UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO

DIVISIÓN ACADÉMICA DE CIENCIAS BÁSICAS

INFERENCIA ESTADÍSTICA EN PRESENCIA DE CENSURA Y TRUNCAMIENTO

T E S I S

QUE PARA OBTENER EL GRADO DE

MAESTRO EN CIENCIAS
EN
MATEMÁTICAS APLICADAS

P R E S E N T A :

HEBER RODRÍGUEZ CARRERA

DIRECTOR DE TESIS :

DR. FIDEL ULÍN MONTEJO

CUNDUACÁN, TAB., MÉXICO AGOSTO, 2018

Carta Autorización

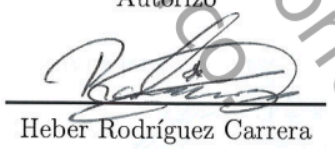
El que suscribe, autoriza por medio del presente escrito a la Universidad Juárez Autónoma de Tabasco para que sea utilizada tanto física como digitalmente la tesis de maestría: **INFERENCIA ESTADÍSTICA EN PRESENCIA DE CENSURA Y TRUNCAMIENTO**, de la cual soy autor y titular de los derechos de autor.

La finalidad del uso por parte de la Universidad Juárez Autónoma de Tabasco de la tesis antes mencionada, será única y exclusivamente para difusión, educación y sin fines de lucro; autorización que se hace de manera enunciativa más no limitativa para subirla a la Red Abierta de Bibliotecas Digitales (RABID) y a cualquier otra red académica con las que la Universidad tenga relación institucional.

Por lo antes manifestado, libero a la Universidad Juárez Autónoma de Tabasco de cualquier reclamación legal que pudiera ejercer respecto al uso y manipulación de la tesis mencionada y para los fines estipulados en éste documento.

Se firma la presente autorización en la ciudad de Villahermosa, Tabasco a los 01 días del mes de Agosto del año 2018.

Autorizó



Heber Rodríguez Carrera
102A10003

UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO
"ESTUDIO EN LA DUDA. ACCIÓN EN LA FE"

División Académica de Ciencias Básicas

UJAT 1958 - 2018 PATRIMONIO DE TABASCO

DIRECCIÓN

28 de junio de 2018

Lic. Heber Rodríguez Carrera
Maestría en Ciencias en Matemáticas Aplicadas
Presente

Por medio del presente y de la manera más cordial, me dirijo a Usted para hacer de su conocimiento que proceda a la impresión del trabajo titulado **"INFERENCIA ESTADÍSTICA EN PRESENCIA DE CENSURA Y TRUNCAMIENTO"** en virtud de que reúne los requisitos para el EXAMEN PROFESIONAL para obtener el grado de Maestro en Ciencias en Matemáticas Aplicadas.

Sin otro particular, reciba un cordial saludo.

A t e n t a m e n t e

Dr. Gerardo Delgadillo Piñón
Director

C.c.p.- Archivo
Dr'GDP/Dr'JGPS/emt

Miembro CUMEX desde 2008 Consorcio de Universidades Mexicanas
UNA ALIANZA DE CALIDAD POR LA EDUCACIÓN SUPERIOR

Km.1 Carretera Cunduacán-Jalpa de Méndez, A.P. 24, C.P. 86690, Cunduacán, Tab., México.
Tel/Fax: (993) 3581500 Ext. 6701 E-Mail: direccion.dacb@ujat.mx

www.ujat.mx

Índice general

A mis padres

Catalina y Juventino

A mis hermanos

Ana María, Omar y Obed

Agradecimientos

A Dios el gran creador del universo.

Al Dr. Fidel Ulín Montejo por haber aceptado ser el director de esta tesis y por su paciencia para guiarme durante todo el desarrollo.

A los profesores de la Academia de Matemáticas de la Universidad Juárez Autónoma de Tabasco por transmitir de manera muy profesional sus conocimientos.

Al Dr. Edilberto Nájera Rangel, sus observaciones y sugerencias ayudaron a mejorar significativamente este trabajo.

A mis padres porque ellos me enseñaron a trabajar y me apoyaron en tiempos difíciles.

A mis amigos de la universidad quienes de una u otra forma me apoyaron, gracias por su amistad brindada en cada momento.

INFERENCIA ESTADÍSTICA EN PRESENCIA DE CENSURA Y TRUNCAMIENTO

Capítulo 1

Introducción

La estadística es realmente muy importante en el desarrollo de investigaciones en diversos campos, cualquiera que sea nuestra actividad científica o tecnológica; cada vez son más los profesionales de diferentes disciplinas que requieren de métodos estadísticos como muestreo, simulación, diseño de experimentos, modelación estadística e inferencia, para llevar a cabo recolección, compendio y análisis de datos para su posterior interpretación. En términos más precisos, la estadística es el estudio de los fenómenos aleatorios [36].

El aspecto más importante de la estadística es la obtención de conclusiones sobre una población en estudio, a partir de la información que proporciona una muestra representativa de la misma. Esta muestra es obtenida por observación o experimentación. A este proceso se le define como inferencia estadística [8].

20 En este sentido la ciencia de la estadística tiene, virtualmente, un alcance ilimitado de aplicaciones en un espectro tan amplio de disciplinas y actividades humanas como lo son: la sociología (para comparar el comportamiento de grupos socioeconómicos y culturales y en el estudio de su comportamiento), geografía (natalidad, mortalidad, estructura socio demográfica de la población), psicología (para medir y comparar la conducta, las actitudes, la inteligencia y las aptitudes del hombre), la medicina (grado de eficiencia de un medicamento), las ingenierías (control de calidad), las leyes (en el recuento de conductas como el crimen y el suicidio), la economía (suministra los valores que ayudan a descubrir interrelaciones entre múltiples parámetros macro y microeconómicos), la sismología (en la estimación del riesgo sísmico, predicción sísmica, determinación de magnitudes y cuantificación de incertidumbre), la genética (genética cuantitativa y la genética de poblaciones), la biología (para estudiar las reacciones de las plantas y los animales ante diferentes períodos ambientales y para investigar la herencia), informática y la actuaría (seguros), por mencionar algunas.

19 En muchas situaciones prácticas, se presentan muestras donde no es posible obtener la información de todas las unidades muestrales o no es posible revelar enteramente los

datos de las unidades de análisis debido a factores que escapan del control experimental presentándose la pérdida de información. Estas muestras se denominan "censuradas", en algunas ocasiones se trata de pocos casos censurados, pero en otras las restricciones en el muestreo son severas y se deben de tener en cuenta para realizar el análisis. Así, una observación está censurada cuando solo contiene información parcial sobre la variable a estudiar y cuyas características no se han podido incluir en la muestra. Los tipos de censura que suelen considerarse son: censura por la izquierda, censura por la derecha y censura por intervalo.

En algunas ocasiones ciertas observaciones de una muestra son sistemáticamente excluidas ya que el evento de interés se presenta fuera de nuestro intervalo dado, a dicha muestra se le considera truncada. Los tipos de truncamiento que suelen considerarse son: truncamiento por la izquierda y truncamiento por la derecha.

En este trabajo se muestra el procedimiento de máxima verosimilitud como una herramienta para la inferencia y el análisis estadístico para muestras de datos que incluyan observaciones censuradas por la derecha, truncadas por la izquierda y censuradas por la derecha y truncadas por la izquierda, se describen los modelos probabilísticos más usados en la literatura en estudios donde se observan este tipo de datos, se propone el criterio de razón de verosimilitudes y el criterio de Akaike para seleccionar el mejor modelo probabilístico para una muestra con datos censurados por la derecha, truncados por la izquierda y censurados por la derecha y truncados por la izquierda. Finalmente, se implementan algoritmos bajo la plataforma de programación R para calcular los estimadores de máxima verosimilitud de los parámetros, de los modelos probabilísticos seleccionados.

Capítulo 2

Censura y Truncamiento

2.1. Censura

En un estudio práctico se presentan con frecuencia muestras con observaciones incompletas o donde se tiene pérdida de información en diferentes formas que crean problemas con su análisis. Estas muestras se denominan "censuradas", ya que la censura hace referencia a situaciones en las que hay una pérdida de información en la variable de interés. En algunas ocasiones se trata de pocos casos censurados, pero en otras las restricciones en el muestreo son severas y se deben tener en cuenta para realizar el análisis. Así, muestras censuradas son aquellas para las cuales se conoce y se identifica el número de observaciones cuyas características no se han podido incluir en la muestra.

Los principales tipos de censura que suelen considerarse son: censura por la izquierda, censura por la derecha y censura por intervalo.

2.1.1. Censura por la izquierda

Una observación Y_i obtenida de algún estudio, experimento o fenómeno de interés especifico, es considerada como censurada por la izquierda si es menor que un valor de censura conocido C_l, es decir, se desconoce el valor exacto de la observación y solo se sabe que es menor que C_l, por lo que se registra como un valor desconocido menor que C_l. Si Y_i es mayor o igual a C_l, entonces se reporta su valor numérico.

2.1.2. Censura por la derecha

Una observación Y_i obtenida de algún estudio, experimento o fenómeno de interés especifico, es considerada como censurada por la derecha si es mayor que un valor de censura conocido C_r, es decir, se desconoce el valor exacto de la observación y solo se sabe que esta es mayor que C_r. El valor exacto de la observación solo es conocido cuando Y_i es menor o igual a C_r y en este caso se registra su valor exacto.

2.2. Datos truncados

Una segunda característica de muchos estudios, principalmente en estudios de supervivencia, y que a veces se confunde con la censura, es la truncación. La truncación se produce cuando el evento de interés es observado dentro de un intervalo (T_L, T_R), llamado también ventana de observación. Si el evento de interés sucede fuera de este intevalo, dicha observación será excluida por el investigador. Esto está en contraste con la censura donde hay por lo menos información parcial sobre cada observación [15]. Al realizar un truncamiento, la muestra excluye aquellas observaciones que son censuradas.

Los datos truncados se clasifican en: Datos truncados por la izquierda y datos truncados por la derecha.

2.2.1. Truncamiento por la izquierda

Se dice que una observación Y_i es truncada por la izquierda en T_L si su valor es menor que T_L; esta no se registra. Sin embargo, si su valor es mayor que T_L entonces se registra su valor numérico observado. Es decir, Y_i es observado si y solo si $T_L < Y_i$.

2.2.2. Truncamiento por la derecha

Se dice que una observación Y_i es truncada por la derecha en T_R si su valor es mayor que T_R; de igual modo, no se reporta. Pero cuando es menor que T_R, sí. En otras palabras solo se toman en cuenta aquellas observaciones para cuyo caso se cumple que $Y_i < T_R$.

Capítulo 3

Distribuciones Continuas

3.1. Teoría general de distribuciones continuas

En esta sección presentaremos los conceptos básicos esenciales y la teoría de distribuciones estadísticas continuas. Tales distribuciones son las que se utilizan comunmente para analizar datos en diversas áreas, como en la estadística ambiental para monitorear la calidad del aire y agua, en la física para modelar el comportamiento teórico de diferentes fenómenos aleatorios observables que aparecen en el mundo real, en medicina para modelar los tiempos de vida de un individuo o grupos de individuos, en confiabilidad para modelar las características y fiabilidad de componentes y sistemas, en estadística actuarial para modelizar siniestros con grandes costes o cuantías por reclamación,en control de calidad se utilizan para la representación de resistencia a la tracción, resistencia al impacto, y muchas otras propiedades [2],[4], [18],[10],[39]. Las distribuciones más utilizadas son: exponencial, normal, lognormal, Weibull, Pareto y las distribuciones de valores extremos.

3.1.1. Función de densidad de probabilidad

Definición 3.1.1 *Si Y es una variable aleatoria continua, entonces la función de densidad de probabilidad (fdp) de Y, o bien la densidad de Y, es una función f(y) integrable en todos los reales, que cumple con las condiciones siguientes:*

1. $f(y) \geq 0$ *para toda* y, $-\infty < y < \infty$.
2. $\int_{-\infty}^{\infty} f(y)dy = 1$.
3. *Para cualesquiera reales* a *y* b *tales que* $a \leq b$, *tenemos*

$$P(a \leq y \leq b) = \int_a^b f(y)dy.$$

Así, la probabilidad de que Y tome valores en el intervalo [a,b] es calculada integrando la *fdp* sobre el intervalo de tiempo deseado.

La función de densidad de probabilidad es un *modelo teórico* para la distribución de frecuencia de una población de medidas [40].

3.1.2. Función de distribución acumulada

Definición 3.1.2 *Si Y es una variable aleatoria con función de densidad f(y), se denomina Función de distribución acumulada (fda) de la variable aleatoria Y, o bien la función de distribución de Y, a la función F(y) definida por:*

$$F(y) = P(Y \leq y) = \int_{-\infty}^{y} f(t)dt, \quad para - \infty < y < \infty. \tag{3.1}$$

Esto denota la probabilidad de que la variable aleatoria Y tome valores menores o iguales a y.

La función de distribución tiene las siguientes propiedades [40]:

1. Es una función continua no decreciente para toda y.
2. $F(-\infty) = \lim_{y \to -\infty} F(y) = 0 \quad$ y $\quad F(\infty) = \lim_{y \to \infty} F(y) = 1.$
3. $F(y) \leq F(y')$ para toda $y \leq y'$.

Inversamente, cualquier funcion F(y) que cumpla las propiedades 1, 2 y 3 es una función de distribución acumulada continua [41].

3.1.3. La función cuantil

El p-ésimo cuantil de la variable aleatoria Y, denotado por y_p, está definido como el valor más pequeño de la variable aleatoria Y que satisface la relación

$$F_Y(y_p) \geq p, \quad \text{para} \quad p \in (0,1).$$

Formalmente el p-ésimo cuantil está definido como sigue

$$y_p := \inf_{y \in \mathbb{R}} \{y : F_Y(y) \geq p\}, \quad \text{para} \quad p \in (0,1).$$

En el caso donde F(y) es continua y estrictamente creciente, y_p es la solución única de la ecuación

$$F(y_p) = p. \tag{3.2}$$

El valor y_p es conocido como el $100p$-ésimo percentil.

3.2. Mediana y moda

Definición 3.2.1 *Para cualquier variable aleatoria continua Y, $y_{0.5}$ recibe el nombre de mediana de la distribución de Y, también denotada como Me.*

La mediana es una medida de tendencia central, en el sentido de que es el valor para el cual la distribución de probabilidad se divide en dos partes iguales [8].

Definición 3.2.2 *Para cualquier variable aleatoria continua Y, se define la moda como el valor y_m de Y que maximiza la función de densidad de Y [8]. La moda es la solución de*

$$\frac{df(y)}{dy} = 0 \quad si \quad \frac{d^2 f(y)}{dy^2} < 0.$$

3.3. Valores esperados

El valor esperado o esperanza matemática de una variable aleatoria es un concepto muy importante en el estudio de distribuciones de probabilidad. La esperanza de una variable aleatoria tiene sus orígenes en los juegos de azar, debido a que los apostadores deseaban saber cuál era su esperanza de ganar repetidamente un juego. En este sentido, el valor esperado representa la cantidad de dinero promedio que el jugador está dispuesto a ganar o perder despúes de un número muy grande de apuestas. Este significado también es válido para una variable aleatoria, Es decir, el valor promedio de una variable aleatoria después de un número grande de experimentos, es su valor esperado [8].

Definición 3.3.1 *Sea Y una variable aleatoria continua con función de densidad f(y) y sea g(Y) una función continua excepto en un conjunto de medida cero; entonces el valor esperado o la esperanza matemática de la variable g(Y) está dada por*

$$\mu_{g(Y)} = E[g(Y)] = \int_{-\infty}^{\infty} g(y) f(y) dy.$$

3.3.1. Momentos de una variable aleatoria

Los momentos de una variable aleatoria Y (también es apropiado emplear la frase momentos de la distribución de probabilidad de Y) son los valores esperados de ciertas funciones de Y. Estos forman una colección de medidas descriptivas que pueden emplearse para caracterizar la distribución de probabilidad de Y y especificarla si todos los momentos de Y son conocidos. A pesar de que los momentos de Y pueden definirse alrededor de cualquier punto de referencia, generalmente se definen alrededor del cero o del valor esperado de Y. El uso de los momentos de una variable aleatoria para caracterizar a la distribución de probabilidad es una tarea muy útil. Lo anterior es especialmente cierto en un medio en el que es poco probable que el experimentador conozca la distribución de probabilidad [8].

Definición 3.3.2 *Sea Y una variable aleatoria continua con función de densidad f(y), dado un r entero positivo, el momento r-ésimo con respecto al origen de Y está definido por*

$$\mu_r^{'} = E[Y^r] = \int_{-\infty}^{\infty} y^r f(y) dy.$$

Definición 3.3.3 *El valor esperado o media de una variable aleatoria continua Y con función de densidad f(y), denotada por E[Y], se define como el primer momento con respecto al origen:*

$$\mu_1^{'} = E[Y] = \int_{-\infty}^{\infty} y f(y) dy.$$

La media de una variable aleatoria Y da un valor típico o promedio de los valores de Y y por esta razón la media es una medida de tendencia central [25]. Es común en los libros de estadística hacer referencia a la media de una variable aleatoria Y simplemente como μ.

Definición 3.3.4 *Sea Y una variable aleatoria, dado un r entero positivo, el momento r-ésimo con respecto a la media de Y o simplemente el momento r-ésimo central de Y está definido por*

$$\mu_r = E[(Y-\mu)^r] = \int_{-\infty}^{\infty} (y-\mu)^r f(y) dy.$$

Definición 3.3.5 *La varianza de una variable aleatoria continua Y con función de densidad f(y), denotada por V(Y), se define como el segundo momento con respecto a la media de Y:*

$$V(Y) = E[(Y-\mu)^2] = \int_{-\infty}^{\infty} (y-\mu)^2 f(y) dy.$$

8 La varianza es una medida de dispersión y algunas veces también se denota por σ^2. Una propiedad muy útil de la varianza es la siguiente:

$$E[(Y-\mu)^2] = E[Y^2] - \mu^2.$$

8 Generalmente, es más fácil obtener la varianza usando este resultado en lugar de aplicar directamente la definición. Otra propiedad que es muy útil es la siguiente:

$$V(aY+b) = a^2 V(Y).$$

3.4. Modelos continuos

3.4.1. Parámetros de las distribuciones

En el estudio de la distribución de una muestra los parámetros juegan un papel importante para determinar la distribución poblacional, puesto que la población puede tener la distribución supuesta pero con otros parámetros que no sean los propuestos. De esta forma crece el interés en conocer y estudiar los parámetros de la distribución que se supone tiene la población.

Los parámetros generalmente son del siguiente tipo:

- **Localización**. Este tipo de parámetro generalmente se relaciona con la media o alguna medida de tipo central, se caracteriza por realizar un desplazamiento en una distribución de referencia, que comúnmente se llama distribución estándar.
- **Escala**. Este tipo de parámetro generalmente se relaciona con la desviación estándar o alguna medida de variación, se caracteriza por mostrar la amplitud de la gráfica sobre el eje de las ordenadas o dicho de otra manera, la misma forma que toma la gráfica pero en escala diferente sobre el eje de las ordenadas. Es decir, los parámetros de escala representan una transformación de una distribución de referencia, que comúnmente se llama distribución estándar.
- **Forma o asimetría**. Este tipo de parámetro como su nombre lo indica se relaciona con la forma de la distribución, ya que para algunos valores puede ser creciente y para otros decreciente la distribución en un mismo segmento de estudio.

3.4.2. El Modelo Exponencial

La distribución exponencial es una distribución muy utilizada para modelar tiempos de vida y tiempos de falla en estudios de supervivencia y confiabilidad de componentes electrónicos [22]. En estudios ambientales empieza a cobrar relevancia por la aparición de muestras aleatorias de datos ambientales censurados.

Los resultados teóricos asociados a la distribución exponencial son a menudo menos complicadas, de tal modo que es posible obtener fórmulas explícitas menos complejas. Por esta razón, modelos construidos a partir de variables aleatorias exponenciales se utilizan a veces como una representación aproximada de otros modelos más elaborados para una aplicación particular [27].

Definición 3.4.1 *Se dice que una variable aleatoria Y tiene distribución Exponencial con parámetro θ; denotado por $Y \sim exp(\theta)$, si su función de densidad de probabilidad está dada como*

$$f(y;\theta) = \frac{1}{\theta}e^{-y/\theta}, \quad y > 0, \theta > 0.$$

donde θ es un parámetro de escala.

Su función de distribución acumulada $F(y;\theta)$ se obtiene al integrar $f(y;\theta)$ y está dada como

$$F(y;\theta) = 1 - e^{-y/\theta}, \quad y > 0, \theta > 0.$$

La figura 3.1 muestra las gráficas del modelo exponencial cuando $\theta = 0.5, 1.5, 2.0$

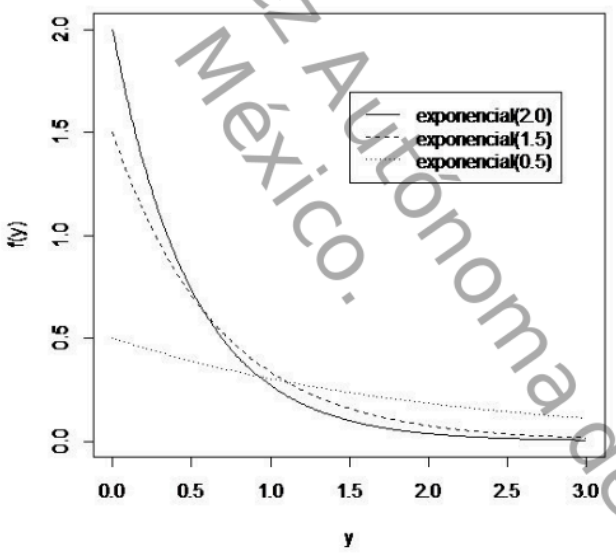


Figura 3.1: Densidades del modelo exponencial para $\theta = 0.5, 1.5, 2.0$.

La moda, media, varianza y mediana para una distribución exponencial son:

$$\begin{array}{ll} \text{moda:} & y_m = 0. \\ \text{media:} & E[Y] = \theta. \\ \text{varianza:} & V(Y) = \theta^2. \\ \text{mediana:} & Me = \theta \ln 2. \end{array} \tag{3.3}$$

Los cuantiles para la función exponencial se obtienen resolviendo (3.2), esto es

$$y_p = -\theta \ln(1-p).$$

3.4.3. El Modelo Lognormal

También conocida como distribución logarítmico-normal. Dicha distribución es muy aplicada en las ciencias ambientales ya que investigadores han reportado que las concentraciones de varios contaminantes tienen distribuciones lognormal. La ubicuidad de la distribución logarítmica normal en los procesos ambientales es sorprendente y no se ha explicado adecuadamente, ya que los procesos comunes en la naturaleza (por ejemplo, el cálculo de la media y el análisis de los errores) suelen dar lugar a las distribuciones que son normales en lugar de lognormal [29].

Algunos ejemplos de la aplicabilidad de la distribución lognormal en los procesos ambientales incluyen material radiactivo en suelos, contaminantes en aire, calidad del aire acondicionado, residuos de metales en ríos, metales pesados en peces, metales en tejidos biológicos y otras concentraciones de residuos radiactivos en tejidos humanos [29],[13].

En la industria la distribución logarítmica normal es un modelo común para tiempos de fracaso. Se ha sugerido que la lognormal es un modelo apropiado para el tiempo de falla causado por un proceso de degradación, con combinaciones al azar de las constantes de velocidad que se combinan múltiples veces. La distribución logarítmica normal es también ampliamente utilizada para describir el tiempo a la fractura de crecimiento de grietas por fatiga de los metales [22].

La aplicación de una transformación logarítmica a los datos puede permitir que estos se aproximen por la distribución normal simétrica, aunque la ausencia de valores negativos puede limitar la validez de este procedimiento [12].

Lo dicho anteriormente es que mediante la transformación inversa se tiene que si Y sigue una distribución lognormal, entonces $X = \ln Y$ sigue una distribución normal. Esta relación, es la que permite al investigador usar la teoría de la distribución normal para analizar datos de la lognormal.

En la estadística actuarial, está distribución es muy útil para modelar siniestros con grandes costes o cuantías por reclamación [10].

Definición 3.4.2 *Una variable aleatoria Y se dice que tiene una distribución lognormal con parámetros μ y σ si su función de densidad de probabilidad está dada por*

$$f(y;\mu,\sigma) = \frac{1}{y\sqrt{2\pi\sigma^2}} e^{-(\ln y-\mu)^2/2\sigma^2}, \qquad y>0, -\infty<\mu<\infty, \sigma>0, \tag{3.4}$$

donde μ y σ^2 son los parámetros de escala y forma de la distribución lognormal, y son la media y la varianza de la variable aleatoria $X = \ln Y$.

La función de distribución acumulada lognormal para Y es,

$$F(y;\mu,\sigma) = \Phi\left(\frac{\ln y - \mu}{\sigma}\right), \qquad y > 0, \tag{3.5}$$

donde $\Phi(.)$ es la función de distribución acumulada normal estándar definida como

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du, \qquad -\infty < z < \infty. \tag{3.6}$$

La figura 3.2 muestra las gráficas del modelo cuando $\mu = 0$ y $\sigma = 0.25, 0.5, 1$.

