UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO

DIVISIÓN ACADÉMICA DE CIENCIAS BÁSICAS

ANÁLISIS DE DATOS AMBIENTALES CENSURADOS

T E S I S

PARA OBTENER EL GRADO DE

MAESTRO EN CIENCIAS EN MATEMÁTICAS APLICADAS

PRESENTA:

ISRAEL JESÚS DAMIÁN FÉLIX

DIRECTOR DE TESIS:

DR. FIDEL ULÍN MONTEJO

CUNDUACÁN, TABASCO, MÉXICO JUNIO 2013

"2013, CENTENARIO LUCTUOSO DE FRANCISCO I. MADERO Y JOSE MA. PINO SUAREZ"

6 de junio de 2013

Mat. Israel Jesús Damián Félix
Pasante de la Maestría en Matemáticas Aplicadas
P r e s e n t e.

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 ***"ANÁLISIS DE DATOS AMBIENTALES CENSURADOS"*** en virtud de que reúne los requisitos para el EXAMEN PROFESIONAL para obtener el grado de Maestría.

Sin otro particular, reciba un cordial saludo.

A t e n t a m e n t e.

DIVISIÓN ACADÉMICA DE CIENCIAS BÁSICAS
DIRECCIÓN.

Dr. Víctor Castellanos Vargas
Director

C.C.P. Expediente del Alumno
Archivo

CARTA DE AUTORIZACIÓN

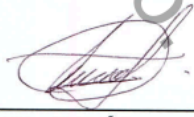
El que suscribe, autoriza por medio del presente escrito a la Universidad Juárez Autónoma de Tabasco para que utilice tanto física como digitalmente la tesis de grado denominada "*ANÁLISIS DE DATOS AMBIENTALES CENSURADOS*", de la cual soy autor y titular de los Derechos de Autor.

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 limitada 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 21 días del mes de junio del año 2013.

AUTORIZO



ISRAEL JESÚS DAMIÁN FÉLIX
072A10001

ANÁLISIS DE DATOS AMBIENTALES CENSURADOS

ISRAEL JESÚS DAMIÁN FÉLIX

Agradecimientos

A Dios por darme la fuerza necesaria y darme la dicha de haber logrado este sueño tan anhelado .

A mis padres, Vicente Damián Hernández y María Isabel Félix Montero por ese amor infinito que me tienen, por su apoyo incondicional, por mantener siempre su confianza en mí y creer que podía lograrlo, porque su fuerza es mi fuerza.

A mis hermanos, Eduardo, Aldo y Lupita por estar siempre conmigo cuando más los necesito, por sus palabras y sus ánimos en mis momentos de debilidad.

A mi asesor, el Dr. Fidel Ulín Montejo, por su dedicación, su paciencia, su apoyo a cada momento pero sobre todo por compartir su sabiduría conmigo y poder concluir este trabajo.

A mi sobrino Axel por ser mi primer sobrino y dar alegría e inspiración a mi alma.

A todos mis amigos(as), tíos(as) y cada persona que haya estado directa o indirectamente en mi crecimiento profesional, por su apoyo, comprensión y cariño.

Cada obstáculo en la vida existe para ser superado; un escalón más hacia el éxito.

Índice general

1. Introducción **4**

2. Datos censurados **7**
2.1. Censura 7
2.2. Censura por la izquierda 8
2.3. Datos no detectados 10
2.4. Enfoques de estimación 12
2.4.1. Sustitución 12
2.4.2. Estimación por máxima verosimilitud 13
2.4.3. Métodos no paramétricos 15

3. Distribuciones de variables aleatorias continuas **16**
3.1. Distribuciones de localización-escala 16
3.2. Modelos continuos de probabilidad 17
3.3. La función cuantil 17
3.4. Algunos modelos importantes 18
3.4.1. Parámetros de las distribuciones 18
3.4.2. Distribución exponencial 19
3.4.3. Distribución lognormal 20
3.4.4. Distribución Weibull 22
3.5. Momentos de una variable aleatoria 24
3.5.1. Valor esperado 24
3.5.2. Momentos y varianza 24
3.5.3. Momentos de una variable aleatoria exponencial 25
3.5.4. Momentos de una variable aleatoria lognormal 26
3.5.5. Momentos de una variable aleatoria Weibull 27

4. Estimación por máxima verosimilitud **29**
4.1. Función de verosimilitud 29
4.2. Estimación para un parámetro 31
4.2.1. Funciones de verosimilitud y log-verosimilitud 31

4.2.2. Función de información 32
4.3. Estimación para dos parámetros 34
4.4. Estimación sobre una muestra aleatoria exponencial censurada por la izquierda 37
4.5. Estimación sobre una muestra aleatoria lognormal censurada por la izquierda 40
4.6. Estimación sobre una muestra aleatoria Weibull censurada por la izquierda 42
4.7. Selección de modelos 44
4.7.1. Criterio de Información de Akaike 44
4.7.2. Prueba de razón de verosimilitudes 45

5. Intervalos de verosimilitud confianza 47
5.1. Aproximación normal 47
5.1.1. Aproximación normal para un solo parámetro 47
5.1.2. Aproximación normal para $\boldsymbol{\theta} = (\theta_1, \theta_2)$ 49
5.2. Matriz de covarianzas para funciones estimadas 49
5.3. Intervalo de confianza para $g = g(\mu, \sigma)$ 50
5.4. Regiones de confianza 50
5.5. Estimación con datos no censurados 51
5.5.1. Intervalo de confianza para la media 51
5.5.2. Intervalo de confianza para la mediana 53
5.6. Estimación con datos censurados 54
5.6.1. Intervalo de confianza de la media 54
5.6.2. Intervalo de confianza de la mediana 55

6. Aplicaciones 56
6.1. Muestra de cobre en Alluvial Fan Zone 59
6.1.1. Estimación muestral de la exponencial 59
6.1.2. Estimación muestral de la lognormal 61
6.1.3. Estimación muestral de la Weibull 63
6.2. Muestra de zinc en Alluvial Fan Zone 66
6.2.1. Estimación muestral de la exponencial 66
6.2.2. Estimación muestral de la lognormal 68
6.2.3. Estimación muestral de la Weibull 70
6.3. Muestra de cobre en Basin-Trough Zone 72
6.3.1. Estimación muestral de la exponencial 72
6.3.2. Estimación muestral de la lognormal 74
6.3.3. Estimación muestral de la Weibull 76
6.4. Muestra de zinc en Basin-Trough Zone 79
6.4.1. Estimación muestral de la exponencial 79

6.4.2. Estimación muestral de la lognormal 81
6.4.3. Estimación muestral de la Weibull 83
6.5. Selección del modelo para los datos 86
6.5.1. Criterio de Información de Akaike 87
6.5.2. Prueba de razón de verosimilitudes 89

7. Conclusión **94**

Bibliografía **96**

Apéndices **100**

A. Algunos resultados estadísticos **101**
A.1. Método delta 101
A.2. Método delta para funciones vectoriales 104

B. Método de Newton-Raphson **106**
B.1. Método de Newton-Raphson univariado 106
B.2. Método de Newton-Raphson bivariado 107

C. Funciones en lenguaje R **109**
C.1. Distribución exponencial 109
C.2. Distribución log normal 110
C.3. Distribución Weibull 110
C.4. Optimización general 111
C.5. Minimización no lineal 112

Capítulo 1

Introducción

El campo de la estadística ambiental es relativamente joven y emplea métodos desarrollados en otras áreas científicas para resolver problemas concernientes al medio ambiente. Algunos de estos problemas se enfocan en el monitoreo de la calidad del aire y agua, calidad de mantos freáticos cercanos a depósitos de desechos tóxicos, aseguramiento de zonas contaminadas, cuantificación en los procesos de remediación, medición del nivel de riesgo de zonas potencialmente contaminadas, etc.

Cuando los procedimientos y equipos químicos no logran cuantificar los niveles de concentración de una sustancia contaminante en una muestra o tejido, se reportan como *datos no detectados* u *observaciones incompletas*, omitiéndose una medida numérica. Esta situación se ha convertido en un problema recurrrente para investigadores que se enfrentan al manejo y análisis de datos ambientales, donde no se puede asumir que todos los datos *no detectados* son ceros; sino más bien, todos los datos no detectados son menores que un *límite de detección* establecido por la precisión del aparato de medición, o por los protocolos de análisis fisicoquímicos. En este trabajo se ilustran procedimientos paramétricos alternativos para una mejor inferencia sobre este tipo de información, considerando la información de origen sin supuestos falsos o erróneos atribuidos al desconocimiento de los métodos de análisis de confiabilidad (o supervivencia) para el manejo y *análisis de datos censurados*.

El análisis de confiabilidad tiene su origen en la dificultad que se presenta al analizar tiempos de vida, donde es común que algunos individuos no puedan ser observados hasta tener una recaída o deceso; que al finalizar un experimento de pruebas de vida no todos los componentes hayan fallado; o que la señal producida por un contaminante sea tan pequeña que los instrumentos de medición no puedan registrarla. Tales observaciones son llamados

datos censurados. Los más comunes son los datos censurados por la derecha; originados por unidades que no fallaron durante el experimento. La censura por intervalos refleja incertidumbre respecto al tiempo exacto en que las unidades fallaron y los datos censurados por la izquierda donde solo se sabe que la falla ocurrió antes de un cierto tiempo. Este trabajo se enfoca precisamente en esta última clasificación de datos, asumiendo que los límites de detección en aparatos y protocolos de medición de contaminantes establecen los límites de censura.

Este trabajo se ha estructurado en 7 capítulos, aquí se desarrolla el Capítulo 1 correspondiente a la Introducción. En el Capítulo 2 se realiza una breve descripción respecto a los distintos tipos de censura mas frecuentes, sus conceptos y definiciones necesarios para una mejor comprensión de este estudio; en un apartado, se muestran algunos antecedentes generales de los trabajos realizados con información censurada por la izquierda, y de igual modo se comentan los diferentes enfoques sobre estimación paramétrica que se han utilizado para la inferencia sobre datos censurados. Aunque la literatura reporta importantes trabajos desde los métodos no-paramétrico, la estimación paramétrica por máxima verosimilitud ha tenido grandes avances en las ciencias ambientales, siendo éste el enfoque sobre el cual se ha desarrollado este trabajo. En el Capítulo 3 se presenta una revisión general sobre algunas distribuciones de localización-escala importantes para datos ambientales censurados; por ejemplo, la distribución exponencial, la distribución lognormal y la distribución Weibull; definidas claramente por su función de distribución, su función de densidad, y por algunas medidas características como los cuantiles, los momentos, la media y la varianza. La teoría que fundamenta los métodos de máxima verosimilitud y la función de verosimilitud misma se ilustran en el Capítulo 4, desde los conceptos, definiciones y resultados claves para el desarrollo de la inferencia estadística requerida. El Capítulo 5 se ha dispuesto para mostrar los elementos necesarios en la estimación por intervalos de cantidades de interes como la media y mediana, a través de la teoría estadística desarrollada con la aproximación normal de la logverosimilitud. En el Capítulo 6 se muestran aplicaciones de la metodología descrita, considerando datos de concentraciones de zinc y cobre en muestras de agua de mantos freáticos de dos áreas geológicas en el Valle de San Joaquín en California, USA; Helsel, Ulín-Montejo, El-Shaarawi y Viveros, en [12, 20,22] han estudiado estos datos remarcando la pertinencia de su análisis desde el enfoque paramétrico bajo máxima verosimilitud, obteniendo inferencias adecuadas y validadas.

Las conclusiones se resumen en el Capítulo 7, de manera condensada

y señalando lo más relevante de la metodología utilizada, las aplicaciones, resultados, selección de modelos y discusiones respectivas. En el Apéndice, se presentan algunos resultados importantes para el proceso de estimacion puntual y por intervalos, como el Método Delta y algunas propiedades de los estimadores de máxima verosimilitud; en el mismo sentido se muestra el Método de Newton-Raphson y por último se describen detalladamente las funciones en lenguaje R que han sido utilizadas para operaciones de optimización y estimación en general.

En síntesis, el presente trabajo inicia con una revisión de los conceptos y fundamentos esenciales con lo concerniente al fenómeno de censura, los modelos estadísticos y probabilísticos más estudiados en datos no detectados de estudios ambientales. Posteriormente puede observarse una descripción detallada de la teoría y métodos de máxima verosimilitud, probando y desarrollando analíticamente las herramientas necesarias en la estimación puntual y por intervalos de la media, mediana para los modelos elegidos; el enfoque uni-paramétrico y bi-paramétrico ha sido considerado y descrito claramente. Luego, con ayuda de una base de datos que la literatura valida y con apoyo computacional que ofrece el lenguaje R, se realizan ejercicios de aplicación de la metodología descrita y de los procedimientos desarrollados en base a ésta, generándose resultados importantes y conclusiones importantes, con los que se adquiere un conocimiento útil e importante para contribuir a la solución de problemas ambientales que involucren datos no detectados en agua, suelo y aire; en particular, generado por sustancias contaminantes que contienen metales pesados o elementos traza, difíciles de cuantificar por sus magnitudes demasiado pequeñas y alto riesgo.

Capítulo 2

Datos censurados

2.1. Censura

Existen dos mecanismos que impiden la observación directa y completa de algunas variables de interés, como son los tiempos de vida, tiempos a la falla o medida de una sustancia; éstos son *la censura* y el truncamiento. Para el fenómeno de censura existen dos tipos: la Censura Tipo I, en la cual los especímenes son observados hasta un tiempo determinado; y la Censura Tipo II, en la cual los especímenes son observados hasta que ocurran un número determinado de fallas o eventos de interés. Los mecanismos de Censura Tipo I más recurrentes son los siguientes:

i) **Censura por la derecha:** Se presenta hasta la última observación que se hace al individuo o espécimen, sin que haya ocurrido el evento deseado. Existen varias razones para ello:
- Que al finalizar el estudio no haya ocurrido el evento.
- Que el individuo haya abandonado el estudio.
- Que ajeno al individuo, haya ocurrido otro evento que imposibilite la ocurrencia del evento que se desea observar.

ii) **Censura por la izquierda:** Es poco común en análisis de supervivencia, se presenta cuando en la primera observación que se realiza sobre el individuo el evento de interés ya ocurrió. En estudios ambientales es muy frecuente al registrar concentraciones muy pequeñas de metales pesados en agua, aire o suelo, utilizando aparatos de medición con precisión limitada.

iii) **Censura por intervalos:** Se tiene cuando solo se sabe que el evento de interés ocurrió entre un instante t_i y un tiempo t_j.

Una descripción general y detallada de los distintos tipos de censura puede verse en las publicaciones [3] y [28].

2.2. Censura por la izquierda

La censura por la izquierda es un tema difícil, y la cuestión de cómo tratarla se presenta a menudo en disciplinas como las ciencias sociales que dependen en gran medida de los datos de encuestas (por ejemplo, en el análisis de la duración del empleo y el desempleo, la duración de la pobreza, la duración de la dependencia del bienestar, etc.). Además, el problema de censurar por la izquierda ocurre en muchas otras disciplinas, como las ciencias ambientales y la epidemiología, donde el comienzo del proceso no se conoce con certeza. Por ejemplo, en la enfermedad de estudios de duración -como el cáncer- , la aparición de la enfermedad a menudo no se conoce (al menos no con certeza) [3].

En estudios ambientales, algunos de los problemas relevantes que requieren el uso de la inferencia estadística son: monitoreo de la calidad del aire y agua, monitoreo de la calidad de mantos freáticos cercanos a depósitos de desechos tóxicos, aseguramiento de zonas contaminadas en proceso de remediación, medición de nivel de riesgo de zonas potencialmente contaminadas, aseguramiento de riesgos en lugares de trabajo, etc. [38]. El campo de la estadística ambiental es relativamente reciente y emplea métodos desarrollados en otras disciplinas para resolver problemas concernientes al medio ambiente.

La evaluación objetiva del riesgo es importante, debido a que grandes concentraciones de contaminantes frecuentemente tienen efectos adversos a la salud, sin embargo, pequeñas concentraciones pueden no ser inocuas. Con la finalidad de conocer el nivel de riesgo de estos contaminantes, se pueden identificar y medir éstos en niveles extraordinariamente bajos. Algunas veces la cantidad presente de un contaminante es tan pequeña para los instrumentos de medición que no pueden ser cuantificadas ni discriminadas [22]. Cuando no es posible cuantificar las concentraciones en una muestra se reportan como *datos no detectados* [31]. Tales observaciones incompletas son llamadas datos censurados por la izquierda. La censura es un punto de referencia que indica hasta donde fue posible tomar una medida de la variable de interés sobre el espécimen. En este caso, el punto de referencia se denomina *límite de detección* (LD). En el mismo sentido, en monitoreos de exposición a contaminantes en centros de trabajo, las mediciones generalmente son valores positivos; sin embargo, pueden tenerse igualmente datos no detectados. Aquí,

las estrategias para asegurar información referente a exposiciones a contaminantes en centros laborales están enfocadas a determinar el nivel medio de exposición, a fin de tomar alguna decisión que evite un riesgo o accidente [15].

Investigadores en diferentes ciencias ambientales han reportado que las mediciones de concentraciones de varias sustancias tienen distribuciones de frecuencias de tipo lognormal o aproximadas a ésta. Es decir, cuando los logaritmos de las concentraciones observadas son graficadas como una distribución de frecuencias, la distribución resultante es aproximadamente normal, sobre la mayor parte del rango observado. Este supuesto se justifica por el hecho de que en muchas situaciones las concentraciones de contaminantes son el resultado de muchas diluciones ligeras [41]. Algunos ejemplos incluyen material radiactivo en suelos, contaminantes en aire, calidad del aire, residuos de metales en ríos y en tejidos biológicos. En estos casos la normalidad se asume después de la log-transformación de los datos y entonces las técnicas basadas en la teoría de la normal son usadas para obtener inferencias acerca de la cantidad de interés. Sin embargo, como se discutió en [13], frecuentemente los resultados necesitan ser reportados como mediciones obtenidas, más que en la escala log-transformada. Por ejemplo, la Agencia de Protección Ambiental de USA (EPA por sus siglas en ingles), requiere que los riesgos sean caracterizados en términos de la concentración media del contaminante. Este requerimiento es parcialmente responsable del énfasis en la concentración media y su inferencia.

Desde una perspectiva diferente, dado que las concentraciones de un contaminante son físicamente aditivas en un sentido útil, éstas pertenecen a la clase de variables extensivas descritas por [10]. A pesar de la forma de la distribución, la concentración media del contaminante tiene una interpretación física, lo que no puede decirse respecto a la log-concentración o alguna otra transformación no lineal. Los datos ambientales pueden presentar características que conducen a problemas de inferencia muy interesantes. La meta es obtener aseveraciones cuantitativas respecto a los parámetros desconocidos, así como regiones de confianza aproximadas usando toda la información paramétrica contenida en la muestra. Un estimador de máxima verosimilitud (EMV) y su varianza son suficientes para especificar o reproducir la función de verosimilitud completa [11].

2.3. Datos no detectados

Las mediciones cuyos resultados sólo se sabe que son mayores o menores a un umbral son llamados datos censurados [31]. Este tipo de mediciones han sido una parte integral de las ciencias sociales, económicas, médicas y la estadística industrial. Existen procedimientos desarrollados recientemente que permiten incorporar los datos censurados en el cálculo de estadísticas descriptivas, regresión y pruebas de hipótesis. Pero esos procedimientos son raramente usados en estudios ambientales, donde los datos censurados son frecuentemente aquellos valores que se encuentran por debajo de los límites de detección. Esos niveles bajos de contaminantes (tales como residuos de metales o compuestos orgánicos) son conocidos con imprecisión y llamados "menores que" o "no detectados". Debido a que bajos niveles son usualmente presentados a la izquierda de un gráfico, los no detectados son etiquetados como çensurados por la izquierda", con valores en algún lugar a la izquierda del límite de detección (LD), pero sin conocer la concentración exacta.

Los datos no detectados son un tema de interés en varias disciplinas y subdisciplinas dentro de las ciencias ambientales, son encontrados en estudios de calidad del aire [43], estudios marinos [24], análisis de aguas residuales [30], lluvia ácida [1], pruebas en bombas de pozos [50], exposición humana a toxinas [42], SIDA y otros estudios de virus [34], contaminantes en cuerpos de animales diversos ([18]; [23]), sedimentos químicos [9], astronomía [25], química de rocas y minerales [36], y en estudios de calidad de agua en ríos y mantos freáticos [19]. De esta manera, el tema de cómo mejorar el análisis estadístico de datos censurados por la izquierda es concerniente a muchos científicos en diversas áreas del conocimiento.

En estudios médicos e industriales son más frecuentes los datos censurados por la derecha, donde sólo se conoce que las observaciones fueron más grandes que un valor establecido [32]. Un ejemplo es el registro del tiempo que la gente vive después de recibir un tratamiento médico. Dos tratamientos pueden ser comparados para medir su efectividad, por ejemplo la mediana del tiempo de sobrevivencia después del tratamiento; esto ocurre en diferentes tiempos para diferentes pacientes, por lo que el tratamiento preferido es el que resulta con mayor expectativa de vida. Para pacientes que aún viven después que el experimento finaliza, los tiempos de supervivencia no son conocidos exactamente, denominándose valores "mayor que" o datos censurados por la derecha.

Pocos estudios en ciencias ambientales han recomendando o realmente

usado técnicas para datos censurados adoptadas de otras disciplinas. Quizá el primero pudo haber sido [36], quien promovió el uso de estimación por máxima verosimilitud para calcular valores medios de datos minerales. En [37] demostraron que las pruebas de rangos y pruebas de comparación de dos grupos para datos censurados, podrían ser usadas para estudios de calidad del agua. [21] Se enfocó en el estudio de técnicas de análisis de supervivencia para pruebas de hipótesis y regresión. Por otro lado, [47] utilizó máxima verosimilitud para regresión de datos censurados de concentraciones. Asimismo, [45] utilizó estimadores Kaplan-Meier para obtener estadísticos descriptivos y caracterizar datos de calidad de agua.

Un antecedente importante es una guía para el uso de técnicas para datos censurados por la izquierda en estudios ambientales, publicado por [2] en el Handbook of Statistics Vol 12. Sin embargo, ni los ejemplos presentados y ni el Handbook han cambiado las prácticas estándar, las cuales están basadas en la omisión o sustitución de los datos no detectados por una fracción de los LD. De modo que ninguno de los métodos sugeridos en esta publicación se usa actualmente en ciencias ambientales.

Por lo general, cuando se tienen datos no detectados, dos métodos demasiado simplistas son usados. El primero consiste en no considerar los datos no detectados, lo cual produce resultados sesgados al eliminar los valores más bajos entre las observaciones. En este caso las pruebas o estadísticos que resultan no se aplican realmente para el total de datos muestreados, sino sólo a una parte de ellos, los de la región más alta de la distribución. Esta práctica es muy frecuente cuando el interés radica únicamente en la detección y en las mediciones de las observaciones detectadas ([29]; [33]). Sin embargo, al comparar entre grupos de datos y entre estudios, los resultados dependerán de los límites de detección en vigor al tiempo en que fueron analizados. Estos resultados pueden diferir erróneamente debido a que las estimaciones son realizadas usando sólo las observaciones detectadas, las cuales estarían alejadas del contexto general de la distribución de probabilidad. Esta práctica hace vulnerable a los autores respecto a sus aseveraciones debido a los sesgos inducidos en el análisis [49].

El segundo método comúnmente usado para manejar datos censurados es asignar fracciones arbitrarias de los LD a cada observación censurada (sustitución o fabricación). En numerosos estudios realizados a lo largo de los años, desde [40], a [7] y otros , y [23] y otros; se ha registrado la sustitución para datos no detectados por un medio del LD. Manuales de procedimientos de al menos tres agencias federales en USA, han recomendado también esta

práctica ([5]; [14]; [39]). Sin embargo, la sustitución puede conducir a indicios no presentes en los datos originales u ocultan una señal que realmente existe.

En un estudio usando simulación para comparar diferentes métodos estadísticos descriptivos de cálculo para datos censurados, [46] reportaron que la sustitución resulta en estimadores sesgados. El trabajo fue realizado para el caso más simple de datos con sólo un límite de detección, al considerar casos con límites de detección múltiples los errores por sustitución se incrementaron.

En este sentido, el manejo inadecuado de los datos censurados es evidencia de un escepticismo general respecto a la información contenida y obtenida a partir de ese tipo de observaciones. En realidad, los datos censurados contienen una gran cantidad de información que puede ser extraída usando métodos eficientes; esta información es tan valiosa como la obtenida de los valores conocidos. Procedimientos eficientes para datos censurados combinan los valores superiores a los límites de detección con la información contenida en la proporción de los datos debajo de estos mismos límites.

2.4. Enfoques de estimación

Acorde a lo anterior, [22] discutió tres enfoques para extraer información de los conjuntos de datos que incluyen datos no detectados: sustitución, estimación por máxima verosimilitud y métodos no paramétricos. El primer enfoque, que se refiere a la sustitución, no es recomendable y solo se discutirá brevemente. Los otros dos métodos son enfoques viables para el análisis de datos censurados y son usados como métodos estándar en otras disciplinas, además de las ciencias ambientales. A continuación se presenta una descripción general de cada enfoque.

2.4.1. Sustitución

Este método de sustitución consiste en calcular estadísticas para datos censurados fabricando valores como función de los límites de detección registrados. Estadísticas descriptivas, pruebas de hipótesis y modelos de regresión se calculan usando esos números fabricados, junto con valores detectados superiores al límite de detección. Los métodos de sustitución han sido usados ampliamente, pero no tienen bases teóricas. Varios estudios han mostrado que el método de sustitución no funciona bien, cuando se comparan con los otros dos tipos de enfoques mencionados anteriormente

([16]; [20]; [46]). La sustitución es usada porque es fácil de hacer, sin embargo, la fabricación de valores no es justificable dado que la elección de dichos valores es arbitraria y los resultados estadísticos difieren dependiendo de los valores elegidos.

2.4.2. Estimación por máxima verosimilitud

La estimación por máxima verosimilitud (MV) ha sido usada muy poco en estudios ambientales. Los estudios realizados por Owen y DeRouen (1980) para calidad del aire y Miesch (1967) para geoquímica son dos de los primeros ejemplos en la literatura. La estimación por MV utilizan tres piezas de información fundamentales: a) datos mayores a los límites de detección, b) proporción de datos por debajo de los límites de detección, y c) una distribución de probabilidad asumida. En este caso se supone que todos los datos, detectados y no detectados, siguen la misma distribución, por ejemplo la log-normal. Los parámetros son estimados de tal forma que se tenga el mejor ajuste a la distribución con los valores detectados y la proporción de los no detectados.

La idea general de los métodos basados en MV es ajustar modelos a los datos, considerando combinaciones de parámetros y modelos para las que la probabilidad de los datos sea grande. Estos métodos pueden ser aplicados con una gran variedad de modelos paramétricos con datos censurados. También es posible ajustar modelos con variables explicatorias (es decir, análisis de regresión). La teoría desarrollada, garantiza que estos métodos son estadísticamente eficientes (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 (Meeker & Escobar 1998) .

Los estimadores de máxima verosimilitud (EMV) son calculados resolviendo una función de verosimilitud $L(\theta)$, donde para una distribución con dos parámetros $\theta = (\theta_1, \theta_2)$, $L(\theta)$ define la verosimilitud de hacer coincidir la distribución de los datos observados. La función $L(\theta)$ aumenta a medida que el ajuste entre la distribución estimada y los datos observados mejora. Los parámetros θ_1 y θ_2 son obtenidos desde una rutina de optimización, eligiendo los valores para maximizar $L(\theta)$. En la práctica es el logaritmo natural de $L(\theta)$ lo que se maximiza; es decir, la "log-verosimilitud", $lnL(\theta)$, que a menudo (aunque no necesariamente) es un número negativo. La maximización se lleva a cabo mediante la solución simultánea del sistema de ecuaciones que se establece de las derivadas parciales de $lnL(\theta)$ con respecto a los dos

parámetros igualados a cero; esto es,

$$\begin{aligned} \frac{\partial[ln\ L(\theta)]}{\partial\theta_1} &= 0, \\ \frac{\partial[ln\ L(\theta)]}{\partial\theta_2} &= 0. \end{aligned} \tag{2.1}$$

La definición precisa para $L(\theta)$ cambiará dependiendo de la distribución asumida y el proceso en estudio (estimación de una media, regresión lineal, etc.). Sin embargo, en cada caso la verosimilitud $L(\theta)$ es el producto de dos componentes: la primera corresponde a las observaciones censuradas y la segunda a las observaciones no censuradas. En la parte no censurada se utiliza la *función de densidad de probabilidad* $f(y)$ *o bien la densidad de* Y, donde Y es una variable aleatoria, que es la ecuación que describe la frecuencia de la observación de los valores individuales de y. En la parte censurada se utiliza $1 - F(y)$, donde $F(y)$ es la *función de distribución acumulada o bien la distribución de* Y, donde Y es una variable aleatoria.

