	0.0503	0.0985	0.1457
	100	0.0060	0.0112	0.0196	0.0482	0.1002	0.1457
	500	0.0059	0.0100	0.0240	0.0520	0.1067	0.1568

continúa...

C	n	α					
		0.005	0.01	0.02	0.05	0.10	0.15
40 %	10	0.0039	0.0068	0.0157	0.0465	0.0934	0.1477
	20	0.0044	0.0082	0.0182	0.0509	0.1024	0.1505
	30	0.0058	0.0125	0.0227	0.0574	0.1006	0.1476
	40	0.0043	0.0096	0.0193	0.0465	0.0952	0.1519
	50	0.0037	0.0083	0.0162	0.0458	0.1006	0.1515
	60	0.0072	0.0133	0.0256	0.0511	0.0936	0.1453
	70	0.0045	0.0087	0.0174	0.0494	0.1002	0.1523
	80	0.0063	0.0121	0.0206	0.0461	0.1027	0.1472
	90	0.0076	0.0114	0.0231	0.0506	0.0978	0.1487
	100	0.0038	0.0082	0.0184	0.0476	0.0940	0.1410
	500	0.0059	0.0098	0.0192	0.0524	0.1021	0.1508
60 %	10	0.0000	0.0036	0.0111	0.0444	0.1029	0.1488
	20	0.0028	0.0078	0.0196	0.0490	0.1007	0.1453
	30	0.0040	0.0097	0.0217	0.0487	0.0996	0.1532
	40	0.0054	0.0118	0.0214	0.0499	0.1031	0.1525
	50	0.0050	0.0090	0.0183	0.0471	0.0922	0.1462
	60	0.0049	0.0100	0.0201	0.0489	0.0987	0.1457
	70	0.0049	0.0115	0.0227	0.0519	0.0985	0.1467
	80	0.0034	0.0073	0.0154	0.0483	0.1042	0.1538
	90	0.0065	0.0105	0.0197	0.0466	0.0994	0.1483
	100	0.0050	0.0126	0.0205	0.0530	0.1041	0.1481
	500	0.0060	0.0112	0.0190	0.0480	0.0995	0.1464
80 %	20	0.0054	0.0121	0.0251	0.0528	0.1008	0.1535
	30	0.0066	0.0131	0.0212	0.0508	0.1011	0.1498
	40	0.0070	0.0116	0.0229	0.0573	0.1050	0.1535
	50	0.0045	0.0086	0.0207	0.0516	0.0975	0.1484
	60	0.0044	0.0091	0.0199	0.0512	0.0970	0.1440
	70	0.0060	0.0100	0.0203	0.0472	0.0932	0.1447
	80	0.0042	0.0080	0.0167	0.0468	0.1023	0.1483
	90	0.0050	0.0124	0.0226	0.0536	0.1031	0.1518
	100	0.0053	0.0106	0.0228	0.0504	0.0967	0.1409
		500	0.0064	0.0131	0.0233	0.0527	0.1007

Tabla 5: Potencia de la prueba para alternativas $Weib(0.5, 1)$ y $Weib(3, 1)$, con $M = 10,000$ réplicas de tamaño n

C	n	$Weib(0.5, 1)$										$Weib(3, 1)$									
		α	0.005	0.01	0.02	0.05	0.10	0.15	0.005	0.01	0.02	0.05	0.10	0.15							
10 %	10	0.1012	0.1511	0.2178	0.3299	0.4601	0.5385	0.1042	0.1532	0.2222	0.3287	0.4612	0.5416								
	20	0.4132	0.5261	0.6304	0.7693	0.8552	0.8968	0.4169	0.5249	0.6354	0.7715	0.8568	0.8979								
	30	0.7866	0.8513	0.9024	0.9538	0.9778	0.9877	0.7993	0.8624	0.9076	0.9530	0.9757	0.9862								
	40	0.9540	0.9745	0.9874	0.9945	0.9978	0.9988	0.9563	0.9739	0.9828	0.9942	0.9971	0.9987								
	50	0.9904	0.9950	0.9976	0.9995	0.9998	1.0000	0.9893	0.9952	0.9977	0.9992	0.9998	1.0000								
	60	0.9990	0.9996	0.9998	1.0000	1.0000	1.0000	0.9984	0.9995	0.9995	0.9998	1.0000	1.0000								
	70	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000	0.9996	1.0000	1.0000	1.0000	1.0000	1.0000								
	80	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000								
	90	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000								
	100	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000								
20 %	500	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000								
	10	0.0802	0.1149	0.1515	0.2717	0.3828	0.4600	0.0779	0.1109	0.1461	0.2600	0.3714	0.4478								
	20	0.3244	0.4240	0.5270	0.6560	0.7632	0.8222	0.3293	0.4232	0.5280	0.6620	0.7696	0.8306								
	30	0.6681	0.7418	0.8175	0.8949	0.9410	0.9653	0.6575	0.7311	0.8080	0.8887	0.9348	0.9560								
	40	0.8836	0.9216	0.9529	0.9741	0.9875	0.9931	0.8788	0.9195	0.9516	0.9769	0.9917	0.9948								
	50	0.9672	0.9810	0.9899	0.9964	0.9986	0.9990	0.9715	0.9838	0.9930	0.9971	0.9985	0.9994								
	60	0.9891	0.9955	0.9978	0.9994	0.9998	1.0000	0.9900	0.9951	0.9972	0.9989	0.9995	0.9998								
	70	0.9972	0.9984	0.9993	0.9999	1.0000	1.0000	0.9980	0.9992	0.9998	1.0000	1.0000	1.0000								
	80	0.9997	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9997	0.9999	1.0000	1.0000	1.0000								
	90	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9999	1.0000	1.0000	1.0000	1.0000								
100	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000									
500	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000									

continúa...