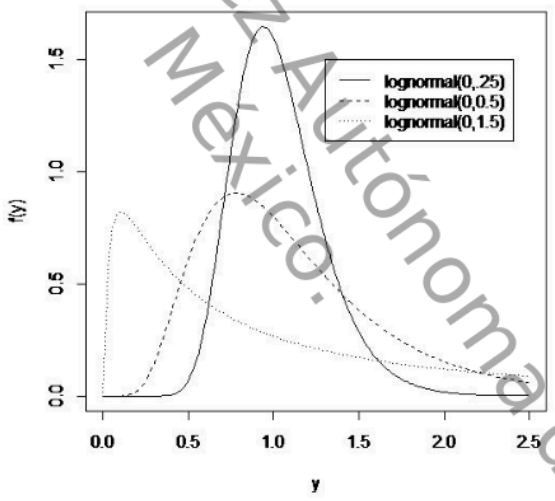


Figura 3.2: Densidades del modelo lognormal para $\mu = 0$ y $\sigma = 0.25, 0.5, 1$.

La moda, media, varianza y mediana para una distribución lognormal son:

$$\begin{array}{ll} \text{moda:} & y_m = e^{\mu-\sigma^2}. \\ \text{media:} & E[Y] = e^{\mu+\frac{\sigma^2}{2}}. \\ \text{varianza:} & V(Y) = e^{\sigma^2}\left(e^{\sigma^2}-1\right)e^{2\mu}. \\ \text{mediana:} & Me = e^{\mu}. \end{array} \tag{3.7}$$

El p-ésimo cuantil lognormal se obtiene resolviendo (3.2) esto es

$$y_p = e^{\mu + z_p \sigma}.$$

Donde z_p es el p-ésimo cuantil normal estandar.

La distribución lognormal de dos parámetros es una representación más realista para las distribuciones de atributos como peso, altura, densidad, que la distribución normal. Esas cantidades no pueden tomar valores negativos, pero la distribución normal asigna una probabilidad positiva a tales eventos, mientras que la distribución lognormal de dos parámetros no lo hace. Además, tomando σ suficientemente pequeña, es posible construir una distribución lognormal muy parecida a alguna distribución normal. De aquí, si una distribución normal es aparentemente adecuada, podría reemplazarse por una distribución lognormal apropiada [38].

3.4.4. El Modelo Weibull

La distribución Weibull es quizás el modelo de distribución más utilizado por su gran variedad de aplicaciones industriales y biomédicas. Su aplicación a la vida o la durabilidad de los productos manufacturados es muy común, se utiliza para modelar los tiempos de vida de un producto o diversos tipos de artículos, tales como componentes del automóvil, aislamiento eléctrico, también se ha utilizado para representar la distribución de la resistencia de ciertos materiales. En medicina, por ejemplo, en los estudios sobre el tiempo de la aparición de tumores en las poblaciones humanas o animales de laboratorio [41],[17].

En la estadística actuarial se usa habitualmente para modelar la compensación de las reclamaciones de los trabajadores. Proporciona una adecuada representación de la distribución de la pérdida [10].

Definición 3.4.3 *Una variable aleatoria Y se dice que tiene distribución Weibull con parámetros θ y β, si su función de densidad de probabilidad está dada por*

$$f(y;\theta,\beta) = \frac{\beta}{\theta^{\beta}} y^{\beta-1} e^{-(y/\theta)^{\beta}}, \qquad y > 0, \theta > 0, \beta > 0. \tag{3.8}$$

El parámetro β es llamado el parámetro de forma y el parámetro θ es llamado el parámetro de escala.

La función de distribución acumulada Weibull es

$$F(y;\theta,\beta) = 1 - e^{-(y/\theta)^{\beta}}, \qquad y > 0. \tag{3.9}$$

Para el caso especial con $\beta = 1$, la distribución de Weibull es la distribución exponencial simple. La distribución de Weibull es muy flexible y ahora ampliamente utilizada, en parte debido a que incluye a la familia de la distribución exponencial.

La figura 3.3 muestra las gráficas de la función de densidad de este modelo cuando $\theta = 1$ y $\beta = 0.5, 1.5, 3.0$, donde θ determina la forma de la distribución y β determina la propagación.

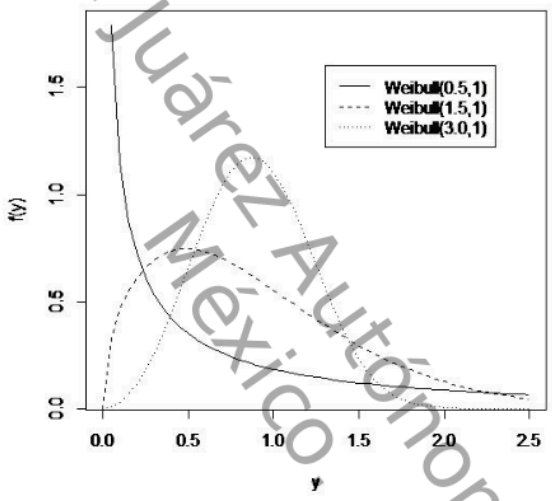


Figura 3.3: Densidades del modelo Weibull para $\theta = 1$ y $\beta = 0.5, 1.5, 3.0$.

La moda para la distribución Weibull, es

$$y_m = \begin{cases} \theta[1 - \left(\frac{1}{\beta}\right)]^{\frac{1}{\beta}} & \text{si } \beta \geq 1, \\ 0 & \text{si } 0 < \beta \leq 1. \end{cases} \tag{3.10}$$

Es decir, y_m es el valor que corresponde al máximo de la función de densidad de probabilidad [41][21].

La media de la distribución Weibull es

$$E[Y] = \theta\Gamma\left[1 + \left(\frac{1}{\beta}\right)\right], \tag{3.11}$$

donde $\Gamma(v) = \int_0^\infty z^{v-1}e^{-z}dz$ es la funcion gamma [34]. Para $v > 0$ entero, se tiene $\Gamma(v) = (v-1)! = (v-1)(v-2)\cdots 2 \cdot 1$.

Y la varianza es,

$$V(Y) = \theta^2\left\{\Gamma\left[1 + \left(\frac{2}{\beta}\right)\right] - \left[\Gamma\left(1 + \left(\frac{1}{\beta}\right)\right)\right]^2\right\}. \tag{3.12}$$

Para el caso de la distribución Weibull el p-ésimo percentil está dado como

$$y_p = \theta\left[-ln(1-p)\right]^{\frac{1}{\beta}}.$$

De lo anterior se tiene que la mediana es:

$$Me = \theta\sqrt[\beta]{\ln 2}. \tag{3.13}$$

La capacidad para describir el aumento o disminución de las tasas de fracaso contribuyó a que la distribución de Weibull sea popular para el análisis de datos de vida.

Los parámetros de la distribución Weibull a veces se expresan de manera diferente. Por ejemplo, si definimos $\lambda = \frac{1}{\theta^\beta}$, entonces se tendría que la función de densidad sería

$$f(y;\lambda,\beta) = \lambda\beta y^{\beta-1}e^{-\lambda y^\beta}. \tag{3.14}$$

y

$$F(y;\lambda,\beta) = 1 - e^{-\lambda y^\beta}. \tag{3.15}$$

3.4.5. El Modelo Pareto

La distribución de Pareto es llamada así en honor al profesor de economía y sociólogo italiano Vilfredo Pareto quien la creo para modelar la distribución de ingresos sobre una población, cuando el ingreso excede cierto límite. La ley de Pareto, tal como fue formulada por él puede enunciarse como sigue [27]:

$$N = Ay^{-\alpha}, \tag{3.16}$$

donde N es el número de personas que tienen un ingreso $\geq y$, A y α son parámetros (α es conocido como la constante de Pareto y es el parámetro de forma). Pareto consideró que

esta ley es universal e inevitable, independientemente de los impuestos y las condiciones sociales y políticas.

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 $y \geq \theta$ es :

$$F_Y = P[Y \geq y] = \left(\frac{\theta}{y}\right)^{\alpha}, \qquad \theta > 0, \alpha > 0, \tag{3.17}$$

donde F_Y es la probabilidad de que el ingreso sea igual o mayor que y y θ representa el ingreso de la gente más pobre. De lo anterior, se obtiene la función de distribución de la variable Y que representa los ingresos, la cual se puede escribir como

$$F(y) = 1 - \left(\frac{\theta}{y}\right)^{\alpha}, \qquad \theta > 0, \alpha > 0, y \geq \theta, \tag{3.18}$$

La distribución Pareto en estadística actuarial es frecuentemente usada para modelar cuantías elevadas [10].

Definición 3.4.4 *Decimos que la variable aleatoria Y tiene distribución de Pareto con parámetros α y θ si su función de densidad de probabilidad está dada por*

$$f(y; \alpha, \theta) = \frac{\alpha\theta^{\alpha}}{y^{\alpha+1}}, \quad \theta > 0, \alpha > 0, y \geq \theta. \tag{3.19}$$

Normalmente θ es un parámetro prefijado de antemano, a partir de los datos iniciales de la aplicación, y es un parámetro de localización. El parámetro α es un parámetro de ajuste, tambien llamado índice de Pareto, cuyo valor se estima o determina a posteriori, 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. Para algunos valores de este parámetro, no existen momentos de esta variable [10].

La moda, media, varianza y mediana de la distribución Pareto están dadas como sigue[12]:

$$\begin{array}{lll} \text{moda:} & y_m = \theta, & \\ \text{media:} & E[Y] = \frac{\alpha\theta}{\alpha-1}, & \alpha > 1 \\ \text{varianza:} & V(Y) = \frac{\alpha\theta^2}{(\alpha-2)(\alpha-1)^2}, & \alpha > 2 \\ \text{mediana:} & Me = \theta\sqrt[\alpha]{2}. & \end{array} \tag{3.20}$$

El p-ésimo percentil pareto se obtiene resolviendo la ecuación (3.2) y está dado por

$$y_p = \theta(1-p)^{-1/\alpha}$$

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, tiempos de servicio y sistemas de espera entre otros [30].

El Modelo Pareto Tipo II.
Como se dijo anteriormente, la distribución Pareto tiene dos parámetros: θ conocido como valor inicial y α conocido como índice de Pareto. En muchas ocasiones es conveniente que el valor inicial sea cero, lo que da lugar a la distribución de Pareto tipo II.

La distribución Pareto de segundo tipo, no es más que la de primer tipo, trasladada al origen. Es decir, si X es una variable aleatoria con distribución Pareto (de primer tipo) y parámetros θ, α, se define como distribución de Pareto de segundo tipo a la distribución de la variable aleatoria $Y = X - \theta$.

De la consideración anterior, se obtiene la función de distribución de la variable Y.

$$F(y;\alpha,\theta) = 1 - \left(\frac{\theta}{y+\theta}\right)^{\alpha}, \qquad y \geq 0.$$

Definición 3.4.5 *Decimos que la variable aleatoria Y tiene distribución de Pareto Tipo II con parámetros α, θ si su función de densidad de probabilidad está dada por*

$$f(y;\alpha,\theta) = \frac{\alpha\theta^{\alpha}}{(y+\theta)^{(\alpha+1)}}, \qquad \theta > 0, \alpha > 0, y \geq 0. \tag{3.21}$$

La moda, media, varianza y mediana para la distribución Pareto tipo II están dadas como sigue [16]:

$$\begin{array}{ll} \text{moda:} & y_m = 0, \\ \text{media:} & E[Y] = \frac{\theta}{(\alpha-1)}, \qquad \alpha > 1 \\ \text{varianza:} & V(Y) = \frac{2\theta^2}{(\alpha-1)(\alpha-2)} - \left[\frac{\theta}{(\alpha-1)}\right]^2, \qquad \alpha > 2 \\ \text{mediana:} & Me = \theta\sqrt[\alpha]{2} - \theta. \end{array} \tag{3.22}$$

El p-ésimo percentil pareto se obtiene resolviendo la ecuación (3.2) y está dado por

$$y_p = \theta(1-p)^{-1/\alpha} - \theta.$$

La figura 3.4 muestra las gráficas de la función de densidad de este modelo cuando $\alpha = 1$ y $\theta = 1, 2, 3$.

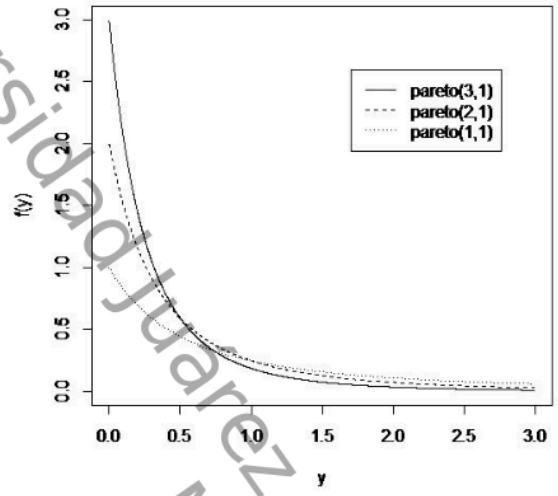


Figura 3.4: Densidades del modelo Pareto tipo II para $\alpha = 1$ y $\theta = 1, 2, 3$.

3.5. Distribuciones continuas truncadas

En algunas ocasiones las variables aleatorias, asociadas a los modelos probabilísticos que se han estudiado, toman valores que corresponden solo a una parte de los que usualmente se trabajan, o bien se delimitan sus valores a la izquierda o a la derecha de los acostumbrados, y entonces se conocen dichas variables como variables aleatorias truncadas cuyas distribuciones están truncadas.

Las distribuciones truncadas se pueden encontrar en un proceso de producción de varias etapas, en el que la inspección se lleva a cabo en cada etapa de producción. Si se pasan a la siguiente etapa solo a los elementos conformes a los límites de calidad, la distribución real es una distribución truncada.

Las pruebas de vida acelerada con muestras censuradas es también un buen ejemplo de distribución truncada. De hecho, el concepto de una distribución truncada juega un papel importante en el análisis de una variedad de procesos de producción, procesos de optimización y mejora de calidad. Las distribuciones truncadas también pueden ser utilizadas como modelos estadísticos de intensidad en el estudio de la heterogeneidad atómica[26].

Una distribución truncada es una parte de una distribución no truncada que está por

encima o por debajo de cierto valor [42].

3.5.1. Función de densidad de probabilidad

Definición 3.5.1 *Si una variable aleatoria Y tiene fdp $f(y)$ y y_0 es una constante, entonces la función de densidad de la variable aleatoria truncada por la izquierda en y_0 está dada por*

$$f_\tau(y|y > y_0) = \frac{f(y)}{P(y > y_0)} = \frac{f(y)}{1 - P(y \leq y_0)} = \frac{f(y)}{1 - F(y_0)} \tag{3.23}$$

3.6. Valores esperados

3.6.1. Momentos de una variable aleatoria truncada

Generalmente estamos interesados en la media y la varianza de la variable aleatoria truncada, ya que son formas de caracterizar la distribución. Los momentos pueden obtenerse utilizando las siguientes fórmulas:

$$E[y|y > y_0] = \int_{y_0}^{\infty} y f_\tau(y|y > y_0)\, dy, \tag{3.24}$$

para la media,

$$V(y|y > y_0) = \int_{y_0}^{\infty} \{y - E[y|y > y_0]\}^2 f_\tau(y|y > y_0)\, dy, \tag{3.25}$$

para la varianza.

Cuando el truncamiento es por debajo, entonces la media de la variable truncada es mayor que la media de la original. Si el truncamiento es por arriba, entonces la media de la variable truncada será menor que la media de la original. El truncamiento reduce la varianza en comparación con la varianza de la distribución no truncada[42].

En adelante usaremos la notación $E_\tau[Y]$ y $V_\tau(Y)$ para referirnos a la media y a la varianza de una variable aleatoria con una distribución truncada, la notación para la función de densidad truncada será simplemente $f_\tau(y)$.

3.7. Modelos continuos truncados

3.7.1. El Modelo Exponencial truncado

Definición 3.7.1 *Se dice que una variable aleatoria Y tiene una distribución de probabilidad Exponencial truncada por la izquierda en y_0 con parámetro θ si y solo si, para $\theta > 0$, la función de densidad de Y es*

$$f_\tau(y) = \frac{e^{-y/\theta}}{\theta e^{-y_0/\theta}}, \qquad 0 < y_0 < y. \tag{3.26}$$

Teorema 3.7.1 *Si Y es una variable aleatoria con distribución exponencial truncada por la izquierda en y_0 y parámetro θ, entonces*

$$E_\tau[Y] = \theta + y_0, \tag{3.27}$$

y

$$V_\tau(Y) = \theta^2. \tag{3.28}$$

Demostración. Sea $U = Y - y_0$. Si $u > 0$ denota un valor arbitrario de U y $y > y_0$ representa un valor arbitrario de Y, entonces $u = y - y_0$, de donde $\frac{dy}{du} = 1$. Por lo tanto

$$f_U(u) = \frac{e^{-(u+y_0)/\theta}}{\theta e^{-y_0/\theta}} \left|\frac{dy}{du}\right| = \frac{e^{-u/\theta} e^{-y_0/\theta}}{\theta e^{-y_0/\theta}} = \frac{1}{\theta} e^{-u/\theta}.$$

Así, $U \sim exp(\theta)$. Luego

$$E[Y] = E[U + y_0] = E[U] + y_0 = \theta + y_0,$$

$$V(Y) = V(U) = \theta^2.$$

■

3.7.2. El Modelo Weibull truncado

Definición 3.7.2 *Se dice que una variable aleatoria Y tiene una distribución de probabilidad Weibull truncada por la izquierda en y_0 con parámetros λ, β si y solo si, para $\lambda > 0$ y $\beta > 0$, la función de densidad de Y es*

$$f_\tau(y) = \frac{\lambda\beta y^{\beta-1} e^{-\lambda y^\beta}}{e^{-\lambda y_0^\beta}}, \qquad 0 < y_0 < y \tag{3.29}$$

Teorema 3.7.2 *Si Y es una variable aleatoria con distribución Weibulll truncada por la izquierda en y_0 con parámetros $\lambda > 0$ y $\beta > 0$, entonces*

$$E_\tau[Y] = \frac{1}{\lambda^{1/\beta} e^{-\lambda y_0^\beta}} \left[\Gamma\left(1 + \frac{1}{\beta}\right) - \gamma\left(1 + \frac{1}{\beta}, \lambda y_0^\beta\right)\right], \tag{3.30}$$

y

$$V_\tau(Y) = \frac{1}{\lambda^{2/\beta} e^{-\lambda y_0^\beta}} \left[\Gamma\left(1 + \frac{2}{\beta}\right) - \gamma\left(1 + \frac{2}{\beta}, \lambda y_0^\beta\right)\right] - (E[Y])^2, \tag{3.31}$$

donde γ es la función gamma incompleta definida como $\gamma(\lambda, z) = \int_0^z u^{\lambda-1} e^{-u} du$.

Demostración. La demostración de la primera parte del teorema se realiza procediendo como sigue

$$\begin{aligned} E_\tau[Y] &= \int_{y_0}^{\infty} y f_\tau(y) dy \\ &= \frac{1}{e^{-\lambda y_0^\beta}} \int_{y_0}^{\infty} \lambda\beta y^\beta e^{-\lambda y^\beta} dy \\ &= \frac{1}{e^{-\lambda y_0^\beta}} \left[\int_0^{\infty} \lambda\beta y^\beta e^{-\lambda y^\beta} dy - \int_0^{y_0} \lambda\beta y^\beta e^{-\lambda y^\beta} dy \right]. \end{aligned}$$

Se resuelve primero la integral $\int_0^\infty \lambda\beta y^\beta e^{-\lambda y^\beta} dy$. Sean $u = \lambda y^\beta$, $du = \lambda\beta y^{\beta-1} dy$, entonces $y = (u/\lambda)^{1/\beta}$ y $dy = du/\lambda\beta y^{\beta-1}$. Así

$$\begin{aligned} \int_0^{\infty} \lambda\beta y^\beta e^{-\lambda y^\beta} dy &= \int_0^{\infty} y e^u du \\ &= \int_0^{\infty} \left(\frac{u}{\lambda}\right)^{1/\beta} e^{-u} du \\ &= \left(\frac{1}{\lambda}\right)^{1/\beta} \int_0^{\infty} u^{(\frac{1}{\beta}+1)-1} e^{-u} du. \end{aligned}$$

Recuerde que $\Gamma(\lambda) = \int_0^\infty t^{\lambda-1} e^{-t} dt$, por lo que

$$\int_0^{\infty} \lambda\beta y^\beta e^{-\lambda y^\beta} dy = \left(\frac{1}{\lambda}\right)^{1/\beta} \Gamma\left(1 + \frac{1}{\beta}\right).$$

Antes de resolver la integral $\int_0^{y_0} \lambda\beta y^\beta e^{-\lambda y^\beta} dy$, se probará de manera general que

$$\int_0^{y_0} \lambda\beta y^{\beta+k} e^{-\lambda y^\beta} dy = \left(\frac{1}{\lambda}\right)^{\frac{k+1}{\beta}} \gamma\left(1 + \frac{k+1}{\beta}, \lambda y_0^\beta\right)$$

Utilizando cambio de variables, sean $u = \lambda y^\beta$, $du = \lambda\beta y^{\beta-1} dy$, entonces $y = (u/\lambda)^{1/\beta}$

y $dy = du/\lambda\beta y^{\beta-1}$. Así

$$\begin{aligned}
\int_0^{y_0} \lambda\beta y^{\beta+k} e^{-\lambda y^\beta} dy &= \int_0^{y_0} \lambda\beta y^\beta y^k e^{-\lambda y^\beta} dy \\
&= \int_0^{\lambda y_0^\beta} y^{k+1} e^{-\lambda y^\beta} \lambda\beta y^{\beta-1} dy \\
&= \int_0^{\lambda y_0^\beta} \left[\left(\frac{u}{\lambda}\right)^{1/\beta}\right]^{k+1} e^{-u} du \\
&= \left(\frac{1}{\lambda}\right)^{\frac{k+1}{\beta}} \int_0^{\lambda y_0^\beta} u^{\frac{k+1}{\beta}} e^{-u} du \\
&= \left(\frac{1}{\lambda}\right)^{\frac{k+1}{\beta}} \int_0^{\lambda y_0^\beta} u^{\left(\frac{k+1}{\beta}+1\right)-1} e^{-u} du
\end{aligned}$$

finalmente, por la definición de la función gamma incompleta se tiene que

$$\int_0^{y_0} \lambda\beta y^{\beta+k} e^{-\lambda y^\beta} dy = \left(\frac{1}{\lambda}\right)^{\frac{k+1}{\beta}} \gamma\left(1 + \frac{k+1}{\beta}, \lambda y_0^\beta\right) \tag{3.32}$$

por lo que, de acuerdo a (3.32) con k=0, se tiene que

$$\int_0^{y_0} \lambda\beta y^\beta e^{-\lambda y^\beta} dy = \left(\frac{1}{\lambda}\right)^{1/\beta} \gamma\left(1 + \frac{1}{\beta}, \lambda y_0^\beta\right)$$

.
Por lo tanto

$$E_\tau[Y] = \frac{1}{\lambda^{1/\beta} e^{-\lambda y_0^\beta}} \left[\Gamma\left(1 + \frac{1}{\beta}\right) - \gamma\left(1 + \frac{1}{\beta}, \lambda y_0^\beta\right)\right].$$

Con esto queda demostrado la primer parte del teorema, para demostrar la segunda parte del teorema se procede como sigue. Por definición

$$\begin{aligned}
V_\tau(Y) &= \int_{y_0}^{\infty} \{y - E_\tau[Y]\}^2 f_\tau(y) dy \\
&= \int_{y_0}^{\infty} y^2 f_\tau(y) dy - 2E_\tau[Y] \int_{y_0}^{\infty} y f_\tau(y) dy + (E_\tau[Y])^2 \int_{y_0}^{\infty} f_\tau(y) dy
\end{aligned}$$

Se resuelven estas integrales por separado. Para la primera integral se tiene que

$$\begin{aligned}
\int_{y_0}^{\infty} y^2 f_T(y) dy &= \frac{1}{e^{-\lambda y_0^\beta}} \int_{y_0}^{\infty} \lambda\beta y^{\beta+1} e^{-\lambda y^\beta} dy \\
&= \frac{1}{e^{-\lambda y_0^\beta}} \left[\int_0^{\infty} \lambda\beta y^{\beta+1} e^{-\lambda y^\beta} dy - \int_0^{y_0} \lambda\beta y^{\beta+1} e^{-\lambda y^\beta} dy\right].
\end{aligned}$$

Ahora, $\int_0^\infty \lambda\beta y^{\beta+1}e^{-\lambda y^\beta}dy$ se resuelve como sigue. Sean $u = \lambda y^\beta$, $du = \lambda\beta y^{\beta-1}dy$, entonces $y = (u/\lambda)^{1/\beta}$, $y^2 = (u/\lambda)^{2/\beta}$ y $dy = du/\lambda\beta y^{\beta-1}$, luego

$$\begin{aligned}\int_0^\infty \lambda\beta y^{\beta+1}e^{-\lambda y^\beta}dy &= \int_0^\infty y^2 e^{-\lambda y^\beta}du \\ &= \left(\frac{1}{\lambda}\right)^{2/\beta}\int_0^\infty u^{\left(\frac{2}{\beta}+1\right)-1}e^{-u}du \\ &= \left(\frac{1}{\lambda}\right)^{2/\beta}\Gamma\left(1+\frac{2}{\beta}\right).\end{aligned}$$