En el caso general, $L(\theta)$ puede ser considerado como el producto de tres factores, donde la parte de los datos censurados se divide en dos, uno para los datos censurados por la izquierda y el otro para los datos censurados por la derecha:

$$L(\theta) = \prod_{i=1}^{n_1} f(y_i) \prod_{j=1}^{n_2} F(y_j) \prod_{k=1}^{n_3} [1 - F(y_k)],$$

donde $f(y_i)$ es la densidad de y_i, estimada a partir de las observaciones detectadas, $F(y_j)$ es la distribución de y_j determinada a partir de las observaciones censuradas por la izquierda, y $1 - F(y_k)$ determinada por las observaciones censuradas por la derecha ("mayor que").

Para datos censurados por la izquierda, la función de verosimilitud requiere de dos variables para representar cada observación, y y δ. El valor para la medición, o para el límite de detección, está dado por y, mientras que la variable indicadora δ toma el valor cero (0) si se trata de una observación censurada y uno (1) si es un dato observado. Dicha función es la siguiente

$$L(\theta) = \prod [f(y)]^{\delta_i} \prod [F(y)]^{1-\delta_i}. \tag{2.2}$$

Ahora, los métodos de máxima verosimilitud pueden ser utilizados para realizar una prueba de hipótesis. Una prueba de interés puede estar configurada para determinar si $\theta = \theta_0$, donde θ es el parámetro o la función

paramétrica en cuestión. Este podría ser un coeficiente de la pendiente en regresión, o un estimador de la diferencia entre dos medias poblacionales. En un esquema de pruebas de significancia la hipótesis $\theta = \theta_0$ se contrasta con el EMV para θ, $\hat{\theta}$, usando una prueba de razón de verosimilitud o una prueba de Wald.

Las pruebas de razón de verosimilitudes se basan en el valor de la función de log-verosimilitud, $lnL(\theta)$. La prueba compara las log-verosimilitudes de dos modelos, uno para la condición máximo verosímil donde $\theta = \hat{\theta}$ y el otro por la condición de la hipótesis, $\theta = \theta_0$. El estadístico de prueba se define y se calcula así $-2ln[L(\theta_0) - L(\hat{\theta})]$, resultando en un valor positivo grande si estadísticamente $\theta \neq \theta_0$. Esta diferencia es el estadístico de prueba para la razón de verosimilitud, y es comparado con la distribución ji-cuadrada para producir el *p-valor* para la prueba.

La prueba estadística de Wald toma una forma similar a la *prueba-t* en regresión. El numerador de la prueba estadística es el valor del EMV para el coeficiente $\hat{\theta}$, y el denominador es el error estándar de θ. Su razón es comparada con una distribución normal estándar. La prueba de Wald no se considera tan precisa como lo son las pruebas de razón de verosimilitud, este último se prefiere, aunque la diferencia en los *p-valores* son a menudo pequeños.

2.4.3. Métodos no paramétricos

Los métodos no paramétricos son llamados así, debido a que no se necesita estimar parámetros tal como la media o la desviación estándar de alguna distribución asumida. En vez de ésto, se usa la posición relativa (rangos) de los datos, una reflexión de los percentiles de los datos. Debido a que esos métodos no requieren de un supuesto respecto a alguna distribución de los datos, son llamados algunas veces como métodos de distribución libre. Los métodos no paramétricos han sido una alternativa útil para el análisis de datos censurados; se sabe que los datos no detectados son menores que sus límites de detección, así que tendrán rangos menores. Estos métodos no requieren estimadores de parámetros o de medidas de centralidad o dispersión, pero si requieren de un orden relativo respecto a las distancias entre los datos detectados y no detectados.

Capítulo 3

Distribuciones de variables aleatorias continuas

3.1. Distribuciones de localización-escala

Una variable aleatoria Y pertenece a una familia de distribuciones de localización-escala, si su función de distribución puede ser expresada como

$$P(Y \leq y) = F(y; \mu, \sigma) = \Phi\left(\frac{y-\mu}{\sigma}\right),$$

donde Φ no depende de ningún parámetro desconocido. Aquí se dice que μ es parámetro de localización ($\mu \in \mathbf{R}$) y σ parámetro de escala ($\sigma > 0$). La expresión de arriba muestra que Φ es la función de distribución acumulada de Y cuando $\mu = 0$ y $\sigma = 1$. De igual modo, Φ es la función de distribución acumulada de $(y - \mu)/\sigma$. En particular, la importancia de las distribuciones de localización-escala estriba en lo siguiente:

- Muchos de los modelos de probabilidad que son ampliamente utilizados son distribuciones de localización-escala o similares a éstas. Estas distribuciones incluyen la distribucioes exponencial, normal, Weibull, lognormal, loglogística, logística y de valores extremo.

- La teoría y métodos para la inferencia, utilizando distribuciones de localización-escala, son relativamente simples.

Si en la expresión de arriba, Φ depende de uno o más parámetros desconocidos, entonces Y no es un miembro de la familia de localización-escala; sin embargo, la estructura y notación de familia de localización-escala todavía podrá ser útil.

3.2. Modelos continuos de probabilidad

Definición 3.2.1. Función de densidad de probabilidad . *Si Y es una variable aleatoria continua, entonces la función de densidad de probabilidad de Y o bien, la densidad de Y, es una función no negativa $f_Y(y)$ tal que para dos números cualesquiera, a y b, con $a \leq b$,*

$$P(a \leq Y \leq b) \quad = \quad \int_a^b f_Y(y)dy.$$

En el límite cuando $a \longrightarrow -\infty$, $b \longrightarrow \infty$, la integral anterior vale 1. Así, la probabilidad de que Y tome valores en el intervalo $[a, b]$ es calculada integrando la densidad de Y sobre este intervalo.

Definición 3.2.2. Función de distribución acumulada. *La Función de distribución acumulada de una variable aleatoria Y o bien, la distribución de Y, se define como*

$$F(y) = P(Y \leq y).$$

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

$F(y)$ es una función continua, no negativa y monótona no decreciente en los reales, tal que $F(y) \underset{y \to -\infty}{\longrightarrow} 0$ y $F(y) \underset{y \to \infty}{\longrightarrow} 1$.

Ahora, si $f(y)$ es la densidad de Y, entonces su distribución estará dada por

$$F(y) = P(Y \leq y) = \int_{-\infty}^{y} f(t)dt.$$

3.3. La función cuantil

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

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

Formalmente, el *p-ésimo* cuantil está definido como sigue

$$y_p = F_Y^{-1}(p) := \underset{y \in \mathbb{R}_Y}{inf} \{y : F_Y(y) \geq p\}, \quad \text{para} \quad p \in (0, 1).$$

De lo anterior, el $\inf_{y \in \mathbb{R}_Y}$ es solo un mínimo ideal y $\mathbb{R}_Y$ denota el rango de valores de y como un subconjunto de $\mathbb{R}$, el cual puede ser un subconjunto discreto o un continuo[48].

Esta definición sugiere que en el caso donde la función de distribución acumalada es continua y estrictamente creciente, y_p es única y está definida por

$$F(y_p) \quad = \quad p.$$

El valor de p es conocido como el *p-ésimo percentil* y el valor y_p el cuantil correspondiente.

3.4. Algunos modelos importantes

Al enfrentar problemas provenientes de estudios ambientales que involucran datos no detectados, la literatura indica una revisión de algunos modelos importantes que consideran censura, sesgo y asimetría en los datos y sus distribuciones de probabilidad. Entre ellos, los modelos exponencial, lognormal y Weibull resultan ser los primeros.

Antes de iniciar a revisar las distribuciones, veamos brevemente los parámetros de este tipo de familias.

3.4.1. Parámetros de las distribuciones

En estos estudios, los parámetros de la distribución de probabilidad de la muestra observada juegan un papel importante para determinar la distribución poblacional, ya que un error en la estimación de los parámetros conlleva a una indeterminación de los modelos de probabilidad asumidos. De esta forma es necesario conocer y analizar los parámetros de la distribución propuesta.

En las distribuciones de localización-escala los parámetros definen características importantes de los modelos. Así los parámetros se clasifican como sigue:

- **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.

- **Asimetría.** Este tipo de parámetro se relaciona con la forma del histograma y la densidad del modelo de probabilidad en cuestión.

3.4.2. Distribución exponencial

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 [26].

La distribución exponencial es una distribución muy utilizada para modelar tiempos de vida y tiempos a la falla en estudios de supervivencia y confiabilidad de componentes electrónicos [35]. En estudios ambientales empieza a cobrar relevancia por la aparición de muestras aleatorias de datos ambientales censuradas.

Definición 3.4.1. *Se dice que una variable aleatoria Y tiene distribución exponencial con parámetros θ, γ; denotado por $Y \sim exp(\theta, \gamma)$, si su densidad es*

$$f(y;\theta,\gamma) = \frac{1}{\theta} exp\left(-\frac{y-\gamma}{\theta}\right), \quad \theta > 0, \quad y > \gamma. \tag{3.1}$$

Con lo cual su distribución de probabilidad $F(y)$ se obtiene integrando $f(y;\theta,\gamma)$

$$F(y;\theta,\gamma) = 1 - exp\left(-\frac{y-\gamma}{\theta}\right), \quad \theta > 0, \quad y > \gamma, \tag{3.2}$$

donde θ es el parámetro de escala y γ su parámetro de localización.

A partir de aquí y para mayor simplicidad, denotaremos esta distribución y su densidad como $F(y)$ y $f(y)$, respectivamente.

Si $\gamma = 0$, (3.1) y (3.2) se convierten en la densidad y la distribución de un modelo exponencial uniparamétrico frecuentemente utilizado; denotándose como $Y \sim exp(\theta)$. La Figura 3.1 muestra las gráficas de la función de densidad de este modelo cuando $\theta = 0.5, 1.5, 2.0$.

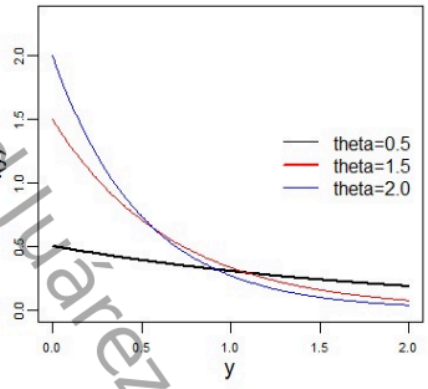


Figura 3.1: Densidades del modelo exponencial para $\gamma = 0$ y $\theta = 0.5, 1.5, 2.0$.

El cuantil y_p para la distribución exponencial se obtiene a partir de (3.2) como sigue:

$$\begin{aligned} 1 - exp\left(-\frac{y_p - \gamma}{\theta}\right) &= p \\ -\frac{y_p - \gamma}{\theta} &= ln(1-p) \\ y_p &= \gamma - \theta ln(1-p). \end{aligned} \tag{3.3}$$

3.4.3. Distribución lognormal

La distribución lognormal es llamada en algunas ocasiones distribución **antilognormal** [26], debido a que no es la distribución del logaritmo de una variable aleatoria normal sino de una exponencial; sin embargo, el nombre "lognormal" es más utilizado en diversas áreas y campos del conocimiento.

La distribución lognormal es un modelo muy utilizado para estudios de tiempos de vida y tiempos a la falla. En estudios ambientales cobra relevancia debido a su relación con el modelo normal, ya por todos conocidos [35].

Definición 3.4.2. *Se dice que una variable aletoria Y sigue una distribución lognormal con parámetros μ, σ; indicado por $Y \sim LN(\mu, \sigma)$, si la densidad*

es

$$\begin{aligned} f(y;\mu,\sigma) &= \frac{1}{\sigma y}\phi_{nor}\left(\frac{ln\ y-\mu}{\sigma}\right) \\ &= \frac{1}{\sqrt{2\pi}\sigma y}exp\left[-\frac{1}{2}\left(\frac{ln\ y-\mu}{\sigma}\right)^2\right], \quad y>0. \end{aligned} \tag{3.4}$$

Así, su distribución es

$$\begin{aligned} F(y;\mu,\sigma) &= \Phi_{nor}\left(\frac{ln\ y-\mu}{\sigma}\right) \\ &= \int_0^y \frac{1}{\sqrt{2\pi}\sigma\nu}exp\left[-\frac{1}{2}\left(\frac{ln\ \nu-\mu}{\sigma}\right)^2\right]d\nu, \quad y>0, \end{aligned} \tag{3.5}$$

donde ϕ_{nor} y Φ_{nor} son las funciones de densidad y de distribución de una variable aleatoria normal estándar, respectivamente.

En lo sucesivo se denotará la función de distribución Φ_{nor} y su densidad ϕ_{nor} como Φ y ϕ, respectivamente (esta notación se redefinirá en función del modelo de probabilidad que se esté abordando). La Figura 3.2 muestra los gráficos de las densidades una variable aleatoria lognormal para $\mu = 0$ y $\sigma = 0.25, 0.5, 1.5$.

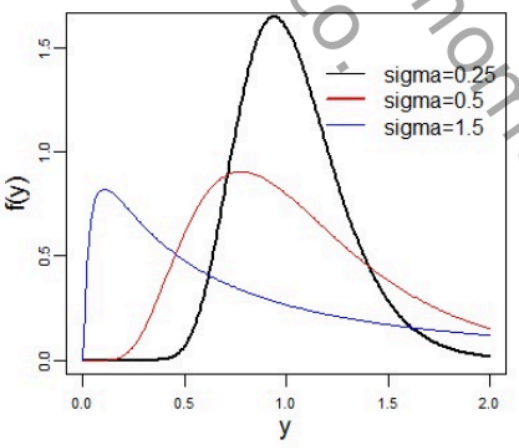


Figura 3.2: Densidades del modelo lognormal para $\mu = 0$ y $\sigma = 0.25, 0.5, 1.5$.

La función cuantil de la distribución lognormal se obtiene fácilmente a

partir de (3.5), observe

$$\begin{aligned} \Phi\left(\frac{ln\, y_p - \mu}{\sigma}\right) &= p \\ ln\, y_p &= \mu + \sigma\Phi^{-1}(p) \\ y_p &= exp\left[\mu + \sigma\Phi^{-1}(p)\right]. \end{aligned} \tag{3.6}$$

3.4.4. Distribución Weibull

La distribución Weibull fue nombrada en honor al físico sueco Waloddi Weibull, quien en 1939 la utilizó para representar la distribución de la resistencia de materiales, y en 1951 extendió su uso para una amplia variedad de aplicaciones. La relación estrecha, que Weibull demostró entre los datos observados y sus predichos con este modelo fue bastante impresionante [26].

Definición 3.4.3. *Una variable aleatoria Y sigue una distribución Weibull con parámetros β, η; expresado comúnmente por $Y \sim Weib(\beta, \eta)$, si la densidad es*

$$f(y;\eta,\beta) = \frac{\beta}{\eta}\left(\frac{y}{\eta}\right)^{\beta-1} exp\left[-\left(\frac{y}{\eta}\right)^{\beta}\right], \quad y > 0. \tag{3.7}$$

Así, su distribución es

$$F(y;\eta,\beta) = 1 - exp\left[-\left(\frac{y}{\eta}\right)^{\beta}\right], \quad y > 0. \tag{3.8}$$

En esta expresión $\beta > 0$ es un parámetro de forma y $\eta > 0$ es un parámetro de escala.

Es conveniente utilizar una parametrización alternativa para la distribución Weibull, la cual está basada en la relación entre la distribución Weibull y el modelo de Valor Extremo Mínimo (SEV por sus siglas en inglés). En particular, si una variable aleatoria Y sigue una distribución Weibull, entonces $X = ln\, Y \sim SEV(\mu, \sigma)$, donde $\sigma = 1/\beta$ es el parámetro de escala y $\mu = ln\, \eta$ es el parámetro de localización. De este modo, si Y tiene una distribución Weibull, esta se denotará por $Y \sim Weib(\mu, \sigma)$. Por lo tanto las funciones de distribución y de densidad del modelo Weibull dadas en (3.7) y (3.8) pueden reescribirse de la siguiente manera

$$f(y;\eta,\beta) = \frac{1}{\sigma y}\phi_{sev}\left[\frac{ln\, y - \mu}{\sigma}\right], \quad y > 0, \tag{3.9}$$

$$F(y;\eta,\beta) = \Phi_{sev}\left[\frac{ln\, y - \mu}{\sigma}\right], \quad y > 0, \tag{3.10}$$

donde $\Phi_{sev}(z) = 1 - exp[-exp(z)]$ y $\phi_{sev}(z) = exp[z - exp(z)]$ son la distribución y la densidad, respectivamente, de una variable aleatoria SEV ($\mu = 0,\ \sigma = 1$).

Consecuentemente se denotará las funciones de distribución y de densidad de la distribución weibull como ${}^{w}\Phi$ y ${}^{w}\phi$, respectivamente. La Figura 3.3 muestra las gráficas de algunas funciones de densidad para distintos valores del parámetro de forma, β.

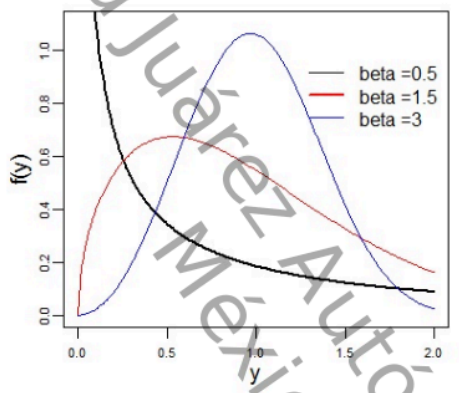


Figura 3.3: Densidades del modelo Weibull para $\eta = 1$ y $\beta = 0.5, 1.5, 3.0$.

El cuantil del modelo Weibull es $y_p = \eta[-ln(1-p)]^{1/\beta}$. Utilizando la estandarización expresada en (3.10) el cuantil y_p se obtiene como sigue

$$\begin{aligned} {}^{w}\Phi\left(\frac{ln\ y_p - \mu}{\sigma}\right) &= p \\ ln\ y_p &= \mu + \sigma\ {}^{w}\Phi^{-1}(p) \\ y_p &= exp\left[\mu + \sigma\ {}^{w}\Phi^{-1}(p)\right]. \end{aligned} \tag{3.11}$$

Puede observarse una estructura similar al cuantil obtenido para la distribución lognormal, la diferencia radica en la elección de la distribución asumida.

3.5. Momentos de una variable aleatoria

3.5.1. Valor esperado

El valor esperado de una variable aleatoria puede entenderse como una medida de centralidad, ponderando los valores de esta variable de acuerdo a su distribución de probabilidad, esperándose obtener una cantidad que resuma el valor típico o esperado para una variable aleatoria.

Definición 3.5.1. *Sea Y una variable aleatoria continua con función de densidad $f(y)$. La* **media** *o* **valor esperado** *de una variable aleatoria $g(Y)$ está dada por*

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

Teorema 3.5.1. *Sea Y una variable aleatoria y sean a, b y c constantes. Entonces para cualesquiera función $g_1(Y)$ y $g_2(Y)$ cuyos valores esperados existen, se tiene*

a. $E[ag_1(Y) + bg_2(Y) + c] = aE[g_1(Y)] + bE[g_2(Y)] + c.$

b. *Si $g_1(y) \geq 0$ para toda y, entonces $E[g_1(Y)] \geq 0$.*

c. *Si $g_1(y) \geq g_2(y)$ para toda y, entonces $E[g_1(Y)] \geq E[g_2(Y)]$.*

d. *Si $a \leq g_1(y) \leq b$ para toda y, entonces $a \leq E[g_1(Y)] \leq b$.*

3.5.2. Momentos y varianza

Los diversos momentos de una variable aleatoria son una clase importante de valores esperados.

Definición 3.5.2. *Sea Y una variable aleatoria, dado un m entero positivo, el m-***ésimo momento de** *Y es:*

$$\mu_m' = E[Y^m].$$

*El m-***ésimo momento central de** *Y, μ_m, es*

$$\mu_m = E[(Y - \mu)^m],$$

donde $\mu = \mu_1' = E[Y]$.

En orden de importancia y después del valor esperado $E[Y]$; el momento más importante es el segundo momento central, mejor conocido como varianza.

Definición 3.5.3. *La* **varianza** *de una variable aleatoria Y coincide con su segundo momento central, $Var[Y] = E[(Y - E[Y])^2]$. La raíz cuadrada positiva de $Var[Y]$ se conoce como la* **desviación estándar** *de Y.*

Teorema 3.5.2. *Si Y es una variable aleatoria con varianza finita, entonces para cualesquiera constantes a y b,*

$$Var[aY + b] = a^2 Var[Y].$$

Demostración. De la definición y el Teorema 3.5.1.a., se tiene

$$\begin{aligned} Var[aY + b] &= E[((aY + b) - E[aY + b])^2] \\ &= E[(aY - aE[Y])^2] \\ &= a^2 E[(Y - E[Y])^2] \\ &= a^2 Var[Y]. \end{aligned}$$

□

Una expresión muy útil para la definición de la varianza es

$$Var[Y] = E[Y^2] - (E[Y])^2, \tag{3.12}$$

la cual puede ser demostrada fácilmente como sigue

$$\begin{aligned} Var[Y] &= E[(Y - E[Y])^2] = E[Y^2] - 2(E[Y])^2 + (E[Y])^2 \\ &= E[Y^2] - (E[Y])^2, \end{aligned}$$

usando el hecho de que $E[YE[Y]] = (E[Y])^2$, dado que $E[Y]$ es una constante.

3.5.3. Momentos de una variable aleatoria exponencial

Teorema 3.5.3. *Sea Y una variable aleatoria exponencial con parámetros θ y γ, y considere m un entero positivo, entonces*

$$E[(Y - \gamma)^m] = m!\theta^m.$$

Demostración.

$$\begin{aligned} E[(Y-\gamma)^m] &= \int_\gamma^\infty \frac{(y-\gamma)^m}{\theta} exp\left(-\frac{y-\gamma}{\theta}\right) dy \\ &= \theta^{m-1}\int_\gamma^\infty \left(\frac{y-\gamma}{\theta}\right)^m exp\left(-\frac{y-\gamma}{\theta}\right) dy. \end{aligned}$$

Haciendo el cambio de variable $u = (y-\gamma)/\theta$ se tiene

$$\begin{aligned} E[(Y-\gamma)^m] &= \theta^m \int_0^\infty u^m e^{-u} du \\ &= -\theta^m (P_m^m + P_{m-1}^m u + \ldots + P_1^m u^{m-1} + P_0^m u^m) e^{-u}\Big|_0^\infty \end{aligned}$$

donde $P_n^m = \frac{m!}{(m-n)!}$. A partir del segundo término, los valores son nulos al evaluarlos en el intervalo dado, por lo tanto

$$\begin{aligned} E[(Y-\gamma)^m] &= -\theta^m P_m^m e^{-u}\Big|_0^\infty \\ &= P_m^m \theta^m \\ &= m!\theta^m. \end{aligned}$$

□

Corolario 3.5.4. *Si $Y \sim exp(\theta, \gamma)$, entonces*

$$E[Y] = \theta + \gamma; \qquad Var[Y] = \theta^2.$$

3.5.4. Momentos de una variable aleatoria lognormal

Teorema 3.5.4. *Sea Y una variable aleatoria lognormal con parámetros μ y σ, y considere m un entero positivo, entonces*

$$E[Y^m] = exp\left[m\mu + \tfrac{m^2\sigma^2}{2}\right].$$

Demostración. De la Definición 3.5.2 el m-ésimo momento de Y es

$$\begin{aligned} E[Y^m] &= \int_0^\infty y^m \frac{1}{\sqrt{2\pi}y\sigma} exp\left[-\frac{1}{2}\left(\frac{ln\, y-\mu}{\sigma}\right)^2\right] dy \\ &= \int_0^\infty \frac{1}{\sqrt{2\pi}y\sigma} exp\,(mln\, y)\, exp\left[-\frac{1}{2}\left(\frac{ln\, y-\mu}{\sigma}\right)^2\right] dy \\ &= \int_0^\infty \frac{1}{\sqrt{2\pi}y\sigma} exp\left[-\frac{(ln\, y)^2 - 2ln\, y(\sigma^2 m+\mu)+\mu^2}{2\sigma^2}\right] dy. \end{aligned}$$

completando el cuadrado en el argumento de la exponencial

$$(ln\ y)^2 - 2ln\ y(\sigma^2 m + \mu) + \mu^2 \;=\; [ln\ y - (\sigma^2 m + \mu)]^2 - \sigma^4 m^2 - 2m\mu\sigma^2,$$

con lo que,

$$\begin{aligned} E[Y^m] &= \int_0^\infty \frac{1}{\sqrt{2\pi}y\sigma} exp\left[-\frac{[ln\ y - (\sigma^2 m + \mu)]^2 - \sigma^4 m^2 - 2m\mu\sigma^2}{2\sigma^2}\right] dy \\ &= exp\left[m\mu + \frac{m^2\sigma^2}{2}\right] \int_0^\infty \frac{1}{\sqrt{2\pi}y\sigma} exp\left[-\frac{1}{2}\left(\frac{ln\ y - (\sigma^2 m + \mu)}{\sigma}\right)^2\right] dy. \end{aligned}$$

Ahora, sea $w = [ln\ y - (\sigma^2 m + \mu)]/\sigma$, entonces $dy = (1/y\sigma)dw$, y

$$\begin{aligned} E[Y^m] &= exp\left[m\mu + \frac{m^2\sigma^2}{2}\right] \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-w^2/2} dw \\ &= exp\left[m\mu + \frac{m^2\sigma^2}{2}\right], \end{aligned}$$

dado que la última integral representa el área bajo una curva de densidad normal estándar y de aquí que sea igual a 1.

□

Corolario 3.5.5. *Si $Y \sim LN(\mu, \sigma)$, entonces*

$$E[Y] = e^{\mu+\sigma^2/2}; \qquad Var[Y] = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1).$$

3.5.5. Momentos de una variable aleatoria Weibull

Teorema 3.5.5. *Sea Y una variable aleatoria Weibull con parámetros β y η, y considere m un entero positivo, entonces el $m-$ésimo momento de Y es*

$$E(Y^m) = \eta^m \Gamma\left(1 + \tfrac{m}{\beta}\right).$$

Demostración. De la Definición 3.4.3, el $m-$ésimo momento de Y es

$$\begin{aligned} E[Y^m] &= \int_0^\infty y^m \frac{\beta}{\eta}\left(\frac{y}{\eta}\right)^{\beta-1} exp\left[-\left(\frac{y}{\eta}\right)^\beta\right] dy \\ &= \eta^m \int_0^\infty \frac{\beta}{\eta}\left(\frac{y}{\eta}\right)^{\beta-1}\left(\frac{y}{\eta}\right)^m exp\left[-\left(\frac{y}{\eta}\right)^\beta\right] dy. \end{aligned}$$

Si $z = \left(\frac{y}{\eta}\right)^{\beta}$, entonces $dz = \frac{\beta}{\eta}\left(\frac{y}{\eta}\right)^{\beta-1} dy$ y tomando $z^{1/\beta} = \frac{y}{\eta}$, se tiene

$$\begin{aligned} E[Y^m] &= \eta^m \int_0^{\infty} z^{m/\beta} e^{-z} dz \\ &= \eta^m \int_0^{\infty} z^{(1+m/\beta)-1} e^{-z} dz \\ &= \eta^m \Gamma\left(1 + \frac{m}{\beta}\right), \end{aligned}$$

donde $\Gamma(k) = \int_0^{\infty} z^{k-1} e^{-z} dz$ corresponde a la función gamma evaluada en k. □

Corolario 3.5.6. *Si $Y \sim Weib(\beta, \eta)$, entonces*

$$E(Y) = \eta\Gamma\left(1 + \tfrac{1}{\beta}\right); \qquad Var(Y) = \eta^2\left[\Gamma\left(1 + \tfrac{2}{\beta}\right) - \Gamma^2\left(1 + \tfrac{1}{\beta}\right)\right].$$

Capítulo 4

Estimación por máxima verosimilitud

El concepto de verosimilitud es crucial para la inferencia estadística en el paradigma de Fisher, quien fue el primero en delinearlo de manera específica y unívoca en 1912. Fisher se dedicó a estudiar sus propiedades y conceptos asociados, como "información" o "suficiencia", aumentando con ello la producción en Estadística Matemática durante varias décadas. Esencialmente, la *verosimilitud* es un indicador de la coherencia entre los modelos estadísticos elegidos para analizar los datos y las observaciones. Da una idea de cuán *verosímiles* o plausibles son las observaciones obtenidas, es decir, mide la "probabilidad" de re-observar la muestra en una hipotética repetición del experimento.

