



"El saber de mis hijos
hará mi grandeza"

UNIVERSIDAD DE SONORA

DIVISIÓN DE CIENCIAS EXACTAS Y NATURALES

Departamento de Matemáticas

Robusticidad en las inferencias sobre el cociente de
medias de variables lognormales

T E S I S

Que para obtener el título de:

Licenciado en Matemáticas

Presenta:

Arling Vázquez Alcántar

Director de Tesis: Dr. José Arturo Montoya Laos

Hermosillo, Sonora, México, 6 de Enero, 2012.

Universidad de Sonora

Repositorio Institucional UNISON



**"El saber de mis hijos
hará mi grandeza"**



Excepto si se señala otra cosa, la licencia del ítem se describe como openAccess

Q A276
V39

R 15 1881



SINODALES

Dr. José Arturo Montoya Laos
Universidad de Sonora

Dr. Oscar Vega Amaya
Universidad de Sonora

M.C. María Teres Robles Alcaráz
Universidad de Sonora

M.C. Martha Cecilia Palafox Duarte
Universidad de Sonora

Agradecimientos

Agradezco a mi familia, principalmente a mi madre Elizabeth Alcántar, a mis hermanos Manuel Servando, Luis Gabriel y Yissell Alejandra. A mi pequeña sobrina Alicia Elizabeth por los buenos momentos que me hace pasar. También agradezo a mi padre Manuel Vázquez y a mi tía Silvia Rojas.

A la familia Romero Robles, en especial a Daniela Romero y María del Rosario Robles. Han sido un gran apoyo para mí.

A mi director de tesis, Dr. José Arturo Montoya Laos, por su esmero y dedicación para que este trabajo se llevara a cabo.

A mis sinodales Dr. José Arturo Montoya Laos, Dr. Oscar Vega Amaya, M.C. María Teres Robles Alcaráz y M.C. Martha Cecilia Palafox Duarte, por su tiempo y por sus observaciones que sirvieron para mejorar notablemente este trabajo.

A Bárbara McGregor, de Kumon Centro de Matemáticas; a Julieta Rosas y Alfredo Cabral de Está Cabral... restaurarte por proporcionarme un lugar entre sus colaboradores.

A Datos y Cifras del Norte A.C. en especial a Juan Pablo Piñeda, Carolina Granados, Milagros Valenzuela, Mayra Burruel, Elena Guardado y Roberto Treviño por acompañarme en estos momentos tan importantes.

A mis compañeros y amigos Adriana Ferrá, Dania Montes, Dora Montes, Sayth Santiago, Juan Carlos Lizaola, José Carlos Mendoza.

A los que además de ser mis compañeros de generación son grandes amigos, Angélica Moreno, Carol Corral, Manuela Tánori y Humberto Ocejo.

Sobre todo, gracias a Dios por darme vida y por acercarme a estas personas.

Índice general

1. La distribución lognormal	9
1.1. Modelo de probabilidad lognormal	10
1.2. Características del modelo lognormal	12
1.3. Relación del modelo lognormal con la familia de Box-Cox	14
2. Conceptos de inferencia estadística	17
2.1. Estimación vía el enfoque de verosimilitud	19
2.1.1. La función de verosimilitud	19
2.1.2. Verosimilitud relativa	25
2.1.3. Regiones de verosimilitud	26
2.1.4. Regiones de verosimilitud-confianza	28
2.1.5. Verosimilitud perfil	30
2.1.6. Ejemplo: Cociente de medias de lognormales	35
2.2. Métodos de valoración de modelos	48
2.2.1. Histograma y densidad estimada	49
2.2.2. Gráfica cuantil-cuantil	50
2.2.3. Uso de la verosimilitud perfil	52
2.2.4. Prueba estadística de hipótesis	54
2.2.5. Pruebas de significancia: <i>p-valor</i>	57
2.3. Selección de modelos	58

2.4. Robusticidad	58
3. Robusticidad en las inferencias: Cociente de medias de lognormales	60
3.1. Análisis de robusticidad	60
3.2. Caso robusto: Ácidos grasos poliinsaturados en especies de rayas.	64
3.2.1. Modelación	65
3.2.2. Valoración del modelo lognormal	66
3.2.3. Selección del modelo	70
3.2.4. Análisis de robusticidad	71
3.2.5. Inferencias sobre el parámetro de interés: Cociente de medias lognormales	73
3.2.6. Interpretación de los resultados	74
3.3. Caso no robusto: Biodisponibilidad de fármacos	75
3.3.1. Análisis de robusticidad	77
4. Conclusiones	81
A. Verosimilitud perfil de λ	83
Bibliografía	86

Prefacio

En general, cuando se utilizan procesos estadísticos para hacer inferencia sobre un parámetro de interés, se plantean de manera implícita o explícita supuestos acerca del modelo. Dichos supuestos no se consideran exactamente verdaderos pero resultan muy útiles en algunas ramas de las matemáticas aplicadas. Para justificarlos se asume un incierto principio de estabilidad conocido como *robusticidad*: un error menor en el modelo produce un error menor en las conclusiones finales. Tal razonamiento no es correcto en todos los casos.

En la Estadística suele ser objeto de interés el problema de hacer inferencia sobre parámetros que permiten comparar ciertas características de dos variables aleatorias. En especial, el problema de hacer inferencia sobre el cociente de medias de dos variables aleatorias independientes con distribución lognormal es interesante y relevante debido a que las inferencias sobre el parámetro de interés, el cociente de medias, pueden depender fuertemente de los valores plausibles de los otros parámetros del modelo, llamados parámetros de estorbo. Además, la distribución lognormal se usa para modelar diversos fenómenos aleatorios en muchas áreas de la ciencia; por ejemplo, la biodisponibilidad de fármacos en farmacología, periodos de latencia en medicina, tiempos de vida de componentes electrónicos en confiabilidad, entre otros.

El objetivo general de este trabajo de tesis es enfatizar la importancia de un análisis de robusticidad en un proceso de inferencia estadística. En particular, se usará la función de verosimilitud perfil para analizar la robusticidad de las inferencias sobre

el cociente de medias de variables aleatorias independientes X y Y con distribución lognormal de parámetros (μ_X, σ_X) y (μ_Y, σ_Y) , $\delta = E(X)/E(Y)$, frente a cambios en los valores plausibles del parámetro de estorbo $\rho = \sigma_X/\sigma_Y$.

Es importante mencionar aquí que la función de verosimilitud perfil se utiliza principalmente para hacer inferencias sobre un parámetro de interés en presencia de otros parámetros considerados de estorbo. Sin embargo, su uso en análisis de robusticidad ha sido explorado muy poco. A continuación se describen los aspectos más importantes abordados en cada capítulo de esta tesis.

En el Capítulo 1 se presenta el modelo lognormal y se comparan algunas características de éste con el modelo normal. Además se muestra que el modelo lognormal forma parte de la familia de densidades de Box-Cox y que tal relación tiene relevancia cuando se desea justificar la elección del modelo lognormal, dentro de todos los posibles miembros de esta familia, para un juego específico de datos.

En el Capítulo 2 se revisan las ideas fundamentales de un proceso de inferencia estadística: modelación, valoración de modelos, selección del modelo, robusticidad, inferencia sobre parámetros de interés e interpretación de los resultados en el contexto del problema. Se definen conceptos importantes para hacer inferencias a través del enfoque de verosimilitud. También se describen herramientas teóricas y heurísticas comúnmente empleadas para valorar modelos. Además, se presentan algunos criterios simples y prácticos para la selección de un modelo. Al final del capítulo, se discute la problemática asociada a la robusticidad en las inferencias.

En el Capítulo 3 se describe un procedimiento novedoso para determinar la robusticidad de las inferencias sobre el cociente de medias de variables aleatorias independientes X y Y con distribución lognormal de parámetros (μ_X, σ_X) y (μ_Y, σ_Y) , $\delta = E(X)/E(Y)$, respecto a cambios en los valores plausibles del parámetro de estorbo $\rho = \sigma_X/\sigma_Y$. Se muestran dos ejemplos con datos reales, uno robusto (comparación de ácidos grasos poliinsaturados en especies de rayas) y uno no robusto (compara-

ción de biodisponibilidad de fármacos). Finalmente, en el Capítulo 4 se presentan las conclusiones generales de este trabajo.

Las principales aportaciones de esta tesis son las siguientes:

- Dar información relevante sobre las etapas de un proceso estadístico de inferencia, modelación, valoración de modelos, selección del modelo, análisis de robusticidad, inferencia sobre parámetros de interés en el contexto del problema e interpretación de los resultados en el contexto del problema, de una forma simple, clara y formal.
- Proponer un método para explorar la robusticidad de las inferencias sobre el cociente de medias de variables aleatorias independientes X y Y con distribución lognormal de parámetros (μ_X, σ_X) y (μ_Y, σ_Y) , $\delta = E(X)/E(Y)$, frente a cambios en los valores plausibles del parámetro de estorbo $\rho = \sigma_X/\sigma_Y$.
- Aplicar el método de análisis de robusticidad propuesto en esta tesis y mostrar inferencias robustas en un caso de estudio con datos reales (comparación de ácidos grasos poliinsaturados en especies de rayas).
- Aplicar el método de análisis de robusticidad propuesto en esta tesis y mostrar inferencias no robustas en un caso de estudio con datos reales (comparación de biodisponibilidad de fármacos).
- Exhibir la importancia del análisis de robusticidad en los procesos de inferencia estadística.

Posibles extensiones inmediatas de este trabajo son:

- Usar la expresión analítica de función de verosimilitud de los parámetros $\delta = E(X)/E(Y)$ y $\rho = \sigma_X/\sigma_Y$ calculada en esta tesis, no antes descrita en la literatura revisada, para hacer inferencias de manera conjunta sobre el vector (δ, ρ) .
- Extender el método de análisis de robusticidad propuesto en esta tesis para el caso de parámetros de interés y de estorbo no escalares.

- Abordar el problema de comparación de medias de variables aleatorias no necesariamente independientes.

Los cálculos computacionales y figuras de esta tesis se realizaron con ayuda del programa *R Development Core Team* (2011).

Capítulo 1

La distribución lognormal

En este capítulo se presenta el modelo lognormal, algunas de sus principales características y la relación de éste con el modelo normal. Además, se muestra que el modelo lognormal es un miembro particular de la familia de densidades de Box y Cox. Generalmente, esto es relevante cuando se desea justificar la elección del modelo lognormal dentro de todos los posibles miembros de esta familia, para un juego específico de datos.

La estructura de este capítulo es la siguiente. En la Sección 1.1 se define una variable aleatoria lognormal y se muestra como obtener su función de densidad en términos de la densidad de una variable aleatoria normal. En la Sección 1.2 se presentan dos características importantes de una variable aleatoria lognormal, la media y la varianza; se enfatiza que a diferencia de la normal, estas cantidades comparten los parámetros del modelo y podrá observarse la influencia de éstos en la función de densidad. Finalmente, en la Sección 1.3 se presenta la familia de transformaciones propuesta por Box y Cox y la correspondiente familia de densidades, que dependen de un parámetro λ . De hecho, la densidad lognormal es un caso particular de tal familia de densidades cuando $\lambda = 0$.

1.1. Modelo de probabilidad lognormal

El objetivo de un modelo de probabilidad es describir la variabilidad e incertidumbre inherente de casi cualquier faceta de la realidad a través de una abstracción matemática. Un modelo de probabilidad permite manipular artificialmente aspectos de la realidad, con el objeto de obtener respuestas. En una aplicación de la estadística, existen pues, dos mundos: el de la realidad, y el del modelo de probabilidad que lo representa. En el mundo real típicamente existe formulada alguna pregunta de interés. En la medida en que el modelo de probabilidad represente adecuadamente a la realidad, las respuestas que se obtengan con el modelo serán también apegadas a la realidad.

En distintas disciplinas científicas, con frecuencia se asume que las variaciones aleatorias de diversos fenómenos pueden describirse con un modelo de probabilidad normal, conocido por su simetría con respecto a la media. Sin embargo, ciertos fenómenos aleatorios, al ser observados y registrados, generalmente producen conjuntos de datos no negativos que presentan una evidente falta de simetría con respecto a su media muestral. Algunos ejemplos son: periodo de latencia de enfermedades infecciosas, concentración de productos químicos en el ambiente, tiempo de supervivencia después de un diagnóstico de cancer, entre otros. Limpert, Stahel y Abbt (2001) muestra que a lo largo de la historia, tales situaciones y muchas otras más han sido modeladas con una distribución lognormal.

Una variable aleatoria con distribución lognormal es, en su forma más sencilla, aquella variable cuyo logaritmo natural tiene distribución normal. La génesis de la distribución lognormal fue divulgada por Galton (1879) quien atribuye a McAlister el desarrollo matemático de ésta.

Definición 1 *Una variable aleatoria Y se distribuye lognormal con parámetros $\mu \in \mathbb{R}$ y $\sigma > 0$ si y sólo si $X = \ln(Y)$ se distribuye normal con media μ y varianza σ^2 .*

La función de densidad de Y es

$$f_Y(y; \mu, \sigma) = \frac{1}{y\sigma\sqrt{2\pi}} \exp\left\{-\frac{[\ln(y) - \mu]^2}{2\sigma^2}\right\} I_{(0, \infty)}(y), \quad (1.1)$$

donde $I_{(0, \infty)}(y)$ es la función indicadora de y en el intervalo $(0, \infty)$.

La función de densidad lognormal dada en (1.1) se puede obtener de manera muy simple con el siguiente teorema.

Teorema 2 *Sea X una variable aleatoria con función de densidad $f_X(x)$ y $Y = g(X)$, donde g es una función monótona. Sea $\mathcal{X} = \{x : f_X(x) > 0\}$ y $\mathcal{Y} = \{y : y = g(x), \text{ para algún } x \in \mathcal{X}\}$ el espacio muestral de X y Y , respectivamente. Supóngase que $f_X(x)$ es continua en \mathcal{X} y $g^{-1}(x)$ tiene derivada continua en \mathcal{X} . Entonces la función de densidad de Y es*

$$f_Y(y) = \begin{cases} f_X[g^{-1}(y)] \left| \frac{d}{dy} g^{-1}(y) \right|, & \text{si } y \in \mathcal{Y}, \\ 0, & \text{en otro caso.} \end{cases}$$

Para detalles de la demostración de este teorema véase Casella y Berger (2002, pág. 51)

Así, cuando X es una variable aleatoria normal con media μ y varianza σ^2 , y $g(x) = \exp(x)$ entonces del Teorema 2 se sigue fácilmente que la variable aleatoria $Y = \exp(X)$ tiene la función de densidad dada en (1.1).

La función de distribución de una variable aleatoria lognormal, $F_Y(y; \mu, \sigma)$, se puede definir en términos de la función de distribución de una variable aleatoria normal estándar. Como $Y = \exp(X)$ sigue una distribución lognormal, cuando X es normal con media μ y varianza σ^2 , y la función logaritmo natural es una función monótona entonces se tiene la siguiente cadena de igualdades:

$$\begin{aligned} F_Y(y; \mu, \sigma) &= P(Y \leq y) \\ &= P[\exp(X) \leq y] \\ &= P[X \leq \ln(y)]. \end{aligned} \quad (1.2)$$

Estandarizando la expresión dada en (1.2) se sigue que

$$F_Y(y; \mu, \sigma) = P \left[\frac{X - \mu}{\sigma} \leq \frac{\ln(y) - \mu}{\sigma} \right]. \quad (1.3)$$

Finalmente, puesto que la variable $(X - \mu)/\sigma$ sigue una distribución normal estándar (aplicación del Teorema 2), de (1.3) se deduce que

$$F_Y(y; \mu, \sigma) = \Phi \left[\frac{\ln(y) - \mu}{\sigma} \right], \quad (1.4)$$

donde $\Phi(z)$ es la función de distribución de una variable normal estándar.

1.2. Características del modelo lognormal

Dos características importantes de la distribución de probabilidad de una variable aleatoria Y son los primeros dos momentos, la media o valor esperado $E(Y)$ y la varianza $Var(Y)$. Para el caso lognormal, se tiene que

$$E(Y) = \exp \left(\mu + \frac{1}{2}\sigma^2 \right) \quad (1.5)$$

y

$$Var(Y) = \exp(2\mu + 2\sigma^2) - \exp(2\mu + \sigma^2) \quad (1.6)$$

Nótese que tanto la media como la varianza dadas en (1.5) y (1.6) dependen de los dos parámetros de la distribución lognormal, μ y σ . Esto contrasta con el caso normal donde μ representa la media y σ^2 la varianza. Obsérvese además que en el caso normal se tiene que la media, la moda y la mediana coinciden; sin embargo, para el caso lognormal estas tres cantidades tienen expresiones matemáticas diferentes. Para mayores detalles sobre estas cantidades y otros momentos de la distribución lognormal véase Evans, Hastings y B.Peacock (2000).

Para especificar completamente la densidad lognormal $f_Y(y; \mu, \sigma)$ dada en (1.1) se necesitan dos parámetros, $\mu = E[\ln(Y)]$ y $\sigma^2 = Var[\ln(Y)]$. A continuación se presentan dos figuras, Figura 1.1 y Figura 1.2, donde se muestra el grado en el que influyen

los parámetros μ y σ en la forma de la función de densidad lognormal. La Figura 1.1 muestra la gráfica de la función de densidad de una variable aleatoria lognormal considerando diferentes valores de μ y manteniendo fijo $\sigma = 1$. Con elección particular de μ y σ puede verificarse que μ no es un parámetro de localización como sucede en el modelo normal.

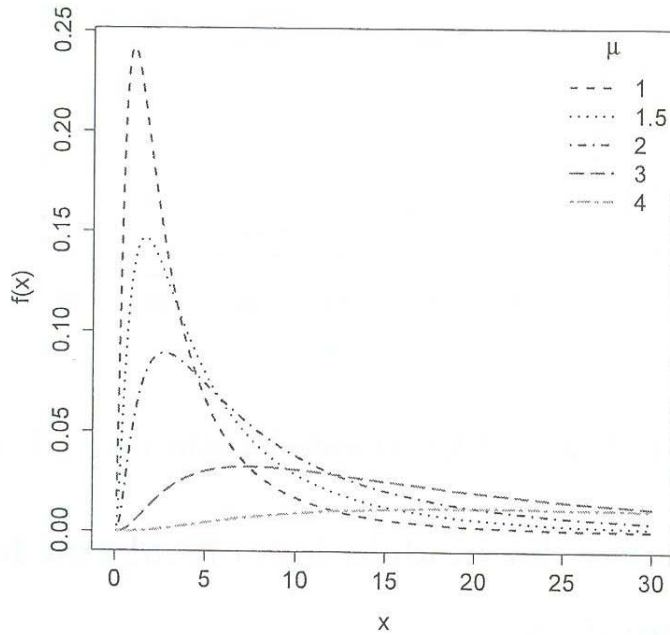


Figura 1.1: Densidad de una variable lognormal con $\sigma = 1$ fijo.