Para resolver la integral $\int_0^{y_0} \lambda\beta y^{\beta+1}e^{-\lambda y^\beta}dy$, usamos (3.32) con k=1 y se obtiene que

$$\int_0^{y_0} \lambda\beta y^{\beta+1}e^{-\lambda y^\beta}dy = \left(\frac{1}{\lambda}\right)^{2/\beta}\gamma\left(1+\frac{2}{\beta}, \lambda y_0^\beta\right),$$

así

$$\int_{y_0}^\infty y^2 f_\tau(y)dy = \frac{1}{\lambda^{2/\beta}e^{-\lambda y_0^\beta}}\left[\Gamma\left(1+\frac{2}{\beta}\right) - \gamma\left(1+\frac{2}{\beta}, \lambda y_0^\beta\right)\right],$$

tambien se tiene que

$$-2E_\tau[Y]\int_{y_0}^\infty y f_\tau(y)dy = -2(E_\tau[Y])^2$$

y finalmente

$$(E_\tau[Y])^2\int_{y_0}^\infty f_\tau(y)dy = (E_\tau[Y])^2$$

.
Por lo tanto,

$$V_\tau(Y) = \frac{1}{\lambda^{2/\beta}e^{-\lambda y_0^\beta}}\left[\Gamma\left(1+\frac{2}{\beta}\right) - \gamma\left(1+\frac{2}{\beta}, \lambda y_0^\beta\right)\right] - (E_\tau[Y])^2. \tag{3.33}$$

■

3.7.3. El Modelo Pareto tipo II truncado

Definición 3.7.3 *Se dice que una variable aleatoria Y tiene una distribución de probabilidad Pareto tipo II truncada por la izquierda en y_0 con parámetros α, θ si y solo si, para $\alpha > 0$ y $\theta > 0$, la función de densidad de Y es*

$$f_\tau(y) = \frac{\alpha(y_0+\theta)^\alpha}{(y+\theta)^{\alpha+1}}, \qquad 0 < y_0 < y. \tag{3.34}$$

Teorema 3.7.3 *Si Y es una variable aleatoria con distribución Pareto truncada por la izquierda con parámetros $\alpha > 0$ y $\theta > 0$, entonces*

$$E_\tau[Y] = \frac{\alpha y_0 + \theta}{\alpha - 1} \tag{3.35}$$

y

$$V_\tau(Y) = y_0^2 + 2\left[\frac{y_0(y_0+\theta)}{(\alpha-1)} + \frac{(y_0+\theta)^2}{(\alpha-1)(\alpha-2)}\right] - \left[\frac{\alpha y_0 + \theta}{(\alpha-1)}\right]. \tag{3.36}$$

Demostración. La demostración de la primera parte del teorema se realiza procediendo como sigue:

$$\begin{aligned} E_\tau[Y] &= \int_{y_0}^{\infty} y f_T(y) dy \\ &= \int_{y_0}^{\infty} \frac{y\alpha(y_0+\theta)^\alpha}{(y+\theta)^{\alpha+1}} dy \\ &= \alpha(y_0+\theta)^\alpha \int_{y_0}^{\infty} \frac{y}{(y+\theta)^{\alpha+1}} dy. \end{aligned}$$

A continuación se resuelve la integral $\int_{y_0}^{\infty} y/(y+\theta)^{\alpha+1} dy$.

$$\begin{aligned} \int_{y_0}^{\infty} \frac{y}{(y+\theta)^{\alpha+1}} dy &= \int_{y_0}^{\infty} \frac{y+\theta-\theta}{(y+\theta)^{\alpha+1}} dy \\ &= \int_{y_0}^{\infty} \frac{y+\theta}{(y+\theta)^{\alpha+1}} dy - \theta \int_{y_0}^{\infty} \frac{dy}{(y+\theta)^{\alpha+1}} \\ &= \int_{y_0}^{\infty} \frac{dy}{(y+\theta)^{\alpha+1}} + \frac{\theta}{\alpha}(y+\theta)^{-\alpha}\Big|_{y_0}^{\infty} \\ &= -\frac{1}{(\alpha-1)}(y+\theta)^{-\alpha+1}\Big|_{y_0}^{\infty} + \frac{\theta}{\alpha}\left(-\frac{1}{(y_0+\theta)^\alpha}\right) \\ &= \frac{1}{(\alpha-1)(y_0+\theta)^{\alpha-1}} - \frac{\theta}{\alpha(y_0+\theta)^\alpha} \end{aligned}$$

Por lo tanto

$$E_\tau[Y] = \frac{\alpha y_0 + \theta}{\alpha - 1}.$$

Con esto la primera parte del teorema está demostrada, para demostrar la segunda parte del teorema se procede como sigue. Por definición

$$V_\tau(Y) = \int_{y_0}^{\infty} \{y - E_\tau[Y]\}^2 f_\tau(Y) dy.$$

$$\begin{aligned} V_\tau(Y) &= \int_{y_0}^{\infty} y^2 f_\tau(y)dy - 2E_\tau[Y]\int_{y_0}^{\infty} y f_\tau(y)dy + (E_\tau[Y])^2 \int_{y_0}^{\infty} f_\tau(y)dy \\ &= \int_{y_0}^{\infty} y^2 f_\tau(y)dy - 2\left(\frac{\alpha y_0+\theta}{\alpha-1}\right)\int_{y_0}^{\infty} y f_\tau(y)dy + \left(\frac{\alpha y_0+\theta}{\alpha-1}\right)^2 \int_{y_0}^{\infty} f_\tau(y)dy. \end{aligned}$$

Se resuelve por separado cada una de las integrales de la ecuación anterior. Recuerde que

$\int_{y_0}^{\infty} f_\tau(y)dy = 1$, por la propiedad de función de densidad de probabilidad.

$\int_{y_0}^{\infty} y f_\tau(y)dy = E_\tau[Y]$, se probó en la primera parte del teorema.

Por lo que

$$\begin{aligned} V_\tau(Y) &= \int_{y_0}^{\infty} y^2 f_\tau(y)dy - 2\left(\frac{\alpha y_0+\theta}{\alpha-1}\right)^2 + \left(\frac{\alpha y_0+\theta}{\alpha-1}\right)^2 \\ &= \int_{y_0}^{\infty} y^2 f_\tau(y)dy - \left(\frac{\alpha y_0+\theta}{\alpha-1}\right)^2, \end{aligned}$$

por otro lado,

$$\int_{y_0}^{\infty} y^2 f_\tau(y)dy = \alpha(y_0+\theta)^\alpha \int_{y_0}^{\infty} \frac{y^2 dy}{(y+\theta)^{\alpha+1}} \tag{3.37}$$

A continuación se resuelve la integral $\int_{y_0}^{\infty} y^2/(y+\theta)^{\alpha+1}dy$.

$$\begin{aligned} \int_{y_0}^{\infty} \frac{y^2 dy}{(y+\theta)^{\alpha+1}} &= \int_{y_0}^{\infty} \frac{(y+\theta-\theta)^2 dy}{(y+\theta)^{\alpha+1}} \\ &= \int_{y_0}^{\infty} \frac{\left[(y+\theta)^2 - 2\theta(y+\theta)+\theta^2\right]dy}{(y+\theta)^{\alpha+1}} \\ &= \int_{y_0}^{\infty} \frac{dy}{(y+\theta)^{\alpha-1}} - 2\theta\int_{y_0}^{\infty} \frac{dy}{(y+\theta)^{\alpha}} + \theta^2 \int_{y_0}^{\infty} \frac{dy}{(y+\theta)^{\alpha+1}} \\ &= \frac{1}{(\alpha-2)(y_0+\theta)^{\alpha-2}} - \frac{2\theta}{(\alpha-1)(y_0+\theta)^{\alpha-2}} + \frac{\theta^2}{\alpha(y_0+\theta)^{\alpha}} \end{aligned}$$

luego,

$$\begin{aligned} \alpha(y_0+\theta)^\alpha \int_{y_0}^{\infty} \frac{y^2 dy}{(y+\theta)^{\alpha+1}} &= \frac{\alpha(y_0+\theta)^2}{\alpha-2} - \frac{2\theta\alpha(y_0+\theta)}{\alpha-1} + \theta^2 \\ &= \theta^2 + \alpha(y_0+\theta)\left[\frac{y_0(\alpha-1)-\theta(\alpha-3)}{(\alpha-2)(\alpha-1)}\right] \end{aligned}$$

Por lo tanto,

$$V_\tau(Y) = \theta^2 + \alpha(y_0+\theta)\left[\frac{y_0(\alpha-1)-\theta(\alpha-3)}{(\alpha-2)(\alpha-1)}\right] - \left[\frac{\alpha y_0+\theta}{(\alpha-1)}\right]^2.$$

■

3.7.4. El Modelo Lognormal truncado

Definición 3.7.4 *Se dice que una variable aleatoria Y tiene una distribución de probabilidad Lognormal truncada por la izquierda en y_0 con parámetros μ, σ si y solo si, para $\mu > 0$ y $\sigma > 0$, la función de densidad de Y es*

$$f_\tau(y) = \frac{\frac{1}{y\sqrt{2\pi\sigma^2}}e^{-(\ln y-\mu)^2/2\sigma^2}}{1-\Phi(\frac{\ln y_0-\mu}{\sigma})}, \qquad 0 < y_0 < y. \tag{3.38}$$

Teorema 3.7.4 *Si Y es una variable aleatoria con distribución Lognormal truncada por la izquierda en y_0 con parámetros μ, σ, entonces*

$$E_\tau[Y] = \frac{e^{\mu+\frac{\sigma^2}{2}}\Phi(\sigma-a_0)}{\Phi(-a_0)} \tag{3.39}$$

y

$$V_\tau(Y) = \frac{e^{2\mu+2\sigma^2}\Phi(\sigma-a_0)}{\Phi(-a_0)} - \left[\frac{e^{\mu+\frac{\sigma^2}{2}}\Phi(\sigma-a_0)}{\Phi(-a_0)}\right]^2, \tag{3.40}$$

donde $a_0 = \frac{\ln y_0-\mu}{\sigma}$.

Demostración. Lo primero que haremos será demostrar que

$$E_\tau[Y^r] = \frac{e^{r\mu+\frac{r^2\sigma^2}{2}}\Phi(r\sigma-a_0)}{\Phi(-a_0)}.$$

Por definición

$$\begin{aligned} E_\tau[Y^r] &= \int_{y_0}^{\infty} y^r f_\tau(y)dy \\ &= \int_{y_0}^{\infty} \frac{y^r f(y)}{[1-\Phi(a_0)]}dy, \\ &= \frac{1}{[1-\Phi(a_0)]}\left[\int_0^{\infty} y^r f(y)dy - \int_0^{y_0} y^r f(y)dy\right], \end{aligned} \tag{3.41}$$

donde $f(y)$ es la distribución lognormal definida en (3.4). Resolvemos por separado las integrales; de la primera integral se sigue que

$$\int_0^{\infty} y^r f(y)dy = e^{r\mu+\frac{r^2\sigma^2}{2}}, \tag{3.42}$$

esto es el momento r-ésimo de la distribución Lognormal.

Para la segunda integral se tiene que

$$\int_0^{y_0} y^r f(y)dy = \int_0^{y_0} \frac{y^r}{y\sqrt{2\pi}\sigma} e^{-(\ln y-\mu)^2/2\sigma^2} dy.$$

Sea $x = \ln y$, entonces $y = e^x$ y $dx = dy/y$. De lo anterior se sigue que

$$\int_0^{y_0} y^r f(y)dy = \int_{-\infty}^{\ln y_0} \frac{e^{rx}}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} dx.$$

Ahora, sea $z = (x-\mu)/\sigma$, entonces $x = z\sigma + \mu$ y $dx = \sigma dz$, por lo que

$$\begin{aligned}\int_0^{y_0} y^r f(y)dy &= \int_{-\infty}^{a_0} \frac{e^{rz\sigma+r\mu}}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz, \\ &= e^{r\mu+\frac{r^2\sigma^2}{2}} \int_{-\infty}^{a_0} \frac{e^{-\frac{1}{2}(z-r\sigma)^2}}{\sqrt{2\pi}} dz.\end{aligned}$$

Realizamos un cambio de variable, sean $u = z - r\sigma$, entonces $du = dz$, por lo que se sigue que

$$\begin{aligned}\int_0^{y_0} y^r f(y)dy &= e^{r\mu+\frac{r^2\sigma^2}{2}} \int_{-\infty}^{a_0-r\sigma} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du \\ &= e^{r\mu+\frac{r^2\sigma^2}{2}} \Phi(a_0 - r\sigma) \end{aligned} \tag{3.43}$$

Sustituyendo (3.42) y (3.43) en (3.41) se tiene que

$$E_\tau[Y^r] = \frac{1}{[1-\Phi(a_0)]} \left[e^{r\mu+\frac{r^2\sigma^2}{2}} - e^{r\mu+\frac{r^2\sigma^2}{2}} \Phi(a_0 - r\sigma)\right]$$

Por lo tanto

$$E_\tau[Y^r] = \frac{e^{r\mu+\frac{r^2\sigma^2}{2}} \Phi(r\sigma - a_0)}{\Phi(-a_0)}. \tag{3.44}$$

La primera parte del teorema se sigue de (3.44) con $r = 1$. Por lo tanto

$$E_\tau[Y] = \frac{e^{\mu+\frac{\sigma^2}{2}} \Phi(\sigma - a_0)}{\Phi(-a_0)}.$$

Para demostrar la segunda parte del teorema se procede como sigue. Por definición

$$V_\tau(Y) = \int_{y_0}^{\infty} \{y - E_\tau[Y]\}^2 f_\tau(y)dy$$
$$= \int_{y_0}^{\infty} y^2 f_\tau(y)dy - 2E_\tau[Y] \int_{y_0}^{\infty} y f_\tau(y)dy + (E_\tau[Y])^2 \int_{y_0}^{\infty} f_\tau(y)dy$$

La primera integral se sigue de (3.44) con $r = 2$.

$$\int_{y_0}^{\infty} y^2 f_\tau(y)dy = \frac{e^{2\mu+2\sigma^2}\Phi(\sigma - a_0)}{\Phi(-a_0)},$$

tambien se tiene que

$$-2E_\tau[Y]\int_{y_0}^{\infty} y f_\tau(y)dy = -2(E_\tau[Y])^2$$

y finalmente

$$(E_\tau[Y])^2 \int_{y_0}^{\infty} f_\tau(y)dy = (E_\tau[Y])^2$$

.
Por lo tanto,

$$V_\tau(Y) = \frac{e^{2\mu+2\sigma^2}\Phi(\sigma - a_0)}{\Phi(-a_0)} - \left[\frac{e^{\mu+\frac{\sigma^2}{2}}\Phi(\sigma - a_0)}{\Phi(-a_0)}\right]^2 \tag{3.45}$$

■

Capítulo 4

Conceptos de Inferencia Estadística

La inferencia estadística es el proceso mediante el cual con base en la observación de una muestra aleatoria, obtenemos resultados y conclusiones que luego generalizamos a la población de donde proviene la muestra, bajo el conocimiento de que esta generalización de resultados no es 100 % segura, sino que conlleva un cierto margen de error asociado, error que se cuantifica en términos de probabilidad [32]. La inferencia estadística presenta dos enfoques: La inferencia paramétrica y la inferencia no paramétrica. Un enfoque paramétrico es el que se muestra en este trabajo.

4.1. Inferencia paramétrica.

En el desarrollo de los métodos estadísticos modernos, las primeras técnicas de inferencia que aparecieron fueron aquellas que hicieron suposiciones acerca de la naturaleza de las poblaciones de las cuales se derivaron las observaciones y los datos. Es decir, suponer que los datos siguen una cierta distribución con uno o más parámetros (características poblacionales).

En la práctica con mucha frecuencia no conocemos los valores de los parámetros, por lo cual a partir de una muestra aleatoria pretendemos estimar dichos valores, mientras más grande sea la muestra más exacta será la estimación. Estos métodos estadísticos se llaman paramétricos [8],[22].

Las metodologías paramétricas proporcionan conclusiones de la forma siguiente: "si las suposiciones acerca de la forma de la distribución de la población son válidas, entonces podemos concluir que. . . ". Debido a las suposiciones comunes, tales pruebas se sistematizan fácilmente y son también muy fáciles de enseñar y aplicar [35].

La inferencia paramétrica plantea tres tipos de problemas:

1. Estimación puntual, la cual involucra el uso de los datos muestrales junto con algún estadístico para estimar el valor de un parámetro desconocido (usando máxima verosimilitud).

2. Estimación por intervalos (buscamos un intervalo de confianza).

3. Contrastes de hipótesis donde buscamos contrastar información acerca del parámetro. Las pruebas de hipótesis tienen una fuerte relación con el concepto de estimación.

Ventajas de las pruebas paramétricas:

- Más poder de eficacia.
- Más sensible a los rasgos de los datos.
- Menos posibilidad de error.
- Robustas (dan estimaciones probabilísticas bastante exactas).
- Trabaja con valores reales.

A continuación se enlistan algunos de los conceptos básicos que se utilizan en este trabajo.

4.2. Estimación paramétrica vía verosimilitud

4.2.1. Máxima verosimilitud

Un enfoque que aborda el problema de la estimación de parámetros, dentro de la teoría de estimación parámetrica y que es ampliamente utilizado por su objetivo uso de la información y por proporcionar resultados estadísticamente eficientes es el de máxima verosimilitud.

El método de máxima verosimilitud fue introducido primero por R. A. Fisher, genetista y experto en estadística, en la década de 1920. La mayoría de los expertos en estadística recomiendan este método, porque los estimadores resultantes tienen ciertas propiedades deseables de eficiencia. La idea general de los métodos basados en máxima verosimilitud es ajustar modelos a los datos, considerando combinaciones de parámetros y modelos para los que se maximiza el valor de la probabilidad de la muestra aleatoria observada. Este método puede ser aplicado con una gran variedad de modelos paramétricos con datos observados, datos censurados o truncados. También es posible ajustar modelos con variables explicatorias (análisis de regresión).

La teoría desarrollada, garantiza que estos métodos estadísticos son eficientes para muestras grandes (es decir, proporcionan los estimadores más exactos). Esas propiedades son aproximadas con tamaños de muestras pequeñas y varios estudios han mostrado que estos métodos tienen igual desempeño que otros disponibles [22].

4.2.2. Función de verosimilitud

Toda la inferencia a través de máxima verosimilitud tiene su fundamento en la función de verosimilitud o simplemente verosimilitud, la cual se define como el producto de las densidades evaluadas en cada muestra observada (ya sean datos observados, censurados o truncados); pero como función de los parámetros involucrados en el modelo que haya sido seleccionado para el fenómeno aleatorio de interés [22].

La función de verosimilitud juega un papel fundamental en la inferencia estadística. Su rol principal es inferir sobre los parámetros de la distribución que haya sido elegida para describir mejor al fenómeno aleatorio de interés a partir de una muestra observada.

Definición 4.2.1 *Sea $y_1, ..., y_n$ una muestra de variables aleatorias continuas, independientes e identicamente distribuidas con función de distribución $F(y;\boldsymbol{\theta})$ y densidad $f(y;\boldsymbol{\theta})$, con $\boldsymbol{\theta} \in \mathbb{R}^p$. Se llama función de verosimilitud, al producto de las densidades evaluadas en cada muestra observada, definida como sigue*

$$L(\boldsymbol{\theta}) = L(\boldsymbol{\theta}; y) = \prod_{i=1}^{n} f(y_i;\boldsymbol{\theta}). \tag{4.1}$$

El valor de $\boldsymbol{\theta}$ que maximiza $L(\boldsymbol{\theta})$ provee un estimador de máxima verosimilitud (EMV) y se denota por $\hat{\boldsymbol{\theta}}$.

Definición 4.2.2 *La función de log-verosimilitud o simplemente la log-verosimilitud se define como el logaritmo natural (ln) de la función de verosimilitud dada en (4.1),*

$$\ell(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}; Y) = \ln[L(y;\boldsymbol{\theta})] = \sum_{i=1}^{n} \ln f(y_i;\boldsymbol{\theta}). \tag{4.2}$$

En la práctica, generalmente se maximiza esta función, ya que esta expresión ofrece más versatilidad analítica y sus máximos coinciden con el de $L(\boldsymbol{\theta})$.
La maximización de la logverosimilitud se obtienen de sus derivadas parciales con respecto del vector $\boldsymbol{\theta}$. Esto es, $\hat{\boldsymbol{\theta}}$ surge de la solución simultánea a las ecuaciones,

$$\frac{\partial l(\theta)}{\partial \theta_1} = 0, \qquad \frac{\partial l(\theta)}{\partial \theta_2} = 0, \qquad ..., \qquad \frac{\partial l(\theta)}{\partial \theta_p} = 0. \tag{4.3}$$

Cada observación contribuye a la función de verosimilitud según su naturaleza o probabilidad. La concepción de la función de verosimilitud debe hacerse considerando los siguientes elementos importantes:

1. El modelo de probabilidad asumido (reportado en artículos anteriores o seleccionado de una prueba de bondad de ajuste).
2. Definir bien los tipos de datos que estaremos analizando.
3. Objetivo, enfoque o importancia de la investigación hecha.

Definición 4.2.3 *La función Score $S(\boldsymbol{\theta})$ es definida como la primer derivada de la función de log-verosimilitud con respecto de $\boldsymbol{\theta}$.*
Para el caso unidimensional la función Score está dada por

$$S(\theta) = \ell'(\theta) = \frac{d\ell(\theta)}{d\theta}. \tag{4.4}$$

Para el caso bidimensional la función Score está dada por

$$S(\theta_1, \theta_2) = \begin{bmatrix} S_1(\theta_1, \theta_2) \\ S_2(\theta_1, \theta_2) \end{bmatrix} = \begin{bmatrix} \frac{\partial \ell(\theta_1, \theta_2)}{\partial \theta_1} \\ \frac{\partial \ell(\theta_1, \theta_2)}{\partial \theta_2} \end{bmatrix} \tag{4.5}$$

Definición 4.2.4 *La función de información $I(\boldsymbol{\theta})$ se define como el negativo de la segunda derivada de la log-verosimilitud con respecto de $\boldsymbol{\theta}$.*
Para el caso unidimensional la función de información está dada por

$$I(\theta) = -\ell''(\theta) = -S'(\theta) = \frac{d^2\ell(\theta)}{d\theta^2}. \tag{4.6}$$

Para el caso bidimensional la función de información resulta ser una matriz simétrica de 2×2 que coincide con la matriz de información de Fisher (la cual definiremos posteriormente) y se define de la siguiente manera

$$I(\theta_1, \theta_2) = \begin{bmatrix} I_{11}(\theta_1, \theta_2) & I_{12}(\theta_1, \theta_2) \\ I_{21}(\theta_1, \theta_2) & I_{22}(\theta_1, \theta_2) \end{bmatrix} = \begin{bmatrix} \frac{-\partial^2 \ell(\theta_1, \theta_2)}{\partial \theta_1^2} & \frac{-\partial^2 \ell(\theta_1, \theta_2)}{\partial \theta_1 \partial \theta_2} \\ \frac{-\partial^2 \ell(\theta_1, \theta_2)}{\partial \theta_1 \partial \theta_2} & \frac{-\partial^2 \ell(\theta_1, \theta_2)}{\partial \theta_2^2} \end{bmatrix} \tag{4.7}$$

Plausibilidad

La función de verosimilitud $L(\theta; y)$ permite ordenar la credibilidad o plausibilidad entre los valores de θ a la luz de los datos. Si $L(\theta_1; y) > L(\theta_2; y)$ entonces de (4.1) se sigue que la muestra observada y es más probable cuando el parámetro θ toma el valor θ_1 que cuando toma el valor θ_2. Así, el cociente de verosimilitudes

$$\frac{L(\theta_1; y)}{L(\theta_2; y)},$$

se interpreta como una medida de plausibilidad de θ_1 respecto de θ_2 basada en la muestra observada y. El cociente

$$\frac{L(\theta_1; y)}{L(\theta_2; y)} = k$$

significa que el valor θ_1 es k veces mas plausible que el valor θ_2 en el sentido de que θ_1 hace a la muestra observada k veces más probable de lo que lo hace θ_2.

Estimador de máxima verosimilitud

Definición 4.2.5 *El estimador de máxima verosimilitud (EMV) de $\boldsymbol{\theta}$, es el valor del parámetro en el espacio de parámetros que maximiza la verosimilitud $L(\boldsymbol{\theta})$ o equivalentemente la log-verosimilitud $\ell(\boldsymbol{\theta})$, $\hat{\boldsymbol{\theta}} \in \Omega$, que satisface*

$$L(\hat{\boldsymbol{\theta}}; y) = \sup_{\boldsymbol{\theta} \in \Omega} L(\boldsymbol{\theta}; y). \tag{4.8}$$

El EMV del parámetro $\boldsymbol{\theta}$ es el valor más plausible de $\boldsymbol{\theta}$. Es decir, el EMV $\hat{\boldsymbol{\theta}}$ es el valor de $\boldsymbol{\theta}$ que explica mejor a la muestra observada en el sentido de que maximiza la función de verosimilitud bajo el modelo de probabilidad propuesto para el fenómeno aleatorio de interés.