4.1. Función de verosimilitud

Sea $Y = (Y_1, \ldots, Y_n)$ un vector aleatorio, con densidad conjunta dada por $f(y;\boldsymbol{\theta})$, siendo $\boldsymbol{\theta}$ un vector de parámetros que toma valores en un conjunto $\boldsymbol{\Theta} \subseteq \mathbb{R}^d$. La *función de verosimilitud* o *la verosimilitud* es una función de $\boldsymbol{\theta}$ para cada valor muestral $y = (y_1, y_2, \ldots, y_n)$,

$$L(\boldsymbol{\theta}) = L(y;\boldsymbol{\theta}) \propto f(y;\boldsymbol{\theta}).$$

La verosimilitud brinda la información resumida acerca de $\boldsymbol{\theta}$, basada en el modelo estadístico y en los datos observados. Usualmente es conveniente trabajar con el logaritmo natural de $L(\boldsymbol{\theta})$, denominada la *función de log-verosimilitud* o simplemente *la log-verosimilitud*,

$$\ell(\boldsymbol{\theta}) = \ell(y;\boldsymbol{\theta}) = ln[L(y;\boldsymbol{\theta})].$$

En particular, si el vector de observaciones se obtiene a través de un muestreo aleatorio simple de una densidad univariada $f(y;\boldsymbol{\theta})$, entonces:

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

Convenientemente se asumirá que la log-verosimilitud es regular, en el sentido que tiene derivadas parciales respecto de $\boldsymbol{\theta}$ y sus momentos son finitos. Se denotará a las derivadas de $\ell(\boldsymbol{\theta})$ respecto de una componente r de $\boldsymbol{\theta}$ mediante $\ell_r(y;\boldsymbol{\theta})$; por ejemplo, $\ell_{rr}(y;\boldsymbol{\theta})$ denotará la segunda derivada de ℓ con respecto a $\boldsymbol{\theta}$. Las dos primeras derivadas de $\ell(\boldsymbol{\theta})$ constituyen los componentes básicos en este enfoque de inferencia estadística.

Definición 4.1.1. *La matriz de información de Fisher, también denominada Matriz de Información Esperada, es definida de la siguiente manera*

$$\boldsymbol{I_\theta} = E\left[-\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right]. \tag{4.1}$$

$\boldsymbol{I_\theta}$ es conocido 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 $\boldsymbol{\theta}$

$$\hat{\boldsymbol{I}}_{\boldsymbol{\theta}} = -\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \tag{4.2}$$

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

Definición 4.1.2. *El Estimador de Máxima Verosimilitud (EMV) de $\boldsymbol{\theta}$, denotado por $\hat{\boldsymbol{\theta}}$, se define como el valor que maximiza la verosimilitud $L(\boldsymbol{\theta})$ o equivalentemente la log-verosimilitud $\ell(\boldsymbol{\theta})$. Es decir, $\hat{\boldsymbol{\theta}}$ es un valor en el espacio de parámetros $\boldsymbol{\Theta} \subseteq \mathbb{R}^d$, para el cual $L(\hat{\boldsymbol{\theta}}) = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{Sup}\, L(\boldsymbol{\theta})$.*

Si $\boldsymbol{\Theta}$ es un conjunto abierto, el modelo es suficientemente regular y el EMV $\hat{\boldsymbol{\theta}}$ es finito, entonces puede obtenerse como solución de las *ecuaciones de verosimilitud*, $\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0$. Frecuentemente resulta imposible resolver de manera analítica este sistema de ecuaciones, lo que ha dado origen a una amplia variedad de métodos numéricos, como los métodos basados en Newton-Raphson para aproximar los EMV satisfactoriamente (Ver apéndice B).

4.2. Estimación para un parámetro

Suponga que el modelo de probabilidad ha sido formulado mediante un experimento, y que se trata de un solo parámetro desconocido θ. Se desea utilizar los datos obtenidos experimentalmente para estimar el valor de θ. Sin pérdida de generalidad y abordando aquí el caso discreto, el objetivo de la estimación es determinar cual de los posibles valores de θ son plausibles o probables a la luz de las observaciones.

Los datos observados pueden considerarse como un evento E en el espacio muestral para el modelo de probabilidad. Entonces la probabilidad del evento E puede ser determinado a través del modelo, y en general será una función del parámetro desconocido, $P(E;\theta)$. El EMV de θ es el valor de θ que maximiza la $P(E;\theta)$, denotado por $\hat{\theta}$. Es el valor del parámetro que mejor explica los datos en el sentido de que maximiza la probabilidad de E bajo el modelo.

4.2.1. Funciones de verosimilitud y log-verosimilitud

De acuerdo a [27] *La verosimilitud* de θ puede definirse como sigue

$$L(\theta) = c \cdot P(E;\theta). \tag{4.3}$$

Aquí c es cualquier constante positiva con respecto de θ; esto es, c no depende de θ, aunque pueda ser una función de los datos. Se elige c para obtener una expresión simple para $L(\theta)$.

Por lo general $P(E;\theta)$ y $L(\theta)$ son expresiones confirmados por el producto de expresiones y términos conocidos, por lo que resulta conveniente trabajar con una transformación continua y positiva de $L(\theta)$, a saber *la log-verosimilitud* de L:

$$\ell(\theta) = ln[L(\theta)]. \tag{4.4}$$

Note que, por (4.3),

$$\ell(\theta) = c' + ln[P(E;\theta)],$$

donde $c' = ln\, c$ no es una función de θ.

El valor de θ que maximiza $P(E;\theta)$ también maximizará $L(\theta)$ y $\ell(\theta)$. Así el EMV $\hat{\theta}$ es el valor de θ que maximiza la verosimilitud y la log-verosimilitud. Generalmente es más fácil trabajar con la log-verosimilitud.

4.2.2. Función de información

Para obtener $\hat{\theta}$ es necesario localizar el máximo de $\ell(\theta)$ sobre todos los posibles valores de θ. Por lo general, se realiza mediante la derivación de $\ell(\theta)$ con respecto a θ, fijando la derivada igual a cero y despejando θ. Es posible que este procedimiento pueda producir un mínimo relativo o un punto de inflexión en lugar del máximo deseado, así que es necesario verificar que un máximo ha sido encontrado, comprobando que la segunda derivada es negativa.

La *función de información* $I(\theta)$ se define como el negativo de la segunda derivada de la log-verosimilitud con respecto de θ

$$I(\theta) = -\frac{d^2\ell(\theta)}{d\theta^2}. \tag{4.5}$$

El conjunto Ω de valores posibles de θ es llamado el *espacio paramétrico*. Usualmente Ω es un intervalo de valores reales, y la primera y segunda derivada de $\ell(\theta)$ existen en todos los puntos interiores de Ω. Entonces, si $\hat{\theta}$ es un punto interior de Ω, la primera derivada será cero y la segunda derivada será negativa en $\theta = \hat{\theta}$. Así bajo estas condiciones se tiene

$$\left.\frac{d\ell(\theta)}{d\theta}\right|_{\theta=\hat{\theta}} = 0; \qquad I(\hat{\theta}) > 0. \tag{4.6}$$

Para encontrar $\hat{\theta}$, se determina la raíz de la *ecuación de máxima verosimilitud* $\frac{d\ell(\theta)}{d\theta} = 0$., verificándose, mediante el signo de $I(\hat{\theta})$, que en efecto un máximo relativo ha sido encontrado.

En algunos casos, la ecuación $\frac{d\ell(\theta)}{d\theta} = 0$ puede resolverse algebraicamente produciendo una expresión simple para $\hat{\theta}$. En situaciones más complicadas, será necesario recurrir a algoritmos y métodos numéricos.

En el siguiente ejemplo, se ilustra como se obtiene el EMV para una distribución exponencial con un parámetro desde una muestra censurada por la derecha. Aunque el objetivo de este trabajo no aborda el fenómeno de censura por la derecha, este ejemplo pretende mostrar el proceso y las condiciones que un EMV debe cumplir, satisfaciendo (4.6).

Ejemplo 4.2.1. *Suponga que $y_1, y_2, \ldots, y_m, y_1, y_2, \ldots, y_{n-m}$ son observaciones de una muestra aleatoria de una población con distribución exponencial con parámetro θ, donde m observaciones (representadas por y_i) no están censuradas y $n-m$ observaciones (representadas por y_j) están censuradas por*

la derecha.

Se tiene que $Y \sim exp(\theta)$*, entonces su densidad y la función de distribución de* Y *son*

$$f(y) = \frac{1}{\theta}e^{-y/\theta}, \qquad F(y) = 1 - e^{-y/\theta}, \qquad y > 0,$$

respectivamente. Luego la verosimilitud de θ *se construye así*

$$L(\theta) = c \cdot \prod_{i=1}^{m} f(y_i) \prod_{j=1}^{n-m} P(Y > y_j) = c \cdot \prod_{i=1}^{m} f(y_i) \prod_{j=1}^{n-m} [1 - F(y_j)].$$

Sustituyendo $f(y_i)$ *y* $F(y_j)$ *la verosimilitud resulta en*

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{m} \left(\frac{1}{\theta}e^{-y_i/\theta}\right) \prod_{j=1}^{n-m} e^{-y_j/\theta} \\ &= \frac{1}{\theta^m} exp\left(-\frac{1}{\theta}\sum_{i=1}^{m} y_i\right) exp\left(-\frac{1}{\theta}\sum_{j=1}^{n-m} y_j\right) \\ &= \frac{1}{\theta^m} exp\left[-\frac{1}{\theta}\left(\sum_{i=1}^{m} y_i + \sum_{j=1}^{n-m} y_j\right)\right] \\ &= \frac{1}{\theta^m} exp\left[-\frac{1}{\theta}\left(m\bar{y} + (n-m)\bar{y}\right)\right] \end{aligned}$$

donde $\bar{y} = \frac{1}{m}\sum_{i=1}^{m} y_i$ *e* $\bar{y} = \frac{1}{n-m}\sum_{j=1}^{n-m} y_j$*. Con lo que la log-verosimilitud es*

$$\ell(\theta) = -mln(\theta) - \frac{1}{\theta}\left[m\bar{y} + (n-m)\bar{y}\right]$$

y la primera derivada de ℓ *con respecto de* θ *es:*

$$\frac{d\ell(\theta)}{d\theta} = -\frac{m}{\theta}\left[1 - \frac{m\bar{y} + (n-m)\bar{y}}{m\theta}\right].$$

Resolviendo algebraicamente $d\ell(\theta)/d\theta = 0$ *se obtiene el EMV*

$$\hat{\theta} = \frac{m\bar{y} + (n-m)\bar{y}}{m}.$$

Ahora, para demostrar que $\hat{\theta}$ es un máximo, se analizan las expresiones dadas en (4.6). Entonces como

$$\frac{d\ell(\theta)}{d\theta} = -\frac{m}{\theta}\left[1 - \frac{m\bar{y} + (n-m)\bar{y}}{m\theta}\right]$$

y de igual modo

$$I(\theta) = -\frac{d^2\ell(\theta)}{d\theta^2} = -\frac{m}{\theta^2}\left[1 - \frac{2\left(m\bar{y} + (n-m)\bar{y}\right)}{m\theta}\right]$$

se tiene que

$$\frac{d\ell(\theta)}{d\theta}\Big|_{\theta=\hat{\theta}} = -\frac{m}{\hat{\theta}}\left(1 - \frac{\hat{\theta}}{\hat{\theta}}\right) = 0$$

e

$$I(\hat{\theta}) = -\frac{m}{\hat{\theta}^2}\left(1 - \frac{2\hat{\theta}}{\hat{\theta}}\right) = -\frac{m}{\hat{\theta}}(-1) > 0.$$

Por lo tanto, con lo anterior se concluye que $\hat{\theta}$ es el EMV para el modelo exponencial con un parámetro.

4.3. Estimación para dos parámetros

Suponga que el modelo de probabilidad para un experimento involucra dos parámetros desconocidos, θ_1 y θ_2. La probabilidad de los datos en el evento E será una función de θ_1 y θ_2, y la verosimilitud conjunta será proporcional a esta probabilidad:

$$L(\theta_1, \theta_2) = c \cdot P(E; \theta_1, \theta_2)$$

donde c es positivo y no depende de θ_1 y θ_2. La log-verosimilitud de $L(\theta_1, \theta_2)$ será denotado por $\ell = \ell(\theta_1, \theta_2)$.

El estimador de máxima verosimilitud de (θ_1, θ_2) es el par de valores paramétricos $(\hat{\theta}_1, \hat{\theta}_2)$ que maximiza la probabilidad de los datos. Equivalentemente, $(\hat{\theta}_1, \hat{\theta}_2)$ maximizan $L(\theta_1, \theta_2)$ y $\ell(\theta_1, \theta_2)$.

Ahora, para encontrar $(\hat{\theta}_1, \hat{\theta}_2)$, se resuelve el sistema de ecuaciones simultáneas

$$\frac{\partial \ell(\theta_1,\theta_2)}{\partial \theta_1}=0; \qquad \frac{\partial \ell(\theta_1,\theta_2)}{\partial \theta_2}=0. \tag{4.7}$$

Entonces, la función de información coincide con la matriz de información de Fisher la cual es simétrica y se define de la siguiente manera

$$\boldsymbol{I}(\theta_1,\theta_2) = \begin{bmatrix} I_{11}(\theta_1,\theta_2) & I_{12}(\theta_1,\theta_2) \\ I_{21}(\theta_1,\theta_2) & I_{22}(\theta_1,\theta_2) \end{bmatrix} = \begin{bmatrix} -\dfrac{\partial^2 \ell(\theta_1,\theta_2)}{\partial \theta_1^2} & -\dfrac{\partial^2 \ell(\theta_1,\theta_2)}{\partial \theta_1 \partial \theta_2} \\ -\dfrac{\partial^2 \ell(\theta_1,\theta_2)}{\partial \theta_2 \partial \theta_1} & -\dfrac{\partial^2 \ell(\theta_1,\theta_2)}{\partial \theta_2^2} \end{bmatrix}. \tag{4.8}$$

Para un máximo relativo, la matriz $I(\hat{\theta}_1,\hat{\theta}_2)$ debe ser definida positiva; esto es,

$$\hat{I}_{11}>0; \quad \hat{I}_{22}>0; \quad \hat{I}_{11}\hat{I}_{22}-\hat{I}_{12}^2>0, \tag{4.9}$$

donde $\hat{I}_{ij}=I_{ij}(\hat{\theta}_1,\hat{\theta}_2)$.

Análogamente al Ejemplo 4.2.1, se obtendrá el EMV para una distribución exponencial con dos parámetros con una muestra aleatoria censurada por la derecha. Para ello es necesario recurrir antes a la definición de un estadístico de orden [8].

Definición 4.3.1. *Los estadísticos de orden de una muestra aleatoria $Y_1, Y_2, \ldots, Y_n$ dan los valores de la muestra en orden ascendente. Estos son denotados por $Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)}$.*

Los estadísticos de orden son variables aleatorias que satisfacen $Y_{(1)} \leq Y_{(2)} \leq \ldots \leq Y_{(n)}$. En particular

$$\begin{aligned} Y_{(1)} &= \min_{1\leq i\leq n} Y_i, \\ Y_{(2)} &= \text{el segundo mas pequeño de } Y_i, \\ &\vdots \\ Y_{(n)} &= \max_{1\leq i\leq n} Y_i. \end{aligned}$$

Con lo anterior se ilustrará el siguiente ejemplo que involucra la estimación de dos parámetros.

Ejemplo 4.3.2. *Ahora suponga que se tiene $y_1, y_2, \ldots, y_m, y_1, y_2, \ldots, y_{n-m}$, n observaciones de una muestra aleatoria de una población con distribución exponencial con parámetros θ y γ, donde m observaciones (representadas por*

y_i) no están censuradas y $n-m$ observaciones (representadas por y_j) están censuradas por la derecha.

Empleando (3.1) y (3.2) la verosimilitud para esta muestra aleatoria se define como sigue

$$\begin{aligned} L(\theta,\gamma) &= \prod_{i=1}^{m}\left(\frac{1}{\theta}e^{-(y_i-\gamma)/\theta}\right)\prod_{j=1}^{n-m}\left[e^{-(y_j-\gamma)/\theta}\right] \\ &= \frac{1}{\theta^m}exp\left[-\frac{1}{\theta}\left(\sum_{i=1}^{m}(y_i-\gamma)+\sum_{j=1}^{n-m}(y_j-\gamma)\right)\right] \\ &= \frac{1}{\theta^m}exp\left[-\frac{1}{\theta}\left[m\bar{y}-m\gamma+(n-m)\bar{y}-(n-m)\gamma\right]\right] \\ &= \frac{1}{\theta^m}exp\left[-\frac{1}{\theta}\left[m(\bar{y}-\bar{y})+n(\bar{y}-\gamma)\right]\right] \end{aligned}$$

de donde la log-verosimilitud es

$$\ell(\theta,\gamma) = -mln(\theta)-\frac{1}{\theta}\left[m(\bar{y}-\bar{y})+n(\bar{y}-\gamma)\right].$$

Para obtener $\hat{\theta}$ y $\hat{\gamma}$, serán necesarias obtener las primeras derivadas de ℓ con respecto a cada uno de ellos:

$$\begin{aligned} \frac{\partial\ell(\theta,\gamma)}{\partial\theta} &= -\frac{m}{\theta}\left[1-\frac{m(\bar{y}-\bar{y})+n(\bar{y}-\gamma)}{m\theta}\right] \\ \frac{\partial\ell(\theta,\gamma)}{\partial\gamma} &= \frac{n}{\theta}. \end{aligned}$$

Resolviendo el sistema de ecuaciones $\frac{\partial\ell(\theta,\gamma)}{\partial\theta}=0$ y $\frac{\partial\ell(\theta,\gamma)}{\partial\gamma}=0$ se obtiene que

$$\hat{\theta} = \frac{m(\bar{y}-\bar{y})+n(\bar{y}-\widehat{\gamma})}{m}. \tag{4.10}$$

Para obtener $\hat{\gamma}$ se recurre a los estadísticos de orden puesto que la solución a su derivada parcial no es analítica.

De esta manera si $y_{(1)}, y_{(2)}, \ldots, y_{(n)}$ son los lo valores observados de la muestra aleatoria, entonces se puede asignar $y_{(1)}$ como el EMV para γ, en virtud que es el valor muestral mas cercano a γ, esto es

$$\widehat{\gamma} = y_{(1)}. \tag{4.11}$$

De este modo, y sustituyendo en (4.10) los EMV para los parámetros son los siguientes

$$\begin{aligned} \hat{\theta} &= \frac{1}{m}\left[m(\bar{y}-\bar{y}) + n(\bar{y}-y_{(1)})\right] \\ \widehat{\gamma} &= y_{(1)}. \end{aligned}$$

El criterio que corrobora la obtención del máximo verosímil no puede ser equivalente al criterio de la segunda derivada en virtud de que la ecuación de arriba carece de este recurso analítico. Por ello se recurre a la definición básica de máximo el cual sólo requiere que en efecto todo el resto de los valores del rango generan un valor menor al máximo verosímil que se ha obtenido.

4.4. Estimación sobre una muestra aleatoria exponencial censurada por la izquierda

Considere que $Y \sim exp(\theta, \gamma)$, con funciones de distribución y densidad definidas por (3.1) y (3.2) respectivamente. Entonces para obtener los EMV de θ y γ a través de una muestra aleatoria censurada por la izquierda se empleará las reparametrizaciones $w = (y-\gamma)/\theta$ y $\varpi = (\mathrm{y}-\gamma)/\theta$ representando éstas las observaciones exactas y las observaciones censuradas respectivamente, esto es con la finalidad de ilustrar de forma más simple el proceso de estimación.

Ahora considere una muestra aleatoria $y_1, y_2, \ldots, y_m, y_1, y_2, \ldots, y_{n-m}$, donde m de ellos son observaciones exactas y $n-m$ resultan censuradas por la izquierda.

Por (3.1) y (3.2) se tiene que $F(y) = 1 - e^{-w}$ y $f(y) = \frac{1}{\theta}e^{-w}$ para $\theta > 0,\ y > \gamma$, con lo que la función de verosimilitud para esta muestra aleatoria censurada se construye como sigue

$$\begin{aligned} L(\theta,\gamma) &= \prod_{i=1}^{m} f(w_i) \prod_{j=1}^{n-m} F(\varpi_j) \\ &= \prod_{i=1}^{m} \frac{1}{\theta} e^{-w_i} \prod_{j=1}^{n-m} [1 - e^{-\varpi_j}]. \end{aligned}$$

Y de aquí la log-verosimilitud es la siguiente

$$\begin{aligned}\ell(\theta,\gamma) &= ln[L(\theta,\gamma)] \\ &= ln\left[\prod_{i=1}^{m}\frac{1}{\theta}e^{-w_i}\prod_{j=1}^{n-m}[1-e^{-\varpi_j}]\right] \\ &= -mln\theta-\sum_{i=1}^{m}w_i+\sum_{j=1}^{n-m}ln[1-e^{-\varpi_j}] \\ &= -m(ln\theta+\bar{w})+\sum_{j=1}^{n-m}ln[1-e^{-\varpi_j}].\end{aligned}$$

Luego los estimadores de los parámetros se obtendrán a través de la solución de las ecuaciones generadas por las primeras derivadas de $\ell(\theta,\gamma)$ respecto a cada parámetro e igualadas a cero.

$$\frac{\partial\ell(\theta,\gamma)}{\partial\theta} = -\frac{m(1-\bar{w})}{\theta}+\sum_{j=1}^{n-m}\frac{F_\theta(\varpi_j)}{F(\varpi_j)}, \tag{4.12}$$

$$\frac{\partial\ell(\theta,\gamma)}{\partial\gamma} = -\frac{m}{\theta}+\sum_{j=1}^{n-m}\frac{F_\gamma(\varpi_j)}{F(\varpi_j)}; \tag{4.13}$$

donde $F_\theta=\partial F/\partial\theta$ y $F_\gamma=\partial F/\partial\gamma$.

De lo anterior se observa que para este caso la solución al sistema de ecuaciones no resulta de manera analítica, ya que se tiene unos sistemas de ecuaciones no lineales. Un algoritmo basado en Newton-Raphson y codificado en R se propone en el apéndice B para resolver este problema. En el ejemplo siguiente se muestra su aplicación e implementación.

Ejemplo 4.4.1. *Estimación de los parámetros θ y γ para una muestra aleatoria censurada por la izquierda de una distribución exponencial. Retomando (4.12) y (4.13) se renombran las ecuaciones como sigue*

$$\begin{aligned}\varphi_1(\theta,\gamma) &= -\frac{m(1-\bar{w})}{\theta}+\sum_{j=1}^{n-m}\frac{F_\theta(\varpi_j)}{F(\varpi_j)} \\ \varphi_2(\theta,\gamma) &= \frac{m}{\theta}+\sum_{j=1}^{n-m}\frac{F_\gamma(\varpi_j)}{F(\varpi_j)},\end{aligned}$$

se define $\boldsymbol{\Psi}(\boldsymbol{\theta}) = \boldsymbol{\Psi}(\theta, \gamma) = [\varphi_1(\theta, \gamma), \varphi_2(\theta, \gamma)]^t$.

Entonces la matriz jacobiana para este sistema de ecuaciones es

$$\mathbf{J}(\theta, \gamma) = \begin{bmatrix} \dfrac{\partial}{\partial\theta}\varphi_1(\theta,\gamma) & \dfrac{\partial}{\partial\gamma}\varphi_1(\theta,\gamma) \\ \dfrac{\partial}{\partial\theta}\varphi_2(\theta,\gamma) & \dfrac{\partial}{\partial\gamma}\varphi_2(\theta,\gamma) \end{bmatrix},$$

donde

$$\begin{aligned} \frac{\partial\varphi_1}{\partial\theta} &= \frac{m(1-2\bar{w})}{\theta^2} + \sum_{j=1}^{n-m} \frac{F_{\theta\theta}(\varpi_j)F(\varpi_j) - [F_\theta(\varpi_j)]^2}{[F(\varpi_j)]^2} \\ \frac{\partial\varphi_1}{\partial\gamma} &= \frac{\partial\varphi_2}{\partial\theta} = -\frac{m}{\theta^2} + \sum_{j=1}^{n-m} \frac{F_{\theta\gamma}(\varpi_j)F(\varpi_j) - F_\theta(\varpi_j)F_\gamma(\varpi_j)}{[F(\varpi_j)]^2} \\ \frac{\partial\varphi_2}{\partial\gamma} &= \sum_{j=1}^{n-m} \frac{F_{\gamma\gamma}(\varpi_j)F(\varpi_j) - [F_\gamma(\varpi_j)]^2}{[F(\varpi_j)]^2}. \end{aligned}$$

El proceso numérico para este algoritmo inicia con algunos valores iniciales importantes. En este sentido se proponen $\boldsymbol{\theta}^{(0)} = (\theta^{(0)}, \gamma^{(0)})^t$ *y consecuentemente* $\boldsymbol{\Psi}(\boldsymbol{\theta}^{(0)}) = \boldsymbol{\Psi}(\theta^{(0)}, \gamma^{(0)})^t$, *con lo que*

$$\mathbf{J}(\boldsymbol{\theta}^{(0)}) = \begin{bmatrix} \dfrac{\partial}{\partial\theta}\varphi_1(\theta^{(0)},\gamma^{(0)}) & \dfrac{\partial}{\partial\gamma}\varphi_1(\theta^{(0)},\gamma^{(0)}) \\ \dfrac{\partial}{\partial\theta}\varphi_2(\theta^{(0)},\gamma^{(0)}) & \dfrac{\partial}{\partial\gamma}\varphi_2(\theta^{(0)},\gamma^{(0)}) \end{bmatrix}.$$

Ahora el algoritmo computacional inicia resolviendo el sistema de ecuaciones generadas por $\boldsymbol{h}^{(0)} = -[\mathbf{J}(\boldsymbol{\theta}^{(0)})]^{-1}\boldsymbol{\Psi}(\boldsymbol{\theta}^{(0)})$, *resultando una mejor aproximación denotada por*

$$\boldsymbol{\theta}^{(1)} = \boldsymbol{\theta}^{(0)} + \boldsymbol{h}^{(0)}.$$

Prosiguiendo de manera recursiva se obtiene la siguiente fórmula

$$\hat{\boldsymbol{\theta}}^{(k)} = \boldsymbol{\theta}^{(k-1)} + \boldsymbol{h}^{(k-1)}, \qquad k = 2, 3, \ldots.$$

Así, de manera general, cada vez se obtiene una mejor aproximación al sistema de ecuaciones $\boldsymbol{h}^{(k-1)} = -[\mathbf{J}(\boldsymbol{\theta}^{(k-1)})]^{-1}\boldsymbol{\Psi}(\boldsymbol{\theta}^{(k-1)})$, *donde*

$$\mathbf{J}(\boldsymbol{\theta}^{(k-1)}) = \left[\frac{\partial}{\partial \theta_i} \boldsymbol{\Psi} \right]_{i=1,2}, \quad \boldsymbol{h}^{(k-1)} = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} \quad y \quad \boldsymbol{\Psi}(\boldsymbol{\theta}^{(k-1)}) = \begin{bmatrix} \varphi_1 \\ \varphi_2 \end{bmatrix},$$

donde $\theta_1 = \theta$ *y* $\theta_2 = \gamma$.

Por último un criterio de convergencia basado en la proximidad de dos estimaciones puntuales serán la regla de paro del algoritmo y así se obtendrán las mejores estimaciones cuando $||\hat{\boldsymbol{\theta}}^{(k_i)} - \hat{\boldsymbol{\theta}}^{(k_j)}|| < \epsilon$ *para* $i \neq j$ *y* $\epsilon > 0$, *resultando así*

$$\begin{bmatrix} \hat{\theta}^{(k)} \\ \hat{\gamma}^{(k)} \end{bmatrix} = \begin{bmatrix} \theta^{(k-1)} \\ \gamma^{(k-1)} \end{bmatrix} + \begin{bmatrix} h_1^{(k-1)} \\ h_2^{(k-1)} \end{bmatrix}$$

Dado que este procedimiento permite obtener estimadores con valores aproximados, considerablemente buenos, a partir de aquí se utlizarán procesos análogos en la inferencia sobre los parámetros en los modelos lognormal y Weibull que serán presentados en las secciones subsecuentes.