La Figura 1.2 presenta la función de densidad de una variable lognormal; pero ahora considerando distintos valores de σ y manteniendo fijo $\mu = 2$. Puede observarse una posible relación entre el valor que toma σ con la simetría de la gráfica. Evans *et al.* (2000) menciona que para valores pequeños de σ la distribución se asemeja a una normal. De hecho, si $\sigma = 0.3$ es posible notar cierta simetría en la gráfica.

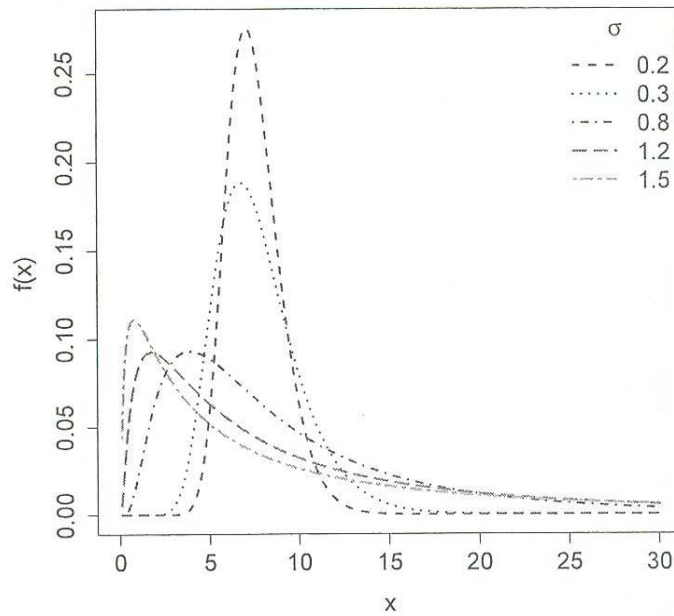


Figura 1.2: Densidad de una variable lognormal con $\mu = 2$ fijo.

1.3. Relación del modelo lognormal con la familia de Box-Cox

Box y Cox (1964) propusieron una familia de modelos definida a través de la transformación de una variable aleatoria X considerando como característica principal que la nueva variable aleatoria $X^{(\lambda)}$ tenga una distribución normal. La familia de transformaciones de Box y Cox es la siguiente:

$$x^{(\lambda)} = \begin{cases} \frac{x^\lambda - 1}{\lambda}, & \text{si } \lambda \neq 0, \\ \ln(x), & \text{si } \lambda = 0, \end{cases} \quad (1.7)$$

donde $x > 0$ y $\lambda \in \mathbb{R}$.

Así, para modelar una variable aleatoria positiva la idea es considerar tentativamente que existen valores de λ tales que $X^{(\lambda)}$ se distribuye normal con media μ y varianza σ^2 , posteriormente deberá validarse esta suposición. Ahora, suponiendo que tal hipótesis se cumple y considerando la transformación dada en (1.7) entonces del Teorema 2 se sigue que la función de densidad de la variable aleatoria X es

$$\begin{aligned} f(x; \lambda, \mu, \sigma) &= \left| \frac{\partial x^{(\lambda)}}{\partial x} \right| \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(x^{(\lambda)} - \mu)^2}{2\sigma^2} \right] I_{(0,\infty)}(x) \\ &= \frac{x^{\lambda-1}}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(x^{(\lambda)} - \mu)^2}{2\sigma^2} \right] I_{(0,\infty)}(x). \end{aligned} \quad (1.8)$$

donde $x > 0$, $\mu \in \mathbb{R}$, $\sigma > 0$ y $\lambda \in \mathbb{R}$.

La función de densidad de Box y Cox dada en (1.8) involucra tres parámetros. Debe tenerse en cuenta que el parámetro λ tiene una interpretación lógica distinta a la que tienen μ y σ . Box y Cox (1964, pág. 213) mencionan que estimar λ puede ser por interés directo en el parámetro o ser un paso preliminar en un proceso de inferencia. En este trabajo se considerará a λ como un parámetro que sugiere matemáticamente el modelo para una muestra dada. De hecho se tendrá un interés especial en el valor $\lambda = 0$ puesto que corresponde a la transformación logaritmo natural de la variable original. Es decir, corresponde al caso en el que $X^{(0)} = \ln(X)$ sigue una distribución normal con media μ y varianza σ^2 , y por lo tanto X sigue una distribución lognormal de parámetros μ y σ (véase Definición 1). Así, resultados de inferencias sobre el parámetro λ serán consideraciones fundamentales para justificar la elección del modelo lognormal dentro de la familia de Box y Cox, para un juego específico de datos.

Es importante recordar aquí que el objetivo general de esta tesis es enfatizar la importancia de un análisis de robusticidad en un proceso de inferencia estadística. En particular, como parte de este proceso, se aborda el problema de hacer inferencia sobre el cociente de medias (parámetro de interés) de variables aleatorias con distribución lognormal. En el siguiente capítulo se darán conceptos básicos para relizar un proceso de inferencia estadística sobre parámetros de interés usando técnicas asociadas con el

enfoque de verosimilitud.

Capítulo 2

Conceptos de inferencia estadística

Los problemas de inferencia estadística usualmente inician con un conjunto de datos e información acerca de la manera en que éstos fueron obtenidos. En la inferencia estadística los datos son considerados valores observados de variables aleatorias. Generalmente, el punto de partida en un proceso de inferencia estadística es el modelado estadístico paramétrico. Es decir, la especificación de un conjunto \mathcal{M} de funciones de densidad de probabilidad paramétricas para la variable aleatoria que representa el fenómeno aleatorio bajo estudio.

Definición 3 (Modelo estadístico paramétrico) *Un modelo estadístico paramétrico \mathcal{M} es un conjunto de funciones de densidad de probabilidad $f(x; \theta)$ indexadas por un parámetro θ que toma valores en un conjunto $\Theta \subset \mathbb{R}^d$,*

$$\mathcal{M} = \{f(x; \theta) | \theta \in \Theta \subset \mathbb{R}^d\},$$

para algún número natural d .

El poder delimitar un conjunto de modelos paramétricos para un problema particular requiere, por parte del investigador, entrenamiento y conocimiento de la disciplina, ya que debe argumentarse las razones por las cuales se incluye cierto modelo, entre

los posibles candidatos a elegir, así como una justificación sobre el porqué descartar algún otro. La especificación del modelo no es arbitraria, sino que está basada en las características físicas del fenómeno bajo estudio y la manera en que los datos fueron obtenidos.

Una vez especificado el modelo para un fenómeno aleatorio de interés observado, en esta tesis se propone realizar los siguientes pasos para completar un proceso de inferencia:

1. Valoración del modelo
2. Selección del modelo
3. Inferencia sobre parámetros de interés en el contexto del problema
4. Interpretación de los resultados en el contexto del problema

Es importante mencionar que en un proceso de inferencia estadística, aún en el caso más simple, se plantean de manera implícita o explícita supuestos acerca del modelo. Dichos supuestos no se consideran exactamente verdaderos pero resultan muy útiles en algunas ramas de las matemáticas aplicadas. Para justificarlos se asume un incierto principio de estabilidad conocido como *robusticidad*: un error menor en el modelo produce un error menor en las conclusiones finales. Tal razonamiento no es correcto en todos los casos. Se sabe que algunos de los procedimientos más comunes son bastante sensibles a ligeros cambios en los supuestos del modelo, Huber y Ronchetti (2009). El concepto de robusticidad será fundamental en este trabajo de tesis. De hecho, aquí se propone incorporar un análisis de robusticidad al proceso de inferencia estadística descrito antes.

A continuación, en la Sección 2.1 se definen algunos conceptos útiles para hacer inferencia estadística vía el enfoque de verosimilitud, en la Sección 2.2 se proporcionan algunas herramientas tanto teóricas como heurísticas para valuar o valorar modelos

estadísticos con base en una muestra observada y en la Sección 2.3 se dan algunos criterios prácticos para la selección de modelos. Por último, en la Sección 2.4 se define el concepto de robusticidad y se enfatiza su importancia en el proceso de inferencia.

2.1. Estimación vía el enfoque de verosimilitud

Un enfoque que aborda el problema de estimación de parámetros y que es ampliamente utilizado por su adecuado uso de la información es el de verosimilitud. En esta sección se presentan conceptos básicos de este enfoque como: la función de verosimilitud, la función de verosimilitud relativa, regiones de verosimilitud, regiones de verosimilitud-confianza y verosimilitud perfil. Todos estos conceptos teóricos serán aplicados más adelante cuando se aborde uno de los problemas centrales de este trabajo, hacer inferencia estadística sobre el cociente de medias de poblaciones con distribución lognormal.

2.1.1. La función de verosimilitud

Fisher (1921), en su artículo sobre el coeficiente de correlación poblacional ρ , define por primera vez el concepto de verosimilitud y explica que, observada la muestra, se puede calcular la verosimilitud para cualquier valor particular de ρ . Define ésta como una cantidad proporcional a la probabilidad de la muestra observada; pero como función del parámetro ρ .

Considerando lo anterior, aquí se define el concepto de función de verosimilitud de la siguiente manera.

Definición 4 (Función de verosimilitud) *Supóngase que $X = (X_1, \dots, X_n)$ es un vector de variables aleatorias y E_{obs} es un evento observado en términos de X con probabilidad $P_X(E_{obs}; \theta)$ que depende de un número finito de parámetros reales desconocidos $\theta = (\theta_1, \dots, \theta_d) \in \Theta \subset \mathbb{R}^d$, donde Θ es el espacio parametral. La verosimilitud de θ es*

una función $L(\theta; E_{obs}) : \Theta \rightarrow [0, 1]$ definida como

$$L(\theta; E_{obs}) = P_X(E_{obs}; \theta). \quad (2.1)$$

Nótese que si $X = (X_1, \dots, X_n)$ es un vector de variables aleatorias discretas con función de probabilidad conjunta $f(x; \theta) : \mathbb{R}^n \rightarrow [0, 1]$ y $x_{obs} = (x_1, \dots, x_n)$ es el evento observado entonces la función de verosimilitud de θ es

$$\begin{aligned} L(\theta; x_{obs}) &= P_X(E_{obs}; \theta) \\ &= f(x_{obs}; \theta) \\ &= f(x_1, \dots, x_n; \theta). \end{aligned}$$

En particular si X_1, \dots, X_n son variables aleatorias discretas independientes e idénticamente distribuidas (i.i.d.) con función de probabilidad $f(x; \theta) : \mathbb{R} \rightarrow [0, 1]$, entonces la verosimilitud de θ es

$$L(\theta; x_{obs}) = \prod_{i=1}^n f(x_i; \theta).$$

Ahora, si X_1, \dots, X_n son variables aleatorias continuas i.i.d. con función de densidad $f(x; \theta) : \mathbb{R} \rightarrow \mathbb{R}^+$ y el evento observado fue $y_i - h \leq X_i \leq y_i + h$, donde y_i y h son valores fijos y conocidos ($h > 0$), entonces la verosimilitud de θ es

$$\begin{aligned} L(\theta; y_{obs}) &= P_X(E_{obs}; \theta) \\ &= \prod_{i=1}^n \int_{y_i-h}^{y_i+h} f(t; \theta) dt \end{aligned} \quad (2.2)$$

En la práctica, la cantidad y_i puede ser considerada como un valor observado de X_i obtenida mediante un instrumento que tiene un mecanismo de medición simétrico y de precisión conocida $h > 0$ para medir la magnitud de X_i .

Nótese que bajo ciertas condiciones la función de densidad de una variable aleatoria continua podrá usarse como una aproximación a la función de verosimilitud. Por el teorema de valor medio para integrales de funciones continuas, la i -ésima integral de

(2.2) es $2hf(y', \theta)$ para algún $y' \in [y_i - h, y_i + h]$. Si $f(y'; \theta) \approx f(y_i; \theta)$ para todo valor de θ entonces

$$\int_{y_i-h}^{y_i+h} f(y; \theta) dy \approx 2hf(y_i; \theta).$$

Si esta aproximación es adecuada para todos los $i \in \{1, \dots, n\}$ y h no depende de θ entonces el producto de integrales dado en (2.2) es aproximadamente proporcional al producto de densidades evaluadas en los valores observados y_i . Todo esto conduce a la siguiente definición práctica de la función de verosimilitud para el caso de muestras observadas de modelos continuos.

Definición 5 (Aproximación continua a la verosimilitud) Sea $X = (X_1, \dots, X_n)$ es un vector de variables aleatorias continuas con función de densidad conjunta $f(x; \theta) : \mathbb{R}^n \rightarrow \mathbb{R}^+$. Si $x_{obs} = (x_1, \dots, x_n)$ es una muestra observada de X entonces la aproximación continua a la función de verosimilitud de θ se define como

$$\begin{aligned} L_a(\theta; x_{obs}) &= c \cdot f(x_{obs}; \theta) \\ &= c \cdot f(x_1, \dots, x_n; \theta) \end{aligned} \tag{2.3}$$

donde $c > 0$ es una constante arbitraria que no depende de θ .

En lo que sigue de esta tesis, la aproximación continua L_a a la función de verosimilitud se denotará con la letra L correspondiente a la función de verosimilitud legítima, a menos que se requiera especificar explícitamente el uso de la aproximación continua.

Si X_1, \dots, X_n son variables aleatorias continuas i.i.d. con función de densidad $f(x; \theta) : \mathbb{R} \rightarrow \mathbb{R}^+$ y $x_{obs} = (x_1, \dots, x_n)$ es una muestra observada entonces la aproximación continua a la función de verosimilitud de θ es

$$L(\theta; x_{obs}) = c \cdot \prod_{i=1}^n f(x_i; \theta),$$

donde c es una constante arbitraria que no depende de θ .

La función de verosimilitud juega un papel fundamental en la Inferencia Estadística. Su rol principal es inferir qué valores del vector de parámetros $\theta = (\theta_1, \dots, \theta_d)$ de la función de densidad de probabilidad $f(x; \theta)$ que haya sido elegida para el fenómeno aleatorio de interés, son razonables a la luz del evento observado (datos). Nótese que esto es particularmente relevante después de un experimento, cuando ya fue observada la muestra.

Plausibilidad

La función de verosimilitud $L(\theta; E_{obs})$ permite ordenar la plausibilidad entre los valores de θ con base en el evento observado E_{obs} . Si $L(\theta_1; E_{obs}) > L(\theta_2; E_{obs})$ entonces de (2.1) se sigue que $P_X(E_{obs}; \theta_1) > P_X(E_{obs}; \theta_2)$. Es decir; el evento observado E_{obs} es más probable cuando el parámetro θ toma el valor θ_1 que cuando toma el valor θ_2 . Así, el cociente de verosimilitudes

$$\frac{L(\theta_1; E_{obs})}{L(\theta_2; E_{obs})} = \frac{P_X(E_{obs}; \theta_1)}{P_X(E_{obs}; \theta_2)},$$

es una medida de la plausibilidad de θ_1 relativa a θ_2 basada en el evento observado E_{obs} . El cociente $L(\theta_1; E_{obs})/L(\theta_2; E_{obs}) = k$ significa que el valor θ_1 es k veces más plausible que el valor θ_2 en el sentido de que θ_1 hace al evento observado k veces más probable de lo que lo hace θ_2 .

Estimador de máxima verosimilitud

La función de verosimilitud $L(\theta; E_{obs})$ proporciona la plausibilidad de cada valor de θ dado el evento observado E_{obs} . Así, un estimador puntual de θ razonable es aquel valor de θ que maximiza $L(\theta; E_{obs})$. Es decir, aquel valor de θ que maximiza la probabilidad de la muestra observada.

Definición 6 (Estimador de máxima verosimilitud) *El estimador de máxima verosimilitud (e.m.v.) de θ es cualquier valor $\hat{\theta}$ en el espacio parametral $\Theta \subset \mathbb{R}^d$, que*

cumple que

$$L(\hat{\theta}; E_{obs}) = \sup_{\theta \in \Theta} L(\theta; E_{obs}).$$

Nótese que $L(\theta; E_{obs}) = P_X(E_{obs}; \theta) \leq 1$ debido a que $P_X(E_{obs}; \theta)$ es una probabilidad como función de θ . Así, el supremo de $L(\theta; E_{obs})$ en $[0, 1]$ existe y es finito. Sin embargo, puede ocurrir que el e.m.v. no exista, o si existe, puede ser que no sea único. De aquí en adelante, en este capítulo se procederá como si existiera un único valor $\hat{\theta} \in \Theta$ que maximiza a la función de verosimilitud $L(\theta; E_{obs})$. Es decir, se supondrá que el e.m.v. existe y es único.

El e.m.v. del parámetro θ es el valor más plausible de θ . Es decir, el e.m.v. $\hat{\theta}$ es el valor de θ que explica mejor al evento observado en el sentido de que maximiza su probabilidad bajo el modelo de probabilidad propuesto para el fenómeno aleatorio de interés.

Para encontrar el e.m.v. de θ es necesario maximizar la función de verosimilitud $L(\theta; E_{obs})$. Sin embargo, usualmente $L(\theta; E_{obs})$ es el producto de varios términos y resulta matemáticamente conveniente (también desde un punto de vista de optimización numérica) trabajar con el logaritmo natural de la función de verosimilitud.

Definición 7 (La función log-verosimilitud) *La función log-verosimilitud se define como*

$$l(\theta; E_{obs}) = \ln [L(\theta; E_{obs})].$$

Nótese que el valor $\hat{\theta}$ que maximiza $L(\theta; E_{obs})$ también maximiza $l(\theta; E_{obs})$ puesto que el logaritmo natural es una transformación monótona.

Invarianza

En algunas ocasiones se utilizan reparametrizaciones de un modelo de probabilidad por conveniencia matemática, conveniencia computacional, interés en un parámetro que es función de otros que aparecen en el modelo, etc. Para estas situaciones la invarianza funcional es una propiedad muy conveniente de la verosimilitud. Significa que, en términos de plausibilidad, cualquier declaración cuantitativa acerca de θ implica la misma declaración cuantitativa acerca de cualquier función uno a uno de θ . Una función $g(\theta)$ se llama función uno a uno si $g(\theta_1) = g(\theta_2)$ implica que $\theta_1 = \theta_2$.

Teorema 8 (Invarianza funcional de la verosimilitud) *Supóngase que $L(\theta; E_{obs})$ es la función de verosimilitud de $\theta \in \Theta$. Sea $g(\theta) : \Theta \rightarrow \Delta$ una función uno a uno. Si*

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

entonces

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

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

Demostración. Se sigue por directa sustitución algebraica. ■

Una consecuencia inmediata del Teorema 8 es la invarianza del e.m.v.