Algunas veces encontrar el valor de $\boldsymbol{\theta}$ que maximiza $L(\boldsymbol{\theta}, y)$ puede ser complicado debido a la forma que toma el producto de las densidades y en consecuencia la verosimilitud. Usualmente es conveniente y válido trabajar con la función de log-verosimilitud de $\boldsymbol{\theta}$ definida en (4.2). Pues muchas veces es más fácil realizar procesos de optimización, cuando la función objetivo es una suma de funciones que cuando la función objetivo es un producto de funciones.

En términos generales, encontrar los valores de $\hat{\boldsymbol{\theta}}$ no siempre es posible hacerlo algebraícamente, por lo cual hay que hacerlo numéricamente. En este trabajo se realizaron programas en el lenguaje R, utilizando la función optimizadora NLM para maximización y UNIROOT para determinar raíces univariadas. Hay que destacar que el software R tiene funciones que permiten estimar los parámetros de varios modelos usando criterios de máxima verosimilitud pero que no incluyen la estimación con datos censurados, estas funciones son FITDISTR y MLE [33].

4.3. Verosimilitud en censura y truncamiento

El diseño de los experimentos en las ciencias ambientales, supervivencia, en la industria y en las otras áreas donde se aplique la estadística implican datos con censura y truncamiento y esto debe ser cuidadosamente considerado en la construcción de funciones de verosimilitud. Un supuesto fundamental es que los datos observados y los datos con censura son independientes. Si no son independientes, entonces se deben invocar técnicas más especializadas [15].

En la construcción de una función de verosimilitud para datos censurados o truncados tenemos que considerar cuidadosamente la información que cada observación nos da. La

función de verosimilitud para los sistemas de todos los tipos de censura y truncamiento se puede escribir mediante la incorporación de los siguientes componentes:

Observaciones exactas	-	$f(y;\theta)$.
Observaciones censuradas por la derecha	-	$1 - F(C_r)$.
Observaciones censuradas por la izquierda	-	$F(C_l)$.
Observaciones truncadas por la izquierda	-	$f(y;\theta)/[1 - F(T_l)]$.
Observaciones truncadas por la derecha	-	$f(y;\theta)/F(T_r)$.

Donde, C_r y C_l son el punto de censura por la derecha y por la izuierda respectivamente. T_r el punto de truncamiento por la derecha y T_l el truncamiento por la izquierda.

4.3.1. Datos censurados por la derecha.

Definición 4.3.1 *Supóngase que $y_1, y_2, ..., y_n$ son observaciones de una muestra aleatoria de tamaño n, donde $y_1, y_2, ..., y_k$ son k datos observados, $y_{k+1}, y_{k+2}, ..., y_n$ son (m=n-k) datos censurados por la derecha en algún valor C_r. La función de verosimilitud se define como:*

$$L(\theta; y) = \prod_{i=1}^{k} f(y_i; \theta) \prod_{i=k+1}^{n} [1 - F(C_r)]. \tag{4.9}$$

4.3.2. Datos truncados por la izquierda.

Definición 4.3.2 *Supóngase que $y_1, y_2, ..., y_n$ son observaciones de una muestra aleatoria de tamaño n, donde $y_1, y_2, ..., y_k$ son k observaciones censuradas por la izquierda en un valor común y_0 y $y_{k+1}, y_{k+2}, ..., y_n$ son (m=n-k) datos observados. Si se trunca por la izquierda en el valor y_0, entonces la función de verosimilitud se define como:*

$$L(\theta; y) = \prod_{i=k+1}^{n} \frac{f(y_i; \theta)}{[1 - F(y_0)]}. \tag{4.10}$$

4.3.3. Datos truncados por la izquierda y censurados por la derecha.

Definición 4.3.3 *Supóngase que $y_1, y_2, ..., y_n$ son observaciones de una muestra aleatoria de tamaño n que se ha truncado por la izquierda en y_0, donde $y_1, y_2, ..., y_k$ son k datos observados y $y_{k+1}, y_{k+2}, ..., y_n$ son (m=n-k) observaciones censuradas por la derecha en algún valor C_r. La función de verosimilitud para este caso se define como:*

$$L(\theta; y) = \frac{\prod_{i=1}^{k} f(y_i; \theta) \prod_{i=k+1}^{n} [1 - F(C_r)]}{\prod_{i=1}^{n} [1 - F(y_0)]}. \tag{4.11}$$

4.3.4. Función de verosimilitud relativa

Para comparar la plausibidad asociada a los diferentes valores de θ con $\hat{\theta}$, podemos hacer uso de la función de verosimilitud relativa.

Definición 4.3.4 *La función de verosimilitud relativa (FVR) es el resultado de estandarizar la función de verosimilitud con respecto a su máximo.* $R(\boldsymbol{\theta}) : \Omega \to [0, 1]$.

Sea $\boldsymbol{\theta} \in \mathbb{R}^p$, es decir, $\boldsymbol{\theta}$ un vector de parámetros reales, la FVR está dada como:

$$R(\boldsymbol{\theta}) = \frac{L(\boldsymbol{\theta})}{\sup_{\boldsymbol{\theta}\in\Omega} L(\boldsymbol{\theta})}, \tag{4.12}$$

donde $L(\boldsymbol{\theta})$ es la función de verosimilitud de $\boldsymbol{\theta}$. Nótese que $0 \leq R(\boldsymbol{\theta}) \leq 1$, para todo valor de $\boldsymbol{\theta}$ en el espacio parametral, de hecho, $R(\hat{\boldsymbol{\theta}}) = 1$. Valores de θ con $R(\boldsymbol{\theta})$ cercanos a uno son muy razonables o creíbles mientras que valores de $\boldsymbol{\theta}$ con $R(\boldsymbol{\theta})$ cercanos a cero son poco creíbles a la luz de los datos.

En la práctica, generalmente es posible encontrar un único valor de $\boldsymbol{\theta}$ en el espacio parametral que maximiza la función de verosimilitud $L(\boldsymbol{\theta})$. Cuando el estimador de máxima verosimilitud de $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}$, existe y es único , entonces la función de verosimilitud relativa se expresa como

$$R(\boldsymbol{\theta}) = \frac{L(\boldsymbol{\theta})}{L(\hat{\boldsymbol{\theta}})}$$

Definición 4.3.5 *El logaritmo de la función de verosimilitud relativa, es conocido como la función de log-verosimilitud relativa y está dado por*

$$r(\boldsymbol{\theta}) = \ln\left[\frac{L(\boldsymbol{\theta})}{L(\hat{\boldsymbol{\theta}})}\right] = \ell(\boldsymbol{\theta}) - \ell(\hat{\boldsymbol{\theta}}) \tag{4.13}$$

La verosimilitud relativa para un valor θ_0 en el caso de $\boldsymbol{\theta}$ unidimensional sería

$$R(\theta_0) = \frac{L(\theta_0)}{L(\hat{\theta})}$$

y se interpreta de la siguiente forma: si $R(\theta_0)$ es cercano a $R(\hat{\theta}) = 1$, esto nos dice que el valor θ_0 hace a la muestra observada casi tan creíble como lo hace el EMV $\hat{\theta}$. En contraste, si $R(\theta_0)$ es cercano a cero, quiere decir que el valor θ_0 es inviable para estimar θ, ya que existen otros valores del parámetro para los cuales la muestra es mucho más creíble [11], [31].

Invarianza funcional

En muchas ocasiones se utilizan reparametrizaciones de un modelo de probabilidad por conveniencia matemática, conveniencia computacional, propio interés de un parámetro

que es función de otro que aparece en el modelo, etc. En estas situaciones, la invarianza funcional es una propiedad muy conveniente de la verosimilitud y significa que, en términos de plausibilidad, cualquier declaración cuantitativa acerca de θ implica una declaración cuantitativa correspondiente acerca de cualquier función uno a uno de θ, $\delta = g(\theta)$, por sustitución directa algebraica $\theta = g^{-1}(\delta)$.

Por ejemplo, si $\theta > 0$ y $\delta = \ln\theta$, entonces la verosimilitud del nuevo parámetro δ es $L(\delta, y) = L[\theta = e^{\delta}; y]$. También $a \leq \theta \leq b$ sí y sólo sí $\ln a \leq \delta \leq \ln b$. Estas dos declaraciones son equivalentes debido a que tienen la misma plausibilidad o incertidumbre.

Teorema 4.3.1 *Supóngase que $L(\theta; y)$ es la función de verosimilitud de $\theta \in \Omega$. Sea $g(\theta) : \Omega \to \triangle$ una función uno a uno. Si*

$$\frac{L(\theta_1; y)}{L(\theta_2; y)} = k$$

entonces

$$\frac{L(\delta_1; y)}{L(\delta_2; y)} = k,$$

donde $\delta_1 = g(\theta_1)$ y $\delta_2 = g(\theta_2)$.

La demostración del teorema anterior se sigue por sustitución directa algebraica.
Una consecuencia inmediata de la propiedad de invarianza de la función de verosimilitud $L(\theta; y)$ es la invarianza del estimador de θ.

Teorema 4.3.2 (Invarianza del EMV [9]) *Si $g(\theta) : \Omega \to \triangle$ es una función uno a uno y $\hat{\theta}$ es el EMV de θ, entonces el EMV de $\delta = g(\theta)$ es $\hat{\delta} = g(\hat{\theta})$.*

La invarianza funcional de la verosimilitud es una propiedad muy útil en la práctica. En muchos casos ocurre que el parámetro θ no es de interés principal sino que lo es otro parámetro que es función de θ. Por otro lado, con frecuencia un cambio en el parámetro puede simetrizar la forma de la función de verosimilitud. Es decir, que $R(\delta; y)$ puede ser más simétrica, o tener una forma aproximadamente más normal, que $R(\theta; y)$. Así, inferencias sobre δ tendrán una estructura más simple, pero matemáticamente equivalente, que aquellas en términos de θ ya que se "acelera" la cercanía a las propiedades asintóticas del EMV $\hat{\delta}$ para la muestra pequeña en cuestión [24].

4.3.5. Matriz de información de Fisher

Ahora definiremos la matriz de información de Fisher, cuya matriz inversa es la matriz de varianzas-covarianzas que utilizaremos para construir los intervalos de confianza basados en las estimaciones de máxima verosimilitud de nuestros parámetros.

Definición 4.3.6 *La matriz de información de Fisher, también denominada Matriz de Información Esperada, para un estimador de máxima verosimilitud* $\hat{\boldsymbol{\theta}}$ *de un vector de parámetros* $\boldsymbol{\theta} \in \mathbb{R}^p$ *es*

$$I_{\boldsymbol{\theta}} = \left[-\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \quad i, j = 1, .., p, \tag{4.14}$$

que es una matriz cuadrada de tamaño $p \times p$. Cada elemento de la matriz es el inverso aditivo ó negativo de la segunda derivada parcial de la función de log-verosimilitud evaluada en el estimador de máxima verosimilitud del modelo de distribución correspondiente.

$I_{\boldsymbol{\theta}}$ es conocida comúnmente como matriz de información de Fisher para $\boldsymbol{\theta}$. Con los datos disponibles, se puede calcular la matriz "local" (o matriz de información) para $\boldsymbol{\theta}$

$$\hat{I}_{\boldsymbol{\theta}} = \left[-\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right], \tag{4.15}$$

donde las derivadas parciales son evaluadas en $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$

La matriz de varianzas-covarianzas del vector aleatorio $\hat{\boldsymbol{\theta}}$ es

$$\Sigma_{\hat{\boldsymbol{\theta}}} = \left[-\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right]^{-1}_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \tag{4.16}$$

que es la inversa de la matriz de información de Fisher. Los elementos diagonales de esta última matriz se corresponden con las varianzas de los estimadores de máxima verosimilitud de cada componente del parámetro.

4.3.6. Intervalos y regiones de verosimilitud

Una manera usual de hacer inferencia sobre un parámetro de interés es a través de intervalos o regiones de estimación. Los intervalos de verosimilitud, o de forma más general, de regiones de verosimilitud, indican los valores más plausibles del parámetro a la luz de la muestra observada. Una región de $100p\,\%$ de verosimilitud (una región de $100p\,\%$) de nivel p para θ, se define como

$$IV(p) = \{\theta : R(\theta) \geq p\}, \text{donde} \quad 0 \leq p \leq 1.$$

Frecuentemente la región $100p\,\%$ de verosimilitud consiste de un intervalo de valores reales, y entonces es llamado *intervalo de* $100p\,\%$ *de verosimilitud* IV(p) de nivel p para θ, todo valor de θ en el IV(p) tiene verosimilitud relativa igual o mayor que p, y todo valor de θ afuera, tiene verosimilitud relativa menor. Por lo tanto el IV(p) separa los valores plausibles de θ de los no plausibles a un nivel p [14].

Las regiones o intervalos de verosimilitud pueden ser determinados a partir de un gráfico de $R(\theta)$ o su logaritmo $r(\theta)$, y generalmente es más conveniente trabajar con $r(\theta)$. Los valores extremos de los intervalos de $100p\,\%$ de verosimilitud pueden ser encontrados como raíces de la ecuación $r(\theta) - ln(p) = 0$. Por lo general es necesario resolver esta ecuación numéricamente.

Aproximación normal para los intervalos de verosimilitud-confianza.
Sea $\ell(\theta)$ la función de log-verosimilitud de un parámetro continuo θ con valores posibles en Ω. Sean $S(\theta) = \ell'(\theta)$ y $I_\theta = -\ell''(\theta)$ la primera derivada de la función de log-verosimilitud y la matriz de información, respectivamente. Si suponemos que $\hat{\theta}$ existe y es un punto interior de Ω, y que $\ell(\theta)$ tiene una expansión en series de Taylor alrededor de $\theta = \hat{\theta}$, como sigue

$$\ell(\theta) = \ell(\hat{\theta}) + \frac{\theta - \hat{\theta}}{1!}\ell'(\hat{\theta}) + \frac{(\theta - \hat{\theta})^2}{2!}\ell''(\hat{\theta}) + \frac{(\theta - \hat{\theta})^3}{3!}\ell'''(\hat{\theta}) + ...$$

Dado que $\ell'(\hat{\theta}) = 0$ y considerando $r(\theta) = \ell(\theta) - \ell(\hat{\theta})$, se tiene que

$$r(\theta) = -\frac{1}{2}(\theta - \hat{\theta})^2 I_{\hat{\theta}} + \frac{(\theta - \hat{\theta})^3}{3!}\ell'''(\hat{\theta}) + ... \tag{4.17}$$

Entonces la aproximación normal para $r(\theta)$ está definida como sigue:

$$r_N(\theta) = -\frac{1}{2}(\theta - \hat{\theta})^2 I_{\hat{\theta}} \tag{4.18}$$

si $|\theta - \hat{\theta}|$ es pequeño, el término cúbico y los de grado superior en (4.17) también son muy pequeños, entonces $r(\theta) \approx r_N(\theta)$, es decir,

$$r(\theta) \approx -\frac{1}{2}(\theta - \hat{\theta})^2 I_{\hat{\theta}}.$$

El efecto de aumentar la cantidad de los datos produce con mayor rapidez que la función de verosimilitud alcanze su máximo y que los intervalos de verosimilitud sean más cortos. Así, para una muestra aleatoria suficientemente grande, $|\theta - \hat{\theta}|$ es pequeño y $r_N(\theta)$ dará una buena aproximación para $r(\theta)$ sobre la región entera de valores parámetricos plausibles.

La región de $100p\,\%$ de verosimilitud para θ es el conjunto de valores de θ para el cual $r(\theta) \geq \ln p$. De este modo se tiene que

$$\theta \in \hat{\theta} \pm \sqrt{-2\ln p I_{\hat{\theta}}^{-1}}, \tag{4.19}$$

Este es un intervalo centrado en $\hat{\theta}$ con longitud

$$2\sqrt{-2\ln p I_{\hat{\theta}}^{-1}}.$$

Cuanto más grande sean los valores de $I_{\hat{\theta}}$, más estrecho será el intervalo aproximado y por lo tanto más información se tendrá sobre θ. Ésta es la razón por la cual $I_{\hat{\theta}}$ es llamada la "matriz de información".

Para el caso que $\theta = (\theta_1, \theta_2)$ también podemos tener una aproximación normal para los intervalos de verosimilitud de θ_1 y θ_2. Está aproximación normal se deriva de manera similar como para el caso de tener un solo parámetro [14].

El intervalo de verosimilitud por aproximación normal para θ_1 está dado por

$$r_{max}(\theta_1) \approx -\frac{1}{2}(\theta_1 - \hat{\theta}_1)^2 \left[\hat{I}_{11} - \hat{I}_{12}^2/\hat{I}_{22}\right]. \tag{4.20}$$

Donde, los extremos del intervalo son las raices resultantes de la ecuación anterior.

El intervalo de verosimilitud por aproximación normal para θ_2 está dado por

$$r_{max}(\theta_2) \approx -\frac{1}{2}(\theta_2 - \hat{\theta}_2)^2 \left[\hat{I}_{22} - \hat{I}_{12}^2/\hat{I}_{11}\right]. \tag{4.21}$$

Donde, los extremos del intervalo son las raices resultantes de la ecuación anterior.

4.3.7. Función de verosimilitud relativa máxima

Con frecuencia al trabajar con modelos continuos de dos parámetros como por ejemplo los modelos Weibull, Lognormal, entre otros, lo que interesa es realizar estimaciones de un solo parámetro α o β, Para realizar dichas estimaciones se utiliza el enfoque de verosimilitud relativa máxima.
Si solo estamos interesados en la información concerniente al parámetro β, entonces se utiliza la función de verosimilitud relativa máxima de β, la cual está dada por

$$R_{\text{máx}}(\beta) = \max_{\alpha} R(\alpha, \beta) = R(\hat{\alpha}(\beta), \beta) = \frac{L(\hat{\alpha}(\beta), \beta)}{L(\hat{\alpha}, \hat{\beta})} \tag{4.22}$$

donde, $\hat{\alpha}(\beta)$ es el EMV de α dado β el cual se puede hallar resolviendo la ecuación $S_1(\alpha, \beta) = 0$.

El logaritmo de la FVR máxima es

$$r_{\text{máx}}(\beta) = \ln R(\hat{\alpha}(\beta), \beta) = \ell(\hat{\alpha}(\beta), \beta) - \ell(\hat{\alpha}, \hat{\beta}) \tag{4.23}$$

De lo anterior se obtiene que un intervalo de verosimilitud máximo del $100p\,\%$ para β es el conjunto de todos los valores de β tales que $R_{\text{máx}}(\beta) \geq p$, o equivalentemente $r_{\text{máx}}(\beta) \geq p$.

Similarmente, la FVR máxima de α está dada por

$$R_{\text{máx}}(\alpha) = \max_{\beta} R(\alpha, \beta) = R(\alpha, \hat{\beta}(\alpha)) = \frac{L(\alpha, \hat{\beta}\alpha)}{L(\hat{\alpha}, \hat{\beta})} \tag{4.24}$$

y el logaritmo de la FVR máxima es

$$r_{\text{máx}}(\alpha) = \ln R(\hat{\alpha}(\beta), \beta) = \ell(\alpha, \hat{\beta}(\alpha)) - \ell(\hat{\alpha}, \hat{\beta}) \tag{4.25}$$

4.4. Intervalos de verosimilitud-confianza por aproximación normal

Los intervalos de verosimilitud confianza obtenidos por aproximación normal son fáciles de calcular, en la actualidad son muy utilizados en los paquetes estadísticos comerciales. Utilizando los estimadores de máxima verosimilitud de los parámetros de los modelos estadísticos y con ayuda de la matriz de covarianzas el cálculo de estos intervalos es posible a través de cálculos no tan complejos.

Los intervalos y regiones de confianza por aproximación normal están basadas en una aproximación cuadrática de la log-verosimilitud y son apropiados cuando la logverosimilitud es aproximadamente cuadrática sobre la región de confianza. Con muestras grandes, bajo condiciones de regularidad, la log-verosimilitud es aproximadamente cuadrática y entonces la aproximación normal y los intervalos serán congruentes. El tamaño de muestra requerido para una buena aproximación no es fácilmente elegible debido a que depende del modelo, de la cantidad de información censurada o truncada y de la cantidad de interés en particular. Aún con muestras pequeñas y cantidades significativas de censura, esta aproximación resulta adecuada y funcional, considerando modelos de localización-escala y un número suficiente de datos observados [22].

4.4.1. Intervalos y regiones de confianza aproximados

La aproximación normal de muestras grandes para la distribución de EMVs puede ser usada para construir intervalos (regiones) de confianza aproximados para funciones escalares (vectoriales) de $\boldsymbol{\theta}$ [22]. En particular, una región aproximada del $100(1-\alpha)\,\%$ de confianza para $\boldsymbol{\theta}$ es el conjunto de todos los valores de $\boldsymbol{\theta}$ en el elipsoide

$$\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)' \left(\hat{\Sigma}_{\hat{\boldsymbol{\theta}}}\right)^{-1} \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \leq \chi^2_{(1-\alpha,r)}, \tag{4.26}$$

donde r es la longitud de $\boldsymbol{\theta}$ y $\hat{\Sigma}_{\hat{\boldsymbol{\theta}}}$ es el estimador local de la matriz de covarinzas de $\hat{\boldsymbol{\theta}}$. Esto se conoce como el Método de Wald, conocido también como el método de aproximación normal. Esta región de confianza está basada en la distribución que

resulta, asintóticamente, cuando es evaluado en el valor verdadero de $\boldsymbol{\theta}$, el estadístico de Wald

$$W(\boldsymbol{\theta}) = \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)' \left(\hat{\Sigma}_{\hat{\boldsymbol{\theta}}}\right)^{-1} \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \sim \chi^2_r. \tag{4.27}$$

Más generalmente, sea $g(\boldsymbol{\theta})$ una función vectorial de $\boldsymbol{\theta}$. Entonces, una región aproximada del $100(1-\alpha)\,\%$ de confianza para un conjunto r_1-dimensional $g_1 = g_1(\boldsymbol{\theta})$, de la partición $g = [g_1(\boldsymbol{\theta}), g_2(\boldsymbol{\theta})]$, es el conjunto de los valores de g_1 en el elipsoide

$$(\hat{g}_1 - g_1)' \left(\hat{\Sigma}_{\hat{g}_1}\right) (\hat{g}_1 - g_1) \leq \chi^2_{(1-\alpha,r)}, \tag{4.28}$$

donde $\hat{g}_1 = g_1(\hat{\boldsymbol{\theta}})$ es el EMV de $g_1(\boldsymbol{\theta})$ y $\hat{\Sigma}_{\hat{g}_1}$ es el estimador local de la matriz de covarianza de $\hat{g}_1$. El estimador $\hat{\Sigma}_{\hat{g}_1}$ puede ser obtenido de,

$$\hat{\Sigma}_{\hat{g}} = \left[\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right] \hat{\Sigma}_{\hat{\boldsymbol{\theta}}} \left[\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]. \tag{4.29}$$

Bajo condiciones de regularidad $\hat{\Sigma}_{\hat{\boldsymbol{\theta}}} = \left(\hat{I}_{\boldsymbol{\theta}}\right)^{-1}$ es un estimador consistente de $\Sigma_{\hat{\boldsymbol{\theta}}}$, donde

$$\hat{I}_{\boldsymbol{\theta}} = E\left[-\frac{\partial^2 \ln L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right] \tag{4.30}$$

es la matriz de información observada de Fisher [22]. Esta región de confianza (o intervalo)(4.28) está basado en el resultado distribucional asintótico del subconjunto del estadístico de Wald, cuando es evaluado en el valor verdadero de g_1,

$$W(g_1) = (\hat{g}_1 - g)' \left(\hat{\Sigma}_{\hat{g}_1}\right) (\hat{g}_1 - g). \tag{4.31}$$

Esta región de confianza aproximada (o intervalo) puede ser vista como una aproximación cuadrática para la log-verosimilitud perfil de $g_1(\boldsymbol{\theta})$ en $\hat{g}_1$ [22]. Cuando $r_1 = 1$, $g_1 = g_1(\boldsymbol{\theta})$, es una función escalar de $\boldsymbol{\theta}$; un intervalo de confianza aproximado del $100(1-\alpha)\,\%$ es obtenido de la formula familiar

$$\left[\underset{\sim}{g_1}, \widetilde{g_1}\right] = \hat{g}_1 \pm z_{(1-\alpha/2)} \hat{es}_{\hat{g}_1}, \tag{4.32}$$

donde, $\hat{es}_{\hat{g}_1} = \sqrt{\hat{Var}\left[g_1(\hat{\boldsymbol{\theta}})\right]}$ es el estimador local para el error estándar de $\hat{g}_1$ y $z_{(1-\alpha/2)}$ es el $1-\alpha/2$ cuantil de la distribución normal estándar. Como un caso particular, un intervalo aproximado del $100(1-\alpha)\,\%$ de confianza para θ_i es

$$\left[\underset{\sim}{\theta_i}, \widetilde{\theta_i}\right] = \hat{\theta}_i \pm z_{(1-\alpha/2)} \hat{es}_{\hat{\theta}_i}, \tag{4.33}$$

donde $\hat{es}_{\hat{\theta}_i}$ es la raíz cuadrada de la i-ésima entrada de $\hat{\Sigma}_{\hat{\boldsymbol{\theta}}}$.