4.5. Estimación sobre una muestra aleatoria lognormal censurada por la izquierda

En (3.4) y (3.5) se definieron las funciones de densidad y distribución de una variable aleatoria $Y \sim LN(\mu, \sigma)$, a saber

$$f(y) = \frac{1}{\sigma y} \phi_{nor}\left(\frac{ln\ y - \mu}{\sigma}\right), \quad F(y) = \Phi_{nor}\left(\frac{ln\ y - \mu}{\sigma}\right), \quad y > 0.$$

Para los fines ilustrativos, en esta sección se utilizarán las notaciones $\Phi(z) = \Phi_{nor}\left[(ln\ y - \mu)/\sigma\right]$ y $\phi(z) = \phi_{nor}\left[(ln\ y - \mu)/\sigma\right]$, donde $z = (ln\ y - \mu)/\sigma$ representará las observaciones exactas y $\mathfrak{z} = (ln\ \mathfrak{y} - \mu)/\sigma$ se escribirá para las observaciones censuradas.

El método de estimación inicia con la construcción de la verosimilitud para μ y σ a partir de $y_1, y_2, \ldots, y_m, \mathfrak{y}_1, \mathfrak{y}_2, \ldots, \mathfrak{y}_{n-m}$ una muestra aleatoria censurada lognormal consistiendo de m observaciones exactas y $n - m$

resultan censuradas por la izquierda

$$\begin{aligned} L(\mu,\sigma) &= \prod_{i=1}^{m}\frac{1}{\sigma y_i}\phi(z_i)\prod_{j=1}^{n-m}\Phi(\mathfrak{Z}_j) \\ &= \prod_{i=1}^{m}\frac{1}{\sqrt{2\pi}\sigma y_i}e^{-\frac{1}{2}\left(\frac{ln\ y_i-\mu}{\sigma}\right)^2}\prod_{j=1}^{n-m}\int_0^{\mathfrak{Y}_j}\frac{1}{\sqrt{2\pi}\sigma\nu_j}e^{-\frac{1}{2}\left(\frac{ln\ \nu_j-\mu}{\sigma}\right)^2}d\nu_j. \end{aligned}$$

Con lo que la log-verosimilitud es la siguiente

$$\begin{aligned} \ell(\mu,\sigma) &= ln[L(\mu,\sigma)] \\ &= -mln(\sqrt{2\pi}\sigma)-\sum_{i=1}^{m}ln\ y_i-\frac{1}{2}\sum_{i=1}^{m}\left(\frac{ln\ y_i-\mu}{\sigma}\right)^2+ \\ &\quad \sum_{j=1}^{n-m}ln\left[\int_0^{\mathfrak{Y}_j}\frac{1}{\sqrt{2\pi}\sigma\nu_j}e^{-\frac{1}{2}\left(\frac{ln\ \nu_j-\mu}{\sigma}\right)^2}d\nu_j\right]; \end{aligned}$$

empleando las notaciones y reparametrizaciones siguentes $\Phi(z)$, $ln\tilde{y} = \sum_{i=1}^{m}(ln\ y_i)/m$, $z_i = (ln\ y_i-\mu)/\sigma$, $\ell(\mu,\sigma)$, $\Phi(\mathfrak{Z})$ y $\mathfrak{Z}_j = (ln\ \mathfrak{Y}_j-\mu)/\sigma$ queda reexpresada

$$\ell(\mu,\sigma) = -m\left(ln(\sqrt{2\pi}\sigma)+ln\tilde{y}\right)-\frac{1}{2}\sum_{i=1}^{m}z_i^2+\sum_{j=1}^{n-m}ln\left[\Phi(\mathfrak{Z}_j)\right].$$

Análogamente al procedimiento descrito en el caso exponencial, los estimadores para μ y σ se obtendrán resolviendo el sistema de ecuaciones siguientes

$$\frac{\partial\ell(\mu,\sigma)}{\partial\mu} = \frac{m\bar{z}}{\sigma}+\sum_{j=1}^{n-m}\frac{\Phi_\mu(\mathfrak{Z}_j)}{\Phi(\mathfrak{Z}_j)}=0 \qquad (4.14)$$

$$\frac{\partial\ell(\mu,\sigma)}{\partial\sigma} = \frac{1}{\sigma}\left(\sum_{i=1}^{m}z_i^2-m\right)+\sum_{j=1}^{n-m}\frac{\Phi_\sigma(\mathfrak{Z}_j)}{\Phi(\mathfrak{Z}_j)}=0. \qquad (4.15)$$

Nuevamente se observa que se tiene un sistema de ecuaciones no lineales en μ y σ , por lo que se recurre otra vez a la herramienta numérica basada en Newton-Raphson y descrita en el apéndice B.

El cómputo numérico inicia renombrando (4.14) y (4.15) como sigue

$$\begin{aligned}\varphi_1(\mu,\sigma) &= \frac{m\bar{z}}{\sigma}+\sum_{j=1}^{n-m}\frac{\Phi_\mu(\mathfrak{Z}_j)}{\Phi(\mathfrak{Z}_j)}=0\\ \varphi_2(\mu,\sigma) &= \frac{1}{\sigma}\left(\sum_{i=1}^{m}z_i^2-m\right)+\sum_{j=1}^{n-m}\frac{\Phi_\sigma(\mathfrak{Z}_j)}{\Phi(\mathfrak{Z}_j)}=0.\end{aligned}$$

Y de este modo queda redefinada $\mathbf{J}$:

$$\mathbf{J}(\mu,\sigma) = \begin{bmatrix}\frac{\partial}{\partial\mu}\varphi_1(\mu,\sigma) & \frac{\partial}{\partial\sigma}\varphi_1(\mu,\sigma)\\ \frac{\partial}{\partial\mu}\varphi_2(\mu,\sigma) & \frac{\partial}{\partial\sigma}\varphi_2(\mu,\sigma)\end{bmatrix},$$

en donde

$$\begin{aligned}\frac{\partial\varphi_1}{\partial\mu} &= -\frac{m}{\sigma^2}+\sum_{j=1}^{n-m}\frac{\Phi_{\mu\mu}(\mathfrak{Z}_j)\Phi(\mathfrak{Z}_j)-[\Phi_\mu(\mathfrak{Z}_j)]^2}{[\Phi(\mathfrak{Z}_j)]^2}\\ \frac{\partial\varphi_1}{\partial\sigma} &= \frac{\partial\varphi_2}{\partial\mu}=\frac{2m\bar{z}}{\sigma^2}+\sum_{j=1}^{n-m}\frac{\Phi_{\mu\sigma}(\mathfrak{Z}_j)\Phi(\mathfrak{Z}_j)-\Phi_\mu(\mathfrak{Z}_j)\Phi_\sigma(\mathfrak{Z}_j)}{[\Phi(\mathfrak{Z}_j)]^2}\\ \frac{\partial\varphi_2}{\partial\sigma} &= \frac{1}{\sigma^2}\left(m-\sum_{i=1}^{n}z_i^2\right)+\sum_{j=1}^{n-m}\frac{\Phi_{\sigma\sigma}(\mathfrak{Z}_j)\Phi(\mathfrak{Z}_j)-[\Phi_\sigma(\mathfrak{Z}_j)]^2}{[\Phi(\mathfrak{Z}_j)]^2}.\end{aligned}$$

Por lo tanto, si asignamos una condición inicial para μ y σ como en el Ejemplo 4.4.1 y respetando la regla de paro cuando $||\hat{\boldsymbol{\theta}}^{(k_i)}-\hat{\boldsymbol{\theta}}^{(k_j)}||<\epsilon$ se tiene la ecuación recursiva

$$\begin{bmatrix}\hat{\mu}^{(k)}\\ \hat{\sigma}^{(k)}\end{bmatrix}=\begin{bmatrix}\mu^{(k-1)}\\ \sigma^{(k-1)}\end{bmatrix}+\begin{bmatrix}h_1^{(k-1)}\\ h_2^{(k-1)}\end{bmatrix}. \tag{4.16}$$

con lo cual se obtendrá una mejor aproximación para los estimadores de μ y σ.

4.6. Estimación sobre una muestra aleatoria Weibull censurada por la izquierda

Ahora considere una variable aleatoria $Y \sim Weib(\beta,\eta)$ y sea $y_1,y_2,\ldots,y_m,y_1,y_2,\ldots,y_{n-m}$ una muestra aleatoria, donde m datos son observaciones exactas y $n-m$ resultan censurados por la izquierda. Entonces el

proceso para la obtención de los estimadores de parámetros μ y σ, la función de verosimilitud se define así

$$L(\mu,\sigma) = \prod_{i=1}^{m}\frac{1}{\sigma y_i}\phi_{sev}\left(\frac{ln\ y_i-\mu}{\sigma}\right)\prod_{j=1}^{n-m}\Phi_{sev}\left(\frac{ln\ y_j-\mu}{\sigma}\right),$$

donde $\phi_{sev}(z)$ y $\Phi_{sev}(\mathfrak{z})$ son las funciones de densidad y de distribución expresadas en (3.9) y (3.10), respectivamente.

En este caso resulta importante denotar $\Phi_{sev}(\mathfrak{z}) = {}^{w}\Phi(\mathfrak{z})$ para un mejor desarrollo analítico y algebraico de las ecuaciones necesarias. De este modo la verosimilitud se reescribe como sigue

$$L(\mu,\sigma) = \prod_{i=1}^{m}\frac{1}{\sigma y_i}exp\left[z_i-exp(z_i)\right]\prod_{j=1}^{n-m}\left[1-exp\left(-exp(\mathfrak{z}_j)\right)\right],$$

en este mismo sentido la log-verosimilitud resulta

$$\begin{aligned}\ell(\mu,\sigma) &= -mln\,\sigma-\sum_{i=1}^{m}ln\,y_i+\sum_{i=1}^{m}\left[z_i-exp(z_i)\right]+\sum_{j=1}^{n-m}ln\left[1-exp\left(-exp(\mathfrak{z}_j)\right)\right]\\ &= -m(ln\,\sigma+ln\tilde{y})+\sum_{i=1}^{m}\left[z_i-exp(z_i)\right]+\sum_{j=1}^{n-m}ln\left[{}^{w}\Phi(\mathfrak{z}_j)\right].\end{aligned}$$

Con lo anterior se deriva el sistema de ecuaciones que conlleva a la obtención de $\hat{\mu}$ y $\hat{\sigma}$. Siendo este el siguiente

$$\frac{\partial\ell(\mu,\sigma)}{\partial\mu} = -\frac{1}{\sigma}\sum_{i=1}^{m}(1-e^{z_i})+\sum_{j=1}^{n-m}\frac{{}^{w}\Phi_{\mu}(\mathfrak{z}_j)}{{}^{w}\Phi(\mathfrak{z}_j)}=0 \qquad (4.17)$$

$$\frac{\partial\ell(\mu,\sigma)}{\partial\sigma} = -\frac{1}{\sigma}\sum_{i=1}^{m}(1+z_i-z_ie^{z_i})+\sum_{j=1}^{n-m}\frac{{}^{w}\Phi_{\sigma}(\mathfrak{z}_j)}{{}^{w}\Phi(\mathfrak{z}_j)}=0. \qquad (4.18)$$

Una vez más Newton-Raphson puede ayudar en la solución al sistema de ecuaciones anterior en virtud de su característica no lineal. Con la ayuda de la notación propuesta, (4.17) y (4.18) se reescribe así

$$\begin{aligned}\varphi_1(\mu,\sigma) &= -\frac{1}{\sigma}\sum_{i=1}^{m}(1-e^{z_i})+\sum_{j=1}^{n-m}\frac{{}^{w}\Phi_{\mu}(\mathfrak{z}_j)}{{}^{w}\Phi(\mathfrak{z}_j)}=0\\ \varphi_1(\mu,\sigma) &= -\frac{1}{\sigma}\sum_{i=1}^{m}(1+z_i-z_ie^{z_i})+\sum_{j=1}^{n-m}\frac{{}^{w}\Phi_{\sigma}(\mathfrak{z}_j)}{{}^{w}\Phi(\mathfrak{z}_j)}=0.\end{aligned}$$

Y la matriz jacobiana es

$$\mathbf{J}(\mu,\sigma) = \begin{bmatrix} \frac{\partial}{\partial\mu}\varphi_1(\mu,\sigma) & \frac{\partial}{\partial\sigma}\varphi_1(\mu,\sigma) \\ \frac{\partial}{\partial\mu}\varphi_2(\mu,\sigma) & \frac{\partial}{\partial\sigma}\varphi_2(\mu,\sigma) \end{bmatrix},$$

donde cada entrada de $\mathbf{J}$ queda

$$\begin{aligned}
\frac{\partial\varphi_1}{\partial\mu} &= -\frac{1}{\sigma^2}\sum_{i=1}^{m} e^{z_i} + \sum_{j=1}^{n-m} \frac{{}^w\Phi_{\mu\mu}(\mathfrak{Z}_j)\cdot {}^w\Phi(\mathfrak{Z}_j) - [{}^w\Phi_{\mu}(\mathfrak{Z}_j)]^2}{[{}^w\Phi(\mathfrak{Z}_j)]^2} \\
\frac{\partial\varphi_1}{\partial\sigma} &= \frac{\partial\varphi_2}{\partial\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{m}(1 - e^{z_i} - z_i e^{z_i}) + \sum_{j=1}^{n-m} \frac{{}^w\Phi_{\mu\sigma}(\mathfrak{Z}_j)\cdot {}^w\Phi(\mathfrak{Z}_j) - {}^w\Phi_{\mu}(\mathfrak{Z}_j)\cdot {}^w\Phi_{\sigma}(\mathfrak{Z}_j)}{[{}^w\Phi(\mathfrak{Z}_j)]^2} \\
\frac{\partial\varphi_2}{\partial\sigma} &= \frac{1}{\sigma^2}\sum_{i=1}^{m}\left(1 + 2z_i - 2z_i e^{z_i} - z_i^2 e^{z_i}\right) + \sum_{j=1}^{m} \frac{{}^w\Phi_{\sigma\sigma}(\mathfrak{Z}_j)\cdot {}^w\Phi(\mathfrak{Z}_j) - [{}^w\Phi_{\sigma}(\mathfrak{Z}_j)]^2}{[{}^w\Phi(\mathfrak{Z}_j)]^2}.
\end{aligned}$$

Prosiguiendo de la misma forma que con las estimaciones anteriores, el proceso para hallar $\hat{\mu}$ y $\hat{\sigma}$ resulta del siguiente proceso recursivo

$$\begin{bmatrix} \hat{\mu}^{(k)} \\ \hat{\sigma}^{(k)} \end{bmatrix} = \begin{bmatrix} \mu^{(k-1)} \\ \sigma^{(k-1)} \end{bmatrix} + \begin{bmatrix} h_1^{(k-1)} \\ h_2^{(k-1)} \end{bmatrix}$$

finalizando este proceso cuando $||\hat{\boldsymbol{\theta}}^{(k_i)} - \hat{\boldsymbol{\theta}}^{(k_j)}|| < \epsilon$ para $\epsilon > 0$; resultando así una buena aproximación a los estimadores de dichos parámetros.

4.7. Selección de modelos

4.7.1. Criterio de Información de Akaike

El *Criterio de Información de Akaike* (*AIC* por sus siglas en ingles) es una medida relativa de la bondad del ajuste para un modelo estadístico. Este criterio se basa en el concepto de *entropía de la información*, ofrece una medida relativa de la pérdida de información cuando un determinado modelo se utiliza para describir datos reales. Se puede decir que sirve para ilustrar el sesgo y varianza en el ajuste de un modelo; es decir, entre la precisión y la complejidad del modelo.

Los valores del *AIC* proporcionan una alternativa para la selección de un modelo entre una familia de ellos (modelos de localización-escala). Una situación a considerar bajo este criterio es que no determina una selección

absoluta de algún modelo en particular sino, una medida relativa para elegir el modelo más apropiado entre varios que comparten características o naturaleza comunes.

Si los modelos candidatos no reflejan un buen ajuste, el criterio de Akaike omitirá esta advertencia en virtud de que solo compara de manera relativa, de tal modo que debe tenerse idea de que los candidatos han sido reportados o utilzados eficientemente en estudios de la misma naturaleza.

La regla general para obtener un valor para el AIC es el siguiente

$$AIC = \quad 2k - 2ln(L(\hat{\boldsymbol{\theta}})),$$

donde k es la dimensión de parámetros $\boldsymbol{\theta}$, y $L(\hat{\boldsymbol{\theta}})$ es el valor estimado de la función de verosimilitud.

Para una familia o conjunto de modelos candidatos para el mismo modelo, el modelo elegido será el que contenga el valor mínimo para el AIC.

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 [6].

4.7.2. Prueba de razón de verosimilitudes

En muchas aplicaciones, la hipótesis a probar puede ser formulada como hipótesis relativas a los valores de parámetros desconocidos en el modelo de probabilidad. Un estadístico de prueba D, llamado la *estadística de razón de verosimilitudes*, puede derivarse de la función de log-verosimilitud. Una prueba de significación en el que se utiliza la estadística de razón de verosimilitudes, como estadístico de prueba se llama *prueba de razón de verosimilitudes* [27].

Una manera natural y muy usada de comparar el ajuste relativo de dos modelos alternativos a una matriz de datos es contrastar las verosimilitudes resultantes mediante la prueba de razón de verosimilitudes

$$-2ln\left[\frac{L_1(\theta_0)}{L_2(\hat{\theta})}\right] \quad \sim \quad \chi^2_v,$$

donde $L_2(\hat{\theta})$ es la máxima verosimilitud del modelo general en $\hat{\theta}$; $L_1(\theta_0)$ es la máxima verosimilitud del modelo restringido en θ_0; χ^2_v es una variable

aleatoria χ^2 con v grados de libertad, obteniéndose de la diferencia entre la dimensión de $\hat{\theta}$ y la dimensión de θ_0.

Esta prueba 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

$$\begin{aligned} \chi^2 &= -2ln\left[\frac{L(\hat{\theta}_{Modelo\,1})}{L(\hat{\theta}_{Modelo\,2})}\right] \sim \chi^2_1 \\ &= -2\left[\ell(\hat{\theta}_{Modelo\,1}) - \ell(\hat{\theta}_{Modelo\,2})\right] \sim \chi^2_1. \end{aligned} \tag{4.19}$$

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

$$H: Modelo1 \equiv Modelo2,$$

donde *modelo1* y *modelo2* 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 $(1-\alpha)100\,\%$.

Capítulo 5

Intervalos de verosimilitud confianza

5.1. Aproximación normal

Los intervalos de verosimilitud confianza obtenidas desde la normal son fáciles de calcular y, en la actualidad son muy utilizados en los paquetes estadísticos comerciales. Utilizando los estimadores de máxima verosimilitud de los parámetros del modelo y con ayuda de su matriz de covarianzas el cálculo de estos intervalos son posibles a través de cálculos no tan complejos.

Estos intervalos (regiones) de verosimilitud confianza se basan en una aproximación cuadrática para la log-verosimilitud y son adecuados cuando la log-verosimilitud se aproxima en forma cuadrática sobre el intervalo de confianza. Con muestras grandes, bajo las condiciones usuales de regularidad, la log-verosimilitud es aproximadamente cuadrática, y así la aproximación normal y los intervalos de verosimilitud confianza concordarán armónicamente. El tamaño de muestra necesario para una aproximación adecuada depende del modelo, de la magnitud de censura o truncamiento. 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 [35].

5.1.1. Aproximación normal para un solo parámetro

Sea $\ell(\theta)$ la log-verosimilitud de un parámetro continuo θ con valores en Ω; $\ell'(\theta)$ y $I(\theta) = -\ell''(\theta)$ la primera derivada de la log-verosimilitud y la matriz

de información, respectivamente. Si se asume que $\widehat{\theta}$ existe y es punto interior de Ω, y que $\ell(\theta)$ tiene una expansión en series de Taylor alrededor de $\theta = \widehat{\theta}$, como sigue

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

Ahora, dado que $\ell'(\widehat{\theta}) = 0$ y considerando a $r(\theta) = \ell(\theta) - \ell(\widehat{\theta})$, se tiene que

$$r(\theta) = -\frac{1}{2}(\theta - \widehat{\theta})^2 I(\widehat{\theta}) + \frac{(\theta - \widehat{\theta})^3}{3!}\ell'''(\widehat{\theta}) + \dots. \tag{5.1}$$

Entonces la *aproximación normal* para $r(\theta)$ está definida así

$$r_N(\theta) = -\frac{1}{2}(\theta - \widehat{\theta})^2 I(\widehat{\theta}). \tag{5.2}$$

Si $|\theta - \widehat{\theta}|$ es pequeño, el término cúbico y de grado superior en (5.1) también lo son, entonces $r(\theta) \approx r_N(\theta)$, es decir,

$$r(\theta) \approx -\frac{1}{2}(\theta - \widehat{\theta})^2 I(\widehat{\theta}).$$

Así para una muestra aleatoria de tamaño suficiente, $|\theta - \widehat{\theta}|$ será pequeño y $r_N(\theta)$ dará una buena aproximación a $r(\theta)$ sobre la región entera de valores paramétricos plausibles.

El $100p\%$ de la región 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

$$\theta \in \hat{\theta} \pm \sqrt{-2\, lnp\, I^{-1}(\hat{\theta})} \tag{5.3}$$

como una aproximación para el $100p\%$ del intervalo de verosimilitud confianza. Este es un intervalo centrado en $\hat{\theta}$ con longitud

$$2\sqrt{-2\, lnp\, I^{-1}(\hat{\theta})}.$$

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 llamado la "matriz de información".

5.1.2. Aproximación normal para $\boldsymbol{\theta} = (\theta_1, \theta_2)$

En la subsección anterior se obtuvo la aproximación normal $r_N(\theta) = -(\theta - \hat{\theta})^2 I(\hat{\theta})/2$, para un solo parámetro θ, suprimiendo el término cúbico y los de grado superior en la expansión de Taylor de $\ell(\hat{\theta})$. Ahora para el caso de dos parámetros, se sigue un procedimiento análogo a través de la expansión en series de Taylor para $\ell(\theta_1, \theta_2)$ y considerando la matriz de información $I(\theta_1, \theta_2)$, se obtiene lo siguiente

$$\boldsymbol{r}(\theta_1, \theta_2) \approx -\frac{1}{2}(\theta_1 - \hat{\theta_1})^2 \hat{I}_{11} - \frac{1}{2}(\theta_2 - \hat{\theta_2})^2 \hat{I}_{22} - (\theta_1 - \hat{\theta_1})(\theta_2 - \hat{\theta_2}) \hat{I}_{12}. \quad (5.4)$$

Luego denotándose

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}; \quad \hat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{\theta_1} \\ \hat{\theta_2} \end{bmatrix}; \quad I(\hat{\boldsymbol{\theta}}) = I(\hat{\theta_1}, \hat{\theta_2}) = \begin{bmatrix} \hat{I}_{11} & \hat{I}_{12} \\ \hat{I}_{21} & \hat{I}_{22} \end{bmatrix}$$

la aproximación (5.4) puede ser reescrita matricialmente como sigue

$$\boldsymbol{r}(\boldsymbol{\theta}) \approx \boldsymbol{r}_N(\boldsymbol{\theta}) = -\frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' I(\hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}), \quad (5.5)$$

invocando una generalización para la aproximación normal obtenida en (5.2).

5.2. Matriz de covarianzas para funciones estimadas

Del apéndice B.6.4 en [35] que trata sobre la estimación de la matriz de Varianza-Covarianza para un estimador de máxima verosimilitud se tiene que bajo condiciones típicas de regularidad, la $\widehat{\mathbf{Var}}(\hat{\boldsymbol{\theta}}) = \hat{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1}$ es un estimador consistente de la $\mathbf{Var}(\hat{\boldsymbol{\theta}})$, donde $\hat{\boldsymbol{I}}_{\boldsymbol{\theta}}$ está definido por (4.2). Este estimador se obtiene a través de (4.1), $\boldsymbol{I}_{\boldsymbol{\theta}} = E\left[-\partial^2 \ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'\right]$. La estimación se realiza directamente evaluando $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ en la misma ecuación.

El estimador "local" de la matriz de covarianzas de $\hat{\mathbf{g}} = \mathbf{g}(\hat{\boldsymbol{\theta}})$, donde $\mathbf{g}$ es una función vectorial, que se obtiene sustituyendo de la siguiente forma:

$$\widehat{\mathbf{Var}}(\hat{\boldsymbol{g}}) = \left[\frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]' \hat{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1} \left[\frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right] = \left[\frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]' \widehat{\mathbf{Var}}(\hat{\boldsymbol{\theta}}) \left[\frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right] \quad (5.6)$$

evaluandose en $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. Para el caso donde g es una función escalar y θ un solo parámetro la expresión se reduce a lo siguiente

$$\widehat{\mathrm{Var}}(\hat{\mathrm{g}}) = \left[\frac{\partial \mathrm{g}(\theta)}{\partial \theta}\right]^2 \widehat{\mathrm{Var}}(\hat{\theta}). \tag{5.7}$$

5.3. Intervalo de confianza para $\mathbf{g} = \mathbf{g}(\mu, \sigma)$

Siguiendo la metodología de la sección anterior, un intervalo de verosimilitud confianza para una función con parámetros μ y σ, es decir, $\mathrm{g} = \mathrm{g}(\mu, \sigma)$, puede aproximarse utilizando la distribución normal estándar. De este modo un intervalo de verosimilitud confianza de $100(1-\alpha)\,\%$ para g es

$$[\mathrm{g}_*\ ,\ \mathrm{g}^*] = \hat{\mathrm{g}} \pm z_{(1-\alpha/2)} \widehat{se}_{\hat{\mathrm{g}}}, \tag{5.8}$$

donde $\widehat{se}_{\hat{\mathrm{g}}} = \sqrt{\widehat{\mathrm{Var}}[\ \mathrm{g}(\hat{\mu}, \hat{\sigma})]}$, con $\widehat{\mathrm{Var}}[\ \mathrm{g}(\hat{\mu}, \hat{\sigma})]$ como un caso especial de (5.6), calculándose como sigue

$$\begin{aligned}
\widehat{\mathrm{Var}}[\ \mathrm{g}(\hat{\mu}, \hat{\sigma})] &= \begin{bmatrix} \dfrac{\partial \mathrm{g}}{\partial \mu} & \dfrac{\partial \mathrm{g}}{\partial \sigma} \end{bmatrix} \begin{bmatrix} -\dfrac{\partial^2 \ell(\mu,\sigma)}{\partial \mu^2} & -\dfrac{\partial^2 \ell(\mu,\sigma)}{\partial \mu \partial \sigma} \\ -\dfrac{\partial^2 \ell(\mu,\sigma)}{\partial \sigma \partial \mu} & -\dfrac{\partial^2 \ell(\mu,\sigma)}{\partial \sigma^2} \end{bmatrix}^{-1} \begin{bmatrix} \dfrac{\partial \mathrm{g}}{\partial \mu} \\ \dfrac{\partial \mathrm{g}}{\partial \sigma} \end{bmatrix} \\
&= \begin{bmatrix} \dfrac{\partial \mathrm{g}}{\partial \mu} & \dfrac{\partial \mathrm{g}}{\partial \sigma} \end{bmatrix} \begin{bmatrix} \widehat{Var}(\hat{\mu}) & \widehat{Cov}(\hat{\mu}, \hat{\sigma}) \\ \widehat{Cov}(\hat{\mu}, \hat{\sigma}) & \widehat{Var}(\hat{\sigma}) \end{bmatrix} \begin{bmatrix} \dfrac{\partial \mathrm{g}}{\partial \mu} \\ \dfrac{\partial \mathrm{g}}{\partial \sigma} \end{bmatrix} \\
&= \left(\frac{\partial \mathrm{g}}{\partial \mu}\right)^2 \widehat{Var}(\hat{\mu}) + 2\left(\frac{\partial \mathrm{g}}{\partial \mu}\right)\left(\frac{\partial \mathrm{g}}{\partial \sigma}\right) \widehat{Cov}(\hat{\mu}, \hat{\sigma}) + \left(\frac{\partial \mathrm{g}}{\partial \sigma}\right)^2 \widehat{Var}(\hat{\sigma}).
\end{aligned} \tag{5.9}$$

Las derivadas parciales en (5.9) deben ser evaluadas en $\mu = \hat{\mu}$ y $\sigma = \hat{\sigma}$.