40 %	10	0.0268	0.0424	0.0693	0.1312	0.2025	0.2754	0.0268	0.0400	0.0667	0.1227	0.2015	0.2734
	20	0.1392	0.2031	0.2795	0.4264	0.5419	0.6162	0.1301	0.1937	0.2719	0.4143	0.5345	0.6074
	30	0.3935	0.4774	0.5594	0.6922	0.7813	0.8338	0.4034	0.4885	0.5713	0.6982	0.7846	0.8347
	40	0.6376	0.7105	0.7882	0.8683	0.9182	0.9439	0.6298	0.7017	0.7732	0.8541	0.9124	0.9390
	50	0.7855	0.8446	0.8899	0.9440	0.9709	0.9819	0.7922	0.8491	0.8911	0.9480	0.9721	0.9820
	60	0.9194	0.9446	0.9647	0.9823	0.9907	0.9944	0.9119	0.9386	0.9626	0.9792	0.9891	0.9939
	70	0.9633	0.9752	0.9840	0.9945	0.9973	0.9985	0.9621	0.9767	0.9846	0.9945	0.9974	0.9985
	80	0.9881	0.9916	0.9957	0.9976	0.9989	0.9995	0.9866	0.9909	0.9940	0.9971	0.9987	0.9993
	90	0.9951	0.9968	0.9982	0.9992	0.9996	0.9999	0.9943	0.9968	0.9986	0.9996	0.9999	0.9999
	100	0.9972	0.9990	0.9995	0.9998	0.9999	0.9999	0.9969	0.9986	0.9993	0.9996	0.9997	0.9998
500	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
60 %	10	0.0046	0.0128	0.0268	0.0759	0.1424	0.1917	0.0028	0.0097	0.0228	0.0710	0.1339	0.1787
	20	0.0301	0.0587	0.1003	0.1842	0.2846	0.3588	0.0341	0.0613	0.1062	0.1871	0.2861	0.3600
	30	0.1270	0.1913	0.2693	0.3823	0.4899	0.5614	0.1267	0.1958	0.2623	0.3811	0.4931	0.5666
	40	0.2645	0.3436	0.4372	0.5608	0.6693	0.7323	0.2656	0.3450	0.4361	0.5649	0.6697	0.7364
	50	0.4341	0.5077	0.5896	0.7180	0.8044	0.8524	0.43	0.5087	0.5854	0.7101	0.7948	0.8429
	60	0.5769	0.6609	0.7338	0.8257	0.8920	0.9208	0.5797	0.6602	0.7378	0.8303	0.8930	0.9236
	70	0.6898	0.7687	0.8290	0.8988	0.9395	0.9587	0.6948	0.7695	0.8323	0.9014	0.9382	0.9592
	80	0.7913	0.8477	0.8917	0.9423	0.9699	0.9782	0.7905	0.8452	0.8916	0.9442	0.9714	0.9795
	90	0.8761	0.9094	0.9421	0.9697	0.9842	0.9909	0.8799	0.9141	0.9424	0.9692	0.9861	0.9916
	100	0.9230	0.9500	0.9662	0.9834	0.9918	0.9958	0.9250	0.9528	0.9683	0.9836	0.9916	0.9947
500	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	
80 %	20	0.0073	0.0190	0.0336	0.0712	0.1156	0.1682	0.0083	0.0195	0.0351	0.0700	0.1176	0.1724
	30	0.0176	0.0311	0.0483	0.0993	0.1719	0.2308	0.0199	0.0344	0.0515	0.1045	0.1805	0.2389
	40	0.0456	0.0692	0.1036	0.1862	0.2686	0.3397	0.0474	0.0693	0.1028	0.1844	0.2696	0.3377
	50	0.0704	0.1082	0.1614	0.2556	0.3519	0.4275	0.0705	0.1026	0.1583	0.2584	0.3557	0.4309
	60	0.1161	0.1609	0.2353	0.3465	0.4596	0.5248	0.1192	0.1635	0.2391	0.3513	0.4619	0.5277
	70	0.1970	0.2553	0.3229	0.4465	0.5523	0.6200	0.1862	0.2430	0.3115	0.4363	0.5487	0.6165
	80	0.2588	0.3177	0.4072	0.5361	0.6422	0.7027	0.2655	0.3241	0.4078	0.5364	0.6457	0.7035
	90	0.3454	0.4186	0.5077	0.6287	0.7219	0.7743	0.3481	0.4225	0.5099	0.6309	0.7171	0.7735
	100	0.4022	0.4938	0.5835	0.6833	0.7689	0.8140	0.4048	0.4928	0.5817	0.6850	0.7742	0.8203
	500	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Tabla 6: Potencia de la prueba para alternativas $LNorm(5, 3)$ y $LNorm(1, 1)$, con $M = 10,000$ réplicas de tamaño n

C	n	$LNorm(5, 3)$										$LNorm(1, 1)$									
		α										α									
		0.005	0.01	0.02	0.05	0.10	0.15	0.005	0.01	0.02	0.05	0.10	0.15	0.005	0.01	0.02	0.05	0.10	0.15		
10 %	10	0.0357	0.0591	0.0974	0.1705	0.2737	0.3431	0.0392	0.0610	0.0957	0.1706	0.2784	0.3522	0.0392	0.0610	0.0957	0.1706	0.2784	0.3522		
	20	0.1138	0.1790	0.2674	0.4182	0.5546	0.6433	0.1111	0.1761	0.2623	0.4223	0.5604	0.6443	0.1111	0.1761	0.2623	0.4223	0.5604	0.6443		
	30	0.3320	0.4341	0.5431	0.6905	0.7976	0.8589	0.3241	0.4274	0.5376	0.6917	0.7984	0.8615	0.3241	0.4274	0.5376	0.6917	0.7984	0.8615		
	40	0.5764	0.6802	0.7667	0.8701	0.9279	0.9539	0.5864	0.6849	0.7660	0.8739	0.9297	0.9548	0.5864	0.6849	0.7660	0.8739	0.9297	0.9548		
	50	0.7438	0.8230	0.8853	0.9469	0.9745	0.9860	0.7368	0.8205	0.8818	0.9452	0.9744	0.9851	0.7368	0.8205	0.8818	0.9452	0.9744	0.9851		
	60	0.8716	0.9190	0.9531	0.9790	0.9926	0.9957	0.8706	0.9215	0.9570	0.9829	0.9931	0.9967	0.8706	0.9215	0.9570	0.9829	0.9931	0.9967		
	70	0.9356	0.9644	0.9830	0.9936	0.9985	0.9994	0.9341	0.9604	0.9822	0.9935	0.9979	0.9987	0.9341	0.9604	0.9822	0.9935	0.9979	0.9987		
	80	0.9719	0.9875	0.9932	0.9985	0.9998	0.9999	0.9715	0.9886	0.9940	0.9984	0.9996	0.9999	0.9715	0.9886	0.9940	0.9984	0.9996	0.9999		
	90	0.9910	0.9950	0.9977	0.9997	1.0000	1.0000	0.9896	0.9954	0.9982	0.9993	1.0000	1.0000	0.9896	0.9954	0.9982	0.9993	1.0000	1.0000		
	100	0.9970	0.9990	0.9997	1.0000	1.0000	1.0000	0.9975	0.9989	0.9994	1.0000	1.0000	1.0000	0.9975	0.9989	0.9994	1.0000	1.0000	1.0000		
500	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000		
20 %	10	0.0316	0.0482	0.0704	0.1466	0.2312	0.2999	0.0325	0.0505	0.0733	0.1555	0.2412	0.3114	0.0325	0.0505	0.0733	0.1555	0.2412	0.3114		
	20	0.1022	0.1564	0.2385	0.3652	0.4966	0.5821	0.0991	0.1525	0.2319	0.3648	0.4992	0.5811	0.0991	0.1525	0.2319	0.3648	0.4992	0.5811		
	30	0.2669	0.3480	0.4542	0.6035	0.7242	0.7981	0.2710	0.3493	0.4494	0.6024	0.7325	0.7996	0.2710	0.3493	0.4494	0.6024	0.7325	0.7996		
	40	0.4858	0.5800	0.6875	0.8026	0.8816	0.9208	0.4838	0.5794	0.6877	0.7999	0.8791	0.9183	0.4838	0.5794	0.6877	0.7999	0.8791	0.9183		
	50	0.6884	0.7826	0.8567	0.9178	0.9570	0.9739	0.6847	0.7781	0.8530	0.9169	0.9572	0.9725	0.6847	0.7781	0.8530	0.9169	0.9572	0.9725		
	60	0.8022	0.8705	0.9157	0.9569	0.9818	0.9897	0.8050	0.8747	0.9189	0.9617	0.9823	0.9886	0.8050	0.8747	0.9189	0.9617	0.9823	0.9886		
	70	0.8847	0.9252	0.9605	0.9856	0.9943	0.9970	0.8897	0.9291	0.9634	0.9874	0.9946	0.9975	0.8897	0.9291	0.9634	0.9874	0.9946	0.9975		
	80	0.9559	0.9739	0.9866	0.9957	0.9984	0.9991	0.9553	0.9725	0.9860	0.9951	0.9980	0.9989	0.9553	0.9725	0.9860	0.9951	0.9980	0.9989		
	90	0.9765	0.9887	0.9946	0.9978	0.9988	0.9995	0.9779	0.9889	0.9952	0.9985	0.9996	0.9999	0.9779	0.9889	0.9952	0.9985	0.9996	0.9999		
	100	0.9928	0.9966	0.9973	0.9995	0.9997	0.9999	0.9935	0.9964	0.9981	0.9994	0.9998	1.0000	0.9935	0.9964	0.9981	0.9994	0.9998	1.0000		
500	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000		