4.4.2. Intervalos de confianza para funciones de μ y σ

Un intervalo de confianza aproximado para una función $g_1 = g_1(\mu, \sigma)$, puede estar basada en la distribución aproximada normal estándar $N(0,1)$ de $Z_{\hat{g}_1} = (\hat{g}_1 - g_1)/\hat{es}_{\hat{g}_1}$.

Entonces un intervalo aproximado del $100(1-\alpha)\,\%$ de confianza para g_1 está dado como $[g_{1_I}, g_{1_D}] = \hat{g}_1 \pm z_{(1-\alpha/2)}\hat{es}_{\hat{g}_1}$. Donde, usando un caso especial de (4.29),

$$\hat{es}_{\hat{g}_1} = \left[\left(\frac{\partial g_1}{\partial \mu}\right)^2 \hat{Var}(\hat{\mu}) + 2\left(\frac{\partial g_1}{\partial \mu}\right)\left(\frac{\partial g_1}{\partial \sigma}\right)\hat{Cov}(\hat{\mu}, \hat{\sigma}) + \left(\frac{\partial g_1}{\partial \sigma}\right)^2 \hat{Var}(\hat{\sigma})\right]^{1/2} \tag{4.34}$$

Las derivadas parciales en (4.34) deben ser evaluadas en $(\mu, \sigma) = (\hat{\mu}, \hat{\sigma})$

4.5. Prueba de hipótesis

Definición 4.5.1 *Se llama hipótesis estadística a cualquier afirmación o conjetura referente a los parámetros de una o más poblaciones [9].*

La definición de una hipótesis es bastante general, pero el punto importante es que una hipótesis es una declaración acerca de la población. Note que al establecer una hipótesis siempre existe de forma implícita, otra que se le contrapone.

Definición 4.5.2 *Las dos hipótesis complementarias en un problema de prueba de hipótesis se conocen como la hipótesis nula y la hipótesis alternativa. Denotadas como H_0 y H_1, respectivamente.*

Si θ denota un parámetro de la población, el formato general de la hipótesis nula y alternativa es $H_0 : \theta \in \Phi_0$ y $H_1 : \theta \in \Phi_0^c$ donde Φ_0 es algún subconjunto del espacio de parámetros y Φ_0^c es su complemento.

Probar una hipótesis estadística consiste en buscar evidencias para decidir sobre la aceptación o rechazo de la afirmación realizada.

15
Los procedimientos que facilitan el decidir si una hipótesis se rechaza o no, así como el determinar si las muestras observadas difieren significativamente de los resultados esperados se llaman pruebas de hipótesis, ensayos de significancia o reglas de desición. Su objetivo es decir, en base a una muestra de la población, cuál de las dos hipótesis complementarias es cierta.

Definición 4.5.3 *Un procedimiento de comprobación de hipótesis o prueba de hipótesis con respecto a alguna característica desconocida de la población de interés es una regla que especifica:*

1. *Para cuales valores de la muestra se acepta H_0 como verdadera.*
2. *Para cuales valores de la muestra se rechaza H_0 y se acepta H_1 como verdadera.*

El subconjunto del espacio muestral para el que H_0 será aceptada se llama región de aceptación o región de no rechazo, y la región para la cual H_0 será rechazada se llama región de rechazo o región crítica; son denotadas por R_a y R_r, respectivamente.

Al tomar una desición con respecto a la validez de la hipótesis nula, el investigador está propenso a cometer uno de dos errores posibles.

4.6. Selección de modelos

Encontrar un modelo razonable es importante para describir bien el comportamiento aleatorio del fenomeno de interés. Pawitan (2001), Anderson (2008) presentan varios métodos de comparación de modelos, de los cuales utilizaremos aquí el criterio de razón de verosimilitudes y el criterio de Akaike.

4.6.1. Razón de verosimilitud

El test de razón de verosimilitudes es un contraste de bondad de ajuste entre dos modelos y que nos sirve de herramienta para evaluar si un modelo general ajusta mejor los datos que uno restringido.

Sea $\{y_1, y_2, ..., y_n\}$ una muestra independiente e idénticamente distribuida de una densidad desconocida $f(y, \boldsymbol{\theta})$, suponiendo que la función de densidad depende de un vector de parámetros desconocidos $\boldsymbol{\theta}$ de dimensión p. Sea $\boldsymbol{\theta}_0$ un vector de valores numéricos, uno para cada componente de $\boldsymbol{\theta}$. El estadístico del test de razón de verosimilitudes para la hipótesis $H : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ es dos veces la diferencia entre el máximo de la log-verosimilitud del modelo general en $\hat{\boldsymbol{\theta}}$ y la log-verosimilitud del modelo restringido bajo H. Entonces, un modelo restringido ajusta los datos casi tan bien como el modelo general si asintóticamente,

$$2\left[\ell(\hat{\boldsymbol{\theta}}) - \ell(\boldsymbol{\theta}_0)\right] \sim \chi^2_v. \tag{4.35}$$

χ^2_v es una variable aleatoria χ^2 con v grados de libertad, donde v está en función del número de parámetros desconocidos en el modelo.

El test de razón de verosimilitudes también puede ser usada para elegir el mejor modelo que ajuste los datos entre varios de ellos, comparando las funciones de verosimilitudes estimadas. Así, si dos modelos ajustan bien los datos, se cumple la siguiente relación

$$\chi^2 = 2\left[\ell(\hat{\boldsymbol{\theta}}_{Modelo2}) - \ell(\hat{\boldsymbol{\theta}}_{Modelo1})\right] \tag{4.36}$$

Esto sirve de soporte y descripción para la prueba de hipótesis siguiente

$$H : Modelo1 \equiv Modelo2, \tag{4.37}$$

donde Modelo 1 y Modelo 2 representan dos modelos candidatos para describir la información de la muestra aleatoria. Entonces se rechazará H si $\chi^2 > \chi^2_{1,\alpha}$ con una confiabilidad del $100(1-\alpha)\,\%$.

4.6.2. Criterio de Akaike

El criterio de información de Akaike (AIC por sus siglas en inglés) es una medida relativa de la bondad de ajuste para un modeo estadístico. Este concepto se basa en la entropía de la información, la cual ofrece una medida relativa de la pérdida de información cuando un determinado modelo se utiliza para describir datos reales [1].

Supóngase que $\{y_1, y_2, ..., y_n\}$ es una muestra independiente e idénticamente distribuida de una densidad desconocida $f(y, \boldsymbol{\theta})$ y $\hat{\boldsymbol{\theta}}$ es el EMV de $\boldsymbol{\theta}$ correspondiente. Suponiendo que el vector de parámetros $\boldsymbol{\theta}$ es de dimensión p, se defíne el AIC como:

$$AIC = -2\ln\left(\mathcal{L}(\hat{\boldsymbol{\theta}})\right) + 2p. \tag{4.38}$$

Los valores del AIC proporcionan una alternativa para la selección de un modelo entre una familia de ellos, de tal suerte que el modelo a elegirse será aquel con el menor AIC. La interpretación de (4.38) suele hacerse en el sentido de que el primer término del lado derecho (la log-verosimilitud) es una medida de bondad de ajuste (Anderson) y el segundo término (parámetros) es una medida de la penalidad de la complejidad del modelo ya que un modelo con menos parámetros es preferible a otro con mayor número de parámetros [1]. Por ello, modelos con más parámetros tendrán una penalización mayor.

Una situación a considerar bajo este criterio es que no determina una selección absoluta de algún modelo en particular sino, una medidad relativa para elegir el modelo más apropiado entre varios que comparten características o naturaleza comunes. Si todos los modelos candidatos tienen el mismo número de parámetros, entonces, la utilización del AIC es equivalente a implementar una prueba de razón de verosimilitudes [5].

Capítulo 5

Metodología Desarrollada

5.1. Estimación para muestras censuradas por la derecha.

5.1.1. Modelo Exponencial.

Supóngase que $y_1, y_2, ..., y_n$ es una muestra aleatoria de n observaciones, las cuales proceden de una población con distribución exponencial de parámetro θ, donde $y_1, y_2, ..., y_k$ son k observaciones no censuradas y $y_{k+1}, y_{k+2}, ..., y_n$ son $(m{=}n{-}k)$ observaciones censuradas por la derecha en algún valor C_r. De (4.3.1) la función de verosimilitud para este caso está dada por

$$L(\theta; y) = \prod_{i=1}^{k} \frac{1}{\theta} e^{-y_i/\theta} \prod_{i=k+1}^{n} e^{-C_r/\theta} \tag{5.1}$$

y la función de log-verosimilitud está dada como sigue

$$\ell(\theta; y) = -k \ln \theta - \frac{1}{\theta} \left[\sum_{i=1}^{k} y_i + \sum_{i=k+1}^{n} C_r \right]. \tag{5.2}$$

Luego, la función score y la función de información son, respectivamente

$$S(\theta) = -\frac{k}{\theta} + \frac{1}{\theta^2} \left[\sum_{i=1}^{k} y_i + \sum_{i=k+1}^{n} C_r \right], \tag{5.3}$$

$$I(\theta) = -\frac{k}{\theta^2} + \frac{2}{\theta^3} \left[\sum_{i=1}^{k} y_i + \sum_{i=k+1}^{n} C_r \right] \tag{5.4}$$

Para encontrar $\hat{\theta}$, el EMV de θ resolvemos $S(\theta) = 0$, de este modo

$$\hat{\theta} = \frac{c}{k}. \tag{5.5}$$

Donde $c = \left[\sum_{i=1}^{k} y_i + \sum_{i=k+1}^{n} C_r\right]$

Para verificar que $\hat{\theta}$ es un máximo, es suficiente que cumpla las condiciones siguientes [14]:

$$S(\hat{\theta}) = 0 \quad y \quad I(\hat{\theta}) > 0$$

Se tiene entonces que

$$S(\hat{\theta}) = -\frac{k}{\hat{\theta}} + \frac{c}{\hat{\theta}^2} = -\frac{k}{\hat{\theta}} + \frac{k\hat{\theta}}{\hat{\theta}^2} = -\frac{k}{\hat{\theta}} + \frac{k}{\hat{\theta}} = 0.$$

Por otro lado, se tiene también que

$$I(\hat{\theta}) = -\frac{k}{\hat{\theta}^2} + \frac{2c}{\hat{\theta}^3} = -\frac{k}{\hat{\theta}^2} + \frac{2k\hat{\theta}}{\hat{\theta}^3} = -\frac{k}{\hat{\theta}^2} + \frac{2k}{\hat{\theta}^2} = \frac{k}{\hat{\theta}^2} > 0.$$

Por lo tanto $\hat{\theta}$ es el EMV para el modelo exponencial con datos censurados por la derecha.

Intervalo de verosimilitud-confianza para θ.
La función de log-verosimilitud relativa de θ para este caso es

$$r(\theta) = -k\ln\theta - \frac{c}{\theta} + k\ln\hat{\theta} + \frac{c}{\hat{\theta}}. \tag{5.6}$$

Para encontrar el intervalo de verosimilitud de $100p\,\%$ para θ resolvemos la ecuación

$$r(\theta) - \ln p = 0. \tag{5.7}$$

Esta ecuación, la resolvemos utilizando la función UNIROOT de R. Dicha función nos da como resultado las raices de la ecuación (5.7), donde las raices encontradas serán los extremos del intervalo de verosimilitud para θ. Los valores iniciales que la función UNIROOT necesita los tomamos de la aproximación normal para el intervalo de verosimilitud de θ, la cual está dada de acuerdo a (4.19) por

$$[\underline{\theta}, \widetilde{\theta}] = \hat{\theta} \pm \sqrt{-2\ln p \left(\frac{k}{\hat{\theta}^2}\right)^{-1}}.$$

Intervalo de confianza de aproximación normal.
Un intervalo de $100(1-\alpha)\,\%$ de confianza para θ está dado por

$$\left[\underline{\theta}, \widetilde{\theta}\right] = \hat{\theta} \pm z_{(1-\alpha/2)}\widehat{es}_{\hat{\theta}}. \tag{5.8}$$

donde

$$\widehat{es}_{\hat{\theta}} = \sqrt{\left[-\frac{d^2\ell(\theta)}{d\theta^2}\right]^{-1}} = \sqrt{I(\hat{\theta})^{-1}}$$

y $z_{(p)}$, es el p-cuantil de la distribución normal estándar.

5.1.2. Modelo Weibull

Supóngase que $y_1, y_2, ..., y_n$ es una muestra aleatoria de n observaciones, las cuales proceden de una población con distribución Weibull con parámetros λ y β, donde $y_1, y_2, ..., y_k$ son k observaciones no censuradas y $y_{k+1}, y_{k+2}, ..., y_n$ son (m=n-k) observaciones censuradas por la derecha en algún valor C_r. De (4.3.1) la función de verosimilitud para este caso está dada por

$$L(\lambda, \beta; y) = \prod_{i=1}^{k} \lambda\beta y_i^{\beta-1} e^{-\lambda y_i^{\beta}} \prod_{i=k+1}^{n} e^{-\lambda C_r^{\beta}} \tag{5.9}$$

Y la función de log-verosimilitud está dada como sigue

$$\ell(\lambda, \beta; y) = k \ln \lambda + k \ln \beta + (\beta - 1) \sum_{i=1}^{k} \ln y_i - \lambda \left[\sum_{i=1}^{k} y_i^{\beta} + \sum_{i=k+1}^{n} C_r^{\beta} \right]. \tag{5.10}$$

La función score para el caso de tener dos parámetros de acuerdo a (4.5) está dada por

$$S(\lambda, \beta) = \begin{bmatrix} S_1(\lambda, \beta) \\ S_2(\lambda, \beta) \end{bmatrix}, \tag{5.11}$$

donde,

$$S_1(\lambda, \beta) = \frac{k}{\lambda} - \left[\sum_{i=1}^{k} y_i^{\beta} + \sum_{i=k+1}^{n} C_r^{\beta} \right].$$

$$S_2(\lambda, \beta) = \frac{k}{\beta} + \sum_{i=1}^{k} \ln y_i - \lambda \left[\sum_{i=1}^{k} y_i^{\beta} \ln(y_i) + \sum_{i=k+1}^{n} C_r^{\beta} \ln(C_r) \right].$$

Para encontrar $(\hat{\lambda}, \hat{\beta})$, resolvemos el par de ecuaciones simultáneas

$$S_1(\lambda, \beta) = 0 \tag{5.12}$$

$$S_2(\lambda, \beta) = 0 \tag{5.13}$$

Note que la ecuación (5.12) puede ser resuelta algebraicamente obteniendo así

$$\hat{\lambda}(\beta) = \frac{k}{c}, \text{donde } c = \left[\sum_{i=1}^{k} y_i^{\beta} + \sum_{i=k+1}^{n} C_r^{\beta} \right] \tag{5.14}$$

Este es el estimador de máxima verosimilitud de λ cuando suponemos que conocemos a β. Para obtener $\hat{\beta}$ el estimador de máxima verosimilitud de β, sustituimos λ en (5.13) y resolvemos la ecuación para β, procediendo como sigue a continuación.

Sea,

$$\begin{aligned} g(\beta) &= S_2(\hat{\lambda}(\beta), \beta) \\ &= \frac{k}{\beta} + \sum_{i=1}^{k} \ln y_i - \frac{k}{c}\left[\sum_{i=1}^{k} y_i^{\beta} \ln(y_i) + \sum_{i=k+1}^{n} C_r^{\beta} \ln(C_r)\right]. \end{aligned}$$

La ecuación $g(\beta) = 0$ puede ser resuelta usando el método de Newton.

El parámetro λ no representa una cantidad de interés, y por lo general es preferible trabajar con los parámetros (θ, β) donde $\lambda = \theta^{-\beta}$, así, de la expresión (5.10), obtenemos $\ell(\theta, \beta)$, dada como sigue

$$\ell(\theta, \beta) = -k\beta \ln \theta + k \ln \beta + (\beta - 1) \sum_{i=1}^{k} \ln y_i - \theta^{-\beta} c. \tag{5.15}$$

De la ecuación anterior se obtiene la función de información $I(\theta, \beta)$, la cual nos será de gran utilidad a la hora de obtener los intervalos de verosimilitud y los intervalos por aproximación normal de θ y β, la función de información para este caso de acuerdo a (4.7) se define de la siguiente manera

$$I(\theta, \beta) = \begin{bmatrix} I_{11}(\theta, \beta) & I_{12}(\theta, \beta) \\ I_{21}(\theta, \beta) & I_{22}(\theta, \beta) \end{bmatrix}, \tag{5.16}$$

donde,

$$\begin{aligned} I_{11}(\theta, \beta) &= -\frac{k\beta}{\theta^2} + c\beta(\beta + 1)\theta^{-(\beta+2)}, \\ I_{12}(\theta, \beta) &= \frac{k}{\theta} - \beta\theta^{-(\beta+1)} c'_{\beta} - \theta^{-(\beta+1)} c + c\beta\theta^{-(\beta+1)} \ln \theta, \\ I_{21}(\theta, \beta) &= \frac{k}{\theta} - \beta\theta^{-(\beta+1)} c'_{\beta} - \theta^{-(\beta+1)} c + c\beta\theta^{-(\beta+1)} \ln \theta, \\ I_{22}(\theta, \beta) &= \frac{k}{\beta^2} + c''_{\beta}\theta^{-\beta} - 2c'_{\beta}\theta^{-\beta} \ln \theta + c\theta^{-\beta} \ln^2 \theta. \end{aligned}$$

Se define la siguiente notación $c'_{\beta} = \frac{dc}{d\beta}$ y $c''_{\beta} = \frac{d^2c}{d\beta^2}$

Intervalos de verosimilitud-confianza
Para calcular los intervalos de verosimilitud, haremos uso de la función de logverosimilitud relativa de θ y β, dada por:

$$r(\theta, \beta) = \ell(\theta, \beta) - \ell(\hat{\theta}, \hat{\beta})$$

Intervalo de verosimilitud para β
Para encontrar el intervalo de verosimilitud de $100p\,\%$ para β resolvemos la ecuación

$$r_{max}(\beta) - \ln p = r(\hat{\lambda}, \beta) - \ln p = 0,$$

sustituimos el valor de $\hat{\lambda}(\beta)$ dentro de la ecuación (5.10), con lo cual tenemos que

$$r_{max}(\beta) - \ln p = k \ln k - k \ln c + k \ln \beta + (\beta - 1) \sum_{i=1}^{k} \ln y_i - k - \ell(\hat{\lambda}, \hat{\beta}) - \ln p = 0.$$

La función uniroot de R nos da como resultado las raíces de está ecuación, las cuales serán los extremos del intervalo de verosimilitud para β. Los valores iniciales que la función uniroot necesita los tomamos de los extremos del intervalo dado por (4.21).

Intervalo de verosimilitud para θ

Para encontrar el intervalo de verosimilitud de θ, retomamos la ecuación (5.15). Entonces la función de verosimilitud relativa máxima de θ, está dada por

$$r_{max}(\theta) = r(\theta, \hat{\beta}(\theta))$$

Para encontrar $\hat{\beta}(\theta)$, el EMV de β dado θ, resolvemos la ecuación $S_2(\theta, \beta) = 0$. Para esto requerimos de métodos númericos como el método de Newton. Procederemos como sigue:

$$S_2(\theta, \beta) = -k \ln \theta + \frac{k}{\beta} + \sum_{i=1}^{k} \ln y_i - \theta^{-\beta} c'_\beta + c\theta^{-\beta} \ln \theta.$$

Sea $g(\beta) = S_2(\theta, \beta)$. Luego,

$$g'(\beta) = -\frac{k}{\beta^2} - c''_\theta \theta^{-\beta} + 2c'_\theta \theta^{-\beta} \ln \theta - c\theta^{-\beta} \ln^2 \theta.$$

Queremos encontrar $\hat{\beta}(\theta)$. Luego, por el método de Newton se tiene que

$$\beta_{New} = \beta_{Old} - g(\beta)/g'(\beta).$$

Entonces tenemos que

$$r_{max}(\theta) = r(\theta, \hat{\beta}) = \ell(\theta, \hat{\beta}(\theta)) - \ell(\hat{\theta}, \hat{\beta}).$$

Para encontrar el intervalo de verosimilitud de $100p\,\%$ para θ, sustituimos el valor de $\hat{\beta}(\theta)$ dentro de la ecuación anterior y resolvemos la ecuación

$$r_{max}(\theta) - \ln p = 0.$$

Intervalos de confianza de aproximación normal
Los intervalos de $100(1-\alpha)\,\%$ de confianza para θ y β están dados por

$$\left[\underset{\sim}{\theta}, \tilde{\theta}\right] = \hat{\theta} \pm z_{(1-\alpha/2)}\widehat{es}_{\hat{\theta}}. \tag{5.17}$$

y

$$\left[\underset{\sim}{\beta}, \tilde{\beta}\right] = \hat{\beta} \pm z_{(1-\alpha/2)}\widehat{es}_{\hat{\beta}}. \tag{5.18}$$

donde $z_{(p)}$, es el p-cuantil de la distribución normal estándar, $\widehat{es}_{\hat{\theta}}$ y $\widehat{es}_{\hat{\beta}}$ son las raíces cuadradas de los elementos diagonales de la matriz de varianzas-covarianzas evaluadas en $\hat{\theta}$ y $\hat{\beta}$.

5.1.3. Modelo Lognormal

Supóngase que $y_1, y_2, ..., y_n$ es una muestra aleatoria de n observaciones, las cuales proceden de una población con distribución Lognormal con parámetros μ y σ, donde $y_1, y_2, ..., y_k$ son k observaciones no censuradas y $y_{k+1}, y_{k+2}, ..., y_n$ son $(m{=}n{-}k)$ observaciones censuradas por la derecha en un valor común C_r. De (4.3.1) la función de verosimilitud para este caso está dada por

$$L(\mu, \sigma; y) = \prod_{i=1}^{k} \frac{1}{y_i\sqrt{2\pi\sigma^2}} e^{-(\ln y_i - \mu)^2/2\sigma^2} \prod_{i=k+1}^{n} [1 - \Phi(C_r)] \tag{5.19}$$

Donde $\Phi(.)$ es la función de distribución acumulada normal estándar.
La función de log-verosimilitud está dada como sigue

$$\ell(\mu, \sigma; y) = -k\ln(\sqrt{2\pi\sigma^2}) - \frac{1}{2}\sum_{i=1}^{k}\left(\frac{\ln y_i - \mu}{\sigma}\right)^2 - \sum_{i=1}^{k}\ln y_i + \sum_{i=k+1}^{n}\ln(1 - \Phi(C_r)).$$

Luego, la función score está dada por

$$S(\mu, \sigma) = \begin{bmatrix} S_1(\mu, \sigma) \\ S_2(\mu, \sigma) \end{bmatrix}, \tag{5.20}$$

donde,

$$\begin{aligned} S_1(\mu, \sigma) &= \frac{1}{\sigma}\sum_{i=1}^{k}\left(\frac{\ln y_i - \mu}{\sigma}\right) - \sum_{i=k+1}^{n}\frac{\Phi_\mu(C_r)}{[1 - \Phi(C_r)]}, \\ S_2(\mu, \sigma) &= -\frac{k}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{k}\left(\frac{\ln y_i - \mu}{\sigma}\right)^2 - \sum_{i=k+1}^{n}\frac{\Phi_\sigma(C_r)}{[1 - \Phi(C_r)]}. \end{aligned}$$

En las expresiones anteriores Φ_μ y Φ_σ son las derivadas de $\Phi(.)$ respecto de μ y σ respectivamente.
Para encontrar $(\hat{\mu}, \hat{\sigma})$, resolvemos el par de ecuaciones simultáneas

$$S_1(\mu, \sigma) = 0 \tag{5.21}$$
$$S_2(\mu, \sigma) = 0 \tag{5.22}$$

Podemos observar que lo que se tiene es un sistema de ecuaciones no lineales con respecto de μ y σ, las cuales no son fáciles de resolver de manera algebraica, por lo que esta vez recurriremos a las herramientas numéricas, tales como el algoritmo de Newton-Raphson. En R-project existe la función NLM, la cual tiene implementado ya el algoritmo de Newton y es de fácil manejo (ver apéndice). Con esta función podemos obtener los valores de $(\hat{\mu}, \hat{\sigma})$.