5.4. Regiones de confianza

La aproximación normal para la distribución de los estimadores de máxima verosimilitud puede ser utilizado para obtener regiones de verosimilitud confianza de funciones escalares (vector) de $\boldsymbol{\theta}$. En particular, una

aproximación a una región de confianza de $100(1-\alpha)\,\%$ para $\boldsymbol{\theta}$ es el conjunto de todos los valores de $\boldsymbol{\theta}$ en el elipsoide

$$(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})' \left[\widehat{\mathbf{Var}}(\hat{\boldsymbol{\theta}})\right]^{-1} (\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}) \quad \leq \quad \chi^2_{(1-\alpha;r)}, \tag{5.10}$$

donde r es la longitud de $\boldsymbol{\theta}$. Esto es llamado algunas veces como "método de Wald". Esta región de verosimilitud confianza se basa sobre los resultados distribucionales asintóticos [35], de tal modo que cuando se evalua en el valor de $\boldsymbol{\theta}$, el "estadístico de Wald "

$$\boldsymbol{W(\theta)} = (\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})' \left[\widehat{\mathbf{Var}}(\hat{\boldsymbol{\theta}})\right]^{-1} (\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})$$

sigue una distribución chi-cuadrada con r grados de libertad.

5.5. Estimación con datos no censurados

Sin pérdida de generalidad y para ilustrar la obtención de intervalos de verosimilitud confianza, considere una variable aleatoria Y que sigue una distribución lognormal definidas por (3.4) y (3.5). Entonces se obtiene, por el Corolario 3.5.5 y (3.6) que la media y la mediana de este modelo son

$$\begin{aligned} \mathrm{g}(\mu,\sigma) &= \mathrm{m} = e^{\mu+\sigma^2/2} && (5.11)\\ h(\mu,\sigma) &= \mathrm{M} = e^{\mu}. && (5.12) \end{aligned}$$

Ahora la construcción de un intervalo de verosimilitud confianza para cada una de estas funciones, se realizará a través de los estimadores de máxima verosimilitud y la matriz de covarianzas descrito en la Sección 5.2, en combinación con (5.8).

5.5.1. Intervalo de confianza para la media

Sean $y_1, y_2, \ldots, y_r$, r observaciones no censuradas que siguen una distribución lognormal. Entonces para esta muestra aleatoria la función de verosimilitud estará definida como sigue

$$L(\mu,\sigma) = \frac{\prod_{i=1}^{r} exp\left[-\frac{1}{2}\left(\frac{ln\ y_i-\mu}{\sigma}\right)^2\right]}{(2\pi)^{r/2}\sigma^r \prod_{i=1}^{r} y_i} = \frac{exp\left[-\frac{1}{2}\sum_{i=1}^{r}\left(\frac{ln\ y_i-\mu}{\sigma}\right)^2\right]}{(2\pi)^{r/2}\sigma^r \prod_{i=1}^{r} y_i}$$

de la cual la log-verosimilitud se expresa así

$$\ell(\mu,\sigma) = ln(L(Y)) = -rln(\sigma) - \frac{r}{2}ln(2\pi) - \sum_{i=1}^{r} ln\ y_i - \frac{1}{2}\sum_{i=1}^{r}\left[\frac{ln\ y_i - \mu}{\sigma}\right]^2.$$

Ahora $\hat{\mu}$ y $\hat{\sigma}$ se obtienen a través de la solución del siguiente sistema de ecuaciones

$$\begin{aligned}\frac{\partial \ell(\mu,\sigma)}{\partial \mu} &= \sum_{i=1}^{r}\left[\frac{ln\ y_i - \mu}{\sigma^2}\right] = \frac{1}{\sigma^2}\left(\sum_{i=1}^{r} ln\ y_i - r\mu\right) = 0\\ \frac{\partial \ell(\mu,\sigma)}{\partial \sigma} &= -\frac{r}{\sigma} + \sum_{i=1}^{r}\left[\frac{(ln\ y_i - \mu)^2}{\sigma^3}\right] = 0,\end{aligned}$$

resolviendo de manera analítica, los estimadores de máxima verosimilitud para los parámetros del modelo lognormal son

$$\hat{\mu} = \frac{1}{r}\sum_{i=1}^{r} ln\ y_i \qquad y \qquad \hat{\sigma} = \left[\frac{1}{r}\sum_{i=1}^{r}(ln\ y_i - \hat{\mu})^2\right]^{1/2}.$$

Con lo anterior se obtiene una estimación para la matriz de covarianzas de estos parámetros, obteniéndose como sigue

$$\begin{aligned}\frac{\partial^2 \ell}{\partial \mu^2}\Big|_{(\mu,\sigma)=(\hat{\mu},\hat{\sigma})} &= -\frac{r}{\hat{\sigma}^2}\\ \frac{\partial^2 \ell}{\partial \sigma^2}\Big|_{(\mu,\sigma)=(\hat{\mu},\hat{\sigma})} &= \frac{r}{\hat{\sigma}^2} - \frac{3}{\hat{\sigma}^4}\sum_{i=1}^{r}(ln\ y_i - \hat{\mu})^2 = \frac{r}{\hat{\sigma}^2} - \frac{3r\hat{\sigma}^2}{\hat{\sigma}^4} = -\frac{2r}{\hat{\sigma}^2}\\ \frac{\partial^2 \ell}{\partial \sigma \partial \mu}\Big|_{(\mu,\sigma)=(\hat{\mu},\hat{\sigma})} &= \frac{\partial^2 \ell}{\partial \mu \partial \sigma}\Big|_{(\mu,\sigma)=(\hat{\mu},\hat{\sigma})} = -\frac{2}{\hat{\sigma}^3}\left(r\hat{\mu} - \sum_{i=1}^{r} ln\ y_i\right)\\ &= -\frac{2(r\hat{\mu} - r\hat{\mu})}{\hat{\sigma}^3} = 0\end{aligned}$$

y con esto se tiene la siguiente estimación

$$\widehat{Var}(\hat{\mu},\hat{\sigma}) = \begin{bmatrix} -\frac{\partial^2 \ell}{\partial \mu^2} & -\frac{\partial^2 \ell}{\partial \sigma \partial \mu} \\ -\frac{\partial^2 \ell}{\partial \mu \partial \sigma} & -\frac{\partial^2 \ell}{\partial \sigma^2} \end{bmatrix}^{-1} = \begin{bmatrix} \frac{r}{\hat{\sigma}^2} & 0 \\ 0 & \frac{2r}{\hat{\sigma}^2} \end{bmatrix}^{-1} = \begin{bmatrix} \frac{\hat{\sigma}^2}{r} & 0 \\ 0 & \frac{\hat{\sigma}^2}{2r} \end{bmatrix}.$$

Ahora la construcción del intervalo de verosimilitud confianza para la media $\mathrm{g} = exp(\mu + \sigma^2/2)$ se muestra utilizando las estimaciones obtenidas anteriormente. Primero

$$\left[\frac{\partial \mathrm{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]' = \left[\frac{\partial \mathrm{g}(\mu,\sigma)}{\partial \mu} \quad \frac{\partial \mathrm{g}(\mu,\sigma)}{\partial \sigma}\right]_{(\mu,\sigma)=(\hat{\mu},\hat{\sigma})} = \left[e^{\hat{\mu}+\frac{1}{2}\hat{\sigma}^2} \quad \hat{\sigma}e^{\hat{\mu}+\frac{1}{2}\hat{\sigma}^2}\right]$$

y por (5.6) y (5.9), se obtiene lo siguiente

$$\begin{aligned}
\widehat{Var}[\mathrm{g}(\hat{\mu},\hat{\sigma})] &= \left[\begin{array}{cc} e^{\hat{\mu}+\frac{1}{2}\hat{\sigma}^2} & \hat{\sigma}e^{\hat{\mu}+\frac{1}{2}\hat{\sigma}^2} \end{array}\right] \left[\begin{array}{cc} \frac{\hat{\sigma}^2}{r} & 0 \\ 0 & \frac{\hat{\sigma}^2}{2r} \end{array}\right] \left[\begin{array}{c} e^{\hat{\mu}+\frac{1}{2}\hat{\sigma}^2} \\ \hat{\sigma}e^{\hat{\mu}+\frac{1}{2}\hat{\sigma}^2} \end{array}\right] \\
&= \left(e^{\hat{\mu}+\frac{1}{2}\hat{\sigma}^2}\right)^2 \left(\frac{\hat{\sigma}^2}{r}\right) + 2\hat{\sigma}\left(e^{\hat{\mu}+\frac{1}{2}\hat{\sigma}^2}\right)^2 (0) + \left(\hat{\sigma}e^{\hat{\mu}+\frac{1}{2}\hat{\sigma}^2}\right)^2 \left(\frac{\hat{\sigma}^2}{2r}\right) \\
&= \frac{(2+\hat{\sigma}^2)\left(\hat{\sigma}e^{\hat{\mu}+\frac{1}{2}\hat{\sigma}^2}\right)^2}{2r}.
\end{aligned}$$

Luego, dado que g es una función escalar de (μ,σ) el intervalo de verosimilitud confianza aproximadamente de $100(1-\alpha)\,\%$ se obtendrá de $\widehat{\mathrm{g}} \pm z_{(1-\alpha/2)}\hat{se}_{\widehat{\mathrm{g}}}$ donde $\hat{se}_{\widehat{\mathrm{g}}} = \sqrt{\widehat{Var}[\mathrm{g}(\widehat{\mu},\widehat{\sigma})]}$ y $\widehat{\mathrm{g}} = \mathrm{g}(\hat{\mu},\hat{\sigma})$, teniéndose finalmente la siguiente expresión para el intervalo deseado

$$e^{\hat{\mu}+\frac{1}{2}\hat{\sigma}^2} \pm z_{(1-\alpha/2)}\left(\hat{\sigma}e^{\hat{\mu}+\frac{1}{2}\hat{\sigma}^2}\right)\sqrt{\frac{2+\hat{\sigma}^2}{2r}}.$$

5.5.2. Intervalo de confianza para la mediana

Siguiendo el mismo procedimiento anterior, se puede calcular el intervalo de confianza para la mediana $h(\mu,\sigma) = e^{\mu}$. Así, empleando (5.6) y (5.7), se observa que

$$\begin{aligned}
\widehat{Var}[h(\hat{\mu},\hat{\sigma})] &= \left[\begin{array}{cc} e^{\hat{\mu}} & 0 \end{array}\right] \left[\begin{array}{cc} \frac{\hat{\sigma}^2}{r} & 0 \\ 0 & \frac{\hat{\sigma}^2}{2r} \end{array}\right] \left[\begin{array}{c} e^{\hat{\mu}} \\ 0 \end{array}\right] \\
&= \frac{e^{2\hat{\mu}}\hat{\sigma}^2}{r}.
\end{aligned}$$

Luego como h es una función escalar de (μ,σ) y utilizando (5.8), un intervalo de verosimilitud confianza aproximadamente de $100(1-\alpha)\,\%$ sería el siguiente

$$e^{\hat{\mu}} \pm z_{(1-\alpha/2)}\frac{e^{\hat{\mu}}\hat{\sigma}}{\sqrt{r}}.$$

5.6. Estimación con datos censurados

5.6.1. Intervalo de confianza de la media

Considere una muestra aleatoria censurada $y_1, y_2, \ldots, y_m, y_1, y_2, \ldots, y_{n-m}$ extraída de una población lognormal, donde m datos son observados y $n-m$ datos censurados por la izquierda.

De la Sección 4.5 se tiene que la log-verosimilitud de la lognormal para una muestra aleatoria censurada por la izquierda es la siguiente

$$\ell(\mu,\sigma) = -m\Big(ln(\sqrt{2\pi}\sigma) + ln\tilde{y}\Big) - \frac{1}{2}\sum_{i=1}^{m} z_i^2 + \sum_{j=1}^{n-m} ln\left[\Phi(\mathfrak{Z}_j)\right],$$

y sus derivadas parciales se han definido como

$$\begin{aligned}
\varphi_1(\mu,\sigma) &= \frac{\partial\ell(\mu,\sigma)}{\partial\mu} = \frac{m\bar{z}}{\sigma} + \sum_{j=1}^{n-m}\frac{\Phi_\mu(\mathfrak{Z}_j)}{\Phi(\mathfrak{Z}_j)} \\
\varphi_2(\mu,\sigma) &= \frac{\partial\ell(\mu,\sigma)}{\partial\sigma} = \frac{1}{\sigma}\left(\sum_{i=1}^{m} z_i^2 - m\right) + \sum_{j=1}^{n-m}\frac{\Phi_\sigma(\mathfrak{Z}_j)}{\Phi(\mathfrak{Z}_j)}.
\end{aligned}$$

De la misma sección y el apéndice B.2, a través de (4.16) se tiene que los estimadores $\hat{\mu}$ y $\hat{\sigma}$ son obtenidos de algoritmos numéricos; entonces en forma análoga al procedimiento descrito en la sección anterior se obtiene primeramente

$$\widehat{Var}(\hat{\mu},\hat{\sigma}) = \begin{bmatrix} -\dfrac{\partial^2\ell}{\partial\mu^2} & -\dfrac{\partial^2\ell}{\partial\sigma\partial\mu} \\ -\dfrac{\partial^2\ell}{\partial\mu\partial\sigma} & -\dfrac{\partial^2\ell}{\partial\sigma^2} \end{bmatrix}^{-1} = \begin{bmatrix} -\dfrac{\partial\varphi_1}{\partial\mu} & -\dfrac{\partial\varphi_1}{\partial\sigma} \\ -\dfrac{\partial\varphi_2}{\partial\mu} & -\dfrac{\partial\varphi_2}{\partial\sigma} \end{bmatrix}^{-1} \tag{5.13}$$

y como g denota la media, se tiene lo siguiente

$$\left[\frac{\partial \mathrm{g}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right]' = \left[e^{\hat{\mu}+\frac{1}{2}\hat{\sigma}^2} \quad \hat{\sigma}e^{\hat{\mu}+\frac{1}{2}\hat{\sigma}^2}\right]. \tag{5.14}$$

De este modo sustituyendo (5.13) y (5.14) en (5.6), se tiene que la varianza para la media g es

$$\widehat{Var}[\mathrm{g}(\hat{\mu},\hat{\sigma})] = \left[\; e^{\hat{\mu}+\frac{1}{2}\hat{\sigma}^2} \quad \hat{\sigma}e^{\hat{\mu}+\frac{1}{2}\hat{\sigma}^2} \;\right] \widehat{Var}(\hat{\mu},\hat{\sigma}) \begin{bmatrix} e^{\hat{\mu}+\frac{1}{2}\hat{\sigma}^2} \\ \hat{\sigma}e^{\hat{\mu}+\frac{1}{2}\hat{\sigma}^2} \end{bmatrix}.$$

y utilizando (5.9) se reescribe como sigue

$$\begin{aligned}\widehat{Var}[\mathrm{g}(\hat{\mu},\hat{\sigma})] &= \left(e^{\hat{\mu}+\frac{1}{2}\hat{\sigma}^2}\right)^2 \widehat{Var}(\hat{\mu}) + 2\hat{\sigma}\left(e^{\hat{\mu}+\frac{1}{2}\hat{\sigma}^2}\right)^2 \widehat{Cov}(\hat{\mu},\hat{\sigma}) \\ &\quad + \left(\hat{\sigma}e^{\hat{\mu}+\frac{1}{2}\hat{\sigma}^2}\right)^2 \widehat{Var}(\hat{\sigma})\end{aligned}$$

Por lo tanto un intervalo de verosimilitud confianza de $100(1-\alpha)\,\%$ para la media de la lognormal de una muestra censurada por la izquierda está dada por $\widehat{\mathrm{g}} \pm z_{(1-\alpha/2)}\widehat{se}_{\widehat{\mathrm{g}}}$ donde $\widehat{se}_{\widehat{\mathrm{g}}} = \sqrt{\widehat{Var}[\mathrm{g}(\hat{\mu},\hat{\sigma})]}$ y $\widehat{\mathrm{g}} = \mathrm{g}(\hat{\mu},\hat{\sigma})$, teniéndose finalmente la siguiente expresión para el intervalo deseado

$$e^{\hat{\mu}+\frac{1}{2}\hat{\sigma}^2} \pm z_{(1-\alpha/2)}\sqrt{\widehat{Var}[\mathrm{g}(\hat{\mu},\hat{\sigma})]}$$

5.6.2. Intervalo de confianza de la mediana

Siguiendo el procedimiento anterior, este intervalo se obtiene inicialmente a través de

$$\left[\frac{\partial h(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]' = \begin{bmatrix} e^{\hat{\mu}} & 0 \end{bmatrix}$$

y por (5.7)

$$\widehat{Var}[h(\hat{\mu},\hat{\sigma})] = e^{2\hat{\mu}}\widehat{Var}(\hat{\mu}).$$

Por lo tanto un intervalo de verosimilitud confianza de $100(1-\alpha)\,\%$ para la mediana estaría dado por

$$e^{\hat{\mu}} \pm z_{(1-\alpha/2)}\sqrt{e^{2\hat{\mu}}\widehat{Var}(\hat{\mu})}.$$

Capítulo 6

Aplicaciones

En los últimos años las regulaciones ambientales y las leyes para la conservación del medio ambiente han sido cada vez más estrictas y exigentes en el rigor analítico sobre los datos de contaminantes. Un fenómeno de relevante importancia han sido los datos que resultan menores a la precisión o a los límites de detección de los protocolos y aparatos de medición. En particular se ha vuelto cada vez más necesario asegurar la calidad del agua a través del análisis de datos en sus mediciones de contaminantes, en donde ha sido recurrente el fenómeno de datos no detectados y límites de detección múltiples.

Para este tipo de datos ambientales obtenido de muestras aleatorias censuradas, los métodos estadísticos convencionales han sido poco eficientes; recurriendose entonces a la teoría y métodos estadísticos de la bioestadística, tales como el análisis de supervivencia y la estimación por máxima verosimilitud en presencia de censura. Los modelos más conocidos en el análisis de supervivencia han sido las distribuciones exponencial, lognormal y Weibull los cuales manifiestan comportamientos estadísticos observados tanto en estudios de contaminación y medio ambiente como en análisis de vida en bioestadística. La metodología presentada en las secciones previas serán ilustradas en el análisis de datos de contaminantes de agua en mantos freáticos de dos sitios ampliamente reportados en la literatura como los son las zonas del Valle de San Joaquín, California.

El análisis de datos obtenidos de las referencias anteriores han sido implementado en el software R y de manera simultánea se irá denotando los comandos, funciones y programas utilizados en cada uno de los conjuntos de datos que se presentan a continuación.

Para el modelo exponencial la implementación de R se ha dispuesto como

sigue:

```
xob
xnd

lv<-function(theta)
{
-(sum(dexp(xob-theta[2],theta[1],TRUE))+sum(pexp(xnd-theta[2],
theta[1],TRUE)))
}

estima<-nlm(lv,c(mean(c(log(xob),log(xnd))),sd(c(log(xob),
log(xnd)))))
estima<-optim(c(mean(c(log(xob),log(xnd))),sd(c(log(xob),
log(xnd)))),lv).
```

En la mayoría de los algoritmos que se utilizarán en este trabajo las funciones `nlm` y `optim` permitieron la obtención de los estimadores de máxima verosimilitud mediante optimización numérica (apéndices C.4 y C.5).

Por otro lado si se desea obtener estimadores a través de otros modelos de probabilidad como la distribución lognormal y weibull, el programa de cómputo quedará redefinido como se muestra a continuación:

```
{
-(sum(dlnorm(xob,theta[1],theta[2],TRUE))+sum(plnorm(xnd,
theta[1],theta[2],TRUE)))
}

{
-(sum(dweibull(xob,theta[1],theta[1],TRUE))+sum(pweibull(xnd,
theta[1],theta[1],TRUE)))
}
```

para la distribución lognormal y Weibull respectivamente.

Para su mayor comprensión y conocimiento de estas funciones implementadas en R se han creado los apartados respectivos en los apéndices C.1, C.2 y C.3.

Ahora estos algoritmos son utilizados para el proceso de estimación e inferencia estadística sobre los datos que se reportan a continuación.

Ejemplo 6.0.1. *En Millar y Deverel (1988) se reportan niveles de zinc y cobre en mantos freáticos muestreados en dos áreas geológicas del Valle San Joaquín en Cal., USA. La tabla muestra los datos, donde 20 % son no detectados.*

Cuadro 6.1: Datos de cobre y de zinc

Alluvial Fan Zone				Basin-Trough Zone			
Copper Conc.	Frec.	Zinc Conc.	Frec.	Copper Conc.	Frec.	Zinc Conc.	Frec.
1 ND	4	3 ND	1	1 ND	2	3 ND	1
5 ND	8	10 ND	15	2 ND	2	10 ND	3
10 ND	3	5	1	5 ND	5	3	2
20 ND	2	7	1	10 ND	4	4	2
1	5	8	1	15 ND	1	5	2
2	21	9	1	1	7	6	1
3	6	10	20	2	4	8	1
4	3	11	2	3	8	10	5
5	3	12	1	4	5	11	1
7	3	17	1	5	1	12	2
8	1	18	1	6	2	13	1
9	1	19	1	8	1	14	1
10	1	20	14	9	2	15	1
11	1	23	1	12	1	17	2
12	1	29	1	14	1	20	11
16	1	30	1	15	1	25	1
20	1	33	1	17	1	30	4
	65	40	1	23	1	40	3
		50	1		49	50	2
		620	1			60	2
			67			70	1
						90	1
							50

Las estimaciones correspondientes al ejemplo anterior han sido calculadas a través de R para cada uno de los sitios y datos señalados en el Cuadro 6.1. A final de las siguientes cuatro secciones se resumen en forma de cuadros las estimaciones de los parámetros y modelos utilizadas.

Ahora con fines de ilustrar y facilitar la comprensión del proceso de estimación elaborado metodológicamente en este trabajo, se presenta el proceso de cálculo, programación y obtención de dichos valores.

6.1. Muestra de cobre en Alluvial Fan Zone

6.1.1. Estimación muestral de la exponencial

```
# Concentraciones no censuradas de cobre (Alluvial Fan Zone)

xob=c(1,1,1,1,1,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,3,3,3,
3,3,3,4,4,4,5,5,5,7,7,7,8,9,10,11,12,16,20)

# Concentraciones censuradas de cobre (Alluvial Fan Zone)

xnd=c(1,1,1,1,5,5,5,5,5,5,5,5,10,10,10,20,20)

# Función de log-verosimilitud censurados por la izquierda
```

$[\ell(\theta,\gamma)]$

```
lv<-function(theta)
{
-(sum(dexp(xob-theta[2],theta[1],log=TRUE))+sum(pexp(xnd-
theta[2],theta[1],log.p=TRUE)))
}

# Estimación mediante optimización (nlm) de la verosimilitud

ve<-nlm(lv,c(1.2,.82),hessian = T)
```

0.3428 0.754

```
# Estimación mediante optimización (optim) de la verosimilitud

estima<-optim(c(1.2,.82),lv,hessian = T)
```

0.343 0.754

```
# Matriz de información de Fisher

ginv(estima$hessian)
```

$$\left[\boldsymbol{I}(\hat{\theta},\hat{\gamma})\right] = \begin{array}{c} \\ \texttt{[1,]} \\ \texttt{[2,]} \end{array} \begin{array}{cc} \texttt{[,1]} & \texttt{[,2]} \\ 0.002 & 0.002 \\ 0.002 & 0.016 \end{array}$$

```
# Valores estimados de la media

m<-c(estima$par[1]+estima$par[2])
```

$\left[\mathrm{g}(\hat{\theta},\hat{\gamma})\right]$ =**1.339**

```
# Valores estimados de la mediana

M<-c(estima$par[2]-estima$par[1]*log(2))
```

$\left[h(\hat{\theta},\hat{\gamma})\right]$ =**0.765**

```
# Estimación de la varianza

(estima$par[1])^2
```

$[Var[Y]]$ =**0.118**

```
# Log-verosimilitud estimada

lv(estima$par)
```

119.516

```
# Error estandar para la media

dm<-matrix(c(1,1),ncol=2)
es<-sqrt(dm%*%(ginv(estima$hessian))%*%t(dm))
```

$\left[\widehat{se}_{\widehat{\mathrm{g}}}\right]$ =**0.148**

```
# Error estandar para la mediana
```

```
dM<-matrix(c(log(0.5),1),ncol=2)
esM<-sqrt(dM%*%(ginv(estima$hessian))%*%t(dM))
```

$\left[\widehat{se}_{\widehat{h}}\right]$ =**0.123**

```
# Intervalo de confianza del 95% para la media

Icm<-c(m-1.95*esm,m+1.95*esm)
```

$\left[\widehat{\mathrm{g}} \pm z_{(1-\alpha/2)}\widehat{se}_{\widehat{\mathrm{g}}}\right]$ =**0.808 1.385**

```
# Intervalo de confianza del 95% para la mediana

IcM<-c(M-1.95*esM,M+1.95*esM)
```

$\left[\widehat{h} \pm z_{(1-\alpha/2)}\widehat{se}_{\widehat{h}}\right]$ =**0.277 0.756**

6.1.2. Estimación muestral de la lognormal

```
# Concentraciones no censuradas de cobre (Alluvial Fan Zone)

xob=c(1,1,1,1,1,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,3,3,
3,3,3,3,4,4,4,5,5,5,7,7,7,8,9,10,11,12,16,20)

# Concentraciones censuradas de cobre(Alluvial Fan Zone)

xnd=c(1,1,1,1,5,5,5,5,5,5,5,5,10,10,10,20,20)

# Función de log-verosimilitud censurados por la izquierda

[ℓ(μ,σ)]

lv<-function(theta)
{
-(sum(dlnorm(xob,theta[1],theta[2],log=TRUE))+sum(plnorm(xnd,
theta[1],theta[2],log.p=TRUE)))
}

# Estimación mediante optimización (nlm) de la verosimilitud

vl<-nlm(lv,c(1.2,.82),hessian=T)
```

0.944 0.801

```
# Estimación mediante optimización (optim) de la verosimilitud

estima<-optim(c(1.2,.82),lv,hessian=T))
```

0.944 0.801

```
# Matriz de información de Fisher

ginv(estima$hessian)
```

$$[\boldsymbol{I}(\hat{\mu},\hat{\sigma})] = \begin{array}{c} \\ \texttt{[1,]} \\ \texttt{[2,]} \end{array} \begin{array}{cc} \texttt{[,1]} & \texttt{[,2]} \\ 0.012 & -0.001 \\ -0.001 & 0.007 \end{array}$$

```
# Valores estimados de la media

m<-c(exp(estima$par[1]+0.5*(estima$par[2])^2))
```

$[\mathrm{g}(\hat{\mu},\hat{\sigma})] =$**3.541**

```
# Valores estimados de la mediana

M<-c(exp(estima$par[1]))
```

$[h(\hat{\mu},\hat{\sigma})] =$**2.570**

```
# Estimación de la varianza

exp(2*estima$par[1]+(estima$par[2])^2)*(exp((estima$par[2])^2
-1))
```

$[Var[Y]] =$**11.268**

```
# Logverosimilitud estimada

lv(estima$par)
```

118.634

```
# Error estandar para la media
```

```
dm<-matrix(c(exp(estima$par[1]+(1/2)*(estima$par[2])^2),
estima$par[2]*exp(estima$par[1]+(1/2)*(estima$par[2])^2)),ncol=
2)
es<-sqrt(dm%*%(ginv(estima$hessian))%*%t(dm))
```

$\left[\widehat{se}_{\widehat{\mathrm{g}}}\right]$ =**0.421**

```
# Error estandar para la mediana

dM<-matrix(c(exp(estima$par[1]),0),ncol=2)
esM<-sqrt(dM%*%(ginv(estima$hessian))%*%t(dM))
```

$\left[\widehat{se}_{\widehat{h}}\right]$ =**0.279**

```
# Intervalo de confianza del 95% para la media

Icm<-c(m-1.95*esm,m+1.95*esM)
```

$\left[\widehat{\mathrm{g}} \pm z_{(1-\alpha/2)}\widehat{se}_{\widehat{\mathrm{g}}}\right]$ =**3.253 3.830**

```
# Intervalo de confianza del 95% para la mediana

IcM<-c(M-1.95*esM,M+1.95*esM)
```

$\left[\widehat{h} \pm z_{(1-\alpha/2)}\widehat{se}_{\widehat{h}}\right]$ =**2.026 3.115**

6.1.3. Estimación muestral de la Weibull

```
# Concentraciones no censuradas de cobre (Alluvial Fan Zone)

xob=c(1,1,1,1,1,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,3,3,3,
3,3,3,4,4,4,5,5,5,7,7,7,8,9,10,11,12,16,20)

# Concentraciones censuradas de cobre (Alluvial Fan Zone)

xnd=c(1,1,1,1,5,5,5,5,5,5,5,5,10,10,10,20,20)

# Función de log-verosimilitud censurados por la izquierda
```