Teorema 9 (Invarianza del e.m.v.) *Si $g(\theta) : \Theta \rightarrow \Delta$ es una función uno a uno y $\hat{\theta}$ es el e.m.v. de θ entonces el e.m.v. de $\delta = g(\theta)$ es $\hat{\delta} = g(\hat{\theta})$.*

Demostración. Supóngase que $L(\theta; E_{obs})$ es la función de verosimilitud de θ . Como $\delta = g(\theta)$ es una función uno a uno entonces $\theta = g^{-1}(\delta)$, donde g^{-1} es la función inversa de g . Así, la función de verosimilitud de δ es $L'(\delta; E_{obs}) = L[\theta = g^{-1}(\delta); E_{obs}]$. Entonces,

$$L'(\hat{\delta}; E_{obs}) = \sup_{\delta \in \Delta} L'(\delta; E_{obs}) = \sup_{\delta \in \Delta} L[\theta = g^{-1}(\delta); E_{obs}] = \sup_{\theta \in \Theta} L(\theta; E_{obs}) = L(\hat{\theta}; E_{obs}).$$

Así, $\hat{\theta} = g^{-1}(\hat{\delta})$. Por lo tanto, $\hat{\delta} = g[g^{-1}(\hat{\delta})] = g(\hat{\theta})$. ■

Cabe mencionar aquí que Casella y Berger (2002, pág. 320) muestran que la propiedad de invarianza del e.m.v. se cumple para cualquier función $g(\theta)$, no necesariamente para funciones uno a uno.

2.1.2. Verosimilitud relativa

La función de verosimilitud relativa es una estandarización de la función de verosimilitud $L(\theta; E_{obs})$ respecto a su valor máximo.

Definición 10 (Verosimilitud relativa) *La verosimilitud relativa de θ es una función $R(\theta; E_{obs}) : \Theta \rightarrow [0, 1]$ definida como*

$$R(\theta; E_{obs}) = \frac{L(\theta; E_{obs})}{\sup_{\theta \in \Theta} L(\theta; E_{obs})} \quad (2.4)$$

donde $L(\theta; E_{obs})$ es la función de verosimilitud de θ .

Como el supremo de $L(\theta; E_{obs})$ en $[0, 1]$ existe y es finito entonces la función de verosimilitud relativa existe y se puede emplear para hacer declaraciones cuantitativas acerca del grado sobre el cual valores del parámetro describen mejor el fenómeno aleatorio de interés, con base en el evento observado.

Cuando es posible encontrar un único valor de θ que maximiza $L(\theta; E_{obs})$, entonces la función de verosimilitud relativa (2.4) es una estandarización de $L(\theta; E_{obs})$ respecto a su máximo $L(\hat{\theta}; E_{obs})$. Es decir, si el estimador de máxima verosimilitud de θ , $\hat{\theta}$, existe y es único entonces la función de verosimilitud relativa es

$$R(\theta; E_{obs}) = \frac{L(\theta; E_{obs})}{L(\hat{\theta}; E_{obs})} \quad (2.5)$$

La verosimilitud relativa toma valores entre cero y uno, $0 \leq R(\theta; E_{obs}) \leq 1$, para todo valor de θ en el espacio parametral Θ . La función de verosimilitud relativa dada en (2.5) proporciona la plausibilidad de cualquier valor especificado de θ relativa al e.m.v., con base en el evento observado. Es decir, valores de θ con $R(\theta; E_{obs})$ cercano

a $R(\hat{\theta}; E_{obs}) = 1$ hacen al evento observado casi tan probable como lo hace el e.m.v. $\hat{\theta}$. En contraste, valores de θ con $R(\theta; E_{obs})$ cercanos a cero hacen que la probabilidad del evento observado sea pequeña con respecto a su máxima probabilidad alcanzada en $\hat{\theta}$.

2.1.3. Regiones de verosimilitud

A diferencia de la estimación puntual donde se proporciona un único valor para θ , estimador puntual; en la estimación por regiones o intervalos se busca dar una expresión del tipo " $\theta \in C$ ", donde $C \subset \Theta$ depende de la muestra observada.

Definición 11 (Región de verosimilitud de nivel c) *Supóngase que $R(\theta; E_{obs})$ es la función de verosimilitud relativa de θ . Una región de verosimilitud de nivel de plausibilidad c para θ se define como el conjunto de todos los valores de θ que cumplen que $R(\theta; E_{obs}) \geq c$,*

$$C(c; E_{obs}) = \{\theta | R(\theta; E_{obs}) \geq c\},$$

donde $0 \leq c \leq 1$.

Todo valor de θ en una región de verosimilitud de nivel de plausibilidad c tiene verosimilitud relativa igual o mayor que c , y todo valor de θ afuera, tiene verosimilitud relativa menor. Por tanto, una región de verosimilitud de nivel de plausibilidad c separa los valores plausibles de θ de los no plausibles a un nivel c , Sprott (2000, pág. 14).

Proposición 1 *Si $\{k_m\}_{m \geq 0}$ es una sucesión monótona creciente de números reales, donde $k_m \in [0, 1]$, entonces las regiones de verosimilitud de nivel de plausibilidad $\{k_m\}_{m \geq 0}$ son conjuntos anidados.*

Demostración. Sea $R(\theta; E_{obs})$ la función de verosimilitud relativa de θ . Supóngase que $C(k_m; E_{obs})$ es una región de verosimilitud de nivel de plausibilidad k_m para θ , donde $0 \leq k_m \leq 1$, para todo $m \geq 0$. Sea n cualquier valor de m . Si $C(k_n; E_{obs}) = \emptyset$ entonces $C(k_n; E_{obs}) \subseteq C(k_{n-1}; E_{obs})$. Si $C(k_n; E_{obs}) \neq \emptyset$ y $\theta' \in C(k_n; E_{obs})$ entonces

$R(\theta'; E_{obs}) \geq k_n$. Ahora, si $\{k_m\}_{m \geq 0}$ es una sucesión monótona creciente de números reales en $[0, 1]$ entonces $0 \leq k_{i-1} \leq k_i \leq 1$, para todo $i \geq 1$. Así,

$$R(\theta') \geq k_n \geq k_{n-1}.$$

Es decir, θ' también pertenece a $C(k_{n-1}; E_{obs})$. Entonces, $C(k_n; E_{obs}) \subseteq C(k_{n-1}; E_{obs})$. Al repetir este proceso para $n-1, n-2, \dots, 0$ se tiene que $C(k_{i+1}; E_{obs}) \subseteq C(k_i; E_{obs})$, para $i = n-2, n-3, \dots, 0$. Así, para todo $n \geq 0$ se cumple que

$$C(k_n; E_{obs}) \subseteq C(k_{n-1}; E_{obs}) \subseteq \dots \subseteq C(k_0; E_{obs}),$$

Por lo tanto, $\{C(k_m; E_{obs})\}_{m \geq 0}$ son conjuntos anidados. ■

De la Proposición 1 se sigue que variando c de 0 a 1 se obtiene una familia jerarquizada y anidada de regiones de verosimilitud que converge al e.m.v. $\hat{\theta}$, $R(\hat{\theta}; E_{obs}) = 1$.

Si el parámetro $\theta \in \Theta \subset \mathbb{R}^d$ es unidimensional ($d = 1$), la región de verosimilitud es un intervalo, o la unión de varios intervalos. En particular, cuando una región de verosimilitud es un intervalo se le llama intervalo de verosimilitud.

Definición 12 (Intervalo de verosimilitud de nivel c) *Un intervalo de verosimilitud de nivel c es una región de verosimilitud de nivel c de la forma $C(c; E_{obs}) = [A(c; E_{obs}), B(c; E_{obs})]$, donde $A(c; E_{obs})$ y $B(c; E_{obs})$ son los extremos del intervalo.*

Nótese que generalmente los extremos de un intervalo de verosimilitud de nivel c , $A = A(c; E_{obs})$ y $B = B(c; E_{obs})$, cumplen que $R(A; E_{obs}) = R(B; E_{obs}) = c$. En estos casos, A y B se pueden obtener calculando los puntos donde llegan a intersectarse el eje cartesiano correspondiente a θ y la curva $R(\theta; E_{obs}) - c$.

Un intervalo de verosimilitud por sí solo no es muy informativo y por lo tanto insuficiente para indicar el cambio en la plausibilidad de los valores de θ adentro del intervalo. Al menos, un intervalo de verosimilitud debe estar acompañado siempre del valor del e.m.v. $\hat{\theta}$ para dar alguna idea de la simetría de la función de verosimilitud con

respecto a $\hat{\theta}$ y de cómo cambia la plausibilidad adentro del intervalo. Se recomienda dar al menos $\hat{\theta}$ junto con varios intervalos de verosimilitud de nivel $c = 0.036, 0.15, 0.25$. En lo posible se debe también graficar y analizar la función de verosimilitud relativa completa.

2.1.4. Regiones de verosimilitud-confianza

Supóngase que $X = (X_1, \dots, X_n)$ es un vector de variables aleatorias con función de densidad de probabilidad conjunta $f(x; \theta)$, donde $\theta = (\theta_1, \dots, \theta_d) \in \Theta \subset \mathbb{R}^d$ es un vector de parámetros desconocidos. Si se tiene una muestra observada $x_{obs} = (x_1, \dots, x_n)$ que proviene de la distribución de X con θ fijo en un valor θ_0 , entonces se puede calcular, a partir de ésta muestra observada x_{obs} , una región C para el valor “verdadero” θ_0 . Así, la región C cambia cuando varía la muestra. De hecho, la región C es aleatoria puesto que es función de las variables aleatorias X_1, \dots, X_n . Por lo tanto, C puede algunas veces incluir y en otras no, al valor verdadero θ_0 .

La probabilidad de cobertura de una región C es la probabilidad de que la región C incluya o cubra, el valor verdadero del parámetro, $\theta = \theta_0$. De manera formal, la probabilidad de cobertura de una región se define de la siguiente manera.

Definición 13 (Probabilidad de cobertura) *Sea $X = (X_1, \dots, X_n)$ es un vector de variables aleatorias con función de densidad de probabilidad conjunta $f(x; \theta)$. La probabilidad de cobertura de una región $C = C(X)$ para $\theta = \theta_0$ se define como*

$$PC(\theta_0) = P(\theta_0 \in C; \theta = \theta_0).$$

La probabilidad de cobertura $PC(\theta_0)$ se interpreta como la fracción de veces que la región C incluirá el valor verdadero θ_0 en un número muy grande de repeticiones de la muestra pero con el valor de θ fijo en θ_0 .

Cuando el valor de la probabilidad de cobertura de una región C , $PC(\theta_0)$, es la misma para todo valor del parámetro θ_0 entonces la región C es llamada región de

confianza. Una definición formal de región de confianza es la siguiente.

Definición 14 (Región de confianza) Sea $X = (X_1, \dots, X_n)$ es un vector de variables aleatorias con función de densidad de probabilidad conjunta $f(x; \theta)$. Una región de confianza para θ_0 es una región $C = C(X)$ para $\theta = \theta_0$ cuya probabilidad de cobertura no depende de θ_0 .

Si la región C es una región de verosimilitud de nivel c entonces la probabilidad de cobertura de esta región se puede calcular a través de la distribución de probabilidad de la estadística de la razón de verosimilitud para un θ fijo en θ_0 , $D_n = -2 \ln [R(\theta_0; X)]$. Aquí, $X = (X_1, \dots, X_n)$ es el vector de variables aleatorias con función de densidad de probabilidad conjunta $f(x; \theta)$ empleada en el cálculo de la verosimilitud relativa $R(\theta; X)$. Nótese que al fijar $\theta = \theta_0$ entonces $R(\theta_0; X)$ es una cantidad aleatoria porque es función del vector aleatorio $X = (X_1, \dots, X_n)$. Ahora, como un valor particular θ_0 está en una región de verosimilitud de nivel c sí y sólo sí $R(\theta_0; X) \geq c$, o de forma equivalente, $-2 \ln [R(\theta_0; X)] \leq -2 \ln(c)$, entonces la probabilidad de cobertura de una región de verosimilitud de nivel c para θ , $C(c; X)$, es

$$\begin{aligned} PC(\theta_0) &= P[\theta_0 \in C(c; X); \theta = \theta_0] \\ &= P[D_n \leq -2 \ln(c); \theta = \theta_0]. \end{aligned}$$

Muchas veces es difícil encontrar la distribución de probabilidad exacta de D_n ; sin embargo existe una teoría asintótica que suele dar una aproximación buena a ésta distribución en muchos casos. Bajo algunas condiciones llamadas de regularidad, la estadística de la razón de verosimilitud $D_n = -2 \ln [R(\theta_0; X)]$ converge en distribución a una Ji-cuadrada con d grados de libertad, $\chi_{(d)}^2$, para todo $\theta_0 \in \Theta \subset \mathbb{R}^d$. En este caso, esto equivale a que se tenga que

$$\lim_{n \rightarrow \infty} P(D_n \leq x; \theta = \theta_0) = P[\chi_{(d)}^2 \leq x],$$

para todo $x \geq 0$. Para mayores detalles de la prueba ver Serfling (1980).

Definición 15 (Región de verosimilitud-confianza) Una región de verosimilitud-confianza para cualquier $\theta_0 \in \Theta \subset \mathbb{R}^d$ es una región de verosimilitud que tiene probabilidad de cobertura aproximada $(1 - \alpha)$ para θ_0 .

Si se selecciona

$$c = \exp\left(-\frac{1}{2}q_{d,1-\alpha}\right),$$

donde $q_{d,1-\alpha}$ es el cuantil $(1 - \alpha)$ de una Ji-cuadrada con d grados de libertad, entonces una región de verosimilitud con este nivel de c para θ tiene una probabilidad de cobertura aproximada de $(1 - \alpha)$ para θ_0 . Así, una región de verosimilitud también es una región de confianza y toma consecuentemente el nombre de región de verosimilitud-confianza para θ .

En la Tabla 2.1 los valores 2.706, 3.841 y 6.635 son los cuantiles 0.90, 0.95 y 0.99 de una distribución Ji-cuadrada con un grado de libertad ($d = 1$), respectivamente. Entonces, las regiones de verosimilitud (generalmente intervalos de verosimilitud) de nivel $c = 0.036, 0.15$ y 0.25 tienen probabilidad de cobertura aproximada del 90 %, 95 % y 99 %, respectivamente.

$(1 - \alpha)$	c	$q_{1,1-\alpha}$
0.90	0.258	2.706
0.95	0.146	3.841
0.99	0.036	6.635

Tabla 2.1: Confianza aproximada de regiones de verosimilitud cuando θ es unidimensional.

2.1.5. Verosimilitud perfil

Con frecuencia se tienen modelos estadísticos con varios parámetros y lo que interesa es estimar un parámetro cuando se desconoce todo sobre los demás, que han recibido el

nombre de parámetros de estorbo, ruido o no deseados. El problema de la estimación por separado de parámetros de interés en presencia de los de estorbo es relevante en la estadística puesto que los parámetros de estorbo pueden provocar un impacto dramático en las inferencias acerca de los parámetros de interés.

La función de verosimilitud perfil es un método estadístico muy simple y poderoso que sirve para estimar por separado un parámetro de interés en presencia de parámetros considerados de estorbo. Para este propósito, la función de verosimilitud perfil es mucho más general que otras verosimilitudes tales como la condicional, la marginal o la integrada, las cuales dependen de una estructura especial en el modelo, y por lo tanto son más restrictivas, (Montoya, 2008).

Definición 16 (Función de verosimilitud perfil) Sea $L(\theta; E_{obs})$ la función de verosimilitud de $\theta = (\delta, \lambda) \in \Theta \subset \mathbb{R}^d$, donde $\delta \in \Delta \subset \mathbb{R}^{d_\delta}$ es un parámetro de interés de dimensión $d_\delta \geq 1$, $\lambda \in \Lambda \subset \mathbb{R}^{d_\lambda}$ es considerado un parámetro de estorbo de dimensión $d_\lambda \geq 1$ y $\Theta = \Delta \times \Lambda$ con $d = d_\delta + d_\lambda$. La verosimilitud perfil de δ es una función $L(\delta; E_{obs}) : \Delta \rightarrow [0, 1]$ definida como

$$L_P(\delta; E_{obs}) = \sup_{\lambda \in \Lambda | \delta} L[\theta = (\delta, \lambda); E_{obs}]. \quad (2.6)$$

Como $L[\theta = (\delta, \lambda); E_{obs}]$ es una probabilidad para cualquier valor particular (δ, λ) , entonces el supremo de $L[\theta = (\delta, \lambda); E_{obs}]$ en $[0, 1]$ existe y es finito para todo $\delta \in \Delta$. Así, la verosimilitud perfil existe y se puede usar para ordenar la plausibilidad entre los valores de δ .

Definición 17 (Verosimilitud perfil relativa) La verosimilitud perfil relativa de δ es una función $R_P(\delta; E_{obs}) : \Delta \rightarrow [0, 1]$ definida como

$$R_P(\delta; E_{obs}) = \frac{L_P(\delta; E_{obs})}{\sup_{\delta \in \Delta} L_P(\delta; E_{obs})}$$

donde $L_P(\delta; E_{obs})$ es la función de verosimilitud perfil de δ .

La función de verosimilitud perfil relativa se puede emplear, de la misma forma como se emplea la verosimilitud relativa dada en (2.4) para hacer declaraciones cuantitativas acerca del grado sobre el cual valores del parámetro de interés δ son plausibles a la luz de la muestra observada.

Definición 18 (Estimador de máxima verosimilitud restringido) *El e.m.v. restringido de λ es cualquier valor $\hat{\lambda}(\delta; E_{obs})$ en el espacio parametral Λ , que cumple que*

$$L \left\{ \theta = \left[\delta, \hat{\lambda}(\delta; E_{obs}) \right]; E_{obs} \right\} = \sup_{\lambda \in \Lambda | \delta} L[\theta = (\delta, \lambda); E_{obs}],$$

para cualquier valor especificado de $\delta \in \Delta$.

Si $\hat{\lambda}(\delta; E_{obs})$ existe y es único para cada valor especificado de δ entonces la función de verosimilitud perfil de δ se puede escribir como

$$L_P(\delta; E_{obs}) = L \left\{ \theta = \left[\delta, \hat{\lambda}(\delta; E_{obs}) \right]; E_{obs} \right\}.$$

Más aún, si el e.m.v. de θ existe y es único, $\hat{\theta} = (\hat{\delta}, \hat{\lambda})$, entonces la función de verosimilitud perfil relativa de δ es

$$R_P(\delta; E_{obs}) = \frac{L_P(\delta; E_{obs})}{L(\hat{\theta}; E_{obs})} = \frac{L \left\{ \theta = \left[\delta, \hat{\lambda}(\delta; E_{obs}) \right]; E_{obs} \right\}}{L \left\{ \theta = \left[\hat{\delta}, \hat{\lambda}(\hat{\delta}; E_{obs}) \right]; E_{obs} \right\}}.$$

Cuando δ y λ son parámetros unidimensionales ($d_\delta = d_\lambda = 1$) entonces la función de verosimilitud de ambos parámetros, $L(\delta, \lambda; E_{obs})$, es una superficie en \mathbb{R}^3 cuyo dominio es el plano cartesiano correspondiente al espacio parametral $\Delta \times \Lambda$. Es decir, la verosimilitud es una función real valuada, que está definida para cada pareja $(\delta, \lambda) \in \Delta \times \Lambda$. Así, cuando uno se posiciona en un punto muy distante sobre el eje de estorbo λ , entonces la silueta o perfil que se observa de esta verosimilitud $L(\delta, \lambda; E_{obs})$ es justamente la función de verosimilitud perfil de δ . A manera de ilustración véase la Figura 2.1 donde δ es el parámetro de interés y λ es el parámetro de estorbo.