Para este caso la función de información está dada por

$$I(\mu, \sigma) = \begin{bmatrix} I_{11}(\mu, \sigma) & I_{12}(\mu, \sigma) \\ I_{21}(\mu, \sigma) & I_{22}(\mu, \sigma) \end{bmatrix}, \tag{5.23}$$

donde,

$$\begin{aligned}
I_{11}(\mu, \sigma) &= \frac{k}{\sigma^2} + \sum_{i=k+1}^{n} \frac{\Phi_{\mu\mu}(C_r)[1 - \Phi(C_r)] + [\Phi_\mu(C_r)]^2}{[1 - \Phi(C_r)]^2}, \\
I_{12}(\mu, \sigma) &= \frac{2}{\sigma^2} \sum_{i=1}^{k} \left(\frac{\ln y_i - \mu}{\sigma} \right) + \sum_{i=k+1}^{n} \frac{\Phi_{\mu\sigma}(C_r)[1 - \Phi(C_r)] + \Phi_\mu(C_r)\Phi_\sigma(C_r)}{[1 - \Phi(C_r)]^2}, \\
I_{21}(\mu, \sigma) &= \frac{2}{\sigma^2} \sum_{i=1}^{k} \left(\frac{\ln y_i - \mu}{\sigma} \right) + \sum_{i=k+1}^{n} \frac{\Phi_{\mu\sigma}(C_r)[1 - \Phi(C_r)] + \Phi_\mu(C_r)\Phi_\sigma(C_r)}{[1 - \Phi(C_r)]^2}, \\
I_{22}(\mu, \sigma) &= -\frac{k}{\sigma^2} + \frac{3}{\sigma^2} \sum_{i=1}^{k} \left(\frac{\ln y_i - \mu}{\sigma} \right)^2 + \sum_{i=k+1}^{n} \frac{\Phi_{\sigma\sigma}(C_r)[1 - \Phi(C_r)] + [\Phi_\sigma(C_r)]^2}{[1 - \Phi(C_r)]^2}.
\end{aligned}$$

Se define la siguiente notación $\Phi_{\mu\mu} = \partial^2\Phi/\partial\mu^2$, $\Phi_{\mu\sigma} = \partial^2\Phi/\partial\mu\sigma$ y $\Phi_{\sigma\sigma} = \partial^2\Phi/\partial\sigma^2$.

Intervalos de verosimilitud-confianza
Para calcular los intervalos de verosimilitud, haremos uso de la función de logverosimilitud relativa de μ y σ, dada por:

$$r(\mu, \sigma) = \ell(\mu, \sigma) - \ell(\hat{\mu}, \hat{\sigma})$$

Itervalo de verosimilitud para μ
El intervalo de verosimilitud de $100p\,\%$ para μ se obtiene al resolver la ecuación

$$r_{max}(\mu) - \log p = r(\mu, \sigma(\mu)) - \log p = 0.$$

Los extremos del intervalo son las raíces de esta ecuación.

Itervalo de verosimilitud para σ
El intervalo de verosimilitud de $100p\,\%$ para σ se obtiene al resolver la

$$r_{max}(\sigma) - \log p = r(\mu(\sigma), \sigma) - \log p = 0.$$

Los extremos del intervalo son las raíces de esta ecuación.
Intervalos de confianza de aproximación normal
Los intervalos de $100(1-\alpha)\,\%$ de confianza para μ y σ están dados por

$$\left[\underset{\sim}{\mu}, \widetilde{\mu}\right] = \hat{\mu} \pm z_{(1-\alpha/2)}\widehat{es}_{\hat{\mu}}, \tag{5.24}$$

y

$$\left[\underset{\sim}{\sigma}, \widetilde{\sigma}\right] = \hat{\sigma} \pm z_{(1-\alpha/2)}\widehat{es}_{\hat{\sigma}}, \tag{5.25}$$

donde $z_{(p)}$, es el p-cuantil de la distribución normal estándar, $\widehat{es}_{\hat{\mu}}$ y $\widehat{es}_{\hat{\sigma}}$ son las raíces cuadradas de las estimaciones de los elementos diagonales de la matriz de varianzas-covarianzas.

5.1.4. Modelo Pareto

Supóngase que $y_1, y_2, ..., y_n$ es una muestra aleatoria de n observaciones, las cuales proceden de una población con distribución Pareto de parámetros α, k, donde $y_1, y_2, ..., y_k$ son k observaciones no censuradas y $y_{k+1}, y_{k+2}, ..., y_n$ son (m=n-k) observaciones censuradas por la derecha en algún valor C_r. De (4.3.1) la función de verosimilitud para este caso se construye como sigue:

$$L(\alpha, \theta; y) = \prod_{i=1}^{k} \frac{\alpha\theta^{\alpha}}{y_i^{\alpha+1}} \prod_{i=k+1}^{n} \left(\frac{\theta}{C_r}\right)^{\alpha} \tag{5.26}$$

y la función de log-verosimilitud está dada como sigue

$$\ell(\alpha, \theta; y) = k\ln\alpha + n\alpha\ln\theta - (\alpha+1)\sum_{i=1}^{k}\ln y_i - \alpha\sum_{i=k+1}^{n}\ln C_r \tag{5.27}$$

La función score para el caso de tener dos parámetros de acuerdo a (4.5) está dada por

$$S(\alpha, \theta) = \begin{bmatrix} S_1(\alpha, \theta) \\ S_2(\alpha, \theta) \end{bmatrix}, \tag{5.28}$$

donde,

$$S_1(\alpha,\theta) = \frac{k}{\alpha} + n\ln\theta - \sum_{i=1}^{k}\ln y_i - \sum_{i=k+1}^{n}\ln C_r,$$
$$S_2(\alpha,\theta) = \frac{n\alpha}{\theta}.$$

Para encontrar $(\hat{\alpha},\hat{k})$, resolvemos el par de ecuaciones simultáneas

$$S_1(\alpha,\theta) = 0, \tag{5.29}$$
$$S_2(\alpha,\theta) = 0. \tag{5.30}$$

De las ecuaciones anteriores se tiene que

$$\hat{\theta} = y_{(1)}.$$

$$\hat{\alpha} = \frac{m}{\sum_{i=1}^{k}\ln y_i + \sum_{i=k+1}^{n}\ln C_r - n\ln\hat{\theta}}.$$

Donde $y_{(1)} = min[y_1, y_2, ..., y_n]$ es el primer estadístico de orden muestral [27].

5.2. Estimación para truncamiento por la izquierda

5.2.1. Modelo Exponencial

Supóngase que $y_1, y_2, ..., y_n$ es una muestra aleatoria de n observaciones, las cuales proceden de una población con distribución exponencial de parámetro θ, donde $y_1, y_2, ..., y_k$ son k observaciones censuradas por la izquierda en un valor común y_0 y $y_{k+1}, y_{k+2}, ..., y_n$ son (m=n-k) observaciones no censuradas. Si se trunca por la izquierda en el valor y_0, entonces de (4.3.2) la función de verosimilitud está dada como sigue:

$$L(\theta; y) = \prod_{i=k+1}^{n}\frac{e^{-y_i/\theta}}{\theta e^{-y_0/\theta}} \tag{5.31}$$

La función de log-verosimilitud es

$$\ell(\theta; y) = \frac{my_0}{\theta} - m\ln\theta - \frac{1}{\theta}\sum_{i=k+1}^{n} y_i. \tag{5.32}$$

La función score y la función de información son, respectivamente

$$S(\theta) = \frac{1}{\theta^2}\sum_{i=k+1}^{n} y_i - \frac{my_0}{\theta^2} - \frac{m}{\theta}, \tag{5.33}$$

$$I(\theta) = \frac{2}{\theta^3}\sum_{i=k+1}^{n} y_i - \frac{2my_0}{\theta^3} - \frac{m}{\theta^2}. \tag{5.34}$$

Para encontrar $\hat{\theta}$, resolvemos $S(\theta) = 0$, de este modo

$$\hat{\theta} = \sum_{i=k+1}^{n} \frac{y_i}{m} - y_0. \tag{5.35}$$

El intervalo de confianza para θ, lo obtenemos como en la ecuación (5.8).

5.2.2. Modelo Weibull

Suponga que $y_1, y_2, ..., y_n$ es una muestra aleatoria de n observaciones, las cuales proceden de una población con distribución Weibull con parámetros λ y β, donde $y_1, y_2, ..., y_k$ son k observaciones censuradas por la izquierda en un valor común y_0 y $y_{k+1}, y_{k+2}, ..., y_n$ son (m=n-k) observaciones no censuradas. Si se trunca por la izquierda en y_0, entonces de (4.3.2) la función de verosimilitud está dada como sigue:

$$L(\lambda, \beta; y) = \prod_{i=k+1}^{n} \frac{\lambda\beta y_i^{\beta-1} e^{-\lambda y_i^{\beta}}}{e^{-\lambda y_0^{\beta}}} \tag{5.36}$$

y la función de log-verosimilitud está dada como

$$\ell(\lambda, \beta; y) = m\ln\lambda + m\ln\beta + m\lambda y_0^{\beta} + (\beta - 1)\sum_{i=k+1}^{n} \ln y_i - \lambda \sum_{i=k+1}^{n} {y_i}^{\beta}. \tag{5.37}$$

Luego, la función score está dada por

$$S(\lambda, \beta) = \begin{bmatrix} S_1(\lambda, \beta) \\ S_2(\beta, \beta) \end{bmatrix}, \tag{5.38}$$

donde,

$$\begin{aligned} S_1(\lambda, \beta) &= \frac{m}{\lambda} + m{y_0}^{\beta} - \sum_{i=k+1}^{n} y_i^{\beta}, \\ S_2(\lambda, \beta) &= \frac{m}{\beta} + m\lambda{y_0}^{\beta}\ln y_0 + \sum_{i=k+1}^{n} \ln y_i - \lambda \sum_{i=k+1}^{n} y_i^{\beta} \ln y_i. \end{aligned}$$

El par de estimadores $(\hat{\lambda}, \hat{\beta})$, se obtienen como en la sección anterior, así, para encontrar $\hat{\lambda}$ resolvemos la ecuación $S_1(\lambda, \beta) = 0$ de donde se obtiene que

$$\hat{\lambda}(\beta) = \frac{m}{\sum_{i=k+1}^{n} y_i^{\beta} - my_0^{\beta}}.$$

Este es el estimador de máxima verosimilitud de λ cuando suponemos que conocemos a β. Para obtener $\hat{\beta}$, el estimador de máxima verosimilitud de β, sustituimos $\hat{\lambda}$ en la ecuación $S_2(\lambda, \beta)$ y resolvemos para β como sigue:

Sea,

$$\begin{aligned} g(\beta) &= S_2(\hat{\lambda}(\beta), \beta) \\ &= \frac{m}{\beta} + \frac{m^2 y_0^\beta \ln y_0}{\sum_{i=k+1}^n y_i^\beta - m y_0^\beta} + \sum_{i=k+1}^n \ln y_i - \frac{m \sum_{i=k+1}^n y_i^\beta \ln y_i}{\sum_{i=k+1}^n y_i^\beta - m y_0^\beta} \end{aligned}$$

La ecuación $g(\beta) = 0$ puede ser resuelta vía el método de Newton.
Como ya habíamos mencionado en la sección anterior, el parámetro λ no representa una cantidad de interés, por lo que obtenemos $\ell(\theta, \beta)$ sustituyendo $\lambda = \theta^{-\beta}$ dentro de la expresión $\ell(\lambda, \beta)$. Luego, la ecuación de logverosimilitud queda como sigue

$$\ell(\theta, \beta) = -m\beta \ln \theta + m \ln \beta + m\theta^{-\beta} y_0^\beta + (\beta - 1) \sum_{i=k+1}^n \ln y_i - \theta^{-\beta} \sum_{i=k+1}^n y_i^\beta. \qquad (5.39)$$

La función de información $I(\theta, \beta)$, es obtenida de la ecuación anterior y está dada por

$$I(\theta, \beta) = \begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix}$$

donde,

$$\begin{aligned} I_{11}(\theta, \beta) &= -\frac{m\beta}{\theta^2} - m\beta(\beta+1)\theta^{-(\beta+2)} y_0^\beta + \beta(\beta+1)\theta^{-(\beta+2)} \sum_{i=k+1}^n y_i^\beta, \\ I_{12}(\theta, \beta) &= \frac{m}{\theta} + m\beta\theta^{-(\beta+1)} y_0^\beta \ln y_0 - m\beta\theta^{-(\beta+1)} y_0^\beta \ln \theta + m\theta^{-(\beta+1)} y_0^\beta \\ &\quad -\beta\theta^{-(\beta+1)} \sum_{i=k+1}^n y_i^\beta \ln y_i + \beta\theta^{-(\beta+1)} \ln \theta \sum_{i=k+1}^n y_i^\beta - \theta^{-(\beta+1)} \sum_{i=k+1}^n y_i^\beta, \\ I_{21}(\theta, \beta) &= \frac{m}{\theta} + m\beta\theta^{-(\beta+1)} y_0^\beta \ln y_0 - m\beta\theta^{-(\beta+1)} y_0^\beta \ln \theta + m\theta^{-(\beta+1)} y_0^\beta \\ &\quad -\beta\theta^{-(\beta+1)} \sum_{i=k+1}^n y_i^\beta \ln y_i + \beta\theta^{-(\beta+1)} \ln \theta \sum_{i=k+1}^n y_i^\beta - \theta^{-(\beta+1)} \sum_{i=k+1}^n y_i^\beta, \\ I_{22}(\theta, \beta) &= \frac{m}{\beta^2} - m\theta^{-\beta} y_0^\beta \ln^2 y_0 + 2m\theta^{-\beta} \ln \theta y_0^\beta \ln y_0 - m y_0^\beta \theta^{-\beta} \ln^2 \theta \\ &\quad +\theta^{-\beta} \sum_{i=k+1}^n y_i^\beta \ln^2 y_i - 2\theta^{-\beta} \ln \theta \sum_{i=k+1}^n y_i^\beta \ln y_i + \theta^{-\beta} \ln^2 \theta \sum_{i=k+1}^n y_i^\beta. \end{aligned}$$

Los intervalos de confianza para θ y β se obtienen como en (5.17) y (5.18).

5.2.3. Modelo Lognormal

Suponga que $y_1, y_2, ..., y_n$ es una muestra aleatoria de n observaciones, las cuales proceden de una población con distribución Lognormal con parámetros μ y σ, donde $y_1, y_2, ..., y_k$ son k observaciones censurados por la izquierda en un valor común y_0 y $y_{k+1}, y_{k+2}, ..., y_n$ son (m=n-k) observaciones no censuradas. Si se trunca por la izquierda en y_0, entonces la función de verosimilitud está dada como sigue:

$$\begin{aligned} L(\mu,\sigma;y) &= \prod_{i=k+1}^{n} \frac{f(y_i;\mu,\sigma)}{[1-F(y_0)]} \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma[1-\Phi(y_0)]}\right)^m \prod_{i=k+1}^{n} \frac{1}{y_i} \exp\left[-\frac{1}{2}\left(\frac{\ln y_i-\mu}{\sigma}\right)^2\right] \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma[1-\Phi(y_0)]}\right)^m \exp\left[-\frac{1}{2}\sum_{i=k+1}^{n}\left(\frac{\ln y_i-\mu}{\sigma}\right)^2\right] \prod_{i=k+1}^{n} \frac{1}{y_i}. \end{aligned}$$

La función de log-verosimilitud está dada como sigue

$$\ell(\mu,\sigma;y) = -m\ln(\sqrt{2\pi}\sigma[1-\Phi(y_0)]) - \sum_{i=k+1}^{n}\ln(y_i) - \frac{1}{2}\sum_{i=k+1}^{n}\left(\frac{\ln y_i-\mu}{\sigma}\right)^2. \quad (5.40)$$

Luego, la función Score está dada por

$$S(\mu,\sigma) = \begin{bmatrix} S_1(\mu,\sigma) \\ S_2(\mu,\sigma) \end{bmatrix} = \begin{bmatrix} \partial\ell(\mu,\sigma)/\partial\mu \\ \partial\ell(\mu,\sigma)/\partial\sigma \end{bmatrix}, \quad (5.41)$$

donde,

$$\begin{aligned} S_1(\mu,\sigma) &= \frac{m\Phi_\mu(y_0)}{[1-\Phi(y_0)]} + \frac{1}{\sigma}\sum_{i=k+1}^{n}\left(\frac{\ln y_i-\mu}{\sigma}\right), \\ S_2(\mu,\sigma) &= -\frac{m}{\sigma} + \frac{m\Phi_\sigma(y_0)}{[1-\Phi(y_0)]} + \frac{1}{\sigma}\sum_{i=k+1}^{n}\left(\frac{\ln y_i-\mu}{\sigma}\right)^2. \end{aligned}$$

Para este caso función de información está dada por

$$I(\mu,\sigma) = \begin{bmatrix} I_{11}(\mu,\sigma) & I_{12}(\mu,\sigma) \\ I_{21}(\mu,\sigma) & I_{22}(\mu,\sigma) \end{bmatrix} = \begin{bmatrix} -\frac{\partial^2\ell(\mu,\sigma)}{\partial\mu^2} & -\frac{\partial^2\ell(\mu,\sigma)}{\partial\mu\partial\sigma} \\ -\frac{\partial^2\ell(\mu,\sigma)}{\partial\mu\partial\sigma} & -\frac{\partial^2\ell(\mu,\sigma)}{\partial\sigma^2} \end{bmatrix}, \quad (5.42)$$

donde,

$$\begin{aligned}
I_{11}(\mu,\sigma) &= -\frac{m\Phi_{\mu\mu}(y_0)[1-\Phi(y_0)]+m[\Phi_{\mu}(y_0)]^2}{[1-\Phi(y_0)]^2}+\frac{m}{\sigma^2},\\
I_{12}(\mu,\sigma) &= -\frac{m\Phi_{\mu\sigma}(y_0)[1-\Phi(y_0)]+m\Phi_{\mu}(y_0)\Phi_{\sigma}(y_0)}{[1-\Phi(y_0)]^2}+\frac{2}{\sigma^2}\sum_{i=k+1}^{n}\left(\frac{\ln y_i-\mu}{\sigma}\right),\\
I_{21}(\mu,\sigma) &= -\frac{m\Phi_{\mu\sigma}(y_0)[1-\Phi(y_0)]+m\Phi_{\mu}(y_0)\Phi_{\sigma}(y_0)}{[1-\Phi(y_0)]^2}+\frac{2}{\sigma^2}\sum_{i=k+1}^{n}\left(\frac{\ln y_i-\mu}{\sigma}\right),\\
I_{22}(\mu,\sigma) &= -\frac{m}{\sigma^2}-\frac{m\Phi_{\sigma\sigma}(y_0)[1-\Phi(y_0)]+m[\Phi_{\sigma}(y_0)]^2}{[1-\Phi(y_0)]^2}+\frac{3}{\sigma^2}\sum_{i=k+1}^{n}\left(\frac{\ln y_i-\mu}{\sigma}\right)^2.
\end{aligned}$$

Los intervalos de confianza para μ y σ se obtienen como en (5.49) y (5.50).

5.2.4. Modelo Pareto tipo II

Suponga que $y_1, y_2, ..., y_n$ es una muestra aleatoria de n observaciones, las cuales proceden de una población con distribución Pareto con parámetros α y θ, donde $y_1, y_2, ..., y_k$ son k observaciones censurados por la izquierda en un valor común y_0 y $y_{k+1}, y_{k+2}, ..., y_n$ son (m=n-k) observaciones no censuradas. Si se trunca por la izquierda en y_0, entonces la función de verosimilitud está dada como sigue:

$$\begin{aligned}
L(\alpha,\theta;y) &= \prod_{i=k+1}^{n}\frac{f(y_i;\alpha,\theta)}{[1-F(y_0;\alpha,\theta)]}\\
&= \prod_{i=k+1}^{n}\frac{\frac{\alpha\theta^{\alpha}}{(y_i+\theta)^{\alpha+1}}}{\frac{\theta^{\alpha}}{(y_0+\theta)^{\alpha}}}.
\end{aligned}$$

La función de log-verosimilitud está dada como

$$\ell(\alpha,\theta;y) = m\ln\alpha + m\alpha\ln(y_0+\theta) - (\alpha+1)\sum_{i=k+1}^{n}\ln(y_i+\theta). \tag{5.43}$$

Luego, la función score está dada por

$$S(\alpha,\theta) = \begin{bmatrix} S_1(\alpha,\theta) \\ S_2(\alpha,\theta) \end{bmatrix}, \tag{5.44}$$

donde,

$$\begin{aligned}
S_1(\alpha,\theta) &= \frac{m}{\alpha}+m\ln(y_0+\theta)-\sum_{i=k+1}^{n}\ln(y_i+\theta),\\
S_2(\alpha,\theta) &= \frac{m\alpha}{y_0+\theta}-(\alpha+1)\sum_{i=k+1}^{n}\frac{1}{y_i+\theta}.
\end{aligned}$$

Para encontrar $(\hat{\alpha}, \hat{\beta})$, resolvemos el par de ecuaciones simultáneas

$$S_1(\alpha, \beta) = 0, \tag{5.45}$$
$$S_2(\alpha, \beta) = 0. \tag{5.46}$$

Resolviendo algebraicamente $S_1(\alpha, \beta) = 0$ se tiene que

$$\hat{\alpha}(\theta) = \frac{m}{\sum_{i=k+1}^{n} \ln(y_i + \theta) - m \ln(y_0 + \theta)} = \frac{m}{s}$$

donde, $s = \sum_{i=k+1}^{n} \ln(y_i + \theta) - m \ln(y_0 + \theta)$. Para calcular $\hat{\theta}$, sustituimos $\hat{\alpha}(\theta)$ en $S_2(\alpha, \theta)$ con lo cual obtenemos

$$S_2(\lambda, \beta) = \frac{m^2}{s(y_0 + \theta)} - \frac{m}{s} \sum_{i=k+1}^{n} \frac{1}{(y_i + \theta)} - \sum_{i=k+1}^{n} \frac{1}{(y_i + \theta)}.$$

Sea ahora $g(\theta) = S_2(\hat{\alpha}(\theta), \theta)$. La ecuación $g(\theta) = 0$ puede ser resuelta utilizando el método de Newton.

Antes de calcular los intervalos de confianza, necesitamos obtener la matriz de información, la cual se define de la siguiente manera

$$I(\alpha, \theta) = \begin{bmatrix} I_{11}(\alpha, \theta) & I_{12}(\alpha, \theta) \\ I_{21}(\alpha, \theta) & I_{22}(\alpha, \theta) \end{bmatrix}, \tag{5.47}$$

donde,

$$\begin{aligned}
I_{11}(\alpha, \theta) &= \frac{m}{\alpha^2}, \\
I_{12}(\alpha, \theta) &= -\frac{m}{y_0 + \theta} + \sum_{i=k+1}^{n} \frac{1}{y_i + \theta}, \\
I_{21}(\alpha, \theta) &= -\frac{m}{y_0 + \theta} + \sum_{i=k+1}^{n} \frac{1}{y_i + \theta}, \\
I_{22}(\alpha, \theta) &= \frac{m\alpha}{(y_0 + \theta)^2} - (\alpha + 1) \sum_{i=k+1}^{n} \frac{1}{(y_i + \theta)^2}.
\end{aligned}$$

Intervalos de verosimilitud-confianza

La función de log-verosimilitud relativa de α y θ está dada como

$$r(\alpha, \theta) = m \ln(\alpha) + m\alpha \ln(y_0 + \theta) - (\alpha + 1) \sum_{i=k+1}^{n} \ln(y_i + \theta) - \ell(\hat{\alpha}, \hat{\theta}). \tag{5.48}$$

Intervalo de verosimilitud para θ
Para encontrar el intervalo de verosimilitud de $100p\,\%$ para θ resolvemos la ecuación

$$r_{max}(\theta) - \ln(p) = r(\hat{\alpha}(\theta), \theta) - \ln p = 0,$$

o equivalentemente, sustitiyendo $\hat{\alpha}(\theta)$ dentro de la ecuación (5.48),

$$m \ln\left(\frac{m}{s}\right) + \frac{m^2}{s} \ln(y_0 + \theta) - \left(\frac{m}{s} + 1\right) \sum_{i=1}^{m} \ln(y_i + \theta) - \ell(\hat{\alpha}, \hat{\theta}) - \ln p = 0.$$

Las raíces encontradas son los extremos del intervalo de verosimilitud

Intervalo de verosimilitud para α
Para encontrar el intervalo de verosimilitud de $100p\,\%$ para α resolvemos la ecuación

$$r_{max}(\alpha) - \ln(p) = r(\alpha, \hat{\theta}(\alpha)) - \ln p = 0$$

Para encontrar $\hat{\theta}(\alpha)$, el EMV de θ dado α, debemos resolver la ecuación $S_2(\alpha, \theta) = 0$ y el resultado encontrado mediante métodos numéricos, lo sustituimos dentro de la ecuación. Luego, entonces se tiene que resolver

$$m \ln(\alpha) + m\alpha \ln(y_0 + \hat{\theta}(\alpha)) - (\alpha + 1) \sum_{i=k+1}^{n} \ln(y_i + \hat{\theta}(\alpha)) - \ell(\hat{\alpha}, \hat{\theta}) - \ln p = 0,$$

las raíces de esta ecuación son los extremos del intervalo de verosimilitud para α.