$[\ell(\beta, \eta)]$

```
lv<-function(theta)
{
-(sum(dweibull(xob,theta[1],theta[1],log=TRUE))+sum(pweibull(
xnd,theta[1],theta[1],log.p=TRUE)))
}

# Estimación mediante optimización (nlm) de la verosimilitud

vw<-nlm(lv,c(1.2,4.4))
```

1.125 3.784

```
# Estimación mediante optimización (optim) de la verosimilitud

estima<-optim(c(1.2,4.4),lv)
```

1.125 3.784

```
# Matriz de información de Fisher

ginv(estima$hessian)
```

$$\left[\boldsymbol{I}(\hat{\beta},\hat{\eta})\right] = \begin{array}{c} \\ \texttt{[1,]} \\ \texttt{[2,]} \end{array} \begin{array}{cc} \texttt{[,1]} & \texttt{[,2]} \\ 0.013 & 0.0217 \\ 0.0217 & 0.224 \end{array}$$

```
# Valores estimados de la media

m<-estima$par[2]*gamma(1+1/estima$par[1])
```

$\left[\mathrm{g}(\hat{\beta},\hat{\eta})\right]$ =**3.626**

```
# Valores estimados de la mediana

M<-estima$par[2]*(-log(1/2))^(1/estima$par[1])
```

$\left[h(\hat{\beta},\hat{\eta})\right]$ =**2.732**

```
# Estimación de la varianza

estima$par[2]^2*(gamma(1+2/estima$par[1])-(gamma(1+1/estima$
par[1]))^2)
```

$[Var(Y)]$ =**10.425**

```
# Logverosimilitud estimada

lv(estima$par)
```

124.288

```
# Error estandar para la media

b<-estima$par[1]; n<-estima$par[2]
g<-expression(n*gamma(1+1/b))
dgb<-D(g,"b")); (dg1<-eval(dgb)
dgn<-D(g,"n"));(dg2<-eval(dgn)
dm<-matrix(c(dg1,dg2),ncol=2)
esm<-sqrt(dm%*%(ginv(estima$hessian))%*%t(dm))
```

$\left[\widehat{se}_{\widehat{\mathrm{g}}}\right]$ =**0.420**

```
# Error estandar para la mediana

h<-expression(n*(-log(1/2))^(1/b))
dhb<-D(h,"b")); (dh1<-eval(dhb)
dhn<-D(h,"n"));(dh2<-eval(dhn)
dM<-matrix(c(dh1,dh2),ncol=2)
esM<-sqrt(dM%*%(ginv(estima$hessian))%*%t(dM))
```

$\left[\widehat{se}_{\widehat{h}}\right]$ =**0.387**

```
# Intervalo de confianza del 95% para la media

Icm<-c(m-1.95*esm,m+1.95*esM)
```

$\left[\widehat{\mathrm{g}} \pm z_{(1-\alpha/2)}\widehat{se}_{\widehat{\mathrm{g}}}\right]$ =**2.806 4.445**

```
# Intervalo de confianza del 95% para la mediana

IcM<-c(M-1.95*esM,M+1.95*esM)
```

$\left[\widehat{h} \pm z_{(1-\alpha/2)}\widehat{se}_{\widehat{h}}\right]$ =**1.978 3.486**

6.2. Muestra de zinc en Alluvial Fan Zone

6.2.1. Estimación muestral de la exponencial

```
# Concentraciones no censuradas de zinc (Alluvial Fan Zone)

xob=c(5,7,8,9,10,10,10,10,10,10,10,10,10,10,10,10,10,10,10,10,
10,10,10,10,11,11,12,17,18,19,20,20,20,20,20,20,20,20,20,20,20,
20,20,20,23,29,30,33,40,50,620)

# Concentraciones censuradas de zinc (Alluvial Fan Zone)

xnd=c(3,10,10,10,10,10,10,10,10,10,10,10,10,10,10,10,10)

# Función de log-verosimilitud censurados por la izquierda
```

$[\ell(\theta,\gamma)]$

```
lv<-function(theta)
{
-(sum(dexp(xob-theta[2],theta[1],log=TRUE))+sum(pexp(xnd-
theta[2],theta[1],log.p=TRUE)))
}

# Estimación mediante optimización (nlm) de la verosimilitud

ve<-nlm(lv,c(1.2,.82),hessian = T)
```

0.049 1.975

```
# Estimación mediante optimización (optim) de la verosimilitud

estima<-optim(c(1.2,.82),lv,hessian = T)
```

0.049 1.977

```
# Matriz de información de Fisher

ginv(estima$hessian)
```

$$\left[\boldsymbol{I}(\hat{\theta},\hat{\gamma})\right] = \begin{array}{c} \\ \texttt{[1,]} \\ \texttt{[2,]} \end{array} \begin{array}{c} \texttt{[,1]} \\ 3.926e-05 \\ 0.002 \end{array} \begin{array}{c} \texttt{[,2]} \\ 0.002 \\ 0.937 \end{array}$$

```
# Valores estimados de la media

m<-c(estima$par[1]+estima$par[2])
```

$\left[\mathrm{g}(\hat{\theta},\hat{\gamma})\right]$ =**2.026**

```
# Valores estimados de la mediana

M<-c(estima$par[2]-estima$par[1]*log(2))
```

$\left[h(\hat{\theta},\hat{\gamma})\right]$ =**1.944**

```
# Estimación de la varianza

(estima$par[1])^2
```

$[Var[Y]]$ =**0.002**

```
# Logverosimilitud estimada

lv(estima$par)
```

238.423

```
# Error estandar para la media

dm<-matrix(c(1,1),ncol=2)
es<-sqrt(dm%*%(ginv(estima$hessian))%*%t(dm))
```

$\left[\widehat{se}_{\widehat{\mathrm{g}}}\right]$ =**0.970**

```
# Error estandar para la mediana

dM<-matrix(c(log(0.5),1),ncol=2)
esM<-sqrt(dM%*%(ginv(estima$hessian))%*%t(dM))
```

$\left[\widehat{se}_{\widehat{h}}\right]$ =**0.967**

```
# Intervalo de confianza del 95% para la media

Icm<-c(m-1.95*esm,m+1.95*esm)
```

$\left[\widehat{\mathrm{g}} \pm z_{(1-\alpha/2)}\widehat{se}_{\widehat{\mathrm{g}}}\right]$ =**1.738 2.314**

```
# Intervalo de confianza del 95% para la mediana

IcM<-c(M-1.95*esM,M+1.95*esM)
```

$\left[\widehat{h} \pm z_{(1-\alpha/2)}\widehat{se}_{\widehat{h}}\right]$ =**0.0589 3.829**

6.2.2. Estimación muestral de la lognormal

```
# Concentraciones no censuradas de zinc (Alluvial Fan Zone)

xob=c(5,7,8,9,10,10,10,10,10,10,10,10,10,10,10,10,10,10,10,10,
10,10,10,10,11,11,12,17,18,19,20,20,20,20,20,20,20,20,20,20,20,
20,20,20,23,29,30,33,40,50,620)

# Concentraciones censuradas de zinc (Alluvial Fan Zone)

xnd=c(3,10,10,10,10,10,10,10,10,10,10,10,10,10,10,10,10)

# Función de log-verosimilitud censurados por la izquierda
```

$[\ell(\mu,\sigma)]$

```
lv<-function(theta)
{
-(sum(dlnorm(xob,theta[1],theta[2],log=TRUE))+sum(plnorm(xnd,
theta[1],theta[2],log.p=TRUE)))
}

# Estimación mediante optimización (nlm) de la verosimilitud

vl<-nlm(lv,c(1.2,.82),hessian=T)
```

2.475 0.802

```
# Estimación mediante optimización (optim) de la verosimilitud

estima<-optim(c(1.2,.82),lv,hessian=T)
```

2.475 0.802

```
# Matriz de información de Fisher

ginv(estima$hessian)
```

$$[\boldsymbol{I}(\hat{\mu},\hat{\sigma})] = \begin{array}{c} \\ \texttt{[1,]} \\ \texttt{[2,]} \end{array} \begin{array}{cc} \texttt{[,1]} & \texttt{[,2]} \\ 0.010 & -0.001 \\ -0.001 & 0.007 \end{array}$$

```
# Valores estimados de la media

m<-c(exp(estima$par[1]+0.5*(estima$par[2])^2))
```

$[\mathrm{g}(\hat{\mu}, \hat{\sigma})]$ =**16.380**

```
# Valores estimados de la mediana

M<-c(exp(estima$par[1]))
```

$[h(\hat{\mu}, \hat{\sigma})]$ =**11.876**

```
# Estimación de la varianza

exp(2*estima$par[1]+(estima$par[2])^2)*(exp((estima$par[2])^2
-1))
```

$[Var[Y]]$ =**242.157**

```
# Logverosimilitud estimada


lv(estima$par)
```

213.312

```
# Error estandar para la media

dm<-matrix(c(exp(estima$par[1]+(1/2)*(estima$par[2])^2),
estima$par[2]*exp(estima$par[1]+(1/2)*(estima$par[2])^2)),ncol=
2)
es<-sqrt(dm%*%(ginv(estima$hessian))%*%t(dm))
```

$\left[\widehat{se}_{\widehat{\mathrm{g}}}\right]$ =**1.852**

```
# Error estandar para la mediana

dM<-matrix(c(exp(estima$par[1]),0),ncol=2)
esM<-sqrt(dM%*%(ginv(estima$hessian))%*%t(dM))
```

$\left[\widehat{se}_{\widehat{h}}\right]$ =**1.225**

```
# Intervalo de confianza del 95% para la media

Icm<-c(m-1.95*esm,m+1.95*esM)
```

$\left[\widehat{\mathrm{g}} \pm z_{(1-\alpha/2)}\widehat{se}_{\widehat{\mathrm{g}}}\right]$ =**16.092 16.669**

```
# Intervalo de confianza del 95% para la mediana

IcM<-c(M-1.95*esM,M+1.95*esM)
```

$\left[\widehat{h} \pm z_{(1-\alpha/2)}\widehat{se}_{\widehat{h}}\right]$ =**9.488 14.264**

6.2.3. Estimación muestral de la Weibull

```
# Concentraciones no censuradas de zinc (Alluvial Fan Zone)

xob=c(5,7,8,9,10,10,10,10,10,10,10,10,10,10,10,10,10,10,10,10,
10,10,10,10,11,11,12,17,18,19,20,20,20,20,20,20,20,20,20,20,
20,20,20,20,23,29,30,33,40,50,620)

# Concentraciones censuradas de zinc (Alluvial Fan Zone)

xnd=c(3,10,10,10,10,10,10,10,10,10,10,10,10,10,10,10,10)

# Función de log-verosimilitud censurados por la izquierda
```

$[\ell(\beta,\eta)]$

```
lv<-function(theta)
{
-(sum(dweibull(xob,theta[1],theta[2],log=TRUE))+sum(pweibull(
xnd,theta[1],theta[2],log.p=TRUE)))
}

# Estimación mediante optimización (nlm) de la verosimilitud

vw<-nlm(lv,c(0.8,23.5))
```

0.756 16.834

```
# Estimación mediante optimización (optim) de la verosimilitud

estima<-optim(c(0.8,23.5),lv)
```

0.756 16.834

```
# Matriz de información de Fisher

ginv(estima$hessian)
```

$$\left[\boldsymbol{I}(\hat{\beta},\hat{\eta})\right] = \begin{array}{c} \\ \texttt{[1,]} \\ \texttt{[2,]} \end{array} \begin{array}{cc} \texttt{[,1]} & \texttt{[,2]} \\ 0.004 & 0.069 \\ 0.069 & 8.711 \end{array}$$

```
# Valores estimados de la media

m<-estima$par[2]*gamma(1+1/estima$par[1])
```

$\left[\mathrm{g}(\hat{\beta},\hat{\eta})\right]$ =**19.918**

```
# Valores estimados de la mediana

M<-estima$par[2]*(-log(1/2))^(1/estima$par[1])
```

$\left[h(\hat{\beta},\hat{\eta})\right]$ =**10.365**

```
# Estimación de la varianza

estima$par[2]^2*(gamma(1+2/estima$par[1])-(gamma(1+1/estima$
par[1]))^2)
```

$[Var(Y)]$ =**713.820**

```
# Logverosimilitud estimada

lv(estima$par)
```

231.598

```
# Error estandar para la media

b<-estima$par[1]; n<-estima$par[2]
g<-expression(n*gamma(1+1/b))
dgb<-D(g,"b")); (dg1<-eval(dgb)
dgn<-D(g,"n"));(dg2<-eval(dgn)
dm<-matrix(c(dg1,dg2),ncol=2)
esm<-sqrt(dm%*%(ginv(estima$hessian))%*%t(dm))
```

$\left[\widehat{se}_{\widehat{\mathrm{g}}}\right]$ =**3.233**

```
# Error estandar para la mediana

h<-expression(n*(-log(1/2))^(1/b))
dhb<-D(h,"b")); (dh1<-eval(dhb)
dhn<-D(h,"n"));(dh2<-eval(dhn)
dM<-matrix(c(dh1,dh2),ncol=2)
esM<-sqrt(dM%*%(ginv(estima$hessian))%*%t(dM))
```

$\left[\widehat{se}_{\widehat{h}}\right]$ =**2.011**

```
# Intervalo de confianza del 95% para la media

Icm<-c(m-1.95*esm,m+1.95*esM)
```

$\left[\widehat{\mathrm{g}} \pm z_{(1-\alpha/2)}\widehat{se}_{\widehat{\mathrm{g}}}\right]$ =**13.613 26.221**

```
# Intervalo de confianza del 95% para la mediana

IcM<-c(M-1.95*esM,M+1.95*esM)
```

$\left[\widehat{h} \pm z_{(1-\alpha/2)}\widehat{se}_{\widehat{h}}\right]$ =**6.445 14.286**

6.3. Muestra de cobre en Basin-Trough Zone

6.3.1. Estimación muestral de la exponencial

```
# Concentraciones no censuradas de cobre (Basin-Trough Zone)

xob=c(1,1,1,1,1,1,1,2,2,2,2,3,3,3,3,3,3,3,4,4,4,4,4,5,6,6,8,9,
9,12,14,15,17,23)

# Concentraciones censuradas de cobre (Basin-Trough Zone)

xnd=c(1,1,2,2,5,5,5,5,5,10,10,10,10,15)

# Función de log-verosimilitud censurados por la izquierda

[ℓ(θ,γ)]

lv<-function(theta)
{
-(sum(dexp(xob-theta[2],theta[1],log=TRUE))+sum(pexp(xnd-
theta[2],theta[1],log.p=TRUE)))
}

# Estimación mediante optimización (nlm) de la verosimilitud

ve<-nlm(lv,c(1.2,.82),hessian = T)
```

0.269 0.737

```
# Estimación mediante optimización (optim) de la verosimilitud

estima<-optim(c(1.2,.82),lv,hessian = T)
```

0.269 0.737

```
# Matriz de información de Fisher

ginv(estima$hessian)
```

$$\left[\boldsymbol{I}(\hat{\theta},\hat{\gamma})\right] = \begin{array}{c} \\ \texttt{[1,]} \\ \texttt{[2,]} \end{array} \begin{array}{cc} \texttt{[,1]} & \texttt{[,2]} \\ 0.002 & 0.002 \\ 0.002 & 0.036 \end{array}$$

```
# Valores estimados de la media

m<-c(estima$par[1]+estima$par[2])
```

$\left[g(\hat{\theta},\hat{\gamma})\right]$ =**1.006**

```
# Valores estimados de la mediana

M<-c(estima$par[2]-estima$par[1]*log(2))
```

$\left[h(\hat{\theta},\hat{\gamma})\right]$ =**0.551**

```
# Estimación de la varianza

(estima$par[1])^2
```

$[Var[Y]]$ =**0.072**

```
# Logverosimilitud estimada

lv(estima$par)
```

98.380

```
# Error estandar para la media

dm<-matrix(c(1,1),ncol=2)
es<-sqrt(dm%*%(ginv(estima$hessian))%*%t(dm))
```

$\left[\widehat{se}_{\widehat{\mathrm{g}}}\right]$ =**0.204**

```
# Error estandar para la mediana

dM<-matrix(c(log(0.5),1),ncol=2)
esM<-sqrt(dM%*%(ginv(estima$hessian))%*%t(dM))
```

$\left[\widehat{se}_{\widehat{h}}\right]$ =**0.183**

```
# Intervalo de confianza del 95% para la media

Icm<-c(m-1.95*esm,m+1.95*esm)
```

$\left[\widehat{\mathrm{g}} \pm z_{(1-\alpha/2)}\widehat{se}_{\widehat{\mathrm{g}}}\right]$ =**0.718 1.294**

```
# Intervalo de confianza del 95% para la mediana

IcM<-c(M-1.95*esM,M+1.95*esM)
```

$\left[\widehat{h} \pm z_{(1-\alpha/2)}\widehat{se}_{\widehat{h}}\right]$ =**0.195 0.907**

6.3.2. Estimación muestral de la lognormal

```
# Concentraciones no censuradas de cobre (Basin-Trough Zone)

xob=c(1,1,1,1,1,1,1,2,2,2,2,3,3,3,3,3,3,3,4,4,4,4,4,5,6,6,8,
9,9,12,14,15,17,23)

# Concentraciones censuradas de cobre (Basin-Trough Zone)

xnd=c(1,1,2,2,5,5,5,5,5,10,10,10,10,15)

# Función de log-verosimilitud censurados por la izquierda

[ℓ(μ, σ)]

lv<-function(theta)
{
-(sum(dlnorm(xob,theta[1],theta[2],log=TRUE))+sum(plnorm(xnd,
theta[1],theta[2],log.p=TRUE)))
}

# Estimación mediante optimización (nlm) de la verosimilitud

vl<-nlm(lv,c(1.2,.82),hessian=T)
```

1.033 0.936

```
# Estimación mediante optimización (optim) de la verosimilitud

estima<-optim(c(1.2,.82),lv,hessian=T)
```

1.033 0.936

```
# Matriz de información de Fisher

ginv(estima$hessian)
```

$$[\boldsymbol{I}(\hat{\mu},\hat{\sigma})] = \begin{array}{c} \\ \texttt{[1,]} \\ \texttt{[2,]} \end{array} \begin{array}{cc} \texttt{[,1]} & \texttt{[,2]} \\ 0.022 & -0.003 \\ -0.003 & 0.012 \end{array}$$

```
# Valores estimados de la media

m<-c(exp(estima$par[1]+0.5*(estima$par[2])^2))
```

$[\mathrm{g}(\hat{\mu},\hat{\sigma})]$ =**4.352**

```
# Valores estimados de la mediana

M<-c(exp(estima$par[1]))
```

$[h(\hat{\mu},\hat{\sigma})]$ =**2.810**

```
# Estimación de la varianza

exp(2*estima$par[1]+(estima$par[2])^2)*(exp((estima$par[2])^2)
-1)
```

$[Var[Y]]$ =**26.515**

```
# Logverosimilitud estimada

lv(estima$par)
```

98.409

```
# Error estandar para la media
```

```
dm<-matrix(c(exp(estima$par[1]+(1/2)*(estima$par[2])^2),
estima$par[2]*exp(estima$par[1]+(1/2)*(estima$par[2])^2)),ncol=
2)
es<-sqrt(dm%*%(ginv(estima$hessian))%*%t(dm))
```

$\left[\widehat{se}_{\widehat{\mathrm{g}}}\right]$ =**0.718**

```
# Error estandar para la mediana

dM<-matrix(c(exp(estima$par[1]),0),ncol=2)
(esM<-sqrt(dM%*%(ginv(estima$hessian))%*%t(dM)))
```

$\left[\widehat{se}_{\widehat{h}}\right]$ =**0.413**

```
# Intervalo de confianza del 95% para la media

Icm<-c(m-1.95*esm,m+1.95*esM)
```

$\left[\widehat{\mathrm{g}} \pm z_{(1-\alpha/2)}\widehat{se}_{\widehat{\mathrm{g}}}\right]$ =**4.064 4.641**

```
# Intervalo de confianza del 95% para la mediana

IcM<-c(M-1.95*esM,M+1.95*esM)
```

$\left[\widehat{h} \pm z_{(1-\alpha/2)}\widehat{se}_{\widehat{h}}\right]$ =**2.004 3.615**

6.3.3. Estimación muestral de la Weibull

```
# Concentraciones no censuradas de cobre (Basin-Trough Zone)

xob=c(1,1,1,1,1,1,1,2,2,2,2,3,3,3,3,3,3,3,3,4,4,4,4,4,5,6,6,8,9,
9,12,14,15,17,23)

# Concentraciones censuradas de cobre (Basin-Trough Zone)

xnd=c(1,1,2,2,5,5,5,5,5,10,10,10,10,15)

# Función de log-verosimilitud censurados por la izquierda
```

$[\ell(\mu, \sigma)]$

```
lv<-function(theta)
{
-(sum(dweibull(xob,theta[1],theta[2],log=TRUE))+sum(pweibull(
xnd,theta[1],theta[2],log.p=TRUE)))
}

# Estimación mediante optimización (nlm) de la verosimilitud

vw<-nlm(lv,c(1.1,5.5))
```

1.016 4.380

```
# Estimación mediante optimización (optim) de la verosimilitud

estima<-optim(c(1.1,5.5),lv)
```

1.016 4.380

```
# Matriz de información de Fisher

ginv(estima$hessian)
```

$$\left[\boldsymbol{I}(\hat{\beta},\hat{\eta})\right] = \begin{array}{c} \\ \texttt{[1,]} \\ \texttt{[2,]} \end{array} \begin{array}{cc} \texttt{[,1]} & \texttt{[,2]} \\ 0.015 & 0.036 \\ 0.036 & 0.485 \end{array}$$

```
# Valores estimados de la media

m<-estima$par[2]*gamma(1+1/estima$par[1])
```

$\left[\mathrm{g}(\hat{\beta},\hat{\eta})\right]$ =**4.351**

```
# Valores estimados de la mediana

M<-estima$par[2]*(-log(1/2))^(1/estima$par[1])
```

$\left[h(\hat{\beta},\hat{\eta})\right]$ =**3.054**

```
# Estimación de la varianza

estima$par[2]^2*(gamma(1+2/estima$par[1])-(gamma(1+1/estima$
par[1]))^2)
```

$[Var(Y)]$ =**18.125**

```
# Logverosimilitud estimada

lv(estima$par)
```

18.333

```
# Error estandar para la media

b<-estima$par[1]; n<-estima$par[2]
g<-expression(n*gamma(1+1/b))
dgb<-D(g,"b")); (dg1<-eval(dgb)
dgn<-D(g,"n"));(dg2<-eval(dgn)
dm<-matrix(c(dg1,dg2),ncol=2)
esm<-sqrt(dm%*%(ginv(estima$hessian))%*%t(dm))
```

$\left[\widehat{se}_{\widehat{\mathrm{g}}}\right]$ =**0.633**

```
# Error estandar para la mediana

h<-expression(n*(-log(1/2))^(1/b))
dhb<-D(h,"b")); (dh1<-eval(dhb)
dhn<-D(h,"n"));(dh2<-eval(dhn)
dM<-matrix(c(dh1,dh2),ncol=2)
esM<-sqrt(dM%*%(ginv(estima$hessian))%*%t(dM))
```

$\left[\widehat{se}_{\widehat{h}}\right]$ =**0.556**

```
# Intervalo de confianza del 95% para la media

Icm<-c(m-1.95*esm,m+1.95*esM)
```

$\left[\widehat{\mathrm{g}} \pm z_{(1-\alpha/2)}\widehat{se}_{\widehat{\mathrm{g}}}\right]$ =**3.117 5.585**

```
# Intervalo de confianza del 95% para la mediana

IcM<-c(M-1.95*esM,M+1.95*esM)
```

$\left[\widehat{h} \pm z_{(1-\alpha/2)}\widehat{se}_{\widehat{h}}\right]$ =**1.971 4.137**

6.4. Muestra de zinc en Basin-Trough Zone

6.4.1. Estimación muestral de la exponencial

```
# Concentraciones no censuradas de zinc (Basin-Trough Zone)

xob=c(3,3,4,4,5,5,6,8,10,10,10,10,10,11,12,12,13,14,15,17,17,20,
20,20,20,20,20,20,20,20,20,20,20,25,30,30,30,30,40,40,40,50,50,
60,60,70,90)

# Concentraciones censuradas de zinc (Basin-Trough Zone)

xnd=c(3,10,10,10)

# Función de log-verosimilitud censurados por la izquierda
```

$[\ell(\theta,\gamma)]$

```
lv<-function(theta)
{
-(sum(dexp(xob-theta[2],theta[1],log=TRUE))+sum(pexp(xnd-
theta[2],theta[1],log.p=TRUE)))
}

# Estimación mediante optimización (nlm) de la verosimilitud

ve<-nlm(lv,c(1.2,.82),hessian = T)
```

0.0523 2.536

```
# Estimación mediante optimización (optim) de la verosimilitud

estima<-optim(c(1.2,.82),lv,hessian = T)
```

0.0526 2.646

```
# Matriz de información de Fisher

ginv(estima$hessian)
```

$$\left[\boldsymbol{I}(\hat{\theta},\hat{\gamma})\right] = \begin{array}{c} \\ \texttt{[1,]} \\ \texttt{[2,]} \end{array} \begin{array}{cc} \texttt{[,1]} & \texttt{[,2]} \\ 5.507e-05 & 3.353e-04 \\ 3.353e-04 & 0.127 \end{array}$$

```
# Valores estimados de la media
m<-c(estima$par[1]+estima$par[2])
```

$\left[\mathrm{g}(\hat{\theta},\hat{\gamma})\right]$ =**2.698**

```
# Valores estimados de la mediana
M<-c(estima$par[2]-estima$par[1]*log(2))
```

$\left[h(\hat{\theta},\hat{\gamma})\right]$ =**2.609**

```
# Estimación de la varianza
(estima$par[1])^2
```

$[Var[Y]]$ =**0.003**

```
# Logverosimilitud estimada

lv(estima$par)
```

196.297

```
# Error estandar para la media
dm<-matrix(c(1,1),ncol=2)
es<-sqrt(dm%*%(ginv(estima$hessian))%*%t(dm))
```

$\left[\widehat{se}_{\widehat{\mathrm{g}}}\right]$ =**0.357**

```
# Error estandar para la mediana
dM<-matrix(c(log(0.5),1),ncol=2)
esM<-sqrt(dM%*%(ginv(estima$hessian))%*%t(dM))
```

$\left[\widehat{se}_{\widehat{h}}\right]$ =**0.355**

```
# Intervalo de confianza del 95% para la media
Icm<-c(m-1.95*esm,m+1.95*esm)
```

$\left[\widehat{\mathrm{g}} \pm z_{(1-\alpha/2)}\widehat{se}_{\widehat{\mathrm{g}}}\right]$ =**2.410 2.986**

```
# Intervalo de confianza del 95% para la mediana
IcM<-c(M-1.95*esM,M+1.95*esM)
```

$\left[\widehat{h} \pm z_{(1-\alpha/2)}\widehat{se}_{\widehat{h}}\right]$ =**1.916 3.302**

6.4.2. Estimación muestral de la lognormal

```
# Concentraciones no censuradas de zinc (Basin-Trough Zone)

xob=c(3,3,4,4,5,5,6,8,10,10,10,10,10,11,12,12,13,14,15,17,17,20,
20,20,20,20,20,20,20,20,20,20,20,25,30,30,30,30,40,40,40,50,50,
60,60,70,90)

# Concentraciones censuradas de zinc (Basin-Trough Zone)

xnd=c(3,10,10,10)

# Función de log-verosimilitud censurados por la izquierda

[ℓ(μ,σ)]

lv<-function(theta)
{
-(sum(dlnorm(xob,theta[1],theta[2],log=TRUE))+sum(plnorm(xnd,
theta[1],theta[2],log.p=TRUE)))
}

# Estimación mediante optimización (nlm) de la verosimilitud

vl<-nlm(lv,c(1.2,.82),hessian=T)
```

2.727 0.876

```
# Estimación mediante optimización (optim) de la verosimilitud

estima<-optim(c(1.2,.82),lv,hessian=T)
```

2.727 0.876

```
# Matriz de información de Fisher

ginv(estima$hessian)
```

$$[\boldsymbol{I}(\hat{\mu},\hat{\sigma})] = \begin{array}{c} \\ \texttt{[1,]} \\ \texttt{[2,]} \end{array} \begin{array}{cc} \texttt{[,1]} & \texttt{[,2]} \\ 0.015 & -4.954e-4 \\ -4.954e-4 & 0.008 \end{array}$$

```
# Valores estimados de la media
m<-c(exp(estima$par[1]+0.5*(estima$par[2])^2))
```