continúa...

	10	0.0156	0.0241	0.0416	0.0830	0.1508	0.2156	0.0173	0.0254	0.0435	0.0895	0.1503	0.2142
	20	0.0513	0.0846	0.1300	0.2360	0.3505	0.4247	0.0524	0.0841	0.1291	0.2387	0.3452	0.4229
	30	0.1658	0.2263	0.3003	0.4461	0.5550	0.6369	0.1650	0.2316	0.3037	0.4479	0.5664	0.6415
	40	0.3110	0.3894	0.4850	0.6197	0.7235	0.7869	0.3066	0.3832	0.4782	0.6087	0.7200	0.7849
	50	0.4369	0.5405	0.6283	0.7529	0.8426	0.8868	0.4507	0.5490	0.6347	0.7645	0.8466	0.8881
40 %	60	0.6276	0.7033	0.7806	0.8597	0.9149	0.9420	0.6282	0.6972	0.7750	0.8597	0.9156	0.9429
	70	0.7371	0.7996	0.8558	0.9247	0.9583	0.9722	0.7305	0.7925	0.8464	0.9157	0.9544	0.9701
	80	0.8511	0.8910	0.9201	0.9617	0.9826	0.9889	0.8546	0.8880	0.9226	0.9598	0.9825	0.9888
	90	0.9020	0.9323	0.9587	0.9791	0.9893	0.9948	0.8980	0.9291	0.9554	0.9780	0.9897	0.9944
	100	0.9208	0.9502	0.9717	0.9905	0.9968	0.9982	0.9187	0.9492	0.9743	0.9904	0.9962	0.9980
	500	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	10	0.0055	0.0130	0.0230	0.0673	0.1296	0.1763	0.0044	0.0124	0.0253	0.0687	0.1296	0.1761
	20	0.0198	0.0380	0.0652	0.1293	0.2087	0.2718	0.0152	0.0332	0.0623	0.1236	0.2067	0.2708
	30	0.0552	0.0961	0.1477	0.2424	0.3475	0.4230	0.0552	0.0942	0.1477	0.2360	0.3447	0.4225
	40	0.1202	0.1768	0.2489	0.3709	0.4847	0.5703	0.1163	0.1757	0.2440	0.3633	0.4852	0.5678
	50	0.2062	0.2759	0.3501	0.4880	0.6023	0.6752	0.2016	0.2678	0.3436	0.4859	0.6022	0.6806
	60	0.3037	0.3820	0.4736	0.6109	0.7208	0.7794	0.2983	0.3811	0.4765	0.6128	0.7190	0.7788
60 %	70	0.3971	0.4952	0.5924	0.7215	0.8079	0.8555	0.3986	0.5020	0.5995	0.7248	0.8094	0.8588
	80	0.5078	0.5942	0.6776	0.7941	0.8734	0.9064	0.5095	0.5939	0.6776	0.7934	0.8754	0.9099
	90	0.6105	0.6832	0.7688	0.8558	0.9109	0.9381	0.6226	0.7004	0.7754	0.8586	0.9125	0.9405
	100	0.7022	0.7839	0.8362	0.9063	0.9478	0.9637	0.7012	0.7775	0.8308	0.9039	0.9450	0.9629
	500	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	20	0.0068	0.0157	0.0297	0.0627	0.1089	0.1619	0.0085	0.0178	0.0309	0.0644	0.1133	0.1677
	30	0.0143	0.0241	0.0395	0.0807	0.1427	0.2040	0.0119	0.0212	0.0366	0.0818	0.1434	0.2005
	40	0.0255	0.0415	0.0671	0.1326	0.2016	0.2709	0.0274	0.0410	0.0677	0.1326	0.2026	0.2721
	50	0.0388	0.0654	0.1093	0.1851	0.2712	0.3423	0.0382	0.0656	0.1045	0.1881	0.2706	0.3435
	60	0.0663	0.0969	0.1547	0.2498	0.3518	0.4152	0.0589	0.0911	0.1491	0.2407	0.3477	0.4131
80 %	70	0.1010	0.1404	0.1954	0.3020	0.4114	0.4879	0.1045	0.1466	0.2015	0.3112	0.4189	0.4922
	80	0.1417	0.1857	0.2546	0.3793	0.4974	0.5636	0.1387	0.1813	0.2574	0.3745	0.4885	0.5590
	90	0.1930	0.2517	0.3331	0.4572	0.5644	0.6360	0.1939	0.2513	0.3349	0.4590	0.5648	0.6404
	100	0.2330	0.3097	0.4002	0.5120	0.6204	0.6892	0.2274	0.3041	0.3933	0.5114	0.6183	0.6830
	500	0.9995	0.9996	0.9998	0.9999	1.0000	1.0000	0.9995	0.9996	0.9997	1.0000	1.0000	1.0000

Tabla 7: Potencia de la prueba para alternativas $Ga(3, 1)$ y $Ga(0.8, 1)$, con $M = 10,000$ réplicas de tamaño n