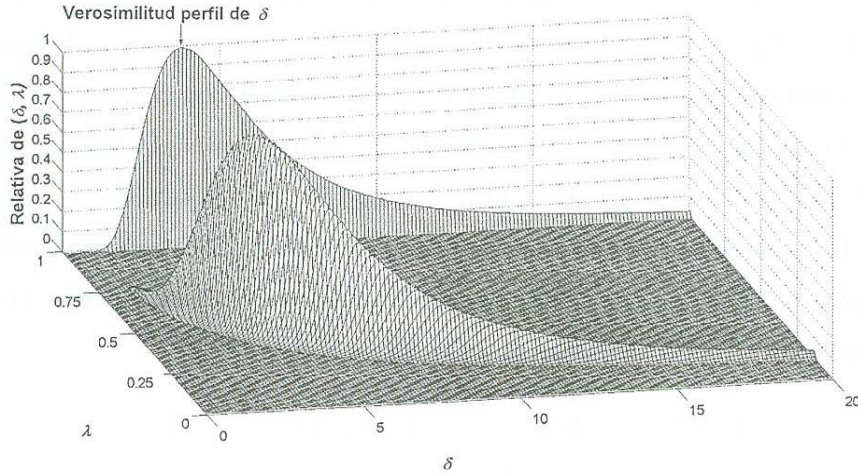


Figura 2.1: Función de verosimilitud perfil de δ .

Cuando $\delta = (\delta_1, \delta_2) \in \Delta = \Delta_1 \times \Delta_2 \subset \mathbb{R}^2$ y $\lambda \in \Lambda \subset \mathbb{R}^{d_\lambda}$, $d_\delta = 2$ y $d_\lambda \geq 1$, entonces la función de verosimilitud perfil de $\delta = (\delta_1, \delta_2)$, $L_P(\delta_1, \delta_2; E_{obs})$, es una superficie en \mathbb{R}^3 cuyo dominio es el plano cartesiano correspondiente al espacio parametral $\Delta = \Delta_1 \times \Delta_2$. Es decir, la verosimilitud perfil es una función real valuada, que está definida para cada pareja $(\delta_1, \delta_2) \in \Delta_1 \times \Delta_2$.

Definición 19 (Región de verosimilitud perfil) *Supóngase que $R_P(\delta; E_{obs})$ es la función de verosimilitud perfil relativa de δ . Una región de verosimilitud perfil de nivel de plausibilidad c para δ se define como el conjunto de todos los valores de δ que cumplen que $R_P(\delta; E_{obs}) \geq c$,*

$$C_P(c; E_{obs}) = \{\delta \mid R_P(\delta; E_{obs}) \geq c\},$$

donde $0 \leq c \leq 1$.

La probabilidad de cobertura de una región de verosimilitud perfil de nivel de plausibilidad c se puede calcular a través de la distribución de probabilidad de la

estadística de la razón de verosimilitud perfil para un δ fijo en δ_0 , $-2 \ln [R_P(\delta_0; X)]$. Aquí, $X = (X_1, \dots, X_n)$ es el vector de variables aleatorias con función de densidad de probabilidad conjunta $f(x; \theta)$ empleada en el cálculo de la verosimilitud perfil relativa $R_P(\delta; X)$. Al fijar $\delta = \delta_0$ entonces $R_P(\delta_0; X)$ es una cantidad aleatoria porque es función del vector aleatorio $X = (X_1, \dots, X_n)$. Un valor particular δ_0 está en una región de verosimilitud perfil de nivel c si y sólo si $R_P(\delta_0; X) \geq c$, o de forma equivalente, $-2 \ln [R_P(\delta_0; X)] \leq -2 \ln(c)$, entonces la probabilidad de cobertura de una región de verosimilitud perfil de nivel c para δ , $C_P(c; X)$, es

$$\begin{aligned} PC(\delta_0) &= P[\delta_0 \in C_P(c; X); \delta = \delta_0] \\ &= P[-2 \ln [R_P(\delta_0; X)] \leq -2 \ln(c); \delta = \delta_0]. \end{aligned}$$

Muchas veces es difícil encontrar la distribución de probabilidad exacta de la estadística de la razón de verosimilitud perfil; sin embargo existe teoría asintótica que suele dar una aproximación buena a ésta distribución en muchos casos. Bajo algunas condiciones de regularidad, $-2 \ln [R_P(\delta_0; X)]$ converge en distribución a una Ji-cuadrada con d_δ grados de libertad para todo $\delta_0 \in \Delta \subset \mathbb{R}^{d_\delta}$,

$$\lim_{n \rightarrow \infty} P\{-2 \ln [R_P(\delta_0; X)] \leq x; \delta = \delta_0\} = P[\chi_{(d_\delta)}^2 \leq x],$$

para todo $x \geq 0$. Para detalles de la prueba véase Serfling (1980).

Definición 20 (Región de verosimilitud perfil-confianza) *Una región de verosimilitud perfil-confianza para cualquier $\delta_0 \in \Delta \subset \mathbb{R}^{d_\delta}$ es una región de verosimilitud perfil que tiene probabilidad de cobertura aproximada $(1 - \alpha)$ para δ_0 .*

Si se selecciona

$$c = \exp\left(-\frac{1}{2}q_{d_\delta, 1-\alpha}\right),$$

donde $q_{\delta, 1-\alpha}$ es el cuantil $(1 - \alpha)$ de una Ji-cuadrada con d_δ grados de libertad, entonces una región de verosimilitud con este nivel de c para δ tiene una probabilidad

de cobertura aproximada de $(1 - \alpha)$ para δ_0 . Así, una región de verosimilitud perfil también es una región de confianza y toma consecuentemente el nombre de región de verosimilitud perfil-confianza para δ .

Cuando δ es un parámetro escalar ($d_\delta = 1$), entonces una región de verosimilitud perfil-confianza para δ puede ser tratada como una región de verosimilitud-confianza de una verosimilitud típica. Así, regiones de verosimilitud (generalmente intervalos de verosimilitud) con niveles de plausibilidad $c = 0.036, 0.15$ y 0.25 tienen probabilidad de cobertura aproximada del 99 %, 95 % y 95 %, respectivamente.

La verosimilitud perfil es importante puesto que es un método formal y simple más fácil de aplicar que otras alternativas para estimar por separado un parámetro de interés en presencia de parámetros de estorbo. En esta tesis, la verosimilitud perfil será usada para este propósito; en particular, para abordar los dos temas centrales de este trabajo que son: hacer inferencias sobre el cociente de medias de variables con distribución lognormal y evaluar la robusticidad.

2.1.6. Ejemplo: Cociente de medias de lognormales

En esta sección se ejemplifican conceptos de verosimilitud definidos anteriormente (algunos serán retomados en el Capítulo 3), utilizando el siguiente escenario estadístico. Sea $X = (X_1, \dots, X_n)$ un vector de variables aleatorias i.i.d. con función de densidad lognormal $f(x; \mu_x, \sigma_x)$, donde $\mu_x \in \mathbb{R}$ y $\sigma_x > 0$. Sea $Y = (Y_1, \dots, Y_m)$ otro vector de variables aleatorias i.i.d. con función de densidad lognormal $f(y; \mu_y, \sigma_y)$, donde $\mu_y \in \mathbb{R}$

y $\sigma_y > 0$, independiente de X . Así, la función de densidad conjunta de X y Y es

$$\begin{aligned}
f(x, y; \mu_x, \mu_y, \sigma_x, \sigma_y) &= \prod_{i=1}^n f(x_i; \mu_x, \sigma_x) \prod_{j=1}^m f(y_j; \mu_y, \sigma_y) \\
&= \prod_{i=1}^n \frac{1}{x_i \sigma_x \sqrt{2\pi}} \exp \left\{ -\frac{[\ln(x_i) - \mu_x]^2}{2\sigma_x^2} \right\} \cdot \\
&\quad \prod_{j=1}^m \frac{1}{y_j \sigma_y \sqrt{2\pi}} \exp \left\{ -\frac{[\ln(y_j) - \mu_y]^2}{2\sigma_y^2} \right\} \\
&= C(x, y) \sigma_x^{-n} \sigma_y^{-m} \cdot \\
&\quad \exp \left\{ -\frac{1}{2\sigma_x^2} \sum_{i=1}^n [\ln(x_i) - \mu_x]^2 \right. \\
&\quad \left. -\frac{1}{2\sigma_y^2} \sum_{j=1}^m [\ln(y_j) - \mu_y]^2 \right\}, \tag{2.7}
\end{aligned}$$

donde $C(x, y) = \left(\prod_{i=1}^n \frac{1}{x_i \sqrt{2\pi}} \right) \left(\prod_{j=1}^m \frac{1}{y_j \sqrt{2\pi}} \right)$ es una función que no depende del vector de parámetros $(\mu_x, \mu_y, \sigma_x, \sigma_y)$. Por otro lado, supóngase que $x_{obs} = (x_1, \dots, x_n)$ y $y_{obs} = (y_1, \dots, y_m)$ son muestras observadas de X y Y , respectivamente.

Función de verosimilitud de $(\mu_x, \mu_y, \sigma_x, \sigma_y)$

La función de verosimilitud del vector de parámetros $(\mu_x, \mu_y, \sigma_x, \sigma_y)$ es

$$\begin{aligned}
L(\mu_x, \mu_y, \sigma_x, \sigma_y; x_{obs}, y_{obs}) &= \sigma_x^{-n} \sigma_y^{-m} \exp \left\{ -\frac{1}{2\sigma_x^2} \sum_{i=1}^n [\ln(x_i) - \mu_x]^2 \right. \\
&\quad \left. -\frac{1}{2\sigma_y^2} \sum_{j=1}^m [\ln(y_j) - \mu_y]^2 \right\}, \tag{2.8}
\end{aligned}$$

Nótese que la verosimilitud (2.8) fue calculada con base en la función de densidad conjunta de X y Y dada en (2.7) y la aproximación continua a la verosimilitud dada en (2.3), usando como constante arbitraria a $c = 1/C(x_{obs}, y_{obs})$.

Función log-verosimilitud de $(\mu_x, \mu_y, \sigma_x, \sigma_y)$

La función log-verosimilitud del vector de parámetros $(\mu_x, \mu_y, \sigma_x, \sigma_y)$ se obtiene aplicando logaritmo natural a la verosimilitud dada en (2.8),

$$\begin{aligned} l(\mu_x, \mu_y, \sigma_x, \sigma_y; x_{obs}, y_{obs}) &= -n \ln(\sigma_x) - m \ln(\sigma_y) - \frac{1}{2\sigma_x^2} \sum_{i=1}^n [\ln(x_i) - \mu_x]^2 \\ &\quad - \frac{1}{2\sigma_y^2} \sum_{j=1}^m [\ln(y_j) - \mu_y]^2. \end{aligned} \quad (2.9)$$

Estimador de máxima verosimilitud de $(\mu_x, \mu_y, \sigma_x, \sigma_y)$

Los puntos críticos correspondientes a la función log-verosimilitud del vector de parámetros $(\mu_x, \mu_y, \sigma_x, \sigma_y)$, dada en (2.9), se calculan resolviendo el siguiente sistema de ecuaciones:

$$0 = \frac{\partial}{\partial \mu_x} l(\mu_x, \mu_y, \sigma_x, \sigma_y; x_{obs}, y_{obs}) = \frac{1}{\sigma_x^2} \sum_{i=1}^n [\ln(x_i) - \mu_x], \quad (2.10)$$

$$0 = \frac{\partial}{\partial \mu_y} l(\mu_x, \mu_y, \sigma_x, \sigma_y; x_{obs}, y_{obs}) = \frac{1}{\sigma_y^2} \sum_{j=1}^m [\ln(y_j) - \mu_y], \quad (2.11)$$

$$0 = \frac{\partial}{\partial \sigma_x} l(\mu_x, \mu_y, \sigma_x, \sigma_y; x_{obs}, y_{obs}) = -\frac{n}{\sigma_x} + \frac{1}{\sigma_x^3} \sum_{i=1}^n [\ln(x_i) - \mu_x]^2, \quad (2.12)$$

$$0 = \frac{\partial}{\partial \sigma_y} l(\mu_x, \mu_y, \sigma_x, \sigma_y; x_{obs}, y_{obs}) = -\frac{m}{\sigma_y} + \frac{1}{\sigma_y^3} \sum_{j=1}^m [\ln(y_j) - \mu_y]^2, \quad (2.13)$$

No es difícil ver que los valores de μ_x , μ_y , σ_x y σ_y que resuelven (2.10), (2.11), (2.12) y (2.13) son

$$\hat{\mu}_x = \frac{1}{n} \sum_{i=1}^n \ln(x_i), \quad (2.14)$$

$$\hat{\mu}_y = \frac{1}{m} \sum_{j=1}^m \ln(y_j), \quad (2.15)$$

$$\hat{\sigma}_x^2 = \frac{1}{n} \sum_{i=1}^n [\ln(x_i) - \hat{\mu}_x]^2, \quad (2.16)$$

$$\hat{\sigma}_y^2 = \frac{1}{m} \sum_{j=1}^m [\ln(y_j) - \hat{\mu}_y]^2, \quad (2.17)$$

respectivamente.

Ahora, como la matriz hessiana de $l(\mu_x, \mu_y, \sigma_x, \sigma_y; x_{obs}, y_{obs})$ en (2.9)

$$\begin{bmatrix} A & 0 & B & 0 \\ 0 & C & 0 & D \\ B & 0 & E & 0 \\ 0 & D & 0 & F \end{bmatrix},$$

donde

$$A = -\frac{n}{\sigma_x^2}, \quad (2.18)$$

$$B = -\frac{2}{\sigma_x^3} \sum_{i=1}^n [\ln(x_i) - \mu_x], \quad (2.19)$$

$$C = -\frac{m}{\sigma_y^2}, \quad (2.20)$$

$$D = -\frac{2}{\sigma_y^3} \sum_{j=1}^m [\ln(y_j) - \mu_y], \quad (2.21)$$

$$E = \frac{n}{\sigma_x^2} - \frac{3}{\sigma_x^4} \sum_{i=1}^n [\ln(x_i) - \mu_x]^2, \quad (2.22)$$

$$F = \frac{m}{\sigma_y^2} - \frac{3}{\sigma_y^4} \sum_{j=1}^m [\ln(y_j) - \mu_y]^2, \quad (2.23)$$

evaluada en $(\hat{\mu}_x, \hat{\mu}_y, \hat{\sigma}_x, \hat{\sigma}_y)$,

$$\begin{bmatrix} -\frac{n}{\hat{\sigma}_x^2} & 0 & 0 & 0 \\ 0 & -\frac{m}{\hat{\sigma}_y^2} & 0 & 0 \\ 0 & 0 & -\frac{2n}{\hat{\sigma}_x^2} & 0 \\ 0 & 0 & 0 & -\frac{2m}{\hat{\sigma}_y^2} \end{bmatrix},$$

es definida negativa, entonces $(\hat{\mu}_x, \hat{\mu}_y, \hat{\sigma}_x, \hat{\sigma}_y)$ es un máximo de la log-verosimilitud. Por lo tanto, el vector $(\hat{\mu}_x, \hat{\mu}_y, \hat{\sigma}_x, \hat{\sigma}_y)$ es el e.m.v. de $(\mu_x, \mu_y, \sigma_x, \sigma_y)$.

Reparametrización: $(\mu_x, \mu_y, \sigma_x, \sigma_y) \leftrightarrow (\delta, \rho, \mu_x, \sigma_y)$

Para los propósitos que se persiguen en esta tesis, se propone la siguiente reparametrización:

$$\begin{aligned} \delta &= \exp \left[(\mu_x - \mu_y) + \frac{1}{2} (\sigma_x^2 - \sigma_y^2) \right], \\ \rho &= \frac{\sigma_x}{\sigma_y}, \\ \mu_x &= \mu_x, \\ \sigma_y &= \sigma_y, \end{aligned}$$

donde δ es el cociente de medias de variables aleatorias con distribución lognormal.

Así, se tiene que

$$\begin{aligned} \mu_x &= \mu_x, \\ \sigma_x &= \sigma_x(\rho, \sigma_y) = \rho\sigma_y \\ \mu_y &= \mu_y(\delta, \rho, \mu_x, \sigma_y) = \mu_x - \ln(\delta) - \frac{1}{2}\sigma_y^2(1 - \rho^2), \\ \sigma_y &= \sigma_y. \end{aligned}$$

Propiedad de invarianza

Por la propiedad de invarianza del e.m.v definida en el Teorema 9, se tiene que los estimadores de máxima verosimilitud de δ y ρ son

$$\hat{\delta} = \exp \left[(\hat{\mu}_x - \hat{\mu}_y) + \frac{1}{2} (\hat{\sigma}_x^2 - \hat{\sigma}_y^2) \right]$$

y

$$\hat{\rho} = \frac{\hat{\sigma}_x}{\hat{\sigma}_y},$$

respectivamente. En (2.14), (2.15), (2.16) y (2.17) se encuentran $\hat{\mu}_x$, $\hat{\mu}_y$, $\hat{\sigma}_x$ y $\hat{\sigma}_y$, los estimadores de máxima verosimilitud de μ_x , μ_y , σ_x y σ_y , respectivamente.