Intervalos de confianza de aproximación normal
Los intervalos de $100(1-\alpha)\,\%$ de confianza para α y θ están dados por

$$\left[\underline{\alpha}, \widetilde{\alpha}\right] = \hat{\alpha} \pm z_{(1-\alpha/2)} \widehat{es}_{\hat{\alpha}}. \tag{5.49}$$

y

$$\left[\underline{\theta}, \widetilde{\theta}\right] = \hat{\theta} \pm z_{(1-\alpha/2)} \widehat{es}_{\hat{\theta}}. \tag{5.50}$$

donde $z_{(p)}$, es el p-cuantil de la distribución normal estándar, $\widehat{es}_{\hat{\alpha}}$ y $\widehat{es}_{\hat{\theta}}$ son las raíces cuadradas de las estimaciones de los elementos diagonales de la matriz de varianzas-covarianzas.

5.3. Estimación para truncamiento por la izquierda y censura por la derecha

5.3.1. Modelo Exponencial

Supóngase que $y_1, y_2, ..., y_n$ es una muestra aleatoria que se ha truncado por la izquierda en y_0 las cuales proceden de una población con distribución exponencial de parámetro

θ, donde $y_1, y_2, ..., y_k$ son k observaciones no censuradas y $y_{k+1}, y_{k+2}, ..., y_n$ son (m=n-k) observaciones censuradas por la derecha en algún valor C_r. La función de verosimilitud para este caso está dada por:

$$L(\theta; y) = \frac{\prod_{i=1}^{k} \frac{1}{\theta} e^{-y_i/\theta} \prod_{i=k+1}^{n} e^{-C_r/\theta}}{\prod_{i=1}^{n} e^{-y_0/\theta}}. \tag{5.51}$$

Luego, la función de log-verosimilitud está dada como sigue

$$\ell(\theta, y) = -k \ln \theta - \frac{1}{\theta} \sum_{i=1}^{k} y_i - \frac{1}{\theta} \sum_{i=k+1}^{n} C_r + \frac{n y_0}{\theta}. \tag{5.52}$$

La función Score y la función de información para este caso son, respectivamente

$$S(\theta) = -\frac{k}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^{k} y_i + \frac{1}{\theta^2} \sum_{i=k+1}^{n} C_r - \frac{n y_0}{\theta^2}, \tag{5.53}$$

$$I(\theta) = -\frac{k}{\theta^2} + \frac{2}{\theta^3} \sum_{i=1}^{k} y_i + \frac{2}{\theta^3} \sum_{i=k+1}^{n} C_r - \frac{2 n y_0}{\theta^3}. \tag{5.54}$$

Para encontrar $\hat{\theta}$ resolvemos $S(\theta) = 0$, de este modo

$$\hat{\theta} = \frac{\sum_{i=1}^{k} y_i + \sum_{i=k+1}^{n} C_r - n y_0}{k} \tag{5.55}$$

es el estimador de máxima verosimilitud de θ para el modelo exponencial con muestras truncadas por la izquierda y censuradas por la derecha.

5.3.2. Modelo Weibull

Supóngase que $y_1, y_2, ..., y_n$ es una muestra aleatoria que se ha truncado por la izquierda en y_0 las cuales proceden de una población con distribución Weibull con parámetros λ y β, donde $y_1, y_2, ..., y_k$ son k observaciones no censuradas y $y_{k+1}, y_{k+2}, ..., y_n$ son (m=n-k) observaciones censuradas por la derecha en algún valor C_r. La función de verosimilitud para este caso está dada por:

$$L(\lambda, \beta; y) = \frac{\prod_{i=1}^{k} \lambda \beta y_i^{\beta-1} e^{-\lambda y_i^{\beta}} \prod_{i=k+1}^{n} e^{-\lambda C_r^{\beta}}}{\prod_{i=1}^{n} e^{-\lambda y_0^{\beta}}}. \tag{5.56}$$

Luego, la función de log-verosimilitud está dada como sigue

$$\ell(\lambda, \beta; y) = k \ln \lambda + k \ln \beta + (\beta - 1) \sum_{i=1}^{k} \ln y_i - \lambda \sum_{i=1}^{k} y_i^{\beta} - \lambda \sum_{i=k+1}^{n} C_r^{\beta} + n \lambda y_0^{\beta}. \tag{5.57}$$

La función score para este caso está dada por

$$S(\lambda,\beta)=\left[\begin{array}{c} S_1(\lambda,\beta) \\ S_2(\lambda,\beta) \end{array}\right]. \tag{5.58}$$

Donde,

$$S_1(\lambda,\beta)=\frac{k}{\lambda}-\sum_{i=i}^{k} y_i^{\beta}-\sum_{i=k+i}^{n} C_r^{\beta}+ny_0^{\beta},$$

$$S_2(\lambda,\beta)=\frac{k}{\beta}+\sum_{i=i}^{k}\ln y_i-\lambda\sum_{i=i}^{k} y_i^{\beta}\ln y_i-\lambda\sum_{i=k+i}^{n} C_r^{\beta}\ln C_r+n\lambda y_0^{\beta}\ln y_0.$$

Luego, $\hat{\lambda}$ se obtiene al resolver $S_1(\lambda,\beta)=0$ y está dado por

$$\hat{\lambda}(\beta)=\frac{k}{\sum_{i=i}^{k} y_i^{\beta}+\sum_{i=k+1}^{n} C_r^{\beta}-ny_0^{\beta}}. \tag{5.59}$$

Para obtener $\hat{\beta}$, procedemos como en el caso para datos censurados o truncados definiendo $g(\beta)=S_2(\hat{\lambda}(\beta),\beta)$ y resolviendo $g(\beta)=0$, para ello haremos uso de el método numérico de Newton-Raphson.

5.3.3. Modelo Pareto II

Supóngase que $y_1, y_2, ..., y_n$ es una muestra aleatoria que se ha truncado por la izquierda en y_0 las cuales proceden de una población con distribución Pareto tipo II con parámetros α y θ, donde $y_1, y_2, ..., y_k$ son k observaciones no censuradas y $y_{k+1}, y_{k+2}, ..., y_n$ son (m=n-k) observaciones censuradas por la derecha en algún valor C_r. La función de verosimilitud para este caso está dada por

$$L(\alpha,\theta;y)=\frac{\prod_{i=i}^{k}\frac{\alpha\theta^{\alpha}}{(y_i+\theta)^{\alpha+1}}\prod_{i=k+i}^{n}\frac{\theta^{\alpha}}{(C_r+\theta)^{\alpha}}}{\prod_{i=i}^{n}\frac{\theta^{\alpha}}{(y_0+\theta)^{\alpha}}}. \tag{5.60}$$

Luego, la función de log-verosimilitud está dada como sigue:

$$\ell(\alpha,\theta;y)=k\ln\alpha-(\alpha+1)\sum_{i=1}^{k}\ln(y_i+\theta)-\alpha\sum_{i=k+1}^{n}\ln(C_r+\theta)+n\alpha\ln(y_0+\theta). \tag{5.61}$$

La función score para este caso está dada por

$$S(\alpha,\theta)=\left[\begin{array}{c} S_1(\alpha,\theta) \\ S_2(\alpha,\theta) \end{array}\right]. \tag{5.62}$$

Donde,

$$S_1(\alpha,\theta) = \frac{k}{\alpha} - \sum_{i=1}^{k} \ln(y_i+\theta) - \sum_{i=k+1}^{n} \ln(C_r+\theta) + n\ln(y_0+\theta),$$

$$S_2(\alpha,\theta) = \frac{n\alpha}{(y_0+\theta)} - (\alpha+1)\sum_{i=1}^{k} \frac{1}{(y_i+\theta)} - \sum_{i=k+1}^{n} \frac{1}{(C_r+\theta)}.$$

Para obtener el EMV de α, resolvemos la ecuación $S_1(\alpha,\theta)=0$, teniendo que

$$\hat{\alpha} = \frac{k}{\sum_{i=1}^{k}\ln(y_i+\theta) + \sum_{i=k+1}^{n}\ln(C_r+\theta) - n\ln(y_0+\theta)} = \frac{k}{s}, \tag{5.63}$$

donde $s = \sum_{i=1}^{k}\ln(y_i+\theta) + \sum_{i=k+1}^{n}\ln(C_r+\theta) - n\ln(y_0+\theta)$. Por otro lado, para obtener el EMV de θ definimos $g(\theta) = S_2(\hat{\alpha}(\theta),\theta)$ y resolvemos $g(\theta)=0$, vía método de Newton-Raphson.

5.4. Algoritmo de implementación

Al resolver los sistemas de ecuaciones para obtener los estimadores de máxima verosimilitud, en la mayoría de los casos, estos son no lineales y no son fáciles de resolver de manera algebraica, por lo que se recurre al método numérico de Newton-Raphson para su solución. El método de Newton-Raphson es eficiente en la solución de sistemas de ecuaciones no lineales, converge muy rápidamente y proporciona una muy buena precisión en los resultados [7].

En R-project existe la función NLM, la cual tiene implementado el algoritmo de Newton y su manejo no es muy difícil, los programas realizados en la plataforma R para la obtención de los estimadores de los parámetros, siguen el siguiente algoritmo:

1. Se introducen los datos observados y censurados.
2. Se define la función de verosimilitud para la distribución deseada.
3. Dar valores iniciales para la utilización del método de Newton-Raphson.
4. Se llama a la función NLM.
5. Salida de resultados. Estimadores y matriz de información.

En el apendice (B.5) se muestra la descripción y el uso de la función NLM.

Capítulo 6

Aplicaciones

La mayoría de las personas hace planes y tiene expectativas acerca del curso que seguirá su vida. Sin embargo, la experiencia nos enseña que los planes no se desarrollan con certeza y algunas veces las expectativas no se realizan ya que los planes se frustran debido a que estaban construidos sobre supuestos irreales, en otras situaciones interfieren circunstancias fortuitas. No es agradable pensar en los riesgos que se corren a diario como una enfermedad, un accidente vehicular, o el robo residencial. Sin embargo, se debe contemplar la posibilidad de que estos riesgos podrían ocurrir, para estos casos un seguro puede significar una valiosa alternativa en caso de emergencia, o bien, para estar preparados ante cualquier tipo de accidentes o enfermedad para resguardar nuestra salud, bienestar y para protegernos en contra de reveses financieros severos.

En el negocio de la oferta de seguros existen tres clases de seguros: los seguros generales, los seguros de vida y los seguros de salud.

6.1. Datos censurados por la derecha.

Hoy en día el aumento en el costo de la atención médica, dificulta que la mayoría de la población tenga acceso a servicios de salud de calidad, sobre todo ante eventos inesperados como lo son los accidentes, por lo que contar con un programa que cubra estas eventualidades resulta necesario e incluso imprescindible, para ayudarnos a solventar los fuertes gastos hospitalarios y médicos que podrían requerirse para recuperar nuestra salud o la de algún ser querido. Un seguro de gastos médicos puede significar una valiosa alternativa en estos casos.

En la tabla siguiente se ilustra una muestra aleatoria de 20 importes totales pagados a trabajadores por reembolso de gastos médicos; aunque se desconoce el esquema de aseguramiento en particular, utilizaremos esta información para ilustrar las aplicaciones que presenta este trabajo.

27	82	115	126	155	161	243	294	340	384
457	680	855	877	974	1 193	1 340	1 884	2 558	15 743

Tabla 6.1: Importes pagados a trabajadores por reembolso ($).

Supóngase que estos datos fueron censurados por la derecha en 250, en un límite de cobertura, así, los datos sobre los cuales se harán inferencia son: 27, 82, 115, 126, 155, 161, 243, +250, +250, +250, +250, +250, +250, +250, +250, +250, +250, +250, +250. Se desea conocer la pérdida media estimada por estos gastos médicos.

Lo primero será identificar un modelo teórico al cual se puedan ajustar bien dichas observaciones, para posteriormente hacer inferencia sobre la pérdida media estimada por estos gastos médicos.

Se proponen los modelos teóricos Exponencial, Weibull y Lognormal.

Mediante el uso del software R-Project, se calculan los estimadores de máxima verosimilitud para los parámetros de dichos modelos. Haciendo uso de la teoría para datos censurados por la derecha, se obtienen los resultados que se muestran en la siguiente tabla.

Modelo	EMV	Media	sd
Exponencial	$\hat{\theta} = 594.143$	594.143	594.143
Weibull	$\hat{\theta} = 459.241$ $\hat{\beta} = 1.363$	420.392	311.86
Lognormal	$\hat{\mu} = 5.990$ $\hat{\sigma} = 1.218$	838.668	1 548.404

Tabla 6.2: EMVs de los parámetros, media y desviación estandar (sd).

Para la selección del modelo se utilizarán dos criterios: criterio de información de Akaike (AIC) y la prueba de razón de verosimilitudes, ambos proveen resultados eficientes.

Para hacer uso de estas pruebas, necesitamos de los siguientes resultados

Modelo	Num. de parámetros	Log-verosimilitud	AIC
Exponencial	1	-51.710	105.420
Weibull	2	-51.373	106.746
Lognormal	2	-59.280	122.560

Tabla 6.3: Valores de la log-verosimilitud y AIC.

Criterio de Akaike.

De la tabla anterior, por el criterio de Akaike las observaciones se ajustan bastante bien por un modelo Exponencial con parámetro estimado $\hat{\theta} = 594.143$, también puede observarse que el modelo Weibull con parámetro de forma $\hat{\beta} = 1.363$ y parámetro de escala $\hat{\theta} = 459.241$ sigue en orden de importancia.

Prueba de razón de verosimilitudes.

Con los resultados de la tabla anterior, y bajo un esquema de pruebas de hipótesis típico, se procede a la selección del mejor modelo mediante comparaciones de par en par. Consideraremos un nivel de confianza de 0.95, sirviendo este de referencia para el comparativo con el cuantil $\chi^2_{.05,1} = 3.84$ considerado. De este modo si la hipótesis se rechaza, quiere decir que el mejor modelo es aquel con la log-verosimilitud más alta de lo contrario ambos modelos ajustarán bien a los datos.

Comparación entre los modelos Exponencial y Weibull.

1. Hipótesis: H:$Modelo_{Exp} \equiv Modelo_{Weib}$

2. Estadístico de prueba:

$$\chi^2 = -2[-51.710 - (-51.373)] = 0.674$$

Se observa que $\chi^2 < \chi^2_{.05,1} = 3.84$, por lo tanto no se rechaza la hipótesis, por lo que ambos modelos se ajustan bien a los datos. Tomamos el modelo Exponencial con parámetro estimado $\hat{\theta} = 594.143$ por su sencillez.

Comparación entre los modelos Exponencial y Lognormal.

1. Hipótesis: H:$Modelo_{Exp} \equiv Modelo_{Logn}$

2. Estadístico de prueba:

$$\chi^2 = -2[-51.710 - (-59.280)] = -15.14$$

Se observa que $\chi^2 < \chi^2_{.05,1} = 3.84$, por lo tanto no se rechaza la hipótesis, por lo que ambos modelos se ajustan bien a los datos. Tomamos el modelo Exponencial con parámetro estimado $\hat{\theta} = 594.143$.

De acuerdo al criterio de Akaike y a la prueba de razón de verosimilitudes, el modelo que mejor se ajusta a los datos es el modelo Exponencial con parámetro $\hat{\theta} = 594.1429$.

6.2. Datos censurados por la derecha, límite inferior constante.

El seguro de autos garantiza la posibilidad económica para hacerle frente a accidentes viales proporcionando tranquilidad para el asegurado y es un medio para proteger el patrimonio de las personas. En el siguiente ejemplo se nos muestran los datos de siniestros en autos compactos en el sureste de México durante 2013. Aunque se desconoce el esquema de aseguramiento en particular, utilizaremos esta información para ilustrar la aplicación cuando se tienen datos censurados por la derecha con un límite inferior constante.

Considerar los valores de la tabla siguiente, los cuales han sido extraidos de la CNSF (Comisión Nacional de Seguros y Fianzas) [19], donde únicamente se tienen reclamaciones por daños parciales y por pérdida total, con una referencia respecto a los límites de cobertura, como el deducible y monto mínimo para pérdida total igual al 5 % y 50 % del valor factura, registrado. Se desea saber cual es el costo promedio del siniestro y la mediana de los costos. Se estudiará este caso bajo el modelo Pareto.

Obs	Tipo de Pérdida	Monto Siniestro	Monto deducible	Valor factura
x_1	Pérdida parcial	\$5 205.23	\$9 373	\$187 460
x_2	Pérdida parcial	\$5 970.88	\$24 595	\$491 900
⋮	⋮	⋮	⋮	⋮
x_{35}	Pérdida parcial	\$192 493.62	\$22 265	\$445 300
y_1	Pérdida total	\$246 564.83	\$14 840	\$296 800
y_2	Pérdida total	\$273 011.9	\$15 095	\$301 900
⋮	⋮	⋮	⋮	⋮
y_{12}	Pérdida total	\$270 853.88	\$27 090	\$541 800

Tabla 6.4: Siniestros en autos compactos en el sureste de México durante 2013

Mediante el uso del sofware R, se calculan los estimados de máxima verosimilitud para los parámetros de dicho modelo.

Modelo	EMV	Media	Mediana
Pareto	$\hat{\alpha} = 1.36$ $\hat{\theta} = 5205.23$	19 664.2	8 665.338

De acuerdo a la teoría de la sección (4.2.2) costruimos intervalos de confianza para la media y la mediana.

El intervalo de confianza de 95 % para la media $\hat{E}$ es

$$(3\ 919.988, 35\ 408.42),$$

y el intervalo de confianza de 95 % para la mediana $\hat{m}$ es

$$(2\ 300.473, 15\ 030.2).$$

6.3. Datos truncados por la izquierda.

Suponga que los valores de la tabla (6.1) han sido truncados por la izquierda en 200, un deducible aplicado. Entonces, los datos a considerar serán las observaciones mayores a 200. De este modo la base de datos se reduce a lo siguiente: 243, 294, 340, 384, 457, 680, 855, 877, 974, 1 193, 1 340, 1 884, 2 558, 15 743.

Queremos encontrar la estimación del costo promedio de estos reembolsos por gastos médicos.

Usamos la teoría para el caso de tener datos truncados por la izquierda, ajustamos los datos a un modelo conocido, para ello proponemos los modelos Exponencial, Weibull, Lognormal y Pareto (II).

La siguiente tabla muestra la estimación de los parámetros de dichos modelos.

Modelo	EMV	Media	sd
Exponencial	$\hat{\theta} = 1\ 787.286$	1 987.286	1 787.286
Weibull	$\hat{\theta} = 120.030$ $\hat{\beta} = 0.350$	1 633.396	4 150.005
Lognormal	$\hat{\mu} = 6.054$ $\hat{\sigma} = 1.494$	1 833.245	4 405.157
Pareto (II)	$\hat{\theta} = 707.98$ $\hat{\alpha} = 1.452$	2 208.361	—

Tabla 6.5: EMVs de los parámetros, media y desviación estandar (sd).

Lo primero que haremos será seleccionar el modelo que mejor se ajuste a los datos, nuevamente haremos uso de el Criterio de Información de Akaike (AIC) y la prueba de razón de verosimilitudes.

Para hacer uso de estas pruebas, necesitamos de los siguientes resultados

Modelo	Num. de parámetros	Logverosimilitud	AIC
Exponencial	1	-118.838	239.677
Weibull	2	-114.102	232.204
Lognormal	2	-113.922	231.844
Pareto (II)	2	-113.776	231.553

Tabla 6.6: Valores de la log-verosimilitud y AIC.

Criterio de Akaike.

De acuerdo a la tabla anterior, por el criterio de Akaike, las observaciones se ajustan bastante bien por un modelo Pareto tipo (II), con parámetros $\hat{\theta} = 707.98$ y $\hat{\alpha} = 1.4521$. El modelo Lognormal con parámetros estimados $\hat{\mu} = 6.054140$ y $\hat{\sigma} = 1.494443$, sería el siguiente modelo en orden de importancia para ajustar nuestro conjunto de datos.

Prueba de razón de verosimilitudes

Se realizan pruebas de hipótesis para seleccionar el mejor modelo mediante comparaciones de par en par. Consideraremos un nivel de confianza de .95, sirviendo esto de referencia para el comparativo con el cuantil $\chi^2_{.05,1} = 3.84$.

Comparación entre los modelos Exponencial y Weibull.

1. Hipótesis: H:$Modelo_{Exp} \equiv Modelo_{Weib}$
2. Estadístico de prueba:

$$\chi^2 = -2[-118.838 - (-114.102)] = 9.472$$

Se observa que $\chi^2 > \chi^2_{.05,1} = 3.84$, por lo tanto se rechaza la hipótesis y el modelo más adecuado resulta ser el modelo Weibulll con parámetros estimados $\hat{\theta} = 120.030$ y $\hat{\beta} = 0.350$.

Comparación entre los modelos Weibull y Lognormal.

1. Hipótesis: H:$Modelo_{Weib} \equiv Modelo_{Logn}$
2. Estadístico de prueba:

$$\chi^2 = -2[-114.102 - (-113.922)] = 0.36$$

Se observa que $\chi^2 < \chi^2_{.05,1} = 3.84$, por lo tanto no se rechaza la hipótesis y ambos modelos se ajustan bien a los datos, elegimos el modelo Weibulll con parámetros estimados $\hat{\theta} = 120.030$ y $\hat{\beta} = 0.350$.

Comparación entre los modelos Weibull y Pareto (II).

1. Hipótesis: H:$Modelo_{Weib} \equiv Modelo_{Paret}$

2. Estadístico de prueba:

$$\chi^2 = -2[-114.102 - (-113.776)] = 0.652$$

Se observa que $\chi^2 < \chi^2_{.05,1} = 3.84$, por lo tanto no se rechaza la hipótesis. Como ambos modelos se ajustan bien a los datos, elegimos el modelo Pareto (II) (también podríamos haber elegido el modelo Weibull).

El criterio de Akaike toma como mejor modelo de ajuste al Pareto (II), la prueba de razón de verosimilitudes nos da como elección los modelos Weibull, Lognormal y Pareto (II), en este caso elegimos el modelo Pareto (II) con parámetros $\hat{\theta} = 707.98$ y $\hat{\alpha} = 1.452$.

6.4. Datos truncados por la izquierda y censurados por la derecha.

Suponga que los valores de la tabla (6.1) han sido truncados por la izquierda en 150 y censurados por la derecha en 1300. Entonces, los datos a considerar serán de este modo los siguiente: 155, 161, 243, 294, 340, 384, 457, 680, 855, 877, 974, 1193, 1300, 1300, 1300, 1300. Se desea conocer la mediana de los costos por reembolso de gastos médicos.

Al igual que en el caso de datos censurados por la derecha y datos truncados por la izquierda, vamos a identificar algún modelo teórico que se ajuste bien a dichas observaciones. Para este caso proponemos cuatro modelos teóricos, los cuales son: el Exponencial, el Weibull, el Lognormal y el Pareto (II).

La siguiente tabla muestra la estimación de los parámetros de dichos modelos.

Modelo	EMV
Exponencial	$\hat{\theta} = 784.417$
Weibull	$\hat{\theta} = 443.831$ $\hat{\beta} = .585$
Lognormal	$\hat{\mu} = 5.874$ $\hat{\sigma} = 1.493$
Pareto (II)	$\hat{\theta} = 747.739$ $\hat{\alpha} = 1.635$

Tabla 6.7: Estimados de máxima verosimilitud de los parámetros.

Lo primero que haremos será seleccionar el modelo que mejor se ajuste a los datos, nuevamente haremos uso de el Criterio de Información de Akaike (AIC) y la prueba de razón de verosimilitudes.

Para hacer uso de estas pruebas, necesitamos de los siguientes resultados

Modelo	Num. de parámetros	Logverosimilitud	AIC
Exponencial	1	-91.979	185.959
Weibull	2	-91.610	187.220
Lognormal	2	-91.713	187.425
Pareto (II)	2	-91.742	187.484

Tabla 6.8: Valores de la log-verosimilitud y AIC.

Criterio de Akaike.

De acuerdo a la tabla anterior, por el criterio de Akaike, las observaciones se podrían ajustar bastante bien por un modelo exponencial, con parámetro $\hat{\theta} = 784.417$. El modelo Weibull con parámetros estimados $\hat{\theta} = 443.831$ y $\hat{\beta} = .585$, sería el siguiente modelo en orden de importancia para ajustarse a nuestro conjunto de datos.

Prueba de razón de verosimilitudes.

Se realizan pruebas de hipótesis para seleccionar el mejor modelo mediante comparaciones de par en par. Consideraremos un nivel de confianza de .95, sirviendo esto de referencia para el comparativo con el cuantil $\chi^2_{.05,1} = 3.84$.

Comparación entre los modelos Exponencial y Weibull.

1. Hipótesis: H:$Modelo_{Exp} \equiv Modelo_{Weib}$
2. Estadístico de prueba:

$$\chi^2 = -2[-91.979 - (-91.610)] = 0.738$$

Se observa que $\chi^2 < \chi^2_{.05,1} = 3.84$, por lo tanto, no se rechaza la hipótesis y ambos modelos se ajustan a los datos, elegimos el modelo Weibull con parámetros estimados $\hat{\theta} = 443.831$ y $\hat{\beta} = .585$.

Comparación entre los modelos Weibull y Lognormal.