C	n	$Ga(3, 1)$											$Ga(0.8, 1)$										
		α											α										
		0.005	0.01	0.02	0.05	0.10	0.15	0.005	0.01	0.02	0.05	0.10	0.15	0.005	0.01	0.02	0.05	0.10	0.15				
10 %	10	0.0655	0.1061	0.1589	0.2536	0.3734	0.4518	0.1169	0.1700	0.2423	0.3519	0.4843	0.5619	0.4598	0.5685	0.6683	0.7964	0.8753	0.9108				
	20	0.2738	0.3742	0.4836	0.6371	0.7536	0.8189	0.8265	0.8797	0.9247	0.9670	0.9843	0.9910	0.9667	0.9818	0.9898	0.9962	0.9987	0.9995				
	30	0.6129	0.7082	0.7949	0.8905	0.9448	0.9671	0.9941	0.9974	0.9983	0.9994	0.9999	1.0000	0.9990	0.9996	0.9998	0.9999	1.0000	1.0000				
	40	0.8625	0.9140	0.9461	0.9777	0.9900	0.9942	0.9990	0.9996	0.9998	0.9999	1.0000	1.0000	0.9998	0.9999	1.0000	1.0000	1.0000	1.0000				
	50	0.9453	0.9685	0.9825	0.9938	0.9981	0.9992	0.9998	0.9999	1.0000	1.0000	1.0000	1.0000	0.9998	0.9999	1.0000	1.0000	1.0000	1.0000				
	60	0.9860	0.9932	0.9969	0.9994	0.9997	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000				
	70	0.9968	0.9985	0.9996	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000				
	80	0.9989	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000				
	90	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000				
	100	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000				
500	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000					
20 %	10	0.0513	0.0768	0.1079	0.2050	0.2974	0.3725	0.0849	0.1197	0.1581	0.2743	0.3853	0.4602	0.3481	0.4519	0.5553	0.6878	0.7923	0.8475				
	20	0.2069	0.2977	0.3989	0.5414	0.6602	0.7343	0.7105	0.7766	0.8427	0.9147	0.9556	0.9736	0.9075	0.9386	0.9654	0.9837	0.9912	0.9953				
	30	0.5053	0.5929	0.6873	0.8005	0.8823	0.9188	0.9776	0.9892	0.9948	0.9978	0.9993	0.9996	0.9776	0.9892	0.9948	0.9978	0.9993	0.9996				
	40	0.7562	0.8238	0.8815	0.9364	0.9662	0.9789	0.9947	0.9978	0.9986	0.9996	0.9997	0.9997	0.9947	0.9978	0.9986	0.9996	0.9996	0.9997				
	50	0.8999	0.9392	0.9667	0.9841	0.9929	0.9961	0.9993	0.9997	1.0000	1.0000	1.0000	1.0000	0.9993	0.9997	1.0000	1.0000	1.0000	1.0000				
	60	0.9579	0.9773	0.9882	0.9953	0.9983	0.9991	0.9998	0.9999	1.0000	1.0000	1.0000	1.0000	0.9998	0.9999	1.0000	1.0000	1.0000	1.0000				
	70	0.9856	0.9922	0.9962	0.9992	0.9997	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000				
	80	0.9974	0.9987	0.9994	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000				
	90	0.9996	0.9998	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000				
	100	0.9998	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000				
500	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000					

continúa...

	10	0.0212	0.0318	0.0520	0.1048	0.1791	0.2447	0.0296	0.0449	0.0699	0.1330	0.2076	0.2783
	20	0.0925	0.1421	0.2023	0.3306	0.4514	0.5331	0.1499	0.2166	0.2930	0.4368	0.5540	0.6270
	30	0.2870	0.3685	0.4547	0.6015	0.6996	0.7640	0.4259	0.5122	0.5959	0.7227	0.8005	0.8463
	40	0.4904	0.5715	0.6657	0.7755	0.8549	0.8934	0.6614	0.7274	0.7938	0.8763	0.9246	0.9482
	50	0.6701	0.7498	0.8123	0.8924	0.9366	0.9571	0.8165	0.8735	0.9109	0.9534	0.9760	0.9868
40 %	60	0.8330	0.8785	0.9174	0.9529	0.9737	0.9841	0.9311	0.9509	0.9698	0.9836	0.9913	0.9944
	70	0.9021	0.9340	0.9540	0.9801	0.9909	0.9939	0.9658	0.9777	0.9871	0.9958	0.9977	0.9993
	80	0.9589	0.9712	0.9807	0.9910	0.9964	0.9986	0.9900	0.9939	0.9969	0.9990	0.9998	0.9999
	90	0.9800	0.9864	0.9924	0.9971	0.9986	0.9995	0.9963	0.9975	0.9988	0.9996	1.0000	1.0000
	100	0.9869	0.9927	0.9966	0.9988	0.9997	1.0000	0.9982	0.9993	0.9994	0.9999	1.0000	1.0000
	500	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	10	0.0043	0.0140	0.0252	0.0690	0.1297	0.1754	0.0049	0.0138	0.0274	0.0741	0.1374	0.1865
	20	0.0277	0.0512	0.0868	0.1590	0.2460	0.3189	0.0365	0.0662	0.1081	0.1919	0.2860	0.3622
	30	0.0879	0.1437	0.2098	0.3195	0.4340	0.4997	0.1372	0.2026	0.2768	0.3987	0.5090	0.5795
	40	0.2018	0.2756	0.3599	0.4878	0.6009	0.6729	0.2819	0.3645	0.4523	0.5800	0.6874	0.7517
	50	0.3232	0.4024	0.4874	0.6176	0.7189	0.7800	0.4481	0.5297	0.6072	0.7248	0.8057	0.8503
60 %	60	0.4599	0.5471	0.6315	0.7469	0.8319	0.8721	0.5959	0.6729	0.7482	0.8362	0.8970	0.9258
	70	0.5815	0.6678	0.7461	0.8390	0.8996	0.9284	0.7189	0.7892	0.8472	0.9128	0.9509	0.9664
	80	0.6923	0.7585	0.8216	0.9019	0.9447	0.9606	0.8132	0.8658	0.9092	0.9556	0.9766	0.9850
	90	0.7985	0.8478	0.9009	0.9449	0.9669	0.9787	0.8954	0.9264	0.9520	0.9723	0.9845	0.9908
	100	0.8530	0.9001	0.9273	0.9650	0.9820	0.9890	0.9346	0.9558	0.9690	0.9843	0.9932	0.9953
	500	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	20	0.0064	0.0164	0.0317	0.0645	0.1161	0.1674	0.0085	0.0200	0.0358	0.0694	0.1214	0.1751
	30	0.0159	0.0262	0.0445	0.0966	0.1714	0.2281	0.0186	0.0327	0.0507	0.1052	0.1806	0.2429
	40	0.0381	0.0608	0.0900	0.1611	0.2456	0.3120	0.0463	0.0681	0.1026	0.1818	0.2657	0.3405
	50	0.0573	0.0863	0.1352	0.2212	0.3116	0.3863	0.0729	0.1098	0.1669	0.2772	0.3715	0.4419
	60	0.0915	0.1295	0.1983	0.3075	0.4161	0.4816	0.1186	0.1641	0.2405	0.3533	0.4660	0.5305
80 %	70	0.1586	0.2115	0.2705	0.3876	0.4965	0.5696	0.2086	0.2663	0.3386	0.4613	0.5690	0.6393
	80	0.2057	0.2584	0.3373	0.4670	0.5865	0.6497	0.2709	0.3316	0.4135	0.5376	0.6477	0.7060
	90	0.2786	0.3427	0.4230	0.5525	0.6501	0.7120	0.3615	0.4338	0.5189	0.6352	0.7272	0.7844
	100	0.3343	0.4233	0.5109	0.6198	0.7169	0.7682	0.4260	0.5135	0.6020	0.6985	0.7791	0.8238
	500	0.9998	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Tabla 8: Potencia estimada para las pruebas basadas en las estadísticas $W^2_{r,n}$ y $A^2_{r,n}$