Función de verosimilitud de $(\delta, \rho, \mu_x, \sigma_y)$

La función de verosimilitud del vector de parámetros $(\delta, \rho, \mu_x, \sigma_y)$ se obtiene reemplazando la reparametrización anterior en la función de verosimilitud de $(\mu_x, \mu_y, \sigma_x, \sigma_y)$ dada en (2.8),

$$\begin{aligned} L(\delta, \rho, \mu_x, \sigma_y; x_{obs}, y_{obs}) &= L[\mu_x, \mu_y = \mu_y(\delta, \rho, \mu_x, \sigma_y), \sigma_x = \sigma_x(\rho, \sigma_y), \sigma_y; x_{obs}, y_{obs}] \\ &= \rho^{-n} \sigma_y^{-(n+m)} \exp \left\{ -\frac{1}{2\rho^2 \sigma_y^2} \sum_{i=1}^n [\ln(x_i) - \mu_x]^2 \right. \\ &\quad \left. - \frac{1}{2\sigma_y^2} \sum_{j=1}^m \left[\ln(y_j) - \mu_x + \ln(\delta) + \frac{\sigma_y^2}{2} (1 - \rho^2) \right]^2 \right\}. \end{aligned} \quad (2.24)$$

Función log-verosimilitud de $(\delta, \rho, \mu_x, \sigma_y)$

La función log-verosimilitud del vector de parámetros $(\delta, \rho, \mu_x, \sigma_y)$ se obtiene aplicando logaritmo natural a la verosimilitud dada en (2.24),

$$\begin{aligned} l(\delta, \rho, \mu_x, \sigma_y; x_{obs}, y_{obs}) &= -n \ln(\rho) - (n+m) \ln(\sigma_y) - \frac{1}{2\rho^2 \sigma_y^2} \sum_{i=1}^n [\ln(x_i) - \mu_x]^2 \\ &\quad - \frac{1}{2\sigma_y^2} \sum_{j=1}^m \left[\ln(y_j) - \mu_x + \ln(\delta) + \frac{\sigma_y^2}{2} (1 - \rho^2) \right]^2. \end{aligned} \quad (2.25)$$

Estimador de máxima verosimilitud restringido de μ_x

Cuando los parámetros δ , ρ y σ están fijos, entonces la log-verosimilitud de $(\delta, \rho, \mu_x, \sigma_y)$ dada en (2.25) es una función de μ_x . En este caso, el punto crítico se obtiene resolviendo

la siguiente ecuación:

$$0 = \frac{\partial}{\partial \mu_x} l(\delta, \rho, \mu_x, \sigma_y; x_{obs}, y_{obs}), \quad (2.26)$$

donde

$$\begin{aligned} \frac{\partial}{\partial \mu_x} l(\delta, \rho, \mu_x, \sigma_y; x_{obs}, y_{obs}) &= \frac{1}{\rho^2 \sigma_y^2} \sum_{i=1}^n [\ln(x_i) - \mu_x] \\ &+ \frac{1}{\sigma_y^2} \sum_{j=1}^m \left[\ln(y_j) - \mu_x + \ln(\delta) + \frac{\sigma_y^2}{2} (1 - \rho^2) \right] \end{aligned} \quad (2.27)$$

Nótese que (2.27) se pueden escribir como

$$\begin{aligned} \frac{\partial}{\partial \mu_x} l(\delta, \rho, \mu_x, \sigma_y) &= \frac{1}{\rho^2 \sigma_y^2} (S_x - n\mu_x) \\ &+ \frac{1}{\sigma_y^2} \left[S_y - m\mu_x + m \ln(\delta) + \frac{m\sigma_y^2}{2} (1 - \rho^2) \right], \end{aligned} \quad (2.28)$$

donde

$$S_x = \sum_{i=1}^n \ln(x_i) \quad (2.29)$$

y

$$S_y = \sum_{j=1}^m \ln(y_j) \quad (2.30)$$

Así, de (2.26) y (2.28) se sigue que

$$\begin{aligned} 0 &= (S_x - n\mu_x) + \rho^2 \left[S_y - m\mu_x + m \ln(\delta) + \frac{m\sigma_y^2}{2} (1 - \rho^2) \right] \\ &= -\mu_x (n + m\rho^2) + S_x + \rho^2 S_y + m\rho^2 \ln(\delta) + \frac{m\sigma_y^2}{2} \rho^2 (1 - \rho^2). \end{aligned} \quad (2.31)$$

Entonces, de (2.31) se tiene que

$$\hat{\mu}_x(\delta, \rho, \sigma_y) = a + b\sigma_y^2, \quad (2.32)$$

donde

$$a = \frac{S_x + \rho^2 S_y + m\rho^2 \ln(\delta)}{n + m\rho^2} \quad (2.33)$$

y

$$b = \frac{m\rho^2(1-\rho^2)}{2(n+m\rho^2)}, \quad (2.34)$$

cumple que

$$\frac{\partial}{\partial \mu_x} l(\delta, \rho, \mu_x, \sigma_y; x_{obs}, y_{obs}) \Big|_{\mu_x = \hat{\mu}_x(\delta, \rho, \sigma_y)} = 0$$

Ahora, como

$$\frac{\partial^2}{\partial \mu_x^2} l(\delta, \rho, \mu_x, \sigma_y; x_{obs}, y_{obs}) \Big|_{\mu_x = \hat{\mu}_x(\delta, \rho, \sigma_y)} = -\frac{n}{\rho^2 \sigma_y^2} - \frac{m}{\sigma_y^2} < 0$$

para cada valor fijo de δ , ρ y σ , se sigue que $\hat{\mu}_x(\delta, \rho, \sigma_y)$ es el e.m.v. de μ_x restringido a δ , ρ y σ .

Estimador de máxima verosimilitud restringido de σ_y

Como $\hat{\mu}_x(\delta, \rho, \sigma_y)$ dado en (2.32) es el e.m.v. restringido de μ_x para cada valor fijo de (δ, ρ, σ_y) , entonces el e.m.v. restringido de σ_y para cada valor fijo de (δ, ρ) , $\hat{\sigma}_y(\delta, \rho)$, se obtiene resolviendo la ecuación

$$0 = \frac{\partial}{\partial \sigma_y} l(\delta, \rho, \sigma_y), \quad (2.35)$$

donde $l(\delta, \rho, \sigma_y) = l(\delta, \rho, \mu_x = \hat{\mu}_x(\delta, \rho, \sigma_y), \sigma_y; x_{obs}, y_{obs})$ es una función que solo depende de σ_y para cada valor fijo de (δ, ρ) ,

$$\begin{aligned} l(\delta, \rho, \sigma_y) &= -n \ln(\rho) - (n+m) \ln(\sigma_y) \\ &\quad - \frac{1}{2\rho^2 \sigma_y^2} \sum_{i=1}^n [\ln(x_i) - \hat{\mu}_x(\delta, \rho, \sigma_y)]^2 \\ &\quad - \frac{1}{2\sigma_y^2} \sum_{j=1}^m \left[\ln(y_j) - \hat{\mu}_x(\delta, \rho, \sigma_y) + \ln(\delta) + \frac{\sigma_y^2}{2} (1-\rho^2) \right]^2, \end{aligned} \quad (2.36)$$

y verificando que

$$\frac{\partial^2}{\partial \sigma_y^2} l(\delta, \rho, \sigma_y) \Big|_{\sigma_y = \hat{\sigma}_y(\delta, \rho)} < 0.$$

Nótese que

$$\begin{aligned}
l(\delta, \rho, \sigma_y) &= -n \ln(\rho) - (n+m) \ln(\sigma_y) \\
&\quad - \frac{1}{2\rho^2\sigma_y^2} \sum_{i=1}^n [\ln(x_i) - a - b\sigma_y^2]^2 \\
&\quad - \frac{1}{2\sigma_y^2} \sum_{j=1}^m \left[\ln(y_j) - a - b\sigma_y^2 + \ln(\delta) + \frac{\sigma_y^2}{2}(1-\rho^2) \right]^2 \\
&= -n \ln(\rho) - (n+m) \ln(\sigma_y) \\
&\quad - \frac{1}{2\rho^2\sigma_y^2} \sum_{i=1}^n [\ln(x_i) - a - b\sigma_y^2]^2 \\
&\quad - \frac{1}{2\sigma_y^2} \sum_{j=1}^m \left\{ \ln(y_j) - a + \ln(\delta) - \left[b - \frac{1}{2}(1-\rho^2) \right] \sigma_y^2 \right\}^2 \\
&= -n \ln(\rho) - (n+m) \ln(\sigma_y) \\
&\quad - \frac{1}{2\rho^2\sigma_y^2} b^2 \sum_{i=1}^n \left[\frac{\ln(x_i) - a}{b} - \sigma_y^2 \right]^2 \\
&\quad - \frac{1}{2\sigma_y^2} \left[b - \frac{1}{2}(1-\rho^2) \right]^2 \sum_{j=1}^m \left[\frac{\ln(y_j) - a + \ln(\delta)}{b - \frac{1}{2}(1-\rho^2)} - \sigma_y^2 \right]^2 \\
&= -n \ln(\rho) - (n+m) \ln(\sigma_y) \\
&\quad - \frac{1}{2\rho^2\sigma_y^2} b^2 \sum_{i=1}^n \left[\frac{\ln(x_i) - a}{b} - \sigma_y^2 \right]^2 \\
&\quad - \frac{1}{2\sigma_y^2} d^2 \sum_{j=1}^m \left[\frac{\ln(y_j) - a + \ln(\delta)}{d} - \sigma_y^2 \right]^2, \tag{2.37}
\end{aligned}$$

donde

$$d = b - \frac{1}{2}(1-\rho^2) = -\frac{n(1-\rho^2)}{2(n+m\rho^2)}, \tag{2.38}$$

véase expresión para b en (2.34). Así,

$$\begin{aligned}
l(\delta, \rho, \sigma_y) &= -n \ln(\rho) - (n+m) \ln(\sigma_y) \\
&\quad - \frac{1}{2\rho^2\sigma_y^2} b^2 \sum_{i=1}^n (A_i - \sigma_y^2)^2 \\
&\quad - \frac{1}{2\sigma_y^2} d^2 \sum_{j=1}^m (B_j - \sigma_y^2)^2, \tag{2.39}
\end{aligned}$$

donde

$$A_i = \frac{\ln(x_i) - a}{b} \quad (2.40)$$

y

$$B_j = \frac{\ln(y_j) - a + \ln(\delta)}{d}, \quad (2.41)$$

para $i = 1, \dots, n$ y $j = 1, \dots, m$. Entonces, de (2.39) se sigue que

$$\begin{aligned} \frac{\partial}{\partial \sigma_y} l(\delta, \rho, \sigma_y) &= -\frac{(n+m)}{\sigma_y} \\ &\quad - \frac{b^2}{2\rho^2} \frac{\partial}{\partial \sigma_y} \left[\frac{1}{\sigma_y^2} \sum_{i=1}^n (A_i - \sigma_y^2)^2 \right] \\ &\quad - \frac{d^2}{2} \frac{\partial}{\partial \sigma_y} \left[\frac{1}{\sigma_y^2} \sum_{j=1}^m (B_j - \sigma_y^2)^2 \right]. \end{aligned} \quad (2.42)$$

No es difícil ver que

$$\frac{\partial}{\partial \sigma_y} \left[\frac{1}{\sigma_y^2} \sum_{i=1}^n (A_i - \sigma_y^2)^2 \right] = \frac{2}{\sigma_y^3} \left(n\sigma_y^4 - \sum_{i=1}^n A_i^2 \right) \quad (2.43)$$

y

$$\frac{\partial}{\partial \sigma_y} \left[\frac{1}{\sigma_y^2} \sum_{j=1}^m (B_j - \sigma_y^2)^2 \right] = \frac{2}{\sigma_y^3} \left(m\sigma_y^4 - \sum_{j=1}^m B_j^2 \right) \quad (2.44)$$

Entonces, reemplazando (2.43) y (2.44) en (2.42) se tiene que

$$\begin{aligned} \frac{\partial}{\partial \sigma_y} l(\delta, \rho, \sigma_y) &= -\frac{(n+m)}{\sigma_y} \\ &\quad - \frac{b^2}{\rho^2 \sigma_y^3} \left(n\sigma_y^4 - \sum_{i=1}^n A_i^2 \right) \\ &\quad - \frac{d^2}{\sigma_y^3} \left(m\sigma_y^4 - \sum_{j=1}^m B_j^2 \right). \end{aligned} \quad (2.45)$$

Así, de (2.35) y (2.45) se sigue que

$$\begin{aligned} 0 &= -(n+m)\rho^2\sigma_y^2 - b^2 \left(n\sigma_y^4 - \sum_{i=1}^n A_i^2 \right) - \rho^2 d^2 \left(m\sigma_y^4 - \sum_{j=1}^m B_j^2 \right) \\ &= -(nb^2 + m\rho^2 d^2) \sigma_y^4 - [(n+m)\rho^2] \sigma_y^2 + \left(b^2 \sum_{i=1}^n A_i^2 + \rho^2 d^2 \sum_{j=1}^m B_j^2 \right). \end{aligned} \quad (2.46)$$

Ahora, de (2.40) y (2.41) se tiene que

$$b^2 \sum_{i=1}^n A_i^2 = b^2 \sum_{i=1}^n \left[\frac{\ln(x_i) - a}{b} \right]^2 = \sum_{i=1}^n [\ln(x_i) - a]^2$$

y

$$d^2 \sum_{j=1}^m B_j^2 = d^2 \sum_{j=1}^m \left[\frac{\ln(y_j) - a + \ln(\delta)}{d} \right]^2 = \sum_{j=1}^m [\ln(y_j) - a + \ln(\delta)]^2.$$

Entonces, la ecuación (2.46) se escribe como

$$0 = (nb^2 + m\rho^2 d^2) \sigma_y^4 + [(n+m)\rho^2] \sigma_y^2 - (M_x + \rho^2 M_y), \quad (2.47)$$

donde

$$M_x = \sum_{i=1}^n [\ln(x_i) - a]^2, \quad (2.48)$$

$$M_y = \sum_{j=1}^m [\ln(y_j) - a + \ln(\delta)]^2, \quad (2.49)$$

y a viene dado en la ecuación (2.33). Así, la ecuación (2.47) se puede reescribir como la siguiente ecuación cuadrática:

$$0 = w_1 Z^2 + w_2 Z + w_3, \quad (2.50)$$

donde

$$Z = \sigma_y^2,$$

$$w_1 = nb^2 + m\rho^2 d^2,$$

$$w_2 = (n+m)\rho^2,$$

$$w_3 = -(M_x + \rho^2 M_y),$$

con $w_1 \geq 0$, $w_2 > 0$ y $w_3 \leq 0$. Aquí se supondrá que $w_3 < 0$ puesto que para dos muestras observadas, $x_{obs} = (x_1, \dots, x_n)$ y $y_{obs} = (y_1, \dots, y_m)$, sería improbable obtener $M_x = M_y = 0$ [véase expresión de M_x y M_y en (2.48) y (2.49), respectivamente].

En el caso de $w_1 = 0$, la solución de (2.50) es

$$-\frac{w_3}{w_2} = \frac{M_x + \rho^2 M_y}{(n+m)\rho^2},$$

Sin embargo, nótese que $w_1 = 0$ si y sólo si $b = d = 0$. Ahora, de (2.34) y (2.38) se sigue que $b = d = 0$ si y sólo si $\rho = 1$. Por lo tanto, la solución de (2.50) para el caso $\rho = 1$ es

$$\hat{\sigma}_y^2(\delta, \rho = 1) = \frac{M_x^* + M_y^*}{n+m}, \quad (2.51)$$

donde

$$M_x^* = \sum_{i=1}^n [\ln(x_i) - a^*]^2, \quad (2.52)$$

$$M_y^* = \sum_{j=1}^m [\ln(y_j) - a^* + \ln(\delta)]^2, \quad (2.53)$$

$$a^* = \frac{S_x + S_y + m \ln(\delta)}{n+m}.$$

En el caso de $w_1 > 0$, las soluciones de (2.50) son

$$\begin{aligned} Z &= \frac{-w_2 \pm \sqrt{w_2^2 - 4w_1w_3}}{2w_1} \\ &= \frac{-(n+m)\rho^2 \pm \sqrt{(n+m)^2\rho^4 - 4(nb^2 + m\rho^2d^2)[-(M_x + \rho^2M_y)]}}{2(nb^2 + m\rho^2d^2)} \\ &= \frac{-(n+m)\rho^2 \pm (n+m)\rho^2 \sqrt{1 + 4\frac{(nb^2 + m\rho^2d^2)}{(n+m)^2\rho^4}(M_x + \rho^2M_y)}}{2(nb^2 + m\rho^2d^2)} \\ &= \frac{(n+m)\rho^2}{2(nb^2 + m\rho^2d^2)} \left[-1 \pm \sqrt{1 + 4\frac{(nb^2 + m\rho^2d^2)}{(n+m)^2\rho^4}(M_x + \rho^2M_y)} \right] \\ &= r \left\{ -1 \pm \sqrt{1 + \frac{2}{r} \left[\frac{M_x + \rho^2M_y}{(n+m)\rho^2} \right]} \right\}, \quad (2.54) \end{aligned}$$

donde

$$r = \frac{(n+m)\rho^2}{2(nb^2 + m\rho^2 d^2)} = 2 \left(\frac{1}{n} + \frac{1}{m} \right) \left[\frac{n+m\rho^2}{(1-\rho^2)^2} \right] \quad (2.55)$$

[véase expresión para b y d en (2.34) y (2.38), respectivamente], con $r > 0$. Ahora, nótese que la solución de (2.50) debe de ser mayor que cero, $Z = \sigma_y^2 > 0$. Además, obsérvese que $w_1 > 0$ para todo $\rho \neq 1$. Por lo tanto, la solución de (2.50) para el caso $\rho \neq 1$ es

$$\hat{\sigma}_y^2(\delta, \rho \neq 1) = r \left\{ -1 + \sqrt{1 + \frac{2}{r} \left[\frac{M_x + \rho^2 M_y}{(n+m)\rho^2} \right]} \right\} > 0,$$

donde las cantidades M_x , M_y y r vienen dadas en (2.48), (2.49) y (2.55), respectivamente.

Entonces,

$$\hat{\sigma}_y^2(\delta, \rho) = \begin{cases} \frac{M_x^* + M_y^*}{n+m}, & \rho = 1 \\ r \left\{ -1 + \sqrt{1 + \frac{2}{r} \left[\frac{M_x + \rho^2 M_y}{(n+m)\rho^2} \right]} \right\}, & \rho \neq 1 \end{cases}, \quad (2.56)$$

donde las cantidades M_x^* , M_y^* , M_x , M_y y r vienen dadas en (2.52), (2.53), (2.48), (2.49) y (2.55), cumple que

$$\left. \frac{\partial}{\partial \sigma_y} l(\delta, \rho, \sigma_y) \right|_{\sigma_y = \hat{\sigma}_y(\delta, \rho)} = 0.$$

Por otro lado, con algunos cálculos simples se obtiene que

$$\frac{\partial^2}{\partial \sigma_y^2} l(\delta, \rho, \sigma_y) = \frac{n+m}{\sigma_y^2} - \frac{3}{\sigma_y^4} \left(\frac{M_x + \rho^2 M_y}{\rho^2} \right) - \frac{(nb^2 + m\rho^2 d^2)}{\rho^2}.$$

Ahora, como

$$\left. \frac{\partial^2}{\partial \sigma_y^2} l(\delta, \rho, \sigma_y) \right|_{\sigma_y = \hat{\sigma}_y(\delta, \rho)} = \begin{cases} -\frac{2(n+m)}{\hat{\sigma}_y^2(\delta, \rho)} < 0, & \rho = 1 \\ -\frac{2(n+m)}{\hat{\sigma}_y^2(\delta, \rho)} - \frac{4(n+m)}{2r} < 0, & \rho \neq 1 \end{cases}, \quad (2.57)$$

para cada valor fijo de δ y ρ , se sigue que $\hat{\sigma}_y(\delta, \rho)$ es el e.m.v. de σ_y restringido a δ y ρ .