$[g(\hat{\mu},\hat{\sigma})]$ =**22.439**

```
# Valores estimados de la mediana
M<-c(exp(estima$par[1]))
```

$[h(\hat{\mu},\hat{\sigma})]$ =**15.289**

```
# Estimación de la varianza
exp(2*estima$par[1]+(estima$par[2])^2)*(exp((estima$par[2])^2)
-1)
```

$[Var[Y]]$ =**581.016**

```
# Logverosimilitud estimada

lv(estima$par)
```

197.596

```
# Error estandar para la media
dm<-matrix(c(exp(estima$par[1]+(1/2)*(estima$par[2])^2),
estima$par[2]*exp(estima$par[1]+(1/2)*(estima$par[2])^2)),ncol=
2)
es<-sqrt(dm%*%(ginv(estima$hessian))%*%t(dm))
```

$\left[\widehat{se}_{\widehat{g}}\right]$ =**3.246**

```
# Error estandar para la mediana
dM<-matrix(c(exp(estima$par[1]),0),ncol=2)
esM<-sqrt(dM%*%(ginv(estima$hessian))%*%t(dM))
```

$\left[\widehat{se}_{\widehat{h}}\right]$ =**1.895**

```
# Intervalo de confianza del 95% para la media
Icm<-c(m-1.95*esm,m+1.95*esM)
```

$\left[\widehat{g} \pm z_{(1-\alpha/2)}\widehat{se}_{\widehat{g}}\right]$ =**22.151 22.727**

```
# Intervalo de confianza del 95% para la mediana
IcM<-c(M-1.95*esM,M+1.95*esM)
```

$\left[\widehat{h} \pm z_{(1-\alpha/2)}\widehat{se}_{\widehat{h}}\right]$ =**11.595 18.984**

6.4.3. Estimación muestral de la Weibull

```
# Concentraciones no censuradas de zinc (Basin-Trough Zone)

xob=c(3,3,4,4,5,5,6,8,10,10,10,10,10,11,12,12,13,14,15,17,17,20,
20,20,20,20,20,20,20,20,20,20,20,25,30,30,30,30,40,40,40,50,50,
60,60,70,90)

# Concentraciones censuradas de zinc (Basin-Trough Zone)

xnd=c(3,10,10,10)

# Función de log-verosimilitud censurados por la izquierda

[ℓ(β,η)]

lv<-function(theta)
{
-(sum(dweibull(xob,theta[1],theta[2],log=TRUE))+sum(pweibull(xnd,
theta[1],theta[2],log.p=TRUE)))
}

# Estimación mediante optimización (nlm) de la verosimilitud

vw<-nlm(lv,c(1.3,25.3))

1.231 23.191

# Estimación mediante optimización (optim) de la verosimilitud

estima<-optim(c(1.3,25.3),lv)

1.231 23.187

# Matriz de información de Fisher

ginv(estima$hessian)
```

$$\left[\boldsymbol{I}(\hat{\beta},\hat{\eta})\right] = \begin{array}{c} \\ \texttt{[1,]} \\ \texttt{[2,]} \end{array} \begin{array}{cc} \texttt{[,1]} & \texttt{[,2]} \\ 0.018 & 0.127 \\ 0.127 & 7.846 \end{array}$$

```
# Valores estimados de la media

m<-estima$par[2]*gamma(1+1/estima$par[1])
```

$\left[\mathrm{g}(\hat{\beta},\hat{\eta})\right]$=**21.672**

```
# Valores estimados de la mediana

M<-estima$par[2]*(-log(1/2))^(1/estima$par[1])
```

$\left[h(\hat{\beta},\hat{\eta})\right]$ =**17.218**

```
# Estimación de la varianza

estima$par[2]^2*(gamma(1+2/estima$par[1])-(gamma(1+1/estima$
par[1]))^2)
```

$[Var(Y)]$ =**313.178**

```
# Logverosimilitud estimada

lv(estima$par)
```

198.001.

```
# Error estandar para la media

b<-estima$par[1]; n<-estima$par[2]
g<-expression(n*gamma(1+1/b))
dgb<-D(g,"b")); (dg1<-eval(dgb)
dgn<-D(g,"n"));(dg2<-eval(dgn)
dm<-matrix(c(dg1,dg2),ncol=2)
esm<-sqrt(dm%*%(ginv(estima$hessian))%*%t(dm))
```

$\left[\widehat{se}_{\widehat{\mathrm{g}}}\right]$ =**2.485**

```
# Error estandar para la mediana

h<-expression(n*(-log(1/2))^(1/b))
dhb<-D(h,"b")); (dh1<-eval(dhb)
dhn<-D(h,"n"));(dh2<-eval(dhn)
dM<-matrix(c(dh1,dh2),ncol=2)
esM<-sqrt(dM%*%(ginv(estima$hessian))%*%t(dM))
```

$\left[\widehat{se}_{\widehat{h}}\right]$=**2.329**

```
# Intervalo de confianza del 95% para la media
```

```
Icm<-c(m-1.95*esm,m+1.95*esM)
```

$\left[\widehat{\mathrm{g}} \pm z_{(1-\alpha/2)}\widehat{se}_{\widehat{\mathrm{g}}}\right]$=**16.826 26.518**

```
# Intervalo de confianza del 95% para la mediana
```

```
IcM<-c(M-1.95*esM,M+1.95*esM)
```

$\left[\widehat{h} \pm z_{(1-\alpha/2)}\widehat{se}_{\widehat{h}}\right]$ =**12.677 21.759**.

En las subsecciones anteriores se han ilustrado los procedimientos analíticos y numéricos para la obtención de las estimaciones de los parámetros en cada uno de los modelos considerados. Han sido de gran ayuda los recursos numéricos que provee el software *R*, por ello las estimaciones realizadas se muestran junto con los comandos y funciones propios de este lenguaje computacional.

A continuación, arreglados en formatos de cuadros, se ilustran y resumen las estimaciones respectivas a los parámetros de cada modelo, junto con sus intervalos de verosimilitud confianza.

Cuadro 6.2: Estimaciones para los parámetros del modelo exponencial

Zona	Elemento	EMV $(\hat{\theta}, \hat{\gamma})$	Media	Mediana
Alluvial Fan	*Cobre*	0.343 0.754	1.097 (0.808,1.385)	0.516 (0.277,0.756)
Alluvial Fan	*Zinc*	0.049 1.977	2.026 (1.738,2.314)	1.944 (0.059,3.829)
Basin-Trough	*Cobre*	0.269 0.737	1.006 (0.718,1.294)	0.551 (0.195,0.907)
Basin-Trough	*Zinc*	0.0526 2.646	2.698 (2.410,2.986)	2.609 (1.916,3.302)

Cuadro 6.3: Estimaciones para los parámetros del modelo lognormal

Zona	Elemento	EMV $(\hat{\mu}, \hat{\sigma})$	Media	Mediana
Alluvial Fan	*Cobre*	0.944 0.801	3.541 (3.253,3.830)	2.570 (2.026,3.115)
Alluvial Fan	*Zinc*	2.474 0.802	16.380 (16.092,16.669)	11.876 (9.488,14.264)
Basin-Trough	*Cobre*	1.033 0.936	4.353 (4.064,4.641)	2.810 (2.004,3.615)
Basin-Trough	*Zinc*	2.727 0.876	22.439 (22.151,22.727)	15.289 (11.595,18.984)

Cuadro 6.4: Estimaciones para los parámetros del modelo Weibull

Zona	Elemento	EMV $(\hat{\beta}, \hat{\eta})$	Media	Mediana
Alluvial Fan	*Cobre*	1.125 3.784	3.626 (2.806,4.445)	2.732 (1.978,3.486)
Alluvial Fan	*Zinc*	0.756 16.834	19.918 (13.613,26.223)	10.365 (6.445,14.286)

estas zonas contaminadas. Para ello será necesario considerar las log-verosimilitutes estimadas en cada modelo, los cuales se ilustran en el siguiente cuadro.

Cuadro 6.5: Log-verosimilitudes estimadas para cada modelo y zona

Zona	Elemento	Exponencial	Lognormal	Weibull
Alluvial Fan	*Cobre*	119.5159	118.6343	124.2878
Alluvial Fan	*Zinc*	238.4226	213.3117	231.5982
Basin-Trough	*Cobre*	98.3801	98.40879	101.0388
Basin-Trough	*Zinc*	196.2972	197.5964	198.0013

Para la selección del modelo se utilizarán dos criterios utilizados bajo modelación con datos censurados, el *Criterio de Información de Akaike* (AIC) y *la prueba de razón de verosimilitudes*, ambos proveen resultados equivalentemente eficientes y su implementación no es difícil.

6.5.1. Criterio de Información de Akaike

Bajo este criterio, la selección del mejor modelo entre dos o más, se realiza utilizando un valor AIC, el cual servirá de referencia y comparativo entre los posibles modelos. Este valor está definido como sigue

$$AIC \quad = \quad 2k - 2ln[L(\hat{\boldsymbol{\theta}})]$$

donde $ln[L(\hat{\boldsymbol{\theta}})]$ no es más que la log-verosimilitud estimada y k el número de parámetros de cada modelo.

Ahora, de las estimaciones obtenidas y dadas en el Cuadro 6.5 se obtienen los siguientes resultados, del cual se observará el mejor modelo en función del valor AIC más pequeño.

Cuadro 6.6: Modelo-AIC para el cobre en Alluvial Fan Zone

Modelo	k	$\ell(\hat{\boldsymbol{\theta}})$	AIC
Exponencial	2	119.5159	-235.0318
Lognormal	2	118.6343	-233.2686
Weibull	2	124.2878	-244.5756

De lo anterior puede observarse que el modelo más adecuado para los datos de concentración de cobre en esta zona es la distribución Weibull con parámetros $\hat{\beta} = 1.125$ y $\hat{\eta} = 3.784$. De esta manera ilustrativa puede notarse que la distribución exponencial con parámetros estimados $\hat{\theta}$ y $\hat{\gamma}$ seguiría en orden de importancia.

Por otro lado, la selección del mejor modelo para las concentraciones de zinc en la misma zona será determinada a través de los resultados en el siguiente cuadro.

Cuadro 6.7: Modelo-AIC para el zinc en Alluvial Fan Zone

Modelo	k	$\ell(\boldsymbol{\theta})$	AIC
Exponencial	2	238.4226	-562.8452
Lognormal	2	213.3117	-422.6234
Weibull	2	231.5982	-459.1964

Para esta zona se observa que el modelo más adecuado para la concentración de zinc es la distribución exponencial con parámetros estimados $\hat{\theta} = 0.0486$ y $\hat{\gamma} = 1.977$. También puede notarse que la distribución Weibull con estimadores $\hat{\beta}$ y $\hat{\eta}$ le continua en orden de importancia.

En forma análoga a los cuadros anteriores se analizará el mejor modelo para la las concentraciones de cobre y zinc en Basin-Trugh Zone. El siguiente cuadro muestra los AIC para la concentración de cobre

Cuadro 6.8: Modelo-AIC para el cobre en Basin-Trough Zone

Modelo	k	$\ell(\boldsymbol{\theta})$	AIC
Exponencial	2	98.3801	-192.7602
Lognormal	2	98.40879	-192.81758
Weibull	2	101.0388	-198.0776

De los resultados anteriores puede notarse que el modelo más adecuado para la concentración de cobre en esta zona es la distribución Weibull con parámetros estimados $\hat{\beta} = 1.016$ y $\hat{\eta} = 4.38$, donde las distribuciones exponencial y lognormal lo siguen en orden de importancia ya que como puede observarse, presentan ambos modelos un AIC casi idénticos.

Continuando con la selección del mejor modelo para esta zona, el siguiente cuadro muestra los AIC para cada modelo sobre el análisis de concentración de zinc.

Cuadro 6.9: Modelo-AIC para el zinc en Basin-Trough Zone

Modelo	k	$\ell(\boldsymbol{\theta})$	AIC
Exponencial	2	196.2972	-388.5944
Logaormal	2	197.5964	-391.1928
Weibull	2	198.0013	-392.0026

Para este último caso se observa que el mejor modelo para los datos de concentración de zinc de esta zona es la distribución Weibull con parámetros estimados $\hat{\beta} = 1.231$ y $\hat{\eta} = 23{,}187$, seguido muy cercanamente de la distribución lognormal.

De todo lo anterior puede aseverarse que los modelos estadísticamente más adecuados, para cada zona y contaminantes, son los que se presentan a continuación con sus respectivos estimadores de máxima verosimilitud $\hat{\theta}_1$ y $\hat{\theta}_2$.

Cuadro 6.10: Selección de modelos en cada zona y sus EMV

Zona	Elemento	Modelo	θ_1	θ_2
Alluvial Fan	*Cobre*	*Weibull*	$\hat{\beta} = 1.125$	$\hat{\eta} = 3.784$
Alluvial Fan	*Zinc*	*Exponencial*	$\hat{\theta} = 0.049$	$\hat{\gamma} = 1.977$
Basin-Trough	*Cobre*	*Weibull*	$\hat{\beta} = 1.016$	$\hat{\eta} = 4.380$
Basin-Troug	*Zinc*	*Weibull*	$\hat{\beta} = 1.231$	$\hat{\eta} = 23.187$

El siguiente cuadro, representa las estimaciones correspondientes a su media, mediana y los intervalos de verosimilitud confianza de estas medidas, de los modelos seleccionados en el cuadro anterior para los contaminantes en sus respectivas zonas.

6.5.2. Prueba de razón de verosimilitudes

En el mismo sentido bajo el cual se seleccionaron los modelos con el criterio anterior, a continuación se hará este mismo ejercicio empleando prueba

Cuadro 6.11: Media, mediana e intervalos de confianza

Zona	Elemento	Modelo	Media	Mediana
Alluvial Fan	*Cobre*	*Weibull*	3.626 (2.806,4.445)	2.732 (1.978,3.486)
Alluvial Fan	*Zinc*	*Exponencial*	2.026 (1.738,2.314)	1.944 (0.059,3.829)
Basin-Trough	*Cobre*	*Weibull*	4.351 (3.117,5.585)	3.054 (1.971,4.137)
Basin-Troug	*Zinc*	*Weibull*	21.672 (16.826,26.518)	17.218 (12.677,21.759)

de razón de verosimilitudes; esto es, con la finalidad de ilustrar este procedimiento estadístico de manera complementaria y buscando coincidir con los modelos seleccionados bajo el Criterio de Información de Akaike representados en el Cuadro 6.10.

Con los mismos resultados del Cuadro 6.5 y bajo un esquema de pruebas de hipótesis típico se procede a la selección del mejor modelo mediante comparaciones de par en par.

Como en toda prueba de hipótesis es necesario un estadístico de prueba; en este caso dicho estadístico ha sido indicado por (4.19). Adicionalmente se considerará 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.

A continuación utilizando los resultados del Cuadro 6.5 se presenta un procedimiento comparativo utilizando las log-verosimilitudes estimadas de cada modelo.

Análisis de concentración de cobre de Alluvial Fan Zone.

Comparación entre los modelos exponencial y la Weibull.

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

2. Estadístico de prueba: $\chi^2 = -2(119.5159 - 124.2878) = 4.7719$.

3. 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 la distribución Weibull con parámetros estimados $\hat{\beta} = 1.125$ y $\hat{\eta} = 3.784$.

Comparación entre los modelos lognormal y Weibull.

1. Hipótesis: $H : Modelo_{lng} \equiv Modelo_{Weib}$.
2. Estadístico de prueba: $\chi^2 = -2(118.6343 - 124.2878) = 11.307$.
3. Observe que $\chi^2 > \chi^2_{.05,1} = 3.84$, por lo tanto se rechaza la hipótesis y el modelo más adecuado es la distribución Weibull.

De lo anterior, se concluye que el modelo que ajusta mejor los datos de concentración cobre para esta zona es la distribución Weibull.

Análisis de concentración de zinc de Alluvial Fan Zone.

Comparación entre los modelos exponencial y lognormal.

1. Hipótesis: $H : Modelo_{exp} \equiv Modelo_{lng}$.
2. Estadístico de prueba: $\chi^2 = -2(213.3117 - 238.4226) = 50.2218$.
3. Observa que $\chi^2 > \chi^2_{.05,1} = 3.84$, por lo tanto se rechaza la hipótesis y el modelo más adecuado es la distribución exponencial con parámetros estimados $\hat{\theta} = 0.049$ y $\hat{\gamma} = 1.977$.

Comparación entre los modelos exponencial y Weibull.

1. Hipótesis: $H : Modelo_{exp} \equiv Modelo_{Weib}$.
2. Estadístico de prueba: $\chi^2 = -2(231.5982 - 238.4226) = 13.6488$
3. Se observa que $\chi^2 > \chi^2_{.05,1} = 3.84$, por lo tanto se rechaza la hipótesis y el modelo adecuado para estos datos es la distribución exponencial con parámetros estimados $\hat{\theta} = 0.049$ y $\hat{\gamma} = 1.977$.

De lo anterior, se concluye que el modelo que mejor ajusta los datos de concentración de zinc para esta zona es la distribución exponencial.

Análisis de la concentración de cobre de Basin-Trough Zone.

Comparación entre los modelos exponencial y Weibull.

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

2. Estadístico de prueba: $\chi^2 = -2(98.3801 - 101.0388 = 5.4832$.

3. Se observa que $\chi^2 > \chi^2_{.05,1} = 3.84$, por lo tanto se rechaza la hipótesis y el modelo que ajusta adecuadamente los datos es la distribución Weibull con parámetros estimados $\hat{\beta} = 1.016$ y $\hat{\eta} = 4.380$.

Comparación entre los modelos lognormal y Weibull.

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

2. Estadístico de prueba: $\chi^2 = -2(98.40879 - 101.0388 = 5.3174$.

3. Observe que $\chi^2 > \chi^2_{.05,1} = 3.84$, por lo tanto se rechaza la hipótesis y el modelo que resulta más adecuado para estos datos es la distribución Weibull con parámetros estimados $\hat{\beta} = 1.016$ y $\hat{\eta} = 4.380$.

Asé, de lo anterior se concluye que el modelo más adecuado para la concentración de zinc en esta zona es la distribución Weibull.

Análisis de concentración de zinc de Basin-Trough Zone.

Comparación entre los modelos exponencial y lognormal.

1. Hipótesis: $H : Mod_{exp} \equiv Mod_{lng}$.

2. Estadístico de prueba: $\chi^2 = -2(196.2972 - 197.5964) = 2.5984$.

3. Observe que $\chi^2 < \chi^2_{.05,1} = 3.84$, por lo tanto no se rechaza la hipótesis y se elige como mejor más adecuado la distribución lognormal con parámetros estimados $\hat{\mu} = 2.727$ y $\hat{\sigma} = 0.876$.

Comparación entre los modelos lognormal y Weibull.

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

2. Estadístico de prueba: $\chi^2 = -2(197.5964 - 198.0013) = 0.8098$.

3. Observe que $\chi^2 < \chi^2_{.05,1} = 3.84$, por lo tanto no se rechaza la hipótesis. De esto, se puede notar la distribución lognormal como la distribución Weibull son igualmente adecuados para el ajuste de estos datos, se elige como modelo la distribución lognormal, en virtud de su versatilidad y su relación con la distribución normal. De manera experimental y bajo la obseración de investigadores en diferentes campos del medio ambiente [41] esto ha sido reportado y fundamentado.

De todo lo anterior, los modelos estadísticamente más adecuados, para cada zona y contaminantes utilizando pruebas de razón de verosimilitudes se presenta a continuación en el siguiente cuadro con sus respectivas estimaciones.

Cuadro 6.12: Selección de modelos en cada zona y sus EMV

Zona	Elemento	Modelo	$\hat{\theta}_1$	$\hat{\theta}_2$
Alluvial Fan	*Cobre*	*Weibull*	$\hat{\beta} = 1.125$	$\hat{\eta} = 3.784$
Alluvial Fan	*Zinc*	*Exponencial*	$\hat{\theta} = 0.049$	$\hat{\gamma} = 1.977$
Basin-Trough	*Cobre*	*Weibull*	$\hat{\beta} = 1.016$	$\hat{\eta} = 4.380$
Basin-Trough	*Zinc*	*Lognormal*	$\hat{\mu} = 2.727$	$\hat{\eta} = 0.876$

El siguiente cuadro, muestra las estimaciones para la media, mediana y los intervalos de verosimilitud confianza de estas medidas, de los modelos seleccionados en el cuadro anterior para los contaminantes en sus respectivas zonas.

Cuadro 6.13: Media, mediana e intervalos de confianza

Zona	Elemento	Modelo	Media	Mediana
Alluvial Fan	*Cobre*	*Weibull*	3.626 (2.806,4.445)	2.732 (1.978,3.486)
Alluvial Fan	*Zinc*	*Exponencial*	2.026 (1.738,2.314)	1.944 (0.059,3.829)
Basin-Trough	*Cobre*	*Weibull*	4.351 (3.117,5.585)	3.054 (1.971,4.137)
Basin-Trough	*Zinc*	*Lognormal*	22.439 (22.151,22.727)	15.289 (11.595,18.984)

Capítulo 7

Conclusión

El uso de la estadística en distintas áreas ha ido adquiriendo cada vez mayor auge, en particular las ciencias naturales, la distribución lognormal y Weibull son algunas de las distribuciones más utilizadas en las ciencias biológicas, comúnmente para el estudio de supervivencia en individuos (datos censurados por la derecha), pero como se pudo observar este tipo de distribuciones también puede ser útil para modelar datos censurados por la izquierda. Para los estudiosos es de gran interés conocer modelos probabilísticos diferentes a los comunes con los que sea posible realizar predicciones confiables sobre datos que en muchas ocasiones por las limitaciones de aparatos de medición no pueden ser registradas.

En este trabajo se mostraron las bases teóricas de tres de las principales distribuciones aplicables en las funciones de datos censurados por la izquierda, la explicación del método de máxima verosimilitud para las estimaciones adecuadas de la media, la mediana, la varianza y los intervalos de confianza para la media, la mediana, todo esto en la forma mas detallada para su análisis.

Sin embargo, en la naturaleza existen fenómenos aleatorios en los cuales no siempre se puede ajustar satisfactoriamente alguna de las distribuciones revisadas en el capítulo dos, es de ahí donde surge la pregunta ¿cuál es la distribución óptima que modela el problema a tratar?, la respuesta a esta pregunta no es inmediata ya que como se pudo observar en el capítulo seis elegir el modelo que mejor ajusta los datos para las concentraciones de cobre y zinc en las zonas Alluvial Fan y Basin-Trough no siempre resulta el mismo modelo. Cabe destacar que para estos grupos de datos la distribución que prevaleció en la mayoría de los ajuestes fue la distribución Weibull. Es interesante mencionar también, que los resultados obtenidos a través de los

métodos, como son el AIC y las pruebas de razón de verosimilitudes concuerdan satisfactoriamente para la selección del mejor modelo.

Para un estudio óptimo de las concentraciones de cobre y zinc resulta conveniente hacer la estimación de los parámetros mediante algún software estadístico ya que facilita de gran manera la realización de cálculos. El ajuste de los datos discutidos en este trabajo se llevó a cabo únicamente con el paquete estadístico R.

La importancia de este trabajo se fundamenta en la fusión de la parte teórica y la parte práctica. El presente trabajo sobre estadística ambiental, pretende motivar a ingenieros, científicos y estudiantes de ingeniería y ciencias, en el análisis de datos censurados; enfatizando los métodos de máxima verosimilitud.

Bibliografía

[1] Ahm H., 1998. Estimating the mean and variance of censored phosphorus concentrations in Florida rainfall: Journal of the American Water Resources Association 34, 583-593.

[2] Akritas, M. G., 1994. Statistical analysis of censored environmental data: Chapter 7 of the Handbook of Statistics, Volume 12, edited by G.P. Patil and C.R. Rao. North-Holland, Amsterdam, 927 pp.

[3] Andersen, P. K., Borgan, Ø., Gill, R. D., Keiding, N. (1993), *Statistical Models Based on Counting Processes*, Springer.

[4] Aris Spanos, 1999. Probability Theory and Statistical Inference: Econometric Modeling with Observational Data. CAMBRIDGE UNIVERSITY PRESS, P. 126.

[5] Army, 1998. Evaluation of dredged material proposed for discharge in wasters of the US-Testinfg manual: Army Corps of Engineers published as EPA-823-B-98-004.

[6] Brockwell, P. J.; Davis, R. A. (2009), *Time Series: Theory and Methods* (2nd ed.), Springer.

[7] Buckley, T.J., J. Liddle, D.L. Ashley, D.C. Paschal, V.W. Burse, L.L. Needham, and G. Akland, 1997. Environmental and biomarker measurements in nine homes in the lower Rio Grande valley: Multiyear results for pesticides, metals, PAHs, and VOCs; Environmental Pollution 23, 705-722.

[8] Casella, G., Berger, R.L. Statistical inference. Second edition, edited by Duxbury, p. 226.

[9] Clarke, J.U., 1998. Estimation of censored data methods to allow statistical comparisons among very small samples with below detection limit observations: Environmental Science and Technology 32, 177-183.

[10] Cox, D.R., and Snell, E. J. 1981. Applied Statistics, New York: Chapman and Hall Inc.

[11] Díaz-Francés, E., and Sprott, D.A. 2000. The use of the likelihood function in the analysis of environmental data. Environmetrics 11, 75-79.

[12] D'Addio, A. C., Rosholm, M. (2002), "The Mobility of French Youth "in" and "out" of the Labour Market", Manuscript.

[13] El-Shaarawi, A. H., and Viveros, R. 1997. Inferences about the mean in log-regression with environmental applications. Environmetrics 8, 569-582.

[14] EPA 1998. Guidance for data quality assessment. Practical methods for data analysis; US Environmental Protection Agency EPA/600/R-96/084.

[15] Frome, E. L., and Watkins, J. W., 2004. Statistical Analysis of Data with Non-Detectable Values, Reports DE-AC05-00OR22725 Oak Ridge National Laboratory, U.S. Department Of Energy, Oak Ridge, Tennessee, http://www.osti.gov/bridge

[16] Gleit, A., 1985. Estimation for small normal data sets with detection limits. Environmental Science and Technology 19, 1201-1206.

[17] Gritz, M. (1993), "The Impact of Training on the Frequency and the Duration of Unployment", *Journal of Econometrics*, n.57, p. 21-25.

[18] Harris, M. L., J.E. Elliot, R. W. Butler, and L. K. Wilson, 2003. Reproductive success and chlorinated hydrocarbon contamination of resident great blue herons from coastal British Columbia, 1977-2000: Environmental Pollution 121, 207-227.

[19] Helsel, D. R., and Gilliom, R. J. 1986. Estimation of distributional parameters for censored trace level water quality data, 2. Verification and applications: Water Resources Research 22, 147-155.

[20] Helsel, D. R., and T. A. Cohn, 1988. Estimation of descriptive statistics for multiply censored water quality data: Water Resources Research 24, 1997-2004.

[21] Helsel, D. R. 1990. Less than obvious: Statistical treatment of data below the detection limit: Environmental Science and Technology 24, pp. 1766-1774.

[22] Helsel, D.R., 2005. Nondetects And Data Analysis, Statistics for Censored Environmental Data, Wiley-Interscience, New York, 251 pp.

[23] Hobbs, K. E., D. C. G. Muir, E. W. Bornb, R. Dietzc, T. Haugd, T. Metcalfee, C. Metcalfee and N. Oien, 2003. Levels and patterns of persistent organochlorines in mink whale (Balaenoptera acutorostrada) stocks from the North Atlantic and European Arctic; Environmental Pollution 121, 239-252.

[24] Huybrechts, T., O. Thas, J. Dewulf, and H. Van Langenhove, 2002. How to estimate moments and quantiles of environmental data sets with non-detected observations? A case study on volatile organic compounds in marine water samples: Journal of hromatography 975, 123-133.

[25] Isobe, T., E. D. Feigelson, and P. I. Nelson, 1986. Statistical methods for astronomical data with upper limits. Astrophysical Journal 306, 490-507.

[26] Johnson, L.N., Kotz, s., & Balakrishnan, N. Continuous Univariate Distributions. Volume 1. Second edition.

[27] Kalbfleisch, J. D. Probability and Statistical Inference. Volume 2: Statistical Inference. Second edition. Springer- Verlag New York Berling Heidelberg Tokyo.

[28] Klein, J.P. y Moeschberger, M.L. (1997). *Survival Analysis: Techniques for Censored and Truncated Data*. N.Y.: Springer-Verlag.