C	n	Weib(0.5, 1)						Weib(3, 1)					
		$W^2_{r,n}$			$A^2_{r,n}$			$W^2_{r,n}$			$A^2_{r,n}$		
		0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
80 %	20	0.001	0.036	0.093	0.189	0.283	0.352	0.158	0.421	0.620	0.023	0.308	0.523
	40	0.081	0.197	0.303	0.364	0.512	0.596	0.545	0.815	0.914	0.373	0.762	0.892
	60	0.173	0.352	0.480	0.499	0.663	0.732	0.788	0.953	0.983	0.710	0.943	0.979
	80	0.291	0.498	0.622	0.628	0.775	0.836	0.924	0.988	0.997	0.895	0.990	0.998
	100	0.405	0.621	0.728	0.720	0.848	0.897	0.977	0.998	1.000	0.969	0.999	1.000
60 %	10	0.024	0.088	0.141	0.194	0.318	0.386	0.149	0.465	0.630	0.001	0.290	0.511
	20	0.130	0.269	0.368	0.372	0.524	0.600	0.538	0.830	0.916	0.309	0.765	0.890
	40	0.398	0.598	0.695	0.679	0.805	0.856	0.951	0.993	0.998	0.922	0.992	0.998
	60	0.626	0.792	0.859	0.842	0.924	0.951	0.998	1.000	1.000	0.997	1.000	1.000
	80	0.773	0.898	0.936	0.926	0.970	0.982	1.000	1.000	1.000	1.000	1.000	1.000
100	0.887	0.960	0.981	0.971	0.991	0.995	1.000	1.000	1.000	1.000	1.000	1.000	
40 %	10	0.133	0.255	0.341	0.325	0.481	0.555	0.338	0.689	0.819	0.070	0.557	0.765
	20	0.347	0.545	0.639	0.583	0.726	0.793	0.817	0.968	0.989	0.693	0.953	0.987
	40	0.707	0.844	0.899	0.870	0.940	0.961	0.999	1.000	1.000	0.998	1.000	1.000
	60	0.888	0.954	0.975	0.970	0.992	0.996	1.000	1.000	1.000	1.000	1.000	1.000
	80	0.965	0.989	0.996	0.993	0.998	0.999	1.000	1.000	1.000	1.000	1.000	1.000
100	0.990	0.998	0.999	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
20 %	10	0.276	0.440	0.530	0.474	0.633	0.702	0.541	0.862	0.937	0.236	0.772	0.906
	20	0.567	0.738	0.815	0.756	0.872	0.913	0.965	0.996	0.999	0.925	0.996	0.999
	40	0.904	0.963	0.979	0.971	0.990	0.994	1.000	1.000	1.000	1.000	1.000	1.000
	60	0.983	0.995	0.998	0.996	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	80	0.997	0.999	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
100	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
10 %	10	0.372	0.534	0.617	0.570	0.711	0.773	0.678	0.922	0.969	0.392	0.857	0.947
	20	0.684	0.823	0.879	0.838	0.919	0.946	0.991	1.000	1.000	0.973	0.999	1.000
	40	0.954	0.982	0.991	0.987	0.996	0.998	1.000	1.000	1.000	1.000	1.000	1.000
	60	0.995	0.999	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	80	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
100	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	

continúa...

C	n	$Ga(3,1)$						$Ga(0.8,1)$					
		$W^2_{r,n}$			$A^2_{r,n}$			$W^2_{r,n}$			$A^2_{r,n}$		
		0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
80%	20	0.088	0.277	0.452	0.010	0.185	0.360	0.006	0.028	0.072	0.027	0.072	0.124
	40	0.300	0.590	0.747	0.174	0.517	0.706	0.009	0.039	0.087	0.042	0.098	0.153
	60	0.487	0.778	0.881	0.398	0.763	0.883	0.010	0.048	0.097	0.043	0.114	0.176
	80	0.676	0.895	0.953	0.619	0.902	0.961	0.016	0.061	0.119	0.050	0.131	0.197
	100	0.814	0.952	0.983	0.793	0.965	0.989	0.020	0.075	0.131	0.059	0.152	0.226
60%	10	0.062	0.267	0.410	0.000	0.151	0.306	0.007	0.042	0.091	0.028	0.085	0.138
	20	0.223	0.518	0.678	0.093	0.428	0.627	0.012	0.050	0.101	0.040	0.101	0.160
	40	0.619	0.857	0.928	0.546	0.862	0.940	0.021	0.071	0.128	0.055	0.131	0.199
	60	0.856	0.968	0.987	0.848	0.977	0.993	0.029	0.094	0.159	0.070	0.165	0.244
	80	0.957	0.995	0.999	0.966	0.997	0.999	0.041	0.127	0.193	0.088	0.206	0.296
40%	100	0.993	1.000	1.000	0.994	1.000	1.000	0.052	0.145	0.230	0.108	0.241	0.329
	10	0.111	0.346	0.504	0.016	0.240	0.429	0.014	0.056	0.104	0.028	0.088	0.144
	20	0.351	0.683	0.814	0.218	0.628	0.795	0.020	0.073	0.129	0.044	0.116	0.184
	40	0.819	0.954	0.982	0.786	0.964	0.989	0.039	0.105	0.166	0.068	0.169	0.246
	60	0.971	0.997	0.999	0.975	0.999	1.000	0.049	0.142	0.218	0.089	0.212	0.298
20%	80	0.996	1.000	1.000	0.999	1.000	1.000	0.065	0.172	0.257	0.115	0.251	0.340
	100	1.000	1.000	1.000	1.000	1.000	1.000	0.079	0.204	0.290	0.143	0.298	0.390
	10	0.152	0.440	0.602	0.042	0.326	0.530	0.020	0.066	0.115	0.034	0.105	0.167
	20	0.443	0.772	0.889	0.348	0.759	0.888	0.027	0.088	0.149	0.051	0.135	0.207
	40	0.918	0.986	0.996	0.930	0.991	0.997	0.048	0.128	0.203	0.089	0.191	0.272
10%	60	0.994	1.000	1.000	0.997	1.000	1.000	0.067	0.178	0.266	0.113	0.256	0.347
	80	1.000	1.000	1.000	1.000	1.000	1.000	0.090	0.221	0.321	0.153	0.313	0.414
	100	1.000	1.000	1.000	1.000	1.000	1.000	0.117	0.267	0.368	0.184	0.364	0.467
	10	0.191	0.492	0.649	0.066	0.374	0.574	0.024	0.075	0.127	0.043	0.112	0.169
	20	0.538	0.834	0.911	0.424	0.817	0.915	0.031	0.092	0.155	0.055	0.141	0.213
10%	40	0.956	0.996	0.999	0.969	0.997	1.000	0.056	0.150	0.229	0.102	0.215	0.302
	60	0.999	1.000	1.000	0.999	1.000	1.000	0.079	0.196	0.285	0.131	0.275	0.369
	80	1.000	1.000	1.000	1.000	1.000	1.000	0.104	0.239	0.338	0.160	0.328	0.431
100	1.000	1.000	1.000	1.000	1.000	1.000	0.128	0.276	0.385	0.196	0.381	0.482	

continúa...