Función de verosimilitud perfil de (δ, ρ)

La función de verosimilitud perfil de (δ, ρ) es [véase la Definición 16]:

$$L_P(\delta, \rho; x_{obs}, y_{obs}) = \sup_{(\mu_x, \sigma_y) \in \mathbb{R} \times \mathbb{R}^+ | (\delta, \rho)} L(\delta, \rho, \mu_x, \sigma_y; x_{obs}, y_{obs}), \quad (2.58)$$

donde $L(\delta, \rho, \mu_x, \sigma_y; x_{obs}, y_{obs})$ es la función de verosimilitud de $(\delta, \rho, \mu_x, \sigma_y)$ dada en (2.24). Ahora, como $\hat{\mu}_x(\delta, \rho, \sigma_y)$ en (2.32) es el e.m.v. de μ_x restringido a δ, ρ y σ_y , se tiene que la función de verosimilitud perfil de (δ, ρ) dada en (2.58) se escribe como

$$L_P(\delta, \rho; x_{obs}, y_{obs}) = \sup_{\sigma_y \in \mathbb{R}^+ | (\delta, \rho)} L(\delta, \rho, \mu_x = \hat{\mu}_x(\delta, \rho, \sigma_y), \sigma_y; x_{obs}, y_{obs}). \quad (2.59)$$

Ahora, como $\hat{\sigma}_y(\delta, \rho)$ en (2.56) es el e.m.v. de σ_y restringido a δ y ρ , se tiene que la función de verosimilitud perfil de (δ, ρ) dada en (2.59) se escribe como

$$L_P(\delta, \rho; x_{obs}, y_{obs}) = L(\delta, \rho, \mu_x = \mu_x^*(\delta, \rho), \sigma_y = \hat{\sigma}_y(\delta, \rho); x_{obs}, y_{obs}), \quad (2.60)$$

donde $\mu_x^*(\delta, \rho) = \hat{\mu}_x(\delta, \rho, \sigma_y = \hat{\sigma}_y(\delta, \rho))$.

En esta sección no se presentan ni verosimilitudes relativas ni regiones de verosimilitud-confianza; sin embargo se pueden calcular a partir de las verosimilitudes dadas aquí. La función de verosimilitud relativa es una estandarización de la función de verosimilitud respecto a su valor máximo, que para las verosimilitudes presentadas en esta sección, se alcanza en su correspondiente estimador de máxima verosimilitud. Las regiones de verosimilitud-confianza se calculan con base en estas relativas [véase Definición 15 y Definición 20].

2.2. Métodos de valoración de modelos

Como ya se mencionó anteriormente, una parte muy importante dentro de un proceso de inferencia estadística es la valoración del modelo elegido para los datos. Existen varios métodos estadísticos para tal propósito. Aquí se describen algunos de ellos.

2.2.1. Histograma y densidad estimada

Un histograma es una representación gráfica de una función de probabilidad de densidad. Según Kotz y Johnson (1982, Vol. 5, pág. 3166) los histogramas presentan un conjunto de datos de manera visual facilitando la comprensión de algunas características generales como valores típicos, variación y forma. En ellos, es posible detectar comportamientos inesperados o valores inusuales en los datos. Además, se consideran una herramienta exploratoria previa al análisis estadístico y modelado puesto que la forma del histograma puede sugerir qué modelo de probabilidad es el más adecuado.

Dada una muestra $x_{obs} = (x_1, \dots, x_n)$ de valores reales, la construcción de un histograma consiste en dividir el eje horizontal \mathbb{R} en K intervalos, I_1, \dots, I_K disjuntos y determinar la cantidad de observaciones contenidas en cada uno. Nótese que si n_k es el número de datos que pertenecen al intervalo I_k , para $k = 1, \dots, K$, entonces $\sum_{k=1}^K n_k = n$.

Un inconveniente de éstos gráficos es la arbitrariedad con la que se determinan sus clases o intervalos. Si se consideran pocas, el histograma tendrá un aspecto demasiado burdo para ser informativo, y si se consideran muchas, entonces el histograma tendrá un aspecto demasiado rugoso como para observar características relevantes de la función de densidad de probabilidad que generó la muestra observada.

Definición 21 (Densidad estimada vía verosimilitud) *Sea X una variable aleatoria con función de densidad $f(x; \theta)$ y $x_{obs} = (x_1, \dots, x_n)$ una muestra observada de X . La densidad estimada de X vía verosimilitud se define como $\hat{f}(x) = f(x; \hat{\theta})$, donde $\hat{\theta}$ el e.m.v. de θ calculado con x_{obs} .*

El histograma cumple aproximarse (Ley Fuerte de los Grandes Números) a la densidad de la variable aleatoria X usada para modelar un fenómeno. Por otro lado, si la densidad de X es una función continua de θ y $\hat{\theta}$ converge casi seguramente a θ_0 entonces la densidad estimada o ajustada $\hat{f}(x) = f(x; \hat{\theta})$ converge casi seguramente a $f(x; \theta_0)$

cuando crece el tamaño de muestra. Pero, si para la estimación de $f(x; \theta_0)$ vía el enfoque de verosimilitud se consideró una función de densidad g diferente a $f(x; \theta)$ (error en la especificación del modelo), entonces no necesariamente ocurrirá que la densidad estimada se parezca a la densidad de X que es desconocida. Así, cuando la densidad estimada sea muy parecida a la forma del histograma entonces esto se considera un ajuste adecuado del modelo a la densidad que originó las observaciones puesto que el histograma se aproxima al modelo verdadero. Por otro lado, si la densidad ajustada no se parece a la forma del histograma se determinará que el modelo no es adecuado para los datos y en consecuencia no es adecuado para el fenómeno aleatorio de interés.

2.2.2. Gráfica cuantil-cuantil

La gráfica cuantil-cuantil es otra herramienta visual comúnmente utilizada para valorar el ajuste de un modelo propuesto para un juego de datos. Supóngase que X una variable aleatoria con función de distribución $F(x; \theta)$, donde θ es un vector de parámetros desconocido, y $x_{obs} = (x_1, \dots, x_n)$ una muestra observada de X . A continuación se presentan algunos conceptos fundamentales para la construcción de una gráfica cuantil-cuantil.

Definición 22 (Estadísticas de orden) *Un vector de estadísticas de orden correspondiente a una muestra observada $x_{obs} = (x_1, \dots, x_n)$ es un vector $[x_{(1)}, \dots, x_{(n)}]$, donde*

$$x_{(1)} = \min_{i=1, \dots, n} \{x_i\} \leq x_{(2)} \leq \dots \leq x_{(n)} = \max_{i=1, \dots, n} \{x_i\}$$

Definición 23 (Distribución empírica) *La distribución empírica es una función*

$F_n(x) : \mathbb{R} \rightarrow [0, 1]$ definida como

$$F_n(x) = \begin{cases} 0, & \text{para } x < x_{(1)}, \\ \frac{i}{n}, & \text{para } x_{(i)} \leq x < x_{(i+1)}, \\ 1, & \text{para } x_{(n)} \leq x, \end{cases}$$

donde $x_{(i)}$ es el i -ésimo elemento del vector de estadísticas de orden correspondiente a una muestra observada $x_{obs} = (x_1, \dots, x_n)$.

Definición 24 (Función cuantil) La función cuantil de una función de distribución F se define como la inversa generalizada de F ,

$$F^{\leftarrow}(t) = \inf \{x \in \mathbb{R} : F(x) \geq t\},$$

para $0 \leq t \leq 1$.

Si F es estrictamente creciente, continua y F^{-1} es su función inversa, entonces $F^{\leftarrow}(t) = F^{-1}(t)$, para todo para $0 < t < 1$.

Definición 25 (Cuantil) El cuantil de probabilidad acumulada p de la función de distribución F se define como $Q(p) = F^{\leftarrow}(p)$, donde F^{\leftarrow} es la función cuantil de F y $0 \leq p \leq 1$.

Definición 26 (Cuantiles empíricos) Los cuantiles empíricos son cuantiles de probabilidad acumulada i/n , $i = 1, \dots, n$, de la función de distribución empírica $F_n(x)$.

Nótese que las estadísticas de orden $x_{(1)}, \dots, x_{(n)}$ de una muestra observada $x_{obs} = (x_1, \dots, x_n)$ son los cuantiles de probabilidad acumulada i/n , $i = 1, \dots, n$, de la función de distribución empírica $F_n(x)$, respectivamente. Si por $Q_E(i)$ se denota al cuantil empírico de probabilidad acumulada i/n , entonces se tiene que $Q_E(i) = x_{(i)}$, para $i = 1, \dots, n$.

Para construir la gráfica cuantil-cuantil deben seguirse los siguientes pasos:

1. Calcular los cuantiles empíricos $Q_E(i) = x_{(i)}$, para $i = 1, \dots, n$.

2. Calcular los valores

$$p_i = \frac{i - 0.5}{n},$$

para $i = 1, \dots, n$

3. Determinar los cuantiles de orden p_i de la distribución teórica representada por la función de distribución $F(x; \theta)$,

$$Q(p_i) = F^{\leftarrow}(p_i; \theta),$$

para $i = 1, \dots, n$. De aquí en adelante se escribirá $Q(p_i) = Q(p_i; \theta)$ para enfatizar que se trata de una cantidad desconocida puesto que depende del parámetro desconocido θ .

4. Calcular

$$\hat{Q}_T(p_i) = Q(p_i; \hat{\theta}) = F^{\leftarrow}(p_i; \hat{\theta}),$$

donde $\hat{\theta}$ es un estimador puntual de θ ; por ejemplo, el e.m.v. de θ .

5. Representar en el plano cartesiano las parejas $[\hat{Q}_T(p_i), Q_E(i)]$, con $i = 1, \dots, n$, junto con la recta $y = x$.

Una vez realizados los pasos anteriores deberá observarse el comportamiento de los puntos alrededor de la recta de 45° . Si los puntos están muy cerca de la recta y serpenteando alrededor de ella, es evidencia de que el modelo no es rechazado por los datos. Por el contrario, cualquier patrón que de evidencia de que los datos se alejan de la recta $y = x$ se interpretará como un mal ajuste del modelo a los datos.

2.2.3. Uso de la verosimilitud perfil

Existen muchas situaciones en las que la familia de distribuciones elegida para modelar un fenómeno aleatorio contiene un parámetro que determina submodelos dentro de

esta familia. En este caso, dicho parámetro se puede considerar de interés y los otros parámetros en la familia de distribuciones como de estorbo. La verosimilitud perfil, descrita en la Sección 2.1, es una herramienta estadística que permite hacer inferencias sobre un parámetro de interés en presencia de otros de estorbo. Así, hacer inferencia sobre este parámetro vía la verosimilitud perfil (gráfica de los valores más plausibles del parámetro de interés e intervalos de verosimilitud) puede ayudar a reconocer o apreciar submodelos adecuados para los datos.

Por ejemplo, la familia de distribuciones de Box-Cox descrita en la Sección 1.3 del Capítulo I, involucra tres parámetros, λ , μ y σ , donde λ es un parámetro que determina submodelos de esta familia. Este es un caso donde λ es un parámetro de interés y (μ, σ) son considerados de estorbo. Así, una cuantificación de los valores plausibles de λ puede ayudar a identificar submodelos adecuados para los datos dentro de esta familia de modelos. De hecho, en esta tesis la plausibilidad del valor $\lambda = 0$ es de especial interés puesto $\lambda = 0$ corresponde a la distribución lognormal de parámetros μ y σ .

En este trabajo se propone utilizar la función de verosimilitud perfil de λ ,

$$L(\lambda; x_{obs}) = \begin{cases} [\hat{\sigma}^2(\lambda)]^{-\frac{n}{2}} (\prod_{i=1}^n x_i)^{\lambda-1} \exp\left(-\frac{n}{2}\right), & \text{si } \lambda \neq 0, \\ [\hat{\sigma}^2(\lambda)]^{-\frac{n}{2}} \exp\left(-\frac{n}{2}\right), & \text{si } \lambda = 0, \end{cases} \quad (2.61)$$

donde $\hat{\sigma}^2(\lambda) = (1/n) \sum_{i=1}^n [x_i^{(\lambda)} - \hat{\mu}(\lambda)]^2$ y $\hat{\mu}(\lambda) = (1/n) \sum_{i=1}^n x_i^{(\lambda)}$ son los estimadores de máxima verosimilitud restringidos de μ y σ para cada valor fijo de λ , para explorar gráficamente la preferencia de los datos $x_{obs} = (x_1, \dots, x_n)$ por algún submodelo de la familia de Box-Cox. Para detalles sobre el cálculo de la verosimilitud perfil de λ dada en (2.61), véase el Apéndice A. En particular, si el valor $\lambda = 0$ tiene plausibilidad alta entonces se puede considerar que el modelo lognormal es un modelo parsimonioso no rechazado por los datos. Para identificar el conjunto de valores plausibles de λ se pueden usar las regiones o intervalos de verosimilitud perfil de niveles de plausibilidad 0.25, 0.15 y 0.036 que tienen una confianza aproximada de 90 %, 95 % y 99 % respectivamente.

2.2.4. Prueba estadística de hipótesis

A continuación se definen algunos conceptos importantes de la estadística paramétrica, relacionados con el problema de pruebas de hipótesis.

Definición 27 (Hipótesis) *Una hipótesis es una suposición acerca de los valores del vector de parámetros de un modelo estadístico paramétrico.*

Como una hipótesis es una suposición fraseada en términos de los parámetros, matemáticamente una hipótesis es un subconjunto de valores del espacio paramétrico. Las hipótesis deben tener existencia previa, aun antes de observar datos. Los valores del parámetro formulados en las hipótesis deben de tener un significado especial que los distinga de los demás valores. Por ejemplo, $\lambda = 0$ en el modelo estadístico de Box y Cox tiene el significado especial de ser el modelo lognormal. De hecho, la toma de datos se concibe precisamente para confrontar una hipótesis preestablecida con la realidad.

Definición 28 (Hipótesis nula H_0) *Es la hipótesis que se desea probar y que se supone cierta.*

Una hipótesis nula se puede interpretar como el estado actual de la naturaleza, o la suposición válida de entrada hasta que los datos sugieran lo contrario.

Definición 29 (Hipótesis alternativa H_1) *Es la hipótesis a favor de la cual se rechaza H_0 .*

Una prueba de hipótesis consiste de examinar evidencia en forma de datos, para dar lugar a una de dos resoluciones posibles: Rechazar H_0 a favor de H_1 , o no rechazar H_0 . Bajo este planteamiento, hay dos tipos de error que se pueden cometer:

Definición 30 (Error de tipo I) *Es el error que se comete cuando se resuelve rechazar H_0 a favor de H_1 siendo que H_0 es cierta.*

Definición 31 (Error de tipo II) *Es el error que se comete cuando se resuelve no rechazar H_0 cuando H_0 es falsa.*

A continuación se define el concepto de prueba estadística de hipótesis.

Definición 32 (Prueba estadística de hipótesis) *Una prueba estadística de hipótesis es una regla que rechaza H_0 si y sólo si $T \in C$, donde T es una estadística y C es una región de posibles valores de T .*

Cuando $T \in C$ se dice que la prueba es significativa y cuando $T \notin C$ se dice que la prueba es no significativa. Nótese que una prueba estadística de hipótesis queda determinada por una estadística de prueba $T = T(X)$ y una región crítica de prueba C . La región crítica no depende de la muestra $X = (X_1, \dots, X_n)$, lo que quiere decir que aún antes de tomar la muestra, la región crítica tiene existencia propia. Los datos intervienen para tomar o no la resolución de rechazar H_0 , lo cual se realiza con la región crítica, al comparar el valor observado de T con el conjunto C . En resumen, una prueba de hipótesis se caracteriza por la pareja (T, C) .

Definición 33 (Prueba estadística de hipótesis de nivel α) *Una prueba estadística de hipótesis (T, C) es de nivel α , para $0 \leq \alpha \leq 1$, si*

$$\sup_{\theta \in \Theta_0} P(T \in C; \theta) \leq \alpha,$$

donde Θ_0 es el subconjunto de valores del espacio paramétrico correspondiente a la hipótesis nula H_0 .

Muchas veces se denota a la región crítica con el subíndice α para indicar que se trata de una prueba de nivel α , C_α .

Definición 34 (Estadística de prueba de la razón de verosimilitud) *La estadística de prueba de la razón de verosimilitud para probar $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_0^c$ se define como*

$$T = -2 \ln \left[\frac{\sup_{\theta \in \Theta_0} L(\theta; E_{obs})}{\sup_{\theta \in \Theta} L(\theta; E_{obs})} \right], \quad (2.62)$$

donde $L(\theta; E_{obs})$ es la función de verosimilitud de θ .

La estadística de prueba de la razón de verosimilitud se distribuye como una Ji-cuadrada con grados de libertad igual que la diferencia entre la cantidad de parámetros del modelo y la cantidad de parámetros del modelo bajo H_0 . Para detalles de la prueba véase Serfling (1980).

En esta tesis se propone realizar pruebas estadísticas de hipótesis de nivel α para hipótesis de la forma $H_0 : \lambda = \lambda_0$ versus $H_1 : \lambda \neq \lambda_0$, donde λ_0 es algún valor particular del parámetro λ de la familia de modelos de Box-Cox, correspondiente a algún submodelo típico dentro de esta familia. En particular, para probar $H_0 : \lambda = 0$ versus $H_1 : \lambda \neq 0$, se propone utilizar la estadística de prueba de la razón de verosimilitud,

$$T = -2 \ln \left[\frac{\sup L(\mu, \sigma; x_{obs})}{\sup L(\mu, \sigma, \lambda; x_{obs})} \right], \quad (2.63)$$

donde $L(\mu, \sigma; x_{obs})$ es la verosimilitud basada en el modelo lognormal y $L(\mu, \sigma, \lambda; x_{obs})$ es la verosimilitud basada en la familia de densidades de Box-Cox dada en la ecuación (A.1) del Apéndice A.

Nótese que $\lambda = 0$ es de especial interés puesto $\lambda = 0$ corresponde a la distribución lognormal de parámetros μ y σ . En este caso, la estadística de la razón de verosimilitud se distribuye (aproximadamente) como una Ji-cuadrada con un grado de libertad. Así, la región crítica de prueba es $C_\alpha = (q_{1,1-\alpha}, \infty)$, donde $q_{1,1-\alpha}$ es el cuantil $(1 - \alpha)$ de una distribución Ji-cuadrada con 1 grado de libertad.

La teoría estadística sobre la que se basa el discurso para pruebas de hipótesis, recibe la denominación de teoría de Neyman-Pearson. Nótese que bajo este enfoque, dada una muestra observada y un nivel α el resultado de la prueba es binario, en el sentido de concluir “rechazar H_0 ” o “no rechazar H_0 ”. Esto puede ser criticable; por ejemplo, si la región crítica de una prueba de hipótesis fuera el conjunto $C_{0.05} = (3.84, \infty)$ y datos en dos situaciones diferentes dieran lugar a valores de la estadística de prueba $D_1 = 3.9$ y $D_2 = 10.5$. La actitud de Neyman-Pearson diría simplemente, en ambos

casos, “rechazar H_0 ”, siendo que es intuitivamente claro que ambas situaciones son diferentes en alguna cualidad. En el segundo caso, se rechaza con mayor fuerza que en el primero, y al decir sólo “rechazar H_0 ” no se involucra esta fuerza de la evidencia en contra de H_0 . Un enfoque que considera la evidencia en contra de una hipótesis es descrito a continuación.