[29] Kolpin, D. W., E. T. Furlong, M. T. Meyer, E. M. Thurman, S. D. Zaugg, L. Barber, and H. T. Buxton, 2002. Pharmaceuticals, hormones, and other organic waste-water contaminants in U. S. streams, 1999-2000: A national reconnaissance; Environmental Science and Technology 36, 1202-1211.

[30] Kroll, C.N. and J.R. Stedinger, 1996. Estimation of moments and quantiles using censored data: Water Resources Research 32, 1005-1012.

[31] Lambert, D., Peterson, B., & Terpenning, I. 1991. Nondetects, Detection Limits, and the Probability of Detection. Journal of the American Statistical Association, Vol. 86, No. 414, 266-277.

[32] Lee, E. T., and J. W. Wang, 2003. Statistical Methods for Survival Data Analysis, Third edition. Wiley, New York, 534 pp.

[33] Lundgren, R. F., and T. J. Lopes, 1999. Occurrence, distribution, and trends of volatile organic compounds in the Ohio River and its major tributaries, 1987-96; U.S. Geological Survey Water-Resources Investigations Report 99-4257, 89 pp.

[34] Lyn, H. S., 2001. Maximum likelihood inference for left-censored HIV RNA data, Statistical in Medicine 20, 33-45.

[35] Meeker, W.Q., and Escobar, L.A. 1998. Statistical Methods for Reliability Data, 450-454. New York: Wiley.

[36] Miesch, A., 1967. Methods of computation for estimating geochemical abundance: U.S. Geological Survey professional Paper 574-B. 15 pp.

[37] Millard, S. P. and S. J. Deverel, 1988. Nonparametric statistical methods for comparing two sites based on data with multiple nondetect limits: Water Resources Research 24, 2087-2098.

[38] Millard, S. P., and Neerchal, N. K. 2001. Environmental Statistics with SPLUS. FL: CRC

[39] Navy 1999. Handbook for statistical analysis of environmental background data, Department of the Navy, Southwest Division, Naval Facilities Engineering Command, San Diego, CA.

[40] Nehls, G.J. and G.G. Akland, 1973, Procedures for handling aerometric data; Journal of the Air Pollution Control Association 23, 180-184.

[41] Ott, W. R. 1995. Environmental Statistics and Data Analysis, FL: CRC Press.

[42] Perkins, J. L., G. N. Cutter, and M.S. Cleveland, 1990, Estimating the mean, variance, and confidence limits from censored (<limit of detection), lognormally-distributed exposure data: American Industrial Hygiene Association Journal 51, 416-419.

[43] Rao, S. T., J. Y. Ku, and K. S. Rao. 1991. Analysis of toxic air contaminant data containing concentrations below the limit if detection: Journal of the Air and Waste Management Association 41, 442-448.

[44] Rosholm, M (2001), "An Analysis of Labour Market Exclusion and (Re-) Inclusion" in *Transitions in the Labour Market*, Disscussion Paper 332, IZA, Bonn.

[45] She, N., 1997. Analyzing censored water quality data using a non-parametric approach: Journal of the American Water Resources Association 33, 615-624.

[46] Singh, A. and J. Nocerino, 2002. Robust estimation of mean and variance using environmental data sets with below detection limit observations; Chemometrics and Intelligent Laboratory Systems 60, 69-86.

[47] Slyman, D. J., A. de Peyster, and R. R. Donohoe, 1994. Hypothesis testing with values below detection limit in environmental studies: Environmental Science and Technology 28, 898-902.

[48] Spanos, A. 1999. Probability Theory and Statistical Inference: Econometric Modeling with Observational Data. CAMBRIDGE UNIVERSITY PRESS, P. 126.

[49] Till, A. E., 2003. Comment on pharmaceuticals hormones, and organic wastewater contaminants in US streams, 1999-2000; Environmental Science and Technology 37, 1052-1053.

[50] Wen, X. H., 1994, Estimation of statistical parameters for censored lognormal hydraulic conductivity measurements: Mathematical Geology 26, 717-731.

Apéndice A

Algunos resultados estadísticos

A.1. Método delta

En la estadística, *el método delta* [35] es un método que se utiliza para derivar una distribución de probabilidad aproximada para una función de un estimador estadístico asintóticamente normal a partir del conocimiento del límite de la varianza del estimador. En términos más generales, el método delta puede ser considerado como un caso bastante general del *teorema del límite central*.

Este apartado muestra como calcular aproximadamente valores esperados, varianzas y covarianzas de estimadores de funciones paramétricas. Sea $g(\boldsymbol{\theta})$ una función de valor real de los parámetros $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_r)'$ y sea $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_r)$ y $g(\hat{\boldsymbol{\theta}})$ el estimador de $\boldsymbol{\theta}$ y $g(\boldsymbol{\theta})$, respectivamente el objetivo es obtener expresiones o expresiones aproximadas para $E[g(\hat{\boldsymbol{\theta}})]$ y $Var[g(\hat{\boldsymbol{\theta}})]$ como una función de $E(\hat{\theta}_i)$, $Var(\hat{\theta}_i)$ y $Cov(\hat{\theta}_i, \hat{\theta}_j)$.

El caso más simple es cuando $g(\hat{\boldsymbol{\theta}})$ es una función lineal de los θ_i, es decir, $g(\hat{\boldsymbol{\theta}}) = a_0 + \sum_{i=1}^{r} a_i\hat{\theta}_i$, donde los a_i son constantes. Para mayor comodidad expresemos $g(\hat{\boldsymbol{\theta}})$ como

$$\begin{aligned} g(\hat{\boldsymbol{\theta}}) &= a_0 + \sum_{i=1}^{r} a_i\hat{\theta}_i + \sum_{i=1}^{r} a_iE(\hat{\theta}_i) - \sum_{i=1}^{r} a_iE(\hat{\theta}_i) \\ &= a_0 + \sum_{i=1}^{r} a_iE(\hat{\theta}_i) + \sum_{i=1}^{r} a_i(\hat{\theta}_i - E(\hat{\theta}_i)) \\ &= b_0 + \sum_{i=1}^{r} b_i(\hat{\theta}_i - E(\hat{\theta}_i)) \end{aligned} \tag{A.1}$$

donde $b_0 \quad a_0 + \sum_{i=1}^{r} a_i E(\hat{\theta}_i)$ y $b_i = a_i$, $i = 1, 2, \ldots, r$.

El valor esperado de $g(\hat{\boldsymbol{\theta}})$ es:

$$\begin{aligned} E[g(\hat{\boldsymbol{\theta}})] &= E\left[b_0 + \sum_{i=1}^{r} b_i(\hat{\theta_i} - E(\hat{\theta_i}))\right] \\ &= E[b_0] + E\left[\sum_{i=1}^{r} b_i(\hat{\theta}_i - E(\hat{\theta}_i))\right] \\ &= b_0 + \sum_{i=1}^{r} E\left[b_i(\hat{\theta}_i - E(\hat{\theta}_i))\right] \\ &= b_0 + \sum_{i=1}^{r} b_i\left[E(\hat{\theta}_i) - E(\hat{\theta}_i)\right] = b_0. \end{aligned}$$

La varianza de $g(\hat{\boldsymbol{\theta}})$ es:

$$\begin{aligned} Var[g(\hat{\boldsymbol{\theta}})] &= Var\left[b_0 + \sum_{i=1}^{r} b_i(\hat{\theta_i} - E(\hat{\theta_i}))\right] \\ &= Var[b_0] + Var\left[\sum_{i=1}^{r} b_i(\hat{\theta}_i - E(\hat{\theta}_i))\right] \\ &= Var\left[\begin{pmatrix} b_1 & b_2 & \ldots & b_r \end{pmatrix} \begin{pmatrix} \hat{\theta}_1 - E(\hat{\theta}_1) \\ \hat{\theta}_2 - E(\hat{\theta}_2) \\ \vdots \\ \hat{\theta}_r - E(\hat{\theta}_r) \end{pmatrix}\right]. \end{aligned}$$

Dado que $Var(AX) = \boldsymbol{A}Var(\boldsymbol{X})\boldsymbol{A}'$, donde $\boldsymbol{A}$ es una matriz 1xr y $\boldsymbol{X}$ una matriz de rx1 apliquemos este resultado a la $Var[g(\hat{\boldsymbol{\theta}})]$, esto es:

$$Var[g(\hat{\boldsymbol{\theta}})] =$$

$$\begin{pmatrix} b_1 & b_2 & \ldots & b_r \end{pmatrix} \begin{pmatrix} var(x_1) & cov(x_1, x_2) & \ldots & cov(x_1, x_r) \\ var(x_2, x_1) & var(x_2) & \ldots & cov(x_2, x_r) \\ \vdots & \vdots & \ddots & \vdots \\ cov(x_r, x_1) & cov(x_r, x_2) & \ldots & var(x_r) \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_r \end{pmatrix}$$

donde

$$\boldsymbol{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_r \end{pmatrix} = \begin{pmatrix} \hat{\theta}_1 - E(\hat{\theta}_1) \\ \hat{\theta}_2 - E(\hat{\theta}_2) \\ \vdots \\ \hat{\theta}_r - E(\hat{\theta}_r) \end{pmatrix}$$

pero observe que

$$\begin{aligned} var(x_i) &= var[\hat{\theta}_i - E(\hat{\theta}_i)] \\ &= E\{[\hat{\theta}_i - E(\hat{\theta}_i)]^2\} - \{E[\hat{\theta}_i - E(\hat{\theta}_i)]\}^2 \\ &= E[\hat{\theta}_i^{\,2} - 2\hat{\theta}_i E(\hat{\theta}_i) + E(\hat{\theta}_i)^2] - [E(\hat{\theta}_i) - E(\hat{\theta}_i)]^2 \\ &= E(\hat{\theta}_i^{\,2}) - 2[E(\hat{\theta}_i)]^2 + [E(\hat{\theta}_i)]^2 \\ &= E(\hat{\theta}_i^{\,2}) - [E(\hat{\theta}_i)]^2 \\ &= var(\hat{\theta}_i) \end{aligned}$$

donde $i = 1, 2, \ldots, r$ y la $cov(x_i, x_j)$ para $i \neq j$ es

$$\begin{aligned} cov(x_i, x_j) &= cov\{[\hat{\theta}_i - E(\hat{\theta}_i)], [\hat{\theta}_j - E(\hat{\theta}_j)]\} \\ &= E\{[\hat{\theta}_i - E(\hat{\theta}_i) - E(\hat{\theta}_i - E(\hat{\theta}_i))][\hat{\theta}_j - E(\hat{\theta}_j) - E(\hat{\theta}_j - E(\hat{\theta}_j))]\} \\ &= E\{[\hat{\theta}_i - E(\hat{\theta}_i)][\hat{\theta}_j - E(\hat{\theta}_j)]\} \\ &= cov(\hat{\theta}_i, \hat{\theta}_j). \end{aligned}$$

Así sustituyendo estos resultados en $Var[g(\hat{\boldsymbol{\theta}})]$, tenemos $Var[g(\hat{\boldsymbol{\theta}})] =$

$$\begin{pmatrix} b_1 & b_2 & \ldots & b_r \end{pmatrix} \begin{pmatrix} var(\hat{\theta}_1) & cov(\hat{\theta}_1, \hat{\theta}_2) & \ldots & cov(\hat{\theta}_1, \hat{\theta}_r) \\ cov(\hat{\theta}_2, \hat{\theta}_1) & var(\hat{\theta}_2) & \ldots & cov(\hat{\theta}_2, \hat{\theta}_r) \\ \vdots & \vdots & \ddots & \vdots \\ cov(\hat{\theta}_r, \hat{\theta}_1) & cov(\hat{\theta}_r, \hat{\theta}_2) & \ldots & var(\hat{\theta}_r) \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_r \end{pmatrix}$$

$$\begin{aligned} &= \begin{matrix} b_1^2 var(\hat{\theta}_1) + b_1 b_2 cov(\hat{\theta}_1, \hat{\theta}_2) + \ldots + b_1 b_r cov(\hat{\theta}_1, \hat{\theta}_r) \\ b_2 b_1 cov(\hat{\theta}_2, \hat{\theta}_1) + b_2^2 var(\hat{\theta}_2) + \ldots + b_2 b_r cov(\hat{\theta}_2, \hat{\theta}_r) \\ \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ b_r b_1 cov(\hat{\theta}_r, \hat{\theta}_1) + b_r b_2 cov(\hat{\theta}_r, \hat{\theta}_2) + \ldots + b_r^2 var(\hat{\theta}_r) \end{matrix} \\ &= \sum_{i=1}^{r} b_i^2 var(\hat{\theta}_i) + \sum_{i=1}^{r} \sum_{j=1}^{r} b_i b_j cov(\hat{\theta}_i, \hat{\theta}_j) \end{aligned}$$

donde $i \neq j$ para la segunda suma de la expresión anterior.

Este procedimiento es conocido como el "método delta" [35].

El método delta, en su esencia, expande una función de una variable aleatoria sobre su media, generalmente con una solo aproximación de Taylor y luego toma la varianza. Por ejemplo, cuando g($\hat{\boldsymbol{\theta}}$) tiene segundas derivadas

parciales continuas con respecto a $\boldsymbol{\theta}$, la expansión en serie de Taylor de primer orden de $\mathrm{g}(\hat{\boldsymbol{\theta}})$ cerca de $\boldsymbol{\mu} = [E(\hat{\theta}_1), \ldots, E(\hat{\theta}_r)]$ está dado por

$$\mathrm{g}(\hat{\boldsymbol{\theta}}) = \mathrm{g}(\hat{\boldsymbol{\mu}}) + \sum_{i=1}^{r} \frac{\partial \mathrm{g}(\boldsymbol{\theta})}{\partial \theta_i} \left[\hat{\theta}_i - E[\hat{\theta}_i]\right], \tag{A.2}$$

donde la derivada parcial de $\mathrm{g}(\boldsymbol{\theta})$ con respecto a θ_i son valores evaluados en $\boldsymbol{\mu}$.

Observe que (A.2) es similar a (A.1) con

$$b_0 = \mathrm{g}(\boldsymbol{\mu}) \text{ y } b_i = \frac{\partial \mathrm{g}(\boldsymbol{\theta})}{\partial \theta_i},\ i = 1, \ldots, r.$$

Consecuentemente

$$\begin{aligned} E[\mathrm{g}(\hat{\boldsymbol{\theta}})] &\approx \mathrm{g}(\boldsymbol{\mu}) \\ Var[\mathrm{g}(\hat{\boldsymbol{\theta}})] &\approx \sum_{i=1}^{r} \left[\frac{\partial \mathrm{g}(\boldsymbol{\theta})}{\partial \theta_i}\right]^2 Var(\hat{\theta}_i) \\ &\quad + \sum_{i=1}^{r} \sum_{j=1}^{r} \left[\frac{\partial \mathrm{g}(\boldsymbol{\theta})}{\partial \theta_i}\right] \left[\frac{\partial \mathrm{g}(\boldsymbol{\theta})}{\partial \theta_j}\right] Cov(\hat{\theta}_i, \hat{\theta}_j) \end{aligned} \tag{A.3}$$

donde $i \neq j$ para la segunda suma de la expresión anterior. Cuando los valores de $\hat{\theta}_i$ no están correlacionados o cuando las covarianzas $Cov(\hat{\theta}_i, \hat{\theta}_j)$, $i \neq j$, son pequeños cuando se compara con las varianzas $Var(\hat{\theta}_i)$, el término anterior de la derecha de (A.3) usualmente se omite de la aproximación.

La misma idea aplica para valores de funciones vectoriales. Por ejemplo, si $\mathrm{g}_1(\boldsymbol{\theta})$ y $\mathrm{g}_2(\boldsymbol{\theta})$ son dos funciones de valores reales entonces

$$\begin{aligned} Cov[\mathrm{g}_1(\hat{\boldsymbol{\theta}}), \mathrm{g}_2(\hat{\boldsymbol{\theta}})] &\approx \sum_{i=1}^{r} \left[\frac{\partial \mathrm{g}_1(\boldsymbol{\theta})}{\partial \theta_i}\right] \left[\frac{\partial \mathrm{g}_2(\boldsymbol{\theta})}{\partial \theta_i}\right] Var(\hat{\theta}_i) \\ &\quad + \sum_{i=1}^{r} \sum_{j=1}^{r} \left[\frac{\partial \mathrm{g}_1(\boldsymbol{\theta})}{\partial \theta_i}\right] \left[\frac{\partial \mathrm{g}_2(\boldsymbol{\theta})}{\partial \theta_j}\right] Cov(\hat{\theta}_i, \hat{\theta}_j) \end{aligned} \tag{A.4}$$

donde $i \neq j$ para la segunda suma de la expresión anterior.

A.2. Método delta para funciones vectoriales

En general para valores de funciones vectoriales $\mathbf{g}(\boldsymbol{\theta})$ de los parámetros tal que todas las segundas derivadas parciales con respecto a los elementos

de $\boldsymbol{\theta}$ son continuas

$$Var[\mathbf{g}(\hat{\boldsymbol{\theta}})] \approx \left[\frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]' Var(\hat{\boldsymbol{\theta}}) \left[\frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]$$

donde $\partial \mathbf{g}(\boldsymbol{\theta})/\partial \boldsymbol{\theta} = [\partial \mathrm{g}_1(\boldsymbol{\theta})/\partial \boldsymbol{\theta}, \partial \mathrm{g}_2(\boldsymbol{\theta})/\partial \boldsymbol{\theta}, \ldots]$ es una matriz de vectores gradientes de las primeras derivadas parciales de $\mathbf{g}(\boldsymbol{\theta})$ con respecto a $\boldsymbol{\theta}$ y

$$Var(\hat{\boldsymbol{\theta}}) = \begin{pmatrix} Var(\hat{\theta}_1) & Cov(\hat{\theta}_1, \hat{\theta}_2) & \ldots & Cov(\hat{\theta}_1, \hat{\theta}_r) \\ Var(\hat{\theta}_2, \hat{\theta}_1) & Var(\hat{\theta}_2) & \ldots & Cov(\hat{\theta}_2, \hat{\theta}_r) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(\hat{\theta}_r, \hat{\theta}_1) & Cov(\hat{\theta}_r, \hat{\theta}_2) & \ldots & Var(\hat{\theta}_r) \end{pmatrix}$$

ambos evaluados en $\boldsymbol{\theta}$.

El método delta puede proporcionar buenas aproximaciones para $E[\mathrm{g}(\hat{\boldsymbol{\theta}})]$ y $Var[\mathrm{g}(\hat{\boldsymbol{\theta}})]$.

Sin embargo, como se indica en los libros de texto más avanzados, uno necesita tener cuidado al aplicar este método porque la adecuación de la aproximación depende de la validez de la aproximación de Taylor y el tamaño del resto en la aproximación. La simulación puede utilizarse para comprobar la adecuación de la aproximación.

Apéndice B

Método de Newton-Raphson

B.1. Método de Newton-Raphson univariado

Supóngase que se desea encontrar una raiz $\widehat{\theta}$ de la ecuación $g(\theta) = 0$. Sea θ^0 un valor paramétrico cercano a $\widehat{\theta}$ y considere la expansión en serie de Taylor de $g(\theta)$ cerca de $\theta = \theta^0$:

$$g(\theta) \;=\; g(\theta^0) + (\theta - \theta^0)g'(\theta^0) + (\theta - \theta^0)^2 \frac{g''(\theta^0)}{2!} + \ldots.$$

Como $|\theta - \theta^0|$ es pequeño, entonces el término cuadrático y los terminos de ordenes mayores serán pequeños, así omitiendo estos términos queda

$$g(\theta) \;\approx\; g(\theta^0) + (\theta - \theta^0)g'(\theta^0).$$

Aproximando $g(\theta)$ por una función lineal de θ que tiene el mismo valor y la pendiente como $g(\theta)$ en $\theta = \theta^0$.

Ya que $g(\widehat{\theta}) = 0$, se tiene

$$g(\theta^0) + (\widehat{\theta} - \theta^0)g'(\theta^0) \;\approx\; 0,$$

y por lo tanto

$$\widehat{\theta} \;\approx\; \theta^0 - \frac{g(\theta^0)}{g'(\theta^0)}.$$

Con el método de Newton, se toma θ^0 como una aproximación preliminar a $\widehat{\theta}$, y entonces se calcula una aproximación estimada θ^1 dado como sigue:

$$\theta^1 \approx \theta^0 - \frac{g(\theta^0)}{g'(\theta^0)}. \tag{B.1}$$

La aproximación estimada es el punto en el cual la aproximación lineal a $g(\theta^0)$ en $\theta = \theta^0$ cruza el eje θ. Ahora tomese θ^1 como la nueva aproximación preliminar y repita el cálculo, esto dará

$$\theta^2 \approx \theta^1 - \frac{g(\theta^1)}{g'(\theta^1)}.$$

Continue este procedimiento hasta $\theta^{k+1} \approx \theta^k$, caso en el cual $g(\theta^k) = 0$ y la raíz haya sido encontrada.

B.2. Método de Newton-Raphson bivariado

El método de Newton para resolver la ecuación con parámetro $g(\theta) = 0$ fue explicado en la apéndice anterior.

En el caso de dos parámetros se necesita resolver un par de ecuaciones simultáneas

$$g(\theta_1, \theta_2) = 0; \quad h(\theta_1, \theta_2) = 0.$$

Sea (θ_1^0, θ_2^0) una estimación preliminar de la raíz $(\widehat{\theta}_1, \widehat{\theta}_2)$, considérese las aproximaciones lineales

$$\begin{aligned} g(\theta_1, \theta_2) &\approx g(\theta_1^0, \theta_2^0) + (\theta_1 - \theta_1^0)\frac{\partial g}{\partial \theta_1} + (\theta_2 - \theta_2^0)\frac{\partial g}{\partial \theta_2} \\ h(\theta_1, \theta_2) &\approx h(\theta_1^0, \theta_2^0) + (\theta_1 - \theta_1^0)\frac{\partial h}{\partial \theta_1} + (\theta_2 - \theta_2^0)\frac{\partial h}{\partial \theta_2} \end{aligned}$$

Aquí las derivadas deben ser evaluadas en $\theta_1 = \theta_1^0$. Estas aproximaciones lineales puede ser derivados truncando la expansión en serie de Taylor de g y h en (θ_1^0, θ_2^0). Tienen los mismos valores y las primeras derivadas de g y h en el punto (θ_1^0, θ_2^0).

Ya que $g(\widehat{\theta}_1, \widehat{\theta}_2) = 0$ y $h(\widehat{\theta}_1, \widehat{\theta}_2) = 0$, se tiene

$$\begin{aligned} (\widehat{\theta}_1 - \theta_1^0)\frac{\partial g}{\partial \theta_1} + (\widehat{\theta}_2 - \theta_2^0)\frac{\partial g}{\partial \theta_2} &= -g(\theta_1^0, \theta_2^0) \\ (\widehat{\theta}_1 - \theta_1^0)\frac{\partial h}{\partial \theta_1} + (\widehat{\theta}_2 - \theta_2^0)\frac{\partial h}{\partial \theta_2} &= -h(\theta_1^0, \theta_2^0). \end{aligned}$$

Resolviendo estas ecuaciones lineales para $\widehat{\theta}_1 - \theta_1^0$ y $\widehat{\theta}_2 - \theta_2^0$ queda

$$\begin{bmatrix} \widehat{\theta}_1 - \theta_1^0 \\ \widehat{\theta}_2 - \theta_2^0 \end{bmatrix} \approx - \begin{bmatrix} \dfrac{\partial g}{\partial \theta_1} & \dfrac{\partial g}{\partial \theta_2} \\ \dfrac{\partial h}{\partial \theta_1} & \dfrac{\partial h}{\partial \theta_2} \end{bmatrix}^{-1} \begin{bmatrix} g(\theta_1^0, \theta_2^0) \\ h(\theta_1^0, \theta_2^0) \end{bmatrix}$$

y por lo tanto

$$\begin{bmatrix} \widehat{\theta}_1 \\ \widehat{\theta}_2 \end{bmatrix} \approx \begin{bmatrix} \theta_1^0 \\ \theta_2^0 \end{bmatrix} - \begin{bmatrix} \dfrac{\partial g}{\partial \theta_1} & \dfrac{\partial g}{\partial \theta_2} \\ \dfrac{\partial h}{\partial \theta_1} & \dfrac{\partial h}{\partial \theta_2} \end{bmatrix}^{-1} \begin{bmatrix} g(\theta_1^0, \theta_2^0) \\ h(\theta_1^0, \theta_2^0) \end{bmatrix}. \tag{B.2}$$

Comenzando con una aproximación preliminar (θ_1^0, θ_2^0), se aplica (B.2) hasta alcanzar una convergencia. Esta generalización del método de Newton es llamado el método de Newton-Raphson.

Apéndice C

Funciones en lenguaje R

C.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 $P[X = x]$, en otro caso, $P[X > x]$.

C.2. Distribución log normal

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 a`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 de la distribución sobre la log sacala con valores por default 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, $P[X > x]$.

C.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`	Parámetros 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, $P[X > x]$.

C.4. Optimización general

Descripción

El uso general de optimización está basado en los algoritmos de Nelder-Mead, quasi-Newton y el gradiente conjugado. Se incluye una opción para el cuadro restringido por la optimización y enfriamiento simulado.

Uso

```
optim(par, fn, gr = NULL, ...,
      method = c("Nelder-Mead", "BFGS", "CG", "L-BFGS-B", "SANN",
      "Brent"),      lower = -Inf, upper = Inf,
      control = list(), hessian = FALSE)
```

Argumentos

`par`	Los valores iniciales de los parámetros a ser optimizados arriba.
`fn`	Una función a ser minimizado (o maximizada), con el primer argumento el vector de parámetros sobre los cual minimización tendrá lugar. Debe devolver un resultado escalar.
`gr`	Una función para devolver el gradiente para los métodos "`BFGS`", "`CG`" y "`L-BFGS-B`". Si es `NULL`, se utilizará una aproximación de diferencias finitas.

	Para el método de "SANN" especifica una función para generar un nuevo punto candidato. Si es NULL se utiliza un núcleo Gaisiano de Markov predeterminado.
...	Más argumentos para pasar a fn y gr.
method	El método a utilizar. Consulte 'Detalles'.
lower, upper	Límites de las variables para el método de "L-BFGS-B", o límites en los que buscar el método "Brent".
control	Una lista de parámetros de control. Consulte 'Detalles'.
hessian	Lógico. ¿Debe devolver una matriz hessiana numéricamente diferenciada?

C.5. Minimización no lineal

Descripción

Esta función lleva a cabo una minimización de la función f utilizando el algoritmo de Newton. Ver referencias para más detalles.

Uso

```
nlm(f, p, ..., hessian = FALSE, typsize = 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 atibutos el gradiente y el hessiano, estos serán utilizados en el cálculo de los valores paramétricos actualizados. En otro caso, las derivadas numéricas son utilizadas. deriv regresa una función con el gradiente adecuado. Este deberia ser una función de un vector de la longitud de 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 de 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 lo 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 la función de optimización de desbordamiento, 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 los dos primeros ocurran, 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 termina.
`check.analyticals`	Un escalar lógico que especifica si los gradientes analíticos y Hessianos, son aplicados, deberá cotejarse con las derivados numéricas en los valores de los parámetros iniciales. Esto puede ayudar a detectar gradientes o hessianos mal formuladas.

ANÁLISIS DE DATOS AMBIENTALES CENSURADOS

ORIGINALITY REPORT

5%
SIMILARITY INDEX

PRIMARY SOURCES

1	colposdigital.colpos.mx:8080 Internet	813 words — 3%
2	runebook.dev Internet	150 words — 1%
3	documents.mx Internet	66 words — < 1%
4	www.ugrad.stat.ubc.ca Internet	54 words — < 1%
5	documents.deq.utah.gov Internet	18 words — < 1%
6	pt.scribd.com Internet	18 words — < 1%
7	cidershop.seesaa.net Internet	15 words — < 1%
8	repositorio.unal.edu.co Internet	14 words — < 1%
9	digital.library.unt.edu Internet	12 words — < 1%

10	Smith, P.N.. "Contaminant exposure in terrestrial vertebrates", Environmental Pollution, 200711 Crossref	11 words — < 1%
11	ddd.uab.cat Internet	10 words — < 1%
12	repository.library.carleton.ca Internet	10 words — < 1%
13	www.coursehero.com Internet	10 words — < 1%

EXCLUDE QUOTES	ON	EXCLUDE SOURCES	OFF
EXCLUDE BIBLIOGRAPHY	ON	EXCLUDE MATCHES	< 10 WORDS