C	n	$LNorm(5,3)$						$LNorm(1,1)$					
		$W^2_{r,n}$			$A^2_{r,n}$			$W^2_{r,n}$			$A^2_{r,n}$		
		0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
80%	20	0.002	0.022	0.067	0.079	0.145	0.202	0.047	0.165	0.304	0.004	0.098	0.216
	40	0.039	0.116	0.197	0.163	0.276	0.361	0.118	0.334	0.490	0.064	0.290	0.466
	60	0.090	0.225	0.330	0.250	0.404	0.494	0.204	0.476	0.638	0.146	0.478	0.653
	80	0.160	0.328	0.446	0.325	0.507	0.598	0.301	0.597	0.746	0.263	0.631	0.790
	100	0.231	0.426	0.544	0.405	0.591	0.683	0.410	0.700	0.821	0.392	0.764	0.882
60%	10	0.024	0.094	0.151	0.156	0.272	0.338	0.029	0.125	0.213	0.000	0.065	0.147
	20	0.154	0.303	0.404	0.342	0.496	0.574	0.055	0.199	0.314	0.020	0.150	0.273
	40	0.455	0.647	0.735	0.649	0.787	0.844	0.139	0.356	0.496	0.111	0.364	0.527
	60	0.703	0.846	0.900	0.835	0.922	0.950	0.235	0.505	0.648	0.241	0.566	0.736
	80	0.845	0.934	0.962	0.927	0.970	0.983	0.349	0.637	0.767	0.391	0.734	0.862
40%	100	0.926	0.974	0.987	0.968	0.990	0.995	0.474	0.746	0.857	0.562	0.851	0.932
	10	0.234	0.375	0.471	0.408	0.559	0.629	0.022	0.103	0.182	0.002	0.060	0.136
	20	0.552	0.722	0.795	0.715	0.830	0.873	0.040	0.162	0.262	0.017	0.136	0.258
	40	0.904	0.959	0.976	0.955	0.981	0.989	0.102	0.278	0.416	0.082	0.317	0.481
	60	0.983	0.995	0.998	0.996	0.999	1.000	0.181	0.419	0.570	0.208	0.522	0.684
20%	80	0.998	1.000	1.000	0.999	1.000	1.000	0.262	0.542	0.689	0.353	0.695	0.829
	100	1.000	1.000	1.000	1.000	1.000	1.000	0.353	0.656	0.788	0.510	0.822	0.912
	10	0.554	0.697	0.764	0.713	0.820	0.864	0.021	0.087	0.161	0.005	0.054	0.122
	20	0.879	0.942	0.963	0.937	0.972	0.980	0.021	0.105	0.204	0.016	0.106	0.205
	40	0.997	0.999	1.000	0.999	1.000	1.000	0.049	0.189	0.317	0.066	0.235	0.386
10%	60	1.000	1.000	1.000	1.000	1.000	1.000	0.094	0.293	0.444	0.125	0.404	0.588
	80	1.000	1.000	1.000	1.000	1.000	1.000	0.131	0.379	0.555	0.216	0.568	0.740
	100	1.000	1.000	1.000	1.000	1.000	1.000	0.201	0.488	0.666	0.330	0.703	0.845
	10	0.727	0.831	0.872	0.837	0.903	0.928	0.020	0.084	0.152	0.009	0.057	0.123
	20	0.960	0.984	0.990	0.982	0.994	0.996	0.026	0.110	0.201	0.015	0.095	0.181
10%	40	1.000	1.000	1.000	1.000	1.000	1.000	0.048	0.180	0.302	0.054	0.213	0.363
	60	1.000	1.000	1.000	1.000	1.000	1.000	0.080	0.258	0.420	0.095	0.343	0.522
	80	1.000	1.000	1.000	1.000	1.000	1.000	0.106	0.348	0.521	0.161	0.488	0.678
10%	100	1.000	1.000	1.000	1.000	1.000	1.000	0.151	0.440	0.618	0.261	0.644	0.805

Tabla 9: Cuantiles de las estadísticas $W_{r,n}^2$ y $A_{r,n}^2$ para probar exponencialidad con parámetro de escala desconocido y origen conocido cuando existe censura por la derecha Tipo II. Tomado de [Stephens \(1986b\)](#)

C	n	Estadística $W_{r,n}^2$			Estadística $A_{r,n}^2$		
		Nivel de significancia			Nivel de significancia		
		0.01	0.05	0.10	0.01	0.05	0.10
80 %	20	0.014	0.018	0.025	0.188	0.232	0.325
	40	0.013	0.017	0.025	0.189	0.241	0.355
	60	0.013	0.017	0.026	0.192	0.244	0.373
	80	0.013	0.017	0.026	0.194	0.249	0.385
	100	0.013	0.017	0.026	0.195	0.252	0.394
60 %	10	0.045	0.055	0.079	0.363	0.445	0.671
	20	0.044	0.056	0.083	0.379	0.477	0.719
	40	0.044	0.056	0.084	0.382	0.485	0.736
	60	0.044	0.056	0.085	0.382	0.493	0.753
	80	0.044	0.056	0.086	0.385	0.496	0.762
	100	0.043	0.056	0.086	0.388	0.497	0.767
40 %	10	0.084	0.104	0.149	0.549	0.684	1.058
	20	0.084	0.106	0.161	0.568	0.721	1.104
	40	0.085	0.109	0.161	0.580	0.733	1.126
	60	0.085	0.109	0.164	0.580	0.730	1.126
	80	0.085	0.109	0.166	0.580	0.731	1.128
	100	0.086	0.109	0.167	0.581	0.733	1.130
20 %	10	0.126	0.156	0.229	0.740	0.929	1.434
	20	0.130	0.167	0.253	0.773	0.979	1.512
	40	0.131	0.167	0.253	0.782	0.985	1.465
	60	0.132	0.168	0.256	0.782	0.989	1.516
	80	0.132	0.169	0.257	0.783	0.991	1.529
	100	0.132	0.169	0.258	0.785	0.993	1.533
10 %	10	0.147	0.182	0.265	0.852	1.059	1.584
	20	0.155	0.194	0.289	0.894	1.117	1.706
	40	0.152	0.193	0.29	0.893	1.117	1.633
	60	0.153	0.195	0.294	0.897	1.127	1.699
	80	0.154	0.196	0.297	0.900	1.132	1.720
	100	0.155	0.196	0.298	0.902	1.135	1.728