2.2.5. Pruebas de significancia: *p-valor*

El concepto de *p-valor* tiene por objeto cuantificar la fuerza con la que se rechaza una hipótesis nula H_0 . Se describe a través de una probabilidad. Tiene la interpretación de ser la probabilidad de haber observado un valor “más extremo” de una estadística de prueba $T = T(X)$ que ya se observó $t = t(x_{obs})$, o bien, la probabilidad de haber rechazado H_0 sólo por azar.

Definición 35 (*p-valor*) Sea $T(X)$ una estadística de prueba y x_{obs} un valor observado de X . El *p-valor* se define como

$$P_{x_{obs}} = \sup_{\theta \in \Theta_0} P_{\theta} [T(X) \geq t(x_{obs})],$$

De esta forma, un *p-valor* “muy chico” denota que los datos contienen mucha evidencia en contra de H_0 . En este sentido de *p-valores*, se puede no hablar de pruebas de hipótesis, sino de pruebas de significancia, donde la cuantificación del concepto abstracto de significancia es el *p-valor*.

En este trabajo de tesis para calcular el *p-valor* se considerará la estadística de prueba (2.63) que se distribuye (aproximadamente) como una Ji-cuadrada con un grado de libertad, χ_1^2 . Así, el *p-valor* se puede calcular a través de la siguiente expresión:

$$P_{x_{obs}} \approx P [\chi_1^2 > t(x_{obs})] = 1 - P [\chi_1^2 \leq t(x_{obs})], \quad (2.64)$$

donde $t(x_{obs})$ es el valor de la estadística (2.63) bajo los datos observados.

2.3. Selección de modelos

La selección de un modelo es, en ocasiones, más difícil que la estimación de los parámetros del modelo mismo y elegir el más apropiado requiere ciertas habilidades que generalmente se obtienen con la experiencia.

La naturaleza de los datos y la forma de su histograma, son aspectos importantes que pueden ayudar a discernir entre modelos, pero, en ocasiones hay modelos de probabilidad continuos bastante flexibles, en cuanto a sus posibles formas y es difícil seleccionar el más apropiado.

Considerar los resultados del proceso de valuación de modelos, descrito en la sección anterior, e incluir análisis adicionales que pueden depender del problema en estudio, contribuyen a la selección del modelo, pero no garantizan su unicidad. Un aspecto importante es el principio de parsimonia, que algunos conocen como Navaja de Occam, ("*Pluralitas non est ponenda sine neccesitate*"); esto es, no se debe incrementar, más allá de lo necesario, el número de entidades requeridas para explicar algo. En lo relativo a modelación, el incluir parámetros innecesarios en un modelo, generalmente provoca una pobre estimación de éstos. Algo que siempre debe recordarse es que ningún modelo es verdadero, pero algunos son más adecuados que otros para explicar la muestra observada.

2.4. Robusticidad

Cuando se realizan inferencias acerca de un parámetro o parámetros de interés, por lo general se plantean de manera implícita o explícita supuestos acerca del modelo y se supone que cambios pequeños en el modelo no afectan de manera significativa las conclusiones finales. Dicho razonamiento puede no ser correcto en todos los casos. Es decir, pueden existir situaciones reales (como las que serán mostradas más adelante) donde las inferencias sobre un parámetro de interés sean bastante sensibles frente a

cambios ligeros en los supuestos del modelo.

Un aspecto importante en un proceso de inferencia estadística es conocer hasta dónde pueden cambiar las inferencias si los supuestos no se satisfacen. Por ello, es de interés preguntarse, ¿cambian las conclusiones al efectuar cambios ligeros en el modelo? Esta pregunta ha dado lugar a un concepto que permite describir la sensibilidad de las inferencias ante variaciones en las suposiciones del modelo. Este concepto se conoce como robusticidad y cuando se cumple implica que las conclusiones no son afectadas por cambios leves en las suposiciones.

En el siguiente capítulo se propone hacer uso de la función de verosimilitud perfil para analizar la robusticidad de las inferencias de un parámetro de interés frente a cambios en los valores plausibles de un parámetro de estorbo. En particular, interesa hacer uso de la función de verosimilitud perfil para analizar la robusticidad de las inferencias sobre el cociente de medias de variables aleatorias independientes X y Y con distribución lognormal de parámetros (μ_X, σ_X) y (μ_Y, σ_Y) , $\delta = E(X)/E(Y)$, frente a cambios en los valores plausibles del parámetro de estorbo $\rho = \sigma_X/\sigma_Y$. Nótese que este problema es el tema central de esta tesis.

Capítulo 3

Robusticidad en las inferencias: Cociente de medias de lognormales

El objetivo principal del capítulo es enfatizar la importancia de un análisis de robusticidad en un proceso de inferencia estadística. Para este propósito, se considera el problema de hacer inferencia sobre el cociente de medias de variables aleatorias independientes X y Y con distribución lognormal de parámetros (μ_x, σ_x) y (μ_y, σ_y) , $\delta = E(X)/E(Y)$, con base en muestras observadas $x_{obs} = (x_1, \dots, x_n)$ y $y_{obs} = (y_1, \dots, y_m)$, respectivamente. Además, se supone que se tiene una hipótesis de interés $H_0 : \delta = \delta_0$.

3.1. Análisis de robusticidad

El concepto particular de robusticidad que se usará en este capítulo es el siguiente:

Si cambios en los valores plausibles de un parámetro de estorbo no cambian la resolución sobre una hipótesis H_0 correspondiente a un parámetro de interés, entonces las inferencias sobre el parámetro de interés son robustas.

En el análisis de robusticidad que se discute aquí, $\delta = E(X)/E(Y)$ es el parámetro de interés y $\rho = \sigma_x/\sigma_y$ será considerado el parámetro de estorbo.

Para determinar la robusticidad de las inferencias sobre el parámetro de interés δ , respecto a cambios en los valores plausibles del parámetro de estorbo ρ , se propone seguir el procedimiento que se describe a continuación:

1. Fijar la hipótesis de interés $H_0 : \delta = \delta_0$.
2. Fijar un nivel de plausibilidad $c_0 = \exp\left(-\frac{1}{2}q_{1,1-\alpha_0}\right)$, donde $q_{1,1-\alpha_0}$ es el cuantil $(1 - \alpha_0)$ de una distribución Ji-cuadrada con un grado de libertad, para una región de verosimilitud perfil del parámetro de interés δ .
3. Fijar un nivel de plausibilidad $c_1 = \exp\left(-\frac{1}{2}q_{1,1-\alpha_1}\right)$, donde $q_{1,1-\alpha_1}$ es el cuantil $(1 - \alpha_1)$ de una distribución Ji-cuadrada con un grado de libertad, para una región de verosimilitud perfil del parámetro de estorbo ρ .
4. Calcular el conjunto $V_\rho(c_1) = \{\rho : R_P(\rho; x_{obs}, y_{obs}) = c_1\}$, donde $R_P(\rho; x_{obs}, y_{obs})$ es la función de verosimilitud perfil relativa de ρ .
5. Calcular $R_P(\delta = \delta_0, \rho = \rho^*; x_{obs}, y_{obs})$, para todo $\rho^* \in V_\rho(c_1)$, donde $R_P(\delta, \rho; x_{obs}, y_{obs})$ es la función de verosimilitud perfil relativa de (δ, ρ) .
6. Calcular $R_P(\delta = \delta_0; x_{obs}, y_{obs})$, donde $R_P(\delta; x_{obs}, y_{obs})$ es la función de verosimilitud perfil relativa de δ .
7. Si el nivel de plausibilidad c_0 cumple que $A < c_0 < B$, donde

$$A = \min\{R_P(\delta = \delta_0; x_{obs}, y_{obs}), \min_{\rho^* \in V_\rho(c_1)} \{R_P(\delta = \delta_0, \rho = \rho^*; x_{obs}, y_{obs})\}\}$$

y

$$B = \max\{R_P(\delta = \delta_0; x_{obs}, y_{obs}), \max_{\rho^* \in V_\rho(c_1)} \{R_P(\delta = \delta_0, \rho = \rho^*; x_{obs}, y_{obs})\}\}$$

entonces se concluye que las inferencias para δ son *no robustas*. En caso contrario, se concluye que las inferencias para δ son *robustas* respecto a cambios en los valores plausibles (a un nivel c_1) del parámetro de estorbo ρ .

Nótese que cuando la inferencia es robusta, puede ocurrir que no se rechace $H_0 : \delta = \delta_0$ debido a que las plausibilidades $R_P(\delta = \delta_0; x_{obs}, y_{obs})$ y $R_P(\delta = \delta_0, \rho = \rho^*; x_{obs}, y_{obs})$, para todo $\rho^* \in V_\rho(c_1)$, son mayores que c_0 ; o que se rechace $H_0 : \delta = \delta_0$ debido a que estas plausibilidades son menores que c_0 . Por otro lado, inferencia no robusta significa que alguna de estas plausibilidades fue mayor que c_0 y alguna otra fue menor, y por lo tanto la resolución de rechazar o no rechazar la hipótesis $H_0 : \delta = \delta_0$ cambia cuando ocurren cambios en los valores plausibles de un parámetro de estorbo.

Para inspeccionar la robusticidad de las inferencias de manera simple, se recomienda graficar en una misma figura: $R_P(\delta; x_{obs}, y_{obs})$ y $R_P(\delta, \rho = \rho^*; x_{obs}, y_{obs})$, para todo $\rho^* \in V_\rho(c_1)$. Además, trazar en la gráfica la línea horizontal $c = c_0$ y la línea vertical $\delta = \delta_0$. Esto con el objetivo de observar no sólo el comportamiento de estas perfiles en $H_0 : \delta = \delta_0$ sino también en toda la región de valores plausibles de δ .

El procedimiento anterior requiere evaluar de forma computacional la función de verosimilitud perfil relativa del parámetro de estorbo ρ , $R_P(\rho; x_{obs}, y_{obs})$, la función de verosimilitud perfil relativa del parámetro de interés δ , $R_P(\delta; x_{obs}, y_{obs})$, y la función de verosimilitud perfil relativa de (δ, ρ) , $R_P(\delta, \rho; x_{obs}, y_{obs})$. A continuación se presenta una expresión computacional simple para cada una de estas funciones:

$$R_P(\rho; x_{obs}, y_{obs}) = \frac{\sup_{\delta \in \mathbb{R}+|\rho} L_P(\delta, \rho; x_{obs}, y_{obs})}{L_P(\delta = \hat{\delta}, \rho = \hat{\rho}; x_{obs}, y_{obs})}, \quad (3.1)$$

$$R_P(\delta; x_{obs}, y_{obs}) = \frac{\sup_{\rho \in \mathbb{R}+|\delta} L_P(\delta, \rho; x_{obs}, y_{obs})}{L_P(\delta = \hat{\delta}, \rho = \hat{\rho}; x_{obs}, y_{obs})}, \quad (3.2)$$

$$R_P(\delta, \rho; x_{obs}, y_{obs}) = \frac{L_P(\delta, \rho; x_{obs}, y_{obs})}{L_P(\delta = \hat{\delta}, \rho = \hat{\rho}; x_{obs}, y_{obs})}, \quad (3.3)$$

donde $L_P(\delta, \rho; x_{obs}, y_{obs})$ es la función de verosimilitud perfil de (δ, ρ) , dada en la ecuación (2.6), y $(\hat{\delta}, \hat{\rho})$ es el e.m.v. de (δ, ρ) ,

$$\hat{\delta} = \exp \left[(\hat{\mu}_x - \hat{\mu}_y) + \frac{1}{2} (\hat{\sigma}_x^2 - \hat{\sigma}_y^2) \right] \quad (3.4)$$

y

$$\hat{\rho} = \frac{\hat{\sigma}_x}{\hat{\sigma}_y}, \quad (3.5)$$

donde

$$\begin{aligned} \hat{\mu}_x &= \frac{1}{n} \sum_{i=1}^n \ln(x_i), \\ \hat{\sigma}_x^2 &= \frac{1}{n} \sum_{i=1}^n [\ln(x_i) - \hat{\mu}_x]^2, \\ \hat{\mu}_y &= \frac{1}{m} \sum_{j=1}^m \ln(y_j), \\ \hat{\sigma}_y^2 &= \frac{1}{m} \sum_{j=1}^m [\ln(y_j) - \hat{\mu}_y]^2, \end{aligned}$$

son los estimadores de máxima verosimilitud de μ_x , σ_x , μ_y y σ_y , respectivamente.

Es importante mencionar que cuando las inferencias son *robustas* entonces procesos de estimación y prueba de hipótesis se pueden realizar a través de la función de verosimilitud perfil del parámetro de interés δ . En caso contrario, se debe tener mucha cuidado con la validéz de las conclusiones derivadas de inferencias sobre δ . En esta situación, se recomienda valorar y solucionar las causas de la no robusticidad antes de realizar cualquier inferencia sobre δ .

A continuación se presentan dos casos de estudio con datos reales donde se ejemplifica el análisis de robusticidad descrito en esta sección. Uno de ellos corresponde a una situación de inferencias robustas y el otro no.

3.2. Caso robusto: Ácidos grasos poliinsaturados en especies de rayas.

Los ácidos grasos poliinsaturados (AGPI) se dividen en dos familias, los AGPI n-3 y los n-6. Las fuentes en las que pueden encontrarse estos ácidos grasos son los aceites de origen vegetal y los aceites de origen animal. Los AGPI son importantes en el ser humano pues facilitan la absorción de vitaminas liposolubles (A, D, E y K), regulan el metabolismo del colesterol e intervienen en la producción de eicosanoides; éstos regulan múltiples procesos celulares como tono vascular y bronquial, coagulación sanguínea, protección gástrica, temperatura corporal, fenómenos inflamatorios e inmunitarios, entre otros. Estudios realizados han mostrado que el consumo de aceite de pescado contribuye a reducir el riesgo de padecer cáncer de mama, diabetes, hipertensión arterial, entre otras. (Mesa, Aguilera y Gil, 2006; Nasiff y Meriño, 2003)

Debido a la alta demanda de aceites de pescado para el consumo humano o la acuicultura, ha sido necesario buscar nuevas fuentes de ellos. En México, en la última década, la captura de raya ha cobrado gran importancia; de ellas solo se aprovecha su carne, un 25%; las víceras y otros órganos, entre ellos el hígado, son desechados en el mar como desperdicio. Al parecer, el hígado podría ser utilizado como fuente alternativa de producción de aceite; desafortunadamente no existe suficiente información que contribuya a obtener un conocimiento más amplio sobre los beneficios y deficiencias de dicho aceite.

Se realizó un estudio con el objetivo general de caracterizar el aceite de hígado de raya de especies procedentes de las capturas comerciales del estado de Sinaloa. Los resultados de este trabajo mostraron que las especies de rayas estudiadas, *U. chillensis*, *U. halleri*, *U. nana*, *R. glaucostigma*, *R. steindachneri*, *N. vermiculatus* y *D. brevis*, tienen un alto contenido de aceite en el hígado, con una composición de ácidos grasos en los que se aprecia una considerable concentración de ácidos poliinsaturados, lo que

le confiere una alta calidad nutricional al producto. Para este estudio, se extrajo el hígado de cada individuo y se registró la cantidad de n-3 y n-6. La Tabla 3.1 muestra estos resultados para las especies *U. chillensis* y *U. halleri*.

U chillensis		U halleri	
6.88	7.87	5.79	5.26
5.88	6.98	7.83	6.00
7.10	6.00	5.00	5.89
7.73	6.52	6.43	5.47
6.37	9.40	6.44	8.00
10.35	6.03	6.99	5.26
7.65	5.71	6.95	8.95
7.27	6.20	5.34	7.17
7.36	5.56	7.66	7.00
5.53	7.93	5.02	9.03

Tabla 3.1: Cantidad de n-3 y n-6 en las especies *U. chillensis* y *U. halleri*.

El objetivo en esta sección es mostrar que las inferencias sobre el cociente de medias poblacionales de AGPI n-6 de las especies *U. chillensis* y *U. halleri* son *robustas* bajo el procedimiento propuesto en la sección anterior. Todo esto se realiza dentro del marco del proceso de inferencia estadística descrito en el Capítulo 2.

3.2.1. Modelación

El fenómeno aleatorio que se desea modelar es la cantidad de AGPI n-6 de dos especies diferentes de rayas, *U. chillensis* y *U. halleri*. Si se supone que X es la variable aleatoria que modela la cantidad de AGPI n-6 de la especie *U. chillensis* y Y es la variable aleatoria que modela la cantidad de AGPI n-6 de la especie *U. halleri*, entonces el problema es proponer una función de densidad de probabilidad conjunta para

(X, Y) . Como se trata de dos especies distintas, parece razonable suponer que existe independencia entre X y Y . Así, el problema de modelación consiste en proponer una función de densidad de probabilidad para X y otra para Y . Ahora, como se trata de dos especies de una misma familia (rayas) entonces se supondrá que la función de densidad de probabilidad de X y de Y pertenecen a un mismo modelo estadístico paramétrico o familia de densidades. Aquí se propone usar la familia de densidades de Box y Cox para modelar los AGPI n-6 de cada especie de raya.

Es importante mencionar que el modelo lognormal es el miembro más simple de la familia de densidades de Box y Cox en el sentido de que este modelo cuenta con sólo dos parámetros. Por tal razón, en la siguiente sección se explorará si el modelo lognormal es adecuado para modelar la cantidad de AGPI n-6 de cada especie de raya bajo estudio, *U. chillensis* y *U. halleri*.

3.2.2. Valoración del modelo lognormal

En el Capítulo 2 se presentó la familia de modelos de Box-Cox, la cual involucra tres parámetros. Además se mencionó que uno de esos parámetros, λ , tiene una interpretación lógica distinta a la que tienen los otros dos, μ y σ ; puesto que puede sugerir matemáticamente el modelo para una muestra dada. De hecho, el valor $\lambda = 0$ corresponde al modelo lognormal. Así, resultados de inferencias sobre el parámetro λ son fundamentales para justificar la elección del modelo lognormal dentro de la familia de Box y Cox, para un juego específico de datos. A continuación se usarán los métodos de valuación de modelos presentados en la Sección 2.2, con el objetivo de valorar el ajuste del modelo lognormal, miembro particular de la familia de densidades de Box y Cox, a los datos de AGPI n-6 dados en la Tabla 3.1.

Uso de la verosimilitud perfil

La Figura 3.1 se muestra la gráfica de la función de verosimilitud perfil de λ , dada en (2.61), junto con un intervalo de verosimilitud-confianza del 95%. Se observa que bajo el modelo de Box y Cox, los datos correspondientes a la especie *U. chillensis* y a la especie *U. halleri* asignan alta plausibilidad al valor $\lambda = 0$. De hecho, el intervalo de verosimilitud-confianza del 95% para λ (nivel de verosimilitud de 0.15) captura el valor de $\lambda = 0$ asociado al modelo lognormal.

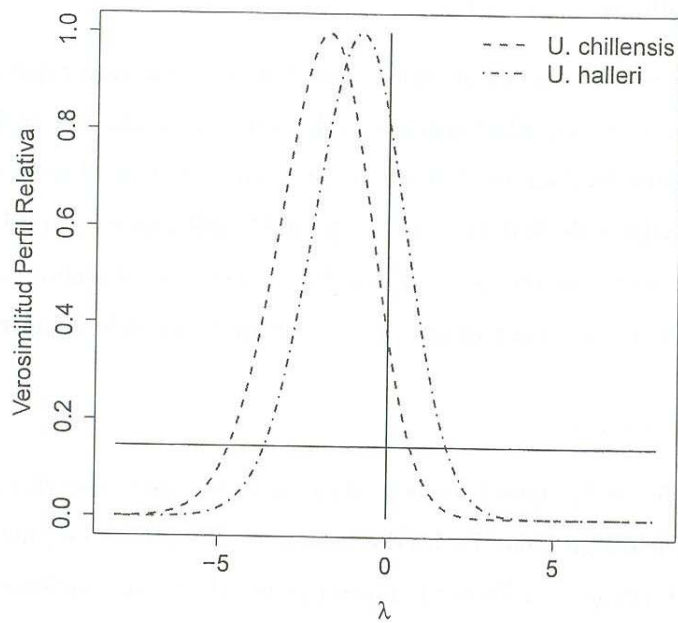


Figura 3.1: Verosimilitud perfil relativa de λ para los datos de AGPI n-6 dados en la Tabla 3.1.

Prueba estadística de hipótesis

Se realizó una prueba de hipótesis con las características descritas en la Sección 2.2.4. La hipótesis nula H_0 es que el modelo lognormal es adecuado. Para los datos de AGPI n-6 de la especie *U. chillensis*, el valor de la estadística de prueba dada en (2.63) es 1.972. Por otro lado, para los datos de AGPI n-6 de la especie *U. halleri*, el valor de la estadística de prueba es 0.394. En ambos casos, la región crítica o región de rechazo de nivel $\alpha = 0.05$ es $C_\alpha = (3.841, \infty)$. Así, en ningún caso se rechaza la hipótesis nula H_0 con una confianza del 95%. Es decir, en ningún caso los datos rechazan el modelo lognormal con una confianza del 95%.

Prueba de significancia: *p-valor*

Nótese que al no rechazar H_0 se espera que el *p-valor* sea mayor que $\alpha = 0.05$. Sin embargo, se procederá a calcularlo para fines de ejemplificar esta metodología. Usando la estadística de prueba dada en (2.63) y la expresión (2.64), se obtiene que el *p-valor* observado con los datos de AGPI n-6 de la especie *U. chillensis* y la especie *U. halleri* es 0.160 y 0.530, respectivamente. Así, como se había previsto, en ambos casos el *p-valor* observado no refleja evidencia contundente en contra del modelo lognormal.

Gráfica cuantil-cuantil

Las Figuras 3.2 y 3.3 muestran la gráfica cuantil-cuantil, descrita en la Sección 2.2.2, bajo el modelo lognormal. En las dos figuras se observa que los puntos serpentean alrededor de la recta $y = x$. Es decir, muestran un ajuste razonablemente bueno del modelo lognormal a los datos.

Histograma y densidad estimada

Las Figuras 3.4 y 3.5 muestran la densidad lognormal ajustada al histograma de frecuencias relativas de los datos de AGPI n-6 de la especie *U. chillensis* y de la especie

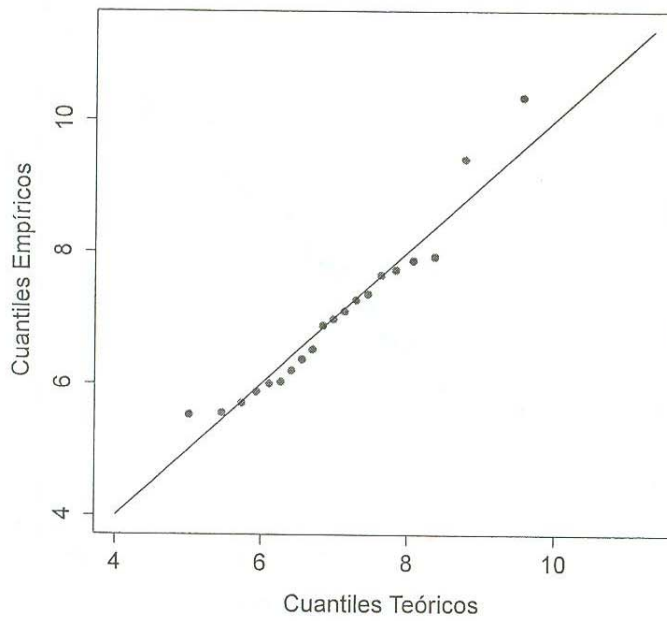


Figura 3.2: Gráfica cuantil-cuantil bajo el modelo lognormal: AGPI n-6 de la especie *U. chillensis*.

U. halleri, respectivamente. Los parámetros μ y σ de estos modelos fueron estimados mediante máxima verosimilitud,

- AGPI n-6 de la especie *U. chillensis*: $\hat{\mu} = 1.934$ y $\hat{\sigma} = 0.165$
- AGPI n-6 de la especie *U. halleri*: $\hat{\mu} = 1.866$ y $\hat{\sigma} = 0.182$.

En ambas figuras se observa que el modelo lognormal se ajusta razonablemente bien al histograma.

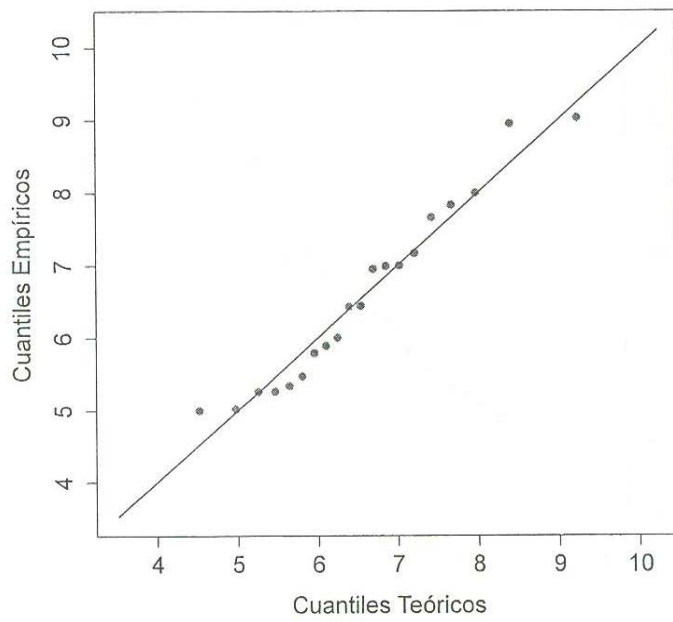


Figura 3.3: Gráfica cuantil-cuantil bajo el modelo lognormal: AGPI n-6 de la especie U. halleri.

3.2.3. Selección del modelo

Considerando los objetivos que se persiguen en esta tesis y la evidencia proporcionada por cada uno de los métodos de valuación de modelos aplicados aquí, se decide seleccionar el modelo lognormal de dos parámetros (por adecuado y parsimonioso) para modelar la cantidad de AGPI n-6 de cada especie de raya bajo estudio, U. chillensis y U. halleri.

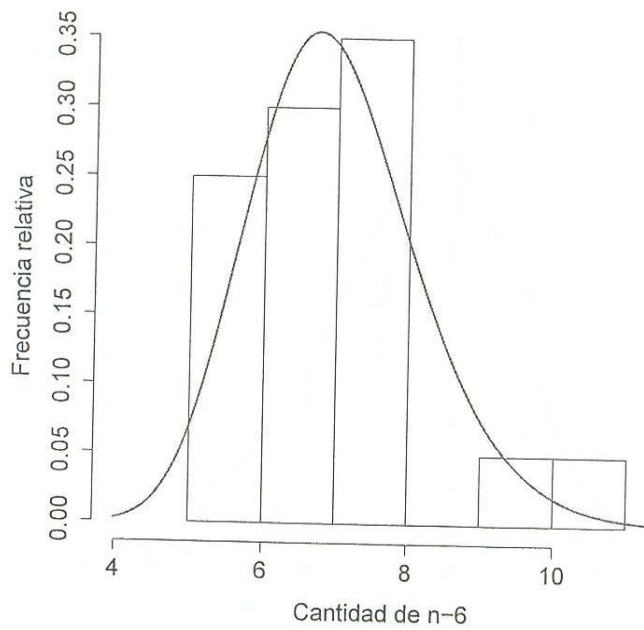


Figura 3.4: Histograma con la densidad lognormal ajustada: AGPI n-6 de la especie *U. chillensis*.

3.2.4. Análisis de robusticidad

En esta sección se realiza un análisis de robusticidad siguiendo el procedimiento descrito en la Sección 3.1. El parámetro de interés es el cociente de medias de variables aleatorias independientes X y Y con distribución lognormal de parámetros (μ_x, σ_x) y (μ_y, σ_y) , $\delta = E(X)/E(Y)$. Los datos que se usarán para este análisis son las muestras observadas de AGPI n-6 de la especie *U. chillensis* (x_{obs}) y de la especie *U. halleri* (y_{obs}); véase la Tabla 3.1.

La hipótesis de interés es $H_0 : \delta = \delta_0$, donde $\delta_0 = 1$. Se fijó un nivel de plausibilidad $c_0 = 0.15$, correspondiente a una región de verosimilitud-confianza perfil del 95 % para

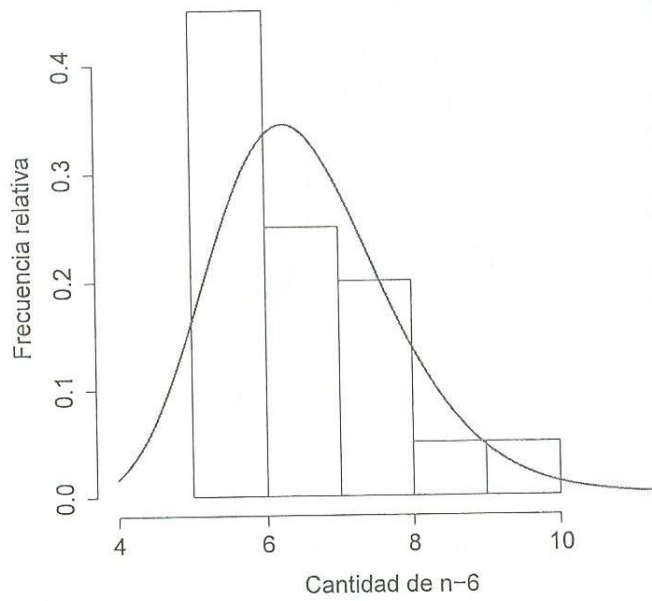


Figura 3.5: Histograma con la densidad lognormal ajustada: AGPI n-6 de la especie U. halleri.

el parámetro de interés δ . Además, se fijó un nivel de plausibilidad $c_1 = 0.15$, correspondiente a una región de verosimilitud-confianza perfil del 95 % para el parámetro de estorbo $\rho = \sigma_x/\sigma_y$. El conjunto de valores plausibles extremos de ρ , calculado con base en la la función de verosimilitud perfil relativa de ρ dada en (3.1), fue $V_\rho(c_1) = \{0.580, 1.414\}$, véase Figura 3.6. Los valores de $R_P(\delta = \delta_0, \rho = \rho^*; x_{obs}, y_{obs})$, donde $R_P(\delta, \rho; x_{obs}, y_{obs})$ es la función de verosimilitud perfil relativa de (δ, ρ) dada en (3.3) y $\rho^* \in V_\rho(c_1)$, fueron: $R_P(\delta = 1, \rho = 0.580; x_{obs}, y_{obs}) = 0.737$ y $R_P(\delta = 1, \rho = 1.414; x_{obs}, y_{obs}) = 0.411$. Por otro lado, el valor de $R_P(\delta = \delta_0; x_{obs}, y_{obs})$, donde $R_P(\delta; x_{obs}, y_{obs})$ es la función de

verosimilitud perfil relativa de δ dada en (3.2), fue $R_P(\delta = 1; x_{obs}, y_{obs}) = 0.513$. Así,

$$\begin{aligned} A &= \min\{R_P(\delta = \delta_0; x_{obs}, y_{obs}), \min_{\rho^* \in V_\rho(c_1)} \{R_P(\delta = \delta_0, \rho = \rho^*; x_{obs}, y_{obs})\}\} \\ &= 0.411 \end{aligned}$$

y

$$\begin{aligned} B &= \max\{R_P(\delta = \delta_0; x_{obs}, y_{obs}), \max_{\rho^* \in V_\rho(c_1)} \{R_P(\delta = \delta_0, \rho = \rho^*; x_{obs}, y_{obs})\}\} \\ &= 0.737. \end{aligned}$$

Ahora, como el nivel de plausibilidad $c_0 = 0.15$ no se encuentra entre A y B entonces se concluye que las inferencias para δ son *robustas* respecto a cambios en los valores plausibles (a un nivel $c_1 = 0.15$) del parámetro de estorbo ρ . La Figura 3.7 muestra la robusticidad de las inferencias de manera gráfica. Se observa que los valores $R_P(\delta = 1, \rho = 0.580; x_{obs}, y_{obs}) = 0.737$, $R_P(\delta = 1; x_{obs}, y_{obs}) = 0.513$ y $R_P(\delta = 1, \rho = 1.414; x_{obs}, y_{obs}) = 0.411$ se encuentran por arriba del valor $c_0 = 0.15$. Es decir, se trata de un caso de inferencias *robustas* donde los datos sugieren no rechazar la hipótesis $H_0 : \delta = 1$ ya que tiene alta plausibilidad.

A continuación se presentan inferencias sobre el parámetro de interés $\delta = E(X)/(Y)$ usando el enfoque de verosimilitud perfil.

3.2.5. Inferencias sobre el parámetro de interés: Cociente de medias lognormales

En la Figura 3.8 se muestra la función de verosimilitud perfil relativa del parámetro de interés δ , donde $\delta = E(X)/(Y)$ es el cociente de medias poblacionales de AGPI n-6 de las especies *U. chillensis* (X) y *U. halleri* (Y). Nótese que se trata de una verosimilitud casi simétrica con respecto a la ubicación del e.m.v. de δ , $\hat{\delta} = 1.067$, marcado con una línea vertical en la gráfica. En la Figura 3.8 también se señalan los extremos del intervalo de verosimilitud-confianza del 95 % para δ , $[0.954, 1.192]$. Nótese

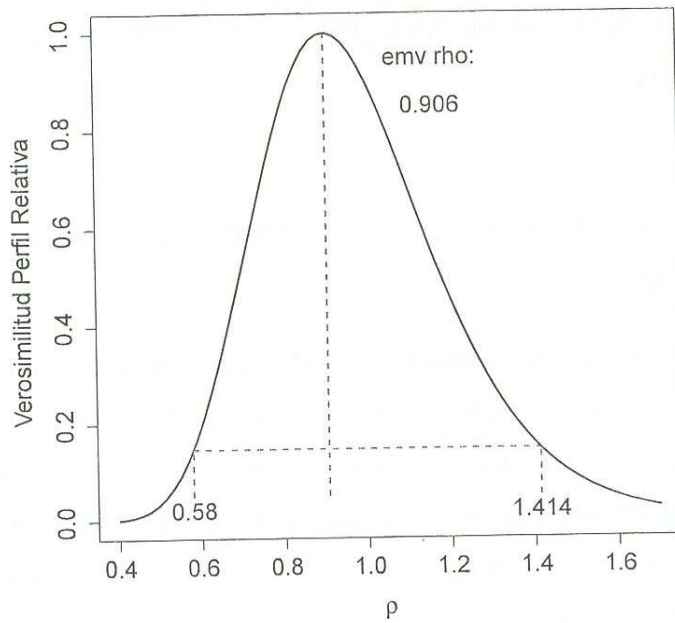


Figura 3.6: Verosimilitud perfil relativa de ρ : AGPI n-6 de la especie *U. chillensis* y *U. halleri*.

que el intervalo de confianza del 95% para δ captura la hipótesis $H_0 : \delta = 1$. Así, es posible concluir que no se tiene evidencia estadística para rechazar $H_0 : \delta = 1$ con una confianza del 95%.

3.2.6. Interpretación de los resultados

La evidencia estadística recolectada indica que no existe diferencia significativa entre la cantidad media de ácidos grasos poliinsaturados n-6 que se obtienen del hígado de la especie de raya *U. chillensis* y la cantidad media de ácidos grasos poliinsaturados n-6 que se obtienen del hígado de la especie de raya *U. halleri*.

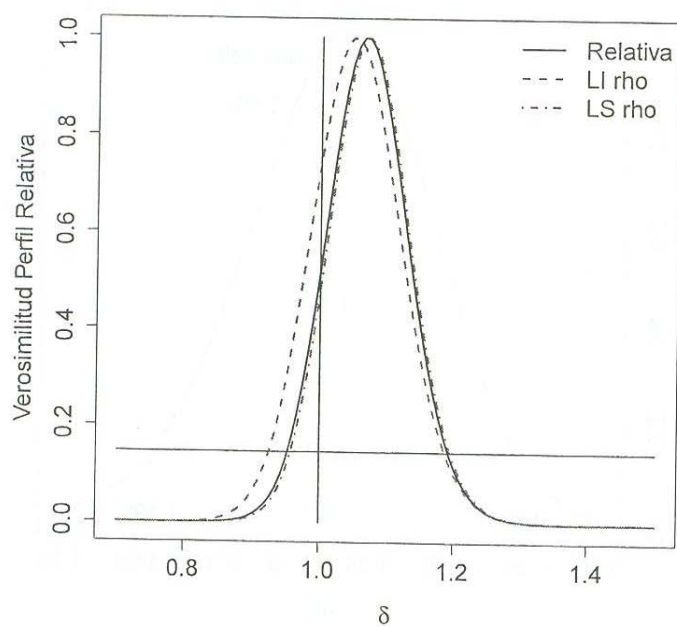


Figura 3.7: Inferencias *robustas* para δ respecto a cambios en los valores plausibles (a un nivel $c_1 = 0.15$) del parámetro de estorbo ρ : AGPI n-6 de la especie *U. chillensis* y *U. halleri*.

3.3. Caso no robusto: Biodisponibilidad de fármacos

La biodisponibilidad de un fármaco es un término farmacocinético que alude a la fracción de la dosis administrada del mismo, que alcanza su diana terapéutica o lo que es lo mismo que llega hasta el tejido sobre el que realiza su actividad. Esta cuantificación, necesaria para dar operatividad al concepto, es prácticamente imposible de hallar en el hombre. Por ello, se toma como valor aproximado la concentración plasmática de dicho