1. Hipótesis: H:$Modelo_{Weib} \equiv Modelo_{Logn}$
2. Estadístico de prueba:

$$\chi^2 = -2[-91.610 - (-91.713)] = -.206$$

Se observa que $\chi^2 < \chi^2_{.05,1} = 3.84$, por lo tanto, ambos modelos se ajustan bien a los datos. Elegimos al modelo Weibull con parámetros estimados $\hat{\theta} = 443.831$ y $\hat{\beta} = .585$.

Comparación entre los modelos Weibull y Pareto (II).

1. Hipótesis: H:$Modelo_{Weib} \equiv Modelo_{Paret}$
2. Estadístico de prueba:

$$\chi^2 = -2[-91.610 - (-91.742)] = -.264$$

Se observa que $\chi^2 < \chi^2_{.05,1} = 3.84$, por lo tanto, ambos modelos se ajustan bien a los datos. Se elige el modelo Weibull cuyos parámetros estimados son $\hat{\theta} = 443.831$ y $\hat{\beta} = .585$.

De acuerdo al criterio de Akaike, el modelo que mejor se ajusta a los datos es el modelo Exponencial con parámetro $\hat{\theta} = 784.4167$, la prueba de razón de verosimilitudes nos indica que cualquiera de los cuatro modelos se ajusta bien a los datos, elegimos el modelo Weibull ya que este generaliza al modelo exponencial y es muy versátil, cuyos parámetros son $\hat{\theta} = 443.831$ y $\hat{\beta} = .585$.

Capítulo 7

Resultados y Conclusiones

7.1. Datos censurados por la derecha.

Costo promedio por reembolso de gastos médicos.
En este ejemplo, aunque no conocemos el esquema de aseguramiento en particular, conocemos el límite de cobertura 250, el cual es un punto de censura por la derecha. Cuando los gastos médicos del asegurado sean mayores al límite de cobertura, los costos excedentes no serán cubiertos por la aseguradora. De acuerdo a la prueba de razón de verosimilitudes y al criterio de Akaike, el modelo que mejor se ajusta a los datos bajo el esquema de censura por la derecha, para este caso particular es el modelo Exponecial con parámetro estimado $\hat{\theta} = 594.14$. Se tiene entonces, que la pérdida media estimada por reembolso de estos gastos médicos es de $\hat{E} = 594.14$ y un intervalo de 95 % de confianza para la media es $(153.996, 1034.29)$.

7.2. Datos censurados por la derecha, límite inferior constante.

Costo promedio del siniestro.
En este ejemplo, la estimación encontrada permite observar de manera más precisa el costo esperado, y su variación significativa al 95 %, para los sinietros individuales que pudieran ocurrirle a cada uno de los asegurados bajo este esquema específico. La consideración de los eventos por pérdidas totales (datos censurados por la derecha), a través de una probabilidad en el análisis, permite complementar y fortalecer los pronósticos futuros realizados al final del periodo anual, donde típicamente un costo promedio se calcula dividiendo el total de costos individuales entre las unidades aseguradas existentes.

Mediana de los costos.
Esta magnitud tiene como finalidad observar una estimación respecto al costo que representa, de manera acumulada, el 50 % de los siniestros ocurridos a las unidades aseguradas. El intervalo obtenido refleja una muy buena estimación para que la aseguradora considere los recursos y reservas proporcionales que permitan garantizar fondos suficientes para atender la demanda y reclamaciones de la mitad (mediana) de los asegurados con eventos de siniestros, en función de sus pólizas y políticas de aseguramiento.

7.3. Datos truncados por la izquierda.

Costo promedio por reembolso de gastos médicos.
Al considerar el enfoque de truncamiento por la izquierda, ya sea por las condiciones propias del fenómeno o como estrategia de análisis, se puede proceder de dos maneras:

- Desplazar los datos a la izquierda, restando el punto de truncamiento de cada observación, y haciendo inferencia convencional para estos nuevos datos.
- Ajustar un modelo para los datos originales, aceptando que no existe información por debajo del punto truncamiento.

En este ejemplo particular, no se desplazan los datos, evitando así, posibles problemas de especificación o definición del modelo, de acuerdo a la prueba de razón de verosimilitudes y al criterio de Akaike, el modelo que mejor se ajusta a los datos es el modelo Pareto (II) con parámetros $\hat{\theta} = 707.98$ y $\hat{\alpha} = 1.4521$. Por lo tanto el pago promedio por reembolso de gastos médicos, con este deducible de 200 es $\hat{m} = 2208.361$.

7.4. Datos truncados por la izquierda y censurados por la derecha.

El truncamiento por la izquierda ocurre cuando la aseguradora aplica un deducible d. Cuando un asegurado tenga una pérdida menor que d, éste decidirá pagar los costos con sus propios recursos sin informar a la aseguradora. Si la pérdida es mayor que d, lo reportará y exigirá los beneficios contratados, en este ejemplo 150 sería el monto del deducible y 1300 sería el límite de cobertura, el cual es un punto de censura por la derecha.

De acuerdo al criterio de Akaike, el modelo que mejor se ajusta a los datos es el modelo Exponencial con parámetro $\hat{\theta} = 784.4167$, la prueba de razón de verosimilitudes nos indica que cualquiera de los cuatro modelos se ajusta bien a los datos, elegimos el modelo Weibull ya que este generaliza al modelo exponencial y es muy versátil, cuyos parámetros son $\hat{\theta} = 443.8309$ y $\hat{\beta} = .5850329$.

Mediana de los Costo por reembolso de gastos médicos.
La mediana de los costos, para este ejemplo partícular es $\hat{m} = 626.35$ un intervalo de 95 % de confianza para la mediana es $(247.7388, 1004.9675)$, de acuerdo a la información del intervalo obtenido, la aseguradora deberá considerar los recursos y reservas proporcionales que permitan garantizar fondos suficientes para atender la demanda y reclamaciones de la mitad (mediana) de los asegurados, en función de sus pólizas y políticas de aseguramiento.

7.5. Conclusiones generales

Sin lugar a dudas, en un futuro no muy lejano, aparecerá algún fenómeno, el cual su estudio será de interés para algunos investigadores, donde podrán presentarse muestras, cuyas observaciones presenten censura, truncamiento, o ambas, dicho fenómeno podrá ser estudiado desde un punto de vista paramétrico. En el presente trabajo se exhorta a hacer uso de la inferencia paramétrica mediante el uso de la metodología de máxima verosimilitud.

Toda la teoría estudiada es presentada en una aplicación, tanto en censura por la derecha, truncamiento por la izquierda y también cuando se presenta el caso de tener un truncamiento por la izquierda y censura por la derecha, los cálculos para obtener los estimadores de máxima verosimilitud de los distintos modelos propuestos, así como para obtener los intervalos de confianza, se obtienen utilizando el software estadístico R, la cual es una herramienta que nos permite estimar los diferentes parámetros y obtener la información necesaria.

Apéndice A

Programas para calcular los EMV

A.1. Programas para muestras censuradas por la derecha

Exponencial con censura por la derecha

```
# programa para estimar el parámetro θ de la distribución exponencial para
datos censurados por la derecha.

# datos observados
xobs=c(27,82,115,126,155,161,243)
# datos censurados por la derecha
xcr=c(250,250,250,250,250,250,250,250,250,250,250,250,250)
data=c(xobs,xcr)
k=length(xobs)

verosi=function(teta)
     {
     ele=sum(dexp(xobs,teta[1],log=T))+sum(pexp(xcr,teta[1],lower.tail=F,log.p=T))
     return(-ele)
     }

p=c(500)
(out=nlm(verosi,p,hessian=T)$estimate)

#se obtiene el valor de θ como:
t=1/out
```

Weibull con censura por la derecha

```
# programa para estimar los parámetros β y θ de la distribución Weibull para
datos censurados por la derecha.
```

```
# datos observados
xobs=c(27,82,115,126,155,161,243)
# datos censurados por la derecha
xcr=c(250,250,250,250,250,250,250,250,250,250,250,250,250)

verosimil=function(p)
     {
     ele=sum(dweibull(xobs,p[1],p[2],log=T))
     +sum(pweibull(xcr,p[1],p[2],lower.tail=F,log.p=T))
     return(-ele)
     }

p=c(1,10)
(out=nlm(verosimil,p,hessian=T)$estimate)

# los valores de β y θ se tienen como:
b=out[1]
t=out[2]
```

Lognormal con censura por la derecha

```
# programa para estimar los parámetro μ y σ de la distribución Lognormal
para datos censurados por la derecha.

# datos observados
xobs=c(27,82,115,126,155,161,243)
# datos censurados por la derecha
xcr=c(250,250,250,250,250,250,250,250,250,250,250,250,250)

verosimil=function(p)
     {
     ele=sum(dlnorm(xobs,p[1],p[2],log=T))
     +sum(plnorm(xcr,p[1],p[2],lower.tail=T,log.p=T))
     return(-ele)
     }

p=c(5.9,1.21)
(out=nlm(verosimil,p,hessian=T)$estimate)

# los valores de μ y σ se tienen como:
mu=out[1]
sigma=out[2]
```

A.2. Programas para muestras truncadas por la izquierda

Exponencial con truncamiento por la izquierda

```
# Datos observados
xt=c(243,294,340,384,457,680,855,877,974,1193,1340,1884,2558,15743)
m=length(xt)
x0=200

verosimil=function(p)
     {
     ele=sum(dexp(xt,p[1],log =T))
     -m*(pexp(x0,p[1],lower=F,log=T))
     return(-ele)
     }

p=c(100)
(thetag=nlm(verosimil,p,hessian=T)$estimate)

#El valor para θ̂ se obtiene como:
t=1/thetag
```

Weibull con truncamiento por la izquierda

```
# Datos observados
xt=c(243,294,340,384,457,680,855,877,974,1193,1340,1884,2558,15743)
m=length(xt)
x0=200

verosimil=function(p)
     {
     ele=sum(dweibull(xt,p[1],p[2],log=T))
     -(m*(pweibull(200,p[1],p[2],lower=F,log=T)))
     return(-ele)
     }

p=c(.1,1)
(thetag=nlm(verosimil,p,hessian=T)$estimate)
# los valores de β y θ se tienen como:
b=thetag[1]
t=thetag[2]
```

Lognormal con truncamiento por la izquierda

```
# Datos observados
xt=c(243,294,340,384,457,680,855,877,974,1193,1340,1884,2558,15743)
m=length(xt)
x0=200

verosimil=function(p)
     {
     ele=sum(dlnorm(xt,p[1],p[2],log=T))
     -(m*(plnorm(200,p[1],p[2],lower=F,log=T)))
     return(-ele)
     }

p=c(200,20)
(thetag=nlm(verosimil,p,hessian=T)$estimate)
# los valores de μ y σ se tienen como:
mu=thetag[1]
sigma=thetag[2]
```

A.3. Programas para muestras truncadas por la izquierda y censura por la derecha

Exponencial con truncamiento por la izquierda y censura por la derecha

```
x0=150
xobs=c(155,161,243,294,340,384,457,680,855,877,974,1193)
xcr=c(1300,1300,1300,1300)
data=c(xobs,xcr)
k=length(xobs)
m=length(xcr)
n=length(data)

#El valor para θ̂ se obtiene como:

t=(sum(data)-(nx0))/k
```

Weibull con truncamiento por la izquierda y censura por la derecha

```
x0=150
xobs=c(155,161,243,294,340,384,457,680,855,877,974,1193)
xcr=c(1300,1300,1300,1300)
data=c(xobs,xcr)
k=length(xobs)
```

```
n=length(data)

verosimil=function(p)
     {
     ele=sum(dweibull(xobs,p[1],p[2],log=T))
     +sum(pweibull(xcr,p[1],p[2],lower.tail=F,log.p=T))
     -(n(pweibull(150,p[1],p[2],lower=F,log=T)))
     return(-ele)
     }
p=c(.5,970)

(out=nlm(verosimil,p,hessian=T)$estimate)

# los valores de β y θ se tienen como:
b=out[1]

t=out[2]
```

Apéndice B

Funciones especiales en R

B.1. Distribución Exponencial

Descripción

Función de densidad, función de distribución, función cuantil y generación aleatoria de la distribución exponencial con razón rate (i.e., mean 1/rate).

Uso

```
dexp(x, rate=1, log=FALSE)
pexp(q, rate=1, lower.tail=TRUE, log.p=FALSE)
qexp(p, rate=1, lower.tail=TRUE, log.p=FALSE)
rexp(n, rate=1)
```

Argumentos

x, q	Vector de cuantiles.
p	Vector de probabilidades.
n	Número de observaciones. Si length(n)>1 la longitud se toma como el número requerido.
rate	Vector de razones.
log, log.p	Lógico; si es TRUE, las probabilidades de p están dadas como log(p).
lower.tail	Lógico; si es TRUE(default), las probabilidades son $P[X \leq x]$, en otro caso $P[X > x]$.

B.2. Distribución Lognormal

Descripción

Función de densidad, función de distribución, función cuantil y generación aleatoria de la distribución log normal cuya media es igual a meanlog y desviación estándar sdlog.

Uso

```
dlnorm(x, meanlog=0, sdlog=1, log=FALSE)
plnorm(q, meanlog=0, sdlog=1, lower.tail=TRUE, log.p=FALSE)
qlnorm(p, meanlog=0, sdlog=1, lower.tail=TRUE, log.p=FALSE)
rlnorm(n, meanlog=0, sdlog=1)
```

Argumentos

x, q	Vector de cuantiles.
p	Vector de probabilidades.
n	Número de observaciones. Si length(n)>1 la longitud se toma como el número requerido.
meanlog, sdlog	Media y desviación estándar por default toma los valores de 0 y 1 respectivamente.
log, log.p	Lógico; si es TRUE, las probabilidades de p están dadas como log(p).
lower.tail	Lógico; si es TRUE(default), las probabilidades son $P[X \leq x]$, en otro caso $P[X > x]$.

B.3. Distribución Weibull

Descripción

Función de densidad, función de distribución, función cuantil y generación aleatoria de la distribución Weibull con parámetros de forma y escala.

Uso

```
dweibull(x, shape, scale=1, log=FALSE)
pweibull(q, shape, scale=1, lower.tail=TRUE, log.p=FALSE)
qweibull(p, shape, scale=1, lower.tail=TRUE, log.p=FALSE)
rweibull(n, shape, scale=1)
```

Argumentos

x, q	Vector de cuantiles.
p	Vector de probabilidades.
n	Número de observaciones. Si length(n)>1 la longitud se toma como el número requerido.
shape, scale	parametros de forma y escala, este último por default es 1.
log, log.p	Lógico; si es TRUE, las probabilidades de p están dadas como log(p).
lower.tail	Lógico; si es TRUE(default), las probabilidades son $P[X \leq x]$, en otro caso $P[X > x]$.

B.4. Distribución Pareto

Descripción

Función de densidad, función de distribución, función cuantil y generación aleatoria de la distribución Pareto con parámetros de localización y escala.

Uso

```
dpareto(x, location, shape, log=FALSE)
ppareto(q, location, shape)
qpareto(p, location, shape)
rpareto(n, location, shape)
```

Argumentos

x, q	Vector de cuantiles.
p	Vector de probabilidades.
n	Número de observaciones. Si length(n)>1 la longitud se toma como el número requerido.
location, shape	alpha y k, parámetros de localización y forma
lower.tail	Lógico; si es TRUE, se regresa el logaritmo de la densidad.

B.5. Minimización no lineal

Descripción

El comando nlm lleva acabo una minimización de la función f utilizando el algoritmo de Newton.

Uso

```
nlm(f,p,...,hessian = FALSE, typesize = rep(1,length(p)),
    fscale = 1, print.level = 0, ndigit = 12,gradtol = 1e-6,
    stepmax = max(1000*sqrt(sum((p/typsize)^2),1000),
    steptol = 1e-6, iterlim = 100,check.analyticals = TRUE)
```

Argumentos

f	La función para ser minimizada. Si el valor de la función tiene un atributo llamado gradiente o ambos atributos el gradiente y el hessiano, estos serán utilizados en el cálculo de los valores paramétricos actualizados. Deriv regresa una función con el gradiente adecuado. Este debería ser una función de un vector de longitud p seguido por cualquier otro argumento especificado por el ... argumento.
p	Valores paramétricos de partida para la minimización.
...	Argumento adicional para f.
hessian	Si es TRUE, el hessiano de f en el mínimo se devuelve.
typsize	Una estimación del tamaño de cada parámetro en el mínimo.
fscale	Una estimación del tamaño de f en el mínimo.
print.level	Este argumento determina el nivel de impresión que se lleva a cabo durante el proceso de minimización. El valor por defecto es 0 significa que no se produce ninguna impresión, un valor de 1 significa que los datos iniciales y finales se imprimen y un valor de 2 significa que toda la información de seguimiento se imprimen.
ndigit	El número de dígitos significativos en la función f.
gradtol	Un escalar positivo dando la tolerancia a la que se considera el gradiente escalado o suficientemente cerca de cero para terminar el algoritmo. El gradiente escalado es una medida del cambio relativo en f en cada dirección de p[i] dividido por el cambio relativo en p[i].

stepmax	Un escalar positivo que da la longitud máxima permitida de un paso escalado. stepmax se utiliza para evitar pasos que causarían que la función de optimización se desborde, para evitar que el algoritmo salga de la zona de interés en el espacio de parámetros, o para detectar la divergencia en el algoritmo. stepmax se elige lo suficientemente pequeño para evitar que lo anterior ocurra, pero debe ser más grande que cualquier paso razonable anticipado.
steptol	Un escalar positivo que establece la longitud mínima admisible de un paso relativo.
iterlim	Un número entero positivo que especifica el número máximo de iteraciones a ser realizadas antes de que el programa se termine.
check.analyticals	Un escalar lógico especifica que si los gradientes analíticos y hessianos, son aplicados, deben ser comprobados con las derivadas numéricas en los valores iniciales de los parámetros. Esto puede ayudar a detectar gradientes o hessianos mal formulados.

Bibliografía

[1] Anderson, D. (2008). *Model Based Inference in the Life Sciences: A Primer On Evidence*. Springer.

[2] Bertram, L. (1971). *Reliability Mathematics: Fundamentals; Practices; Procedures*. McGraw-Hill.

[3] Billingsley, P. (1986). *Probability and Measure*. (Segunda Edición) New York: John Wiley and Sons.

[4] Boland, P. (2007). *Statistical and probabilistic methods in actuarial science*. Chapman and Hall.

[5] Brockwell, P. J. y Davis, R. A. (2009). *Times Series: Theory and Methods*. (Segunda Edición) Springer.

[6] Brufman, J., Urbisaia, H. y Trajtenberg, L. (2006). *Distribución del Ingreso Según Género un Enfoque no Paramétrico*. Cuadernos del CIMBAGE, 8, 129-168.

[7] Burden, R.L. y Faires J.D. (2017). *Análisis Numérico*. (Decima Edición) Cengage Learning.

[8] Canavos, G. (1984). *Applied probability and statistical methods*. Little, Brown Company.

[9] Casella, G. y Berger. R. (2001). *Statistical Inference*. (Segunda Edición) Duxbury Press.

[10] Ferreira, E. y Garin, M. (1997). *Estadística Actuarial: Modelos Estocásticos*. Sarriko-On.

[11] Figueroa, G. (2012). *Las Funciones de Verosimilitud Discretizada y Restringida Perfil en la Inferencia Científica* (Tesis de Doctorado). Universidad de Sonora, Hermosillo, Sonora, México.

[12] Forbes, C., Evans, M., Hastings, N. y Brian P. (2011). *Estatistical Distributions*. (Cuarta Edición) John Wiley and Sons.

[13] Gilbert, R. (1987). *Statistical Methods for Environmental Pollution Monitoring*. New York: John Wiley and Sons.

[14] Kalbfleisch, J. G. (195). *Probability and Statistical Inference*. Vol. 2: Statistical inference. (Segunda Edición) Nueva York: Springer-Verlag.

[15] Klein,J. y Moeschberger, M. (2003). *Survival Analysis Techniques for Censored and Truncated Data*. (Segunda Edición) Springer.

[16] Klugman, S., Panjer, H. y Willmot, G. (2004). *Loss Models From Data to Decisions*. (Segunda Edición) John Wiley and Sons.

[17] Lawless, J. F. (1986) *Statistical Models and Methods for Life Time Data*. New York: John Wiley and Sons.

[18] Lee, E. y Wenyu, J. (2003). *Statistical Methods for Survival Data Analysis*. (Tercera Edición) John Wiley and Sons.

[19] Ley de instituciones de Seguros y Fianzas. (2013). *Comisión Nacional de Seguros y Fianzas*. Diario Oficial de la Federación, 4 de Abril de 2013. URL http://www.cnsf.gob.mx/.

[20] Marques, M. J. (2001). *Estadística Básica un Enfoque No Paramétrico*. México: FES-Zaragoza

[21] Medina, L. A. (2005). *Estimación de máxima verosimilitud en la distribución Weibull para muestras completas y censuradas y su aplicación en el análisis de tiempos de vida* (Tesis Profesional). Universidad de Chapingo, Chapingo Edo. de México.

[22] Meeker, W. Q. y Escobar, L. A. (1998). *Statistical Methods for Reliability Data*. New York: John Wiley and Sons

[23] Millard, S. P. y Neerchal, N. K. (2001). *Environmental Statistics*. New York: CRC Press

[24] Montoya, J. A. (2008). *La verosimilitud perfil en la Inferencia Estadística*. (Tesis de Doctorado). CIMAT, Guanajuato, México.

[25] Murray, R. S. (1976). *Probabilida y Estadística*. (Primera Edición) McGraw-Hill.

[26] Nadarajah, S y Kotz, S. (2006). R Programs for Computing Truncated Distributions. *Journal of statistical software*, 16.

[27] Norman, J., Kotz, S. y Balakrishnan, N. (1994). *Continuous Univariate Distributions*. (Segunda Edición) John Wiley and Sons.

[28] Ossa, E. (1988). *Teoría General del Seguro. La Institución.* Tomo I. Bogotá: Temis.

[29] Ott, W.R. (1990). A physical explanation of the lognormality of pollutant concentrations. *Journal of the Air and Waste Management Association*, 40, 1378-1383.

[30] Ouyang, L.Y. y W, S.J. (1994). Prediction intervals for an ordered observation From a Pareto distribution. *IEEE Transactions on Reliability*, 43(2).

[31] Perez, R. y Correa, J. (2008). Intervalos de confianza vía verosimilitud relativa de los parámetros de la distribución Birnbaum-Saunders. *Revista Colombiana de Estadistica*, 31(1), 85-94.

[32] Quintana, C. (1996). *Elementos de Inferencia Estadística.* Costa Rica: Editorial de la Universidad de Costa Rica.

[33] Ricci, V. (2005). *Fitting Distributions With R* Free Documentation License, http://www.fsf.org/licenses/licenses.html#FDL.

[34] Rivaud, J. J. (2004). La Función Gamma. *Miscelánea Matemática*, 39, 61-84.

[35] Sidney, S. (1995). *Estadística no Paramétrica.* México: Trillas.

[36] Smith, G. (1998). *Introduction to Statistical Reasoning.* WCB/McGraw-Hill.

[37] Spanos, A. (1999). *Probability Theory and Statistical Inference: Econometric Modeling With Observational Data.* Cambridge University Press.

[38] Ulín, F. (2008) *Inferencia sobre datos no detectados de poblaciones lognormales* (Tesis de Doctorado). COLPOS, Texcoco Edo. México

[39] Vic, B. (2004). *Environmental statistics methods and applications.*John Wiley and Sons.

[40] Wackerly, D., Mendenhall, W. y Scheaffer, R. (2010). *Estadística Matemática con Aplicaciones.* (Septima Edición) Cengage Learning.

[41] Wayne, N. (1982). *Applied life data analysis.* John Wiley and Sons.

[42] William, G. (2010). *Econometric analysis.* (Séptima Edición) Pearson Education.

INFERENCIA ESTADÍSTICA EN PRESENCIA DE CENSURA Y TRUNCAMIENTO

ORIGINALITY REPORT

11%
SIMILARITY INDEX

PRIMARY SOURCES

1	colposdigital.colpos.mx:8080 Internet	273 words — 1%
2	runebook.dev Internet	241 words — 1%
3	cimat.repositorioinstitucional.mx Internet	229 words — 1%
4	lic.mat.uson.mx Internet	220 words — 1%
5	www.amestad.mx Internet	201 words — 1%
6	docplayer.es Internet	198 words — 1%
7	archive.org Internet	101 words — 1%
8	pdfcoffee.com Internet	93 words — < 1%
9	bdigital.unal.edu.co Internet	85 words — < 1%

10	www.fernandocoriat.com Internet	78 words — < 1%
11	1library.co Internet	63 words — < 1%
12	edoc.pub Internet	62 words — < 1%
13	ri.ues.edu.sv Internet	45 words — < 1%
14	fr.scribd.com Internet	41 words — < 1%
15	repositorio.ucv.edu.pe Internet	38 words — < 1%
16	www.ugrad.stat.ubc.ca Internet	38 words — < 1%
17	www.coursehero.com Internet	34 words — < 1%
18	www.cib.espol.edu.ec Internet	32 words — < 1%
19	www.ic.fcen.uba.ar Internet	25 words — < 1%
20	www.slideshare.net Internet	22 words — < 1%

EXCLUDE QUOTES ON EXCLUDE SOURCES OFF