Códigos en R

```
#PARA GENERAR LOS CUANTILES EMPÍRICOS DE LA DISTRIBUCIÓN DEL ESTADÍSTICO RN-A
REx1 = function(n,p)
{
#Genera variables aleatorias pareto
pareto=function(beta, theta)
{
u = runif(1, 0, 1)
x = theta/((1-u)^(1/beta))
}

#Valores de los parámetros
beta = 3
theta = 0.5

#Muestra aleatoria Pareto de tamaño n
sample = rep(0,n)
for(i in 1:n)
{
sample[i] = pareto(beta, theta)
}

#Ordena la muestra aleatoria Pareto
orden = sort(sample)

#Censura tipo II. Observaciones
#r es el número total de unidades observadas
#p es el porcentaje de observaciones no censuradas
r = p*n
Observ = rep(0,r)
for(i in 1:r)
{
Observ[i] = orden[i]
}

#Transformación a exponencial dos parámetros
TExp2 = log(Observ)
```

Apéndice

```
#Transformación a exponencial un parámetro
TExp1 = rep(0,r-1)
for(i in 2:r)
{TExp1[i-1] = TExp2[i] - TExp2[1]}

#Valores de las observaciones exactas y frecuencias de datos transformados
obexac = as.numeric(rownames(as.matrix(table(TExp1))))
frec1 = as.vector(table(TExp1))

#Número de individuos en riesgo en cada obexac
k = length(obexac)
Y = rep(0,k)
Y[1] = n-1
for(i in 2:k)
{
Y[i] = Y[i-1] - frec1[i-1]
}
d = frec1

#Estimador N-A
N.A1 = rep(0,k)
for(i in 1:k)
{
N.A1[i] = d[i]/Y[i]
}
Ht = matrix(0, nrow=k, ncol=2)
for(i in 1:k)
{
Ht[i,1] = obexac[i]
Ht[i,2] = sum(N.A1[1:i])
}
colnames(Ht) = c("t", "H(t)")

#Estadístico de correlación basado en el estimador Nelson-Aalen
RNA = cor(Ht[,2], obexac)
RNA
}
n = 15
```

Apéndice

```
p = 0.4
M = 10000
Rtest = rep(0,M)
for(i in 1:M)
{
Rtest[i]=REx1(n,p)
}
alpha = c(0.5,1,2,5,10,15)/100
cuantil = quantile(Rtest, alpha)
cuantil
-----

#PARA ESTIMAR EL TAMAÑO DE LA PRUEBA
REx1 = function(n,p)
{
#Genera variables aleatorias Pareto
pareto=function(beta, theta)
{
u = runif(1, 0, 1)
x = theta/((1-u)^(1/beta))
}

#Valores de los parámetros
beta = 0.5
theta = 100

#Muestra aleatoria Pareto de tamaño n
sample = rep(0,n)
for(i in 1:n)
{
sample[i] = pareto(beta, theta)
}

#Ordena la muestra aleatoria Pareto
orden = sort(sample)

#Censura tipo II. Observaciones
#r es el número total de unidades observadas
```


Apéndice

```
#p es el porcentaje de observaciones no censuradas
r = p*n
Observ = rep(0,r)
for(i in 1:r)
{
Observ[i] = orden[i]
}

#Transformación a exponencial dos parámetros
TExp2 = log(Observ)
#Transformación a exponencial un parámetro
TExp1 = rep(0,r-1)
for(i in 2:r)
{TExp1[i-1] = TExp2[i] - TExp2[1]}

#Valores de las observaciones exactas y frecuencias de datos transformados
obexac = as.numeric(rownames(as.matrix(table(TExp1))))
frec1 = as.vector(table(TExp1))

#Número de individuos en riesgo en cada obexac
k = length(obexac)
Y = rep(0,k)
Y[1] = n-1
for(i in 2:k)
{
Y[i] = Y[i-1] - frec1[i-1]
}
d = frec1

#Estimador N-A
N.A1 = rep(0,k)
for(i in 1:k)
{
N.A1[i] = d[i]/Y[i]
}
Ht = matrix(0, nrow=k, ncol=2)
for(i in 1:k)
{
```

Apéndice

```
Ht[i,1] = obexac[i]
Ht[i,2] = sum(N.A1[1:i])
}
colnames(Ht) = c("t", "H(t)")

#Estadístico de correlación basado en el estimador Nelson-Aalen
RNA = cor(Ht[,2], obexac)
RNA
}
n = 20
p = 0.2
M = 10000
Rtest = rep(0,M)
for(i in 1:M)
{
Rtest[i]=REx1(n,p)
}

#Contar el número de veces que se rechaza H0 siendo cierta
conteo = matrix(0, nrow=M, ncol=6)
errorI = rep(0,6)
cuantiles = c(0.8602075,0.8633920,0.8682409,0.8775143,0.8877058,0.8985424)
#Decisión de rechazar o no rechazar H0
for(i in 1:M)
{
for(j in 1:6)
{
if(Rtest[i]<cuantiles[j])
{conteo[i,j] = c(1)}
}
}

#Tamaño estimado de la prueba
for(i in 1:6)
{
errorI[i] = sum(conteo[,i])/M
}
errorI
```

Apéndice

```
-----

#PARA ESTIMAR LA POTENCIA DE LA PRUEBA RN-A
AltWeib = function(n,p)
{
#Valores de los parámetros de la distribución alternativa
forma = 0.5
escala = 1

#Generar variables aleatorias de la distrib alternativa, en este caso Weibull
weibull=rweibull(n, forma, escala)

#Muestra Weibull ordenada
orden = sort(weibull)

#Censura tipo II. Observaciones
#r es el número total de unidades observadas
#p es el número de observaciones no censurados
r = p*n
Observ = rep(0,r)
for(i in 1:r)
{
Observ[i] = orden[i]
}

#Transformación logaritmo
Tln = log(Observ)
#Doble transformación
T2 = rep(0,r-1)
for(i in 2:r)
{T2[i-1] = Tln[i] - Tln[1]}

#Valores de las observaciones exactas y frecuencias de datos transformados
obexac = as.numeric(rownames(as.matrix(table(T2))))
frec1 = as.vector(table(T2))

#Número de individuos en riesgo en cada obexac
k = length(obexac)
```

Apéndice

```
Y = rep(0,k)
Y[1] = n-1
for(i in 2:k)
{
Y[i] = Y[i-1] - frec1[i-1]
}
d = frec1

#Estimador N-A
N.A1 = rep(0,k)
for(i in 1:k)
{
N.A1[i] = d[i]/Y[i]
}
Ht = matrix(0, nrow=k, ncol=2)
for(i in 1:k)
{
Ht[i,1] = obexac[i]
Ht[i,2] = sum(N.A1[1:i])
}
colnames(Ht) = c("t", "H(t)")

#Estadístico de correlación basado en el estimador Nelson-Aalen
RWeib = cor(Ht[,2], obexac)
RWeib
}
n = 500
p = 0.2
M = 10000
Rtest = rep(0,M)
for(i in 1:M)
{
Rtest[i]=AltWeib(n,p)
}

#Contar el número de veces que se rechaza H0 siendo falsa
conteo = matrix(0, nrow=M, ncol=6)
potencia = rep(0,6)
```

Apéndice

```
cuantiles = c(0.9853892,0.9872664,0.9889218,0.9911080,0.9928475,0.9938726)
#Decisión de rechazar o no rechazar H0
for(i in 1:M)
{
for(j in 1:6)
{
if(Rtest[i]<cuantiles[j])
{conteo[i,j] = c(1)}
}
}

#Potencia estimada de la prueba
for(i in 1:6)
{
potencia[i] = sum(conteo[,i])/M
}
potencia
-----

#PARA ESTIMAR LA POTENCIA DE LA ESTADÍSTICA ANDERSON-DARLING
FA = function(n, r)
{
#Distribución alternativa
#Valores de los parámetros
forma = 3
escala = 1

#Genera variables aleatorias weibull
weib=rweibull(n, forma, escala)

#Muestra weibull ordenada
orden = sort(weib)

#Estimador del parámetro de escala de una exponencial con censura
suma = 0
for(i in 1:r)
{
suma = suma + orden[i]
```

Apéndice

```
}
est = r/(suma + (n-r)*orden[r])

#Calcular Zi, paso 7 del algoritmo 5.1
Z = rep(0,r)
for(i in 1:r)
{
Z[i] = 1-exp(-orden[i]*est)
}

#Estadística Anderson-Darling para censura tipo II
sum2 = 0
sum3 = 0
for(i in 1:r)
{
sum2 = sum2 + ((-1/n)*((2*i)-1)*(log(Z[i])-log(1-Z[i])))
sum3 = sum3 - (2*log(1-Z[i]))
}
A = sum2 + sum3 - (1/n)*(((r-n)^2)*log(1-Z[r])-(r^2)*log(Z[r])+(n^2)*Z[r])
A
}

#Valores
#obs es la proporción de datos no censurados
n = 100
obs = .2
r = n*obs
M = 10000
Atest = rep(0,M)
for(i in 1:M)
{
Atest[i] = FA(n,r)
}

#Contar el número de rechazos de H0 siendo falsa
conteo = matrix(0, nrow=v, ncol=7)
potencia = rep(0,7)
cuantiles = c(0.078,0.126,0.163,0.195,0.252,0.311,0.394)
```

Apéndice

```
for(i in 1:M)
{
for(j in 1:7)
{
if(Atest[i]>cuantiles[j])
{conteo[i,j] = c(1)}
}
}

#Potencia estimada
for(i in 1:7)
{
potencia[i] = sum(conteo[,i])/M
}
potencia
-----

#PARA ESTIMAR LA POTENCIA DE LA ESTADÍSTICA CRAMER-VON MISES
FW = function(n, r)
{
#Distribución alternativa
#Valores de los parámetros
forma = 3
escala = 1

#Genera variables aleatorias Weibull
weib=rweibull(n, forma, escala)

#Muestra weibull ordenada
orden = sort(weib)

#Estimador del parámetro de escala de una exponencial con censura
suma = 0
for(i in 1:r)
{
suma = suma + orden[i]
}
est = r/(suma + (n-r)*orden[r])
```

Apéndice

```
#Calcular Zi, paso 7 del algoritmo 5.1
Z = rep(0,r)
for(i in 1:r)
{
Z[i] = 1-exp(-orden[i]*est)
}

#Estadística Cramer-von Mises para censura tipo II
sum = 0
for(i in 1:r)
{
sum = sum + (Z[i] - (2*i-1)/(2*n))^2
}
W = sum + r/(12*(n^2)) + (n/3)*(Z[r]-(r/n))^3
W
}

#Valores
#obs es la proporción de datos no censurados
n = 100
obs = 0.2
r = n*obs
m = 10000
Wtest = rep(0,M)
for(i in 1:v)
{
Wtest[i] = FW(n,r)
}

#Contar el número de rechazos de H0 siendo falsa
conteo = matrix(0, nrow=M, ncol=7)
potencia = rep(0,7)
cuantiles = c(0.005,0.008,0.011,0.013,0.017,0.020,0.026)
for(i in 1:v)
{
for(j in 1:7)
{
```


Apéndice

```
if(Wtest[i]>cuantiles[j])
{conteo[i,j] = c(1)}
}
}

#Potencia estimada
for(i in 1:7)
{
potencia[i] = sum(conteo[,i])/v
}
potencia
-----

#PARA ESTIMAR LA POTENCIA DE LA ESTADÍSTICA DE INF DE KULLBACK-LEIBLER
KLAlt = function(n,p)
{
#Muestra aleatoria de la distribución alternativa
#Valores de los parámetros
forma = 3
escala = 1

#Genera variables aleatorias weibull
X = rweibull(n,forma,escala)

#Ordenando la muestra
Xorden = sort(X)

#Muestra censurada
r = p*n
Xc = rep(0,r)
for(i in 1:r)
{
Xc[i] = Xorden[i]
}

#Estimador de la función H barra
m = 0
if(5<=r && r<=7){m=2}
```

Apéndice

```
if(8<=r && r<=15){m=3}
if(16<=r && r<=30){m=4}
if(31<=r && r<=40){m=5}
if(41<=r && r<=50){m=6}
suma1 = 0
suma2 = 0
suma3 = 0
for(i in 1:m)
{
suma1 = suma1 + log((n/(2*m))*(Xc[i+m]-Xc[1]))
}
for(i in (m+1):(r-m))
{
suma2 = suma2 + log((n/(2*m))*(Xc[i+m]-Xc[i-m]))
}
for(i in (r-m+1):r)
{
suma3 = suma3 + log((n/(2*m))*(Xc[r]-Xc[i-m]))
}
suma = suma1 + suma2 + suma3
H = (1/n)*suma - (1-(r/n))*log(1-(r/n))

#Estadístico de prueba para  $\exp(1/\beta)$  basado en la información de K-L
T = -H + (r/n)*(log((1/r)*(sum(Xc)+((n-r)*Xc[r])))+1)
T
}

#Valores
#p es la proporción de unidades observadas
p = 0.8
n = 50
M = 10000
Ttest = rep(0,M)
for(i in 1:v)
{
Ttest[i]=KLAlt(n,p)
}
}
```

Apéndice

```
#Contar el número de rechazos de H0 siendo falsa
conteo = matrix(0, nrow=M, ncol=4)
potencia = rep(0,4)
cuantiles = c(0.1565,0.1742,0.1899,0.2097)
for(i in 1:M)
{
for(j in 1:4)
{
if(Ttest[i]>cuantiles[j])
{conteo[i,j] = c(1)}
}
}

#Estimador de la potencia
for(i in 1:4)
{
potencia[i] = sum(conteo[,i])/M
}
potencia
-----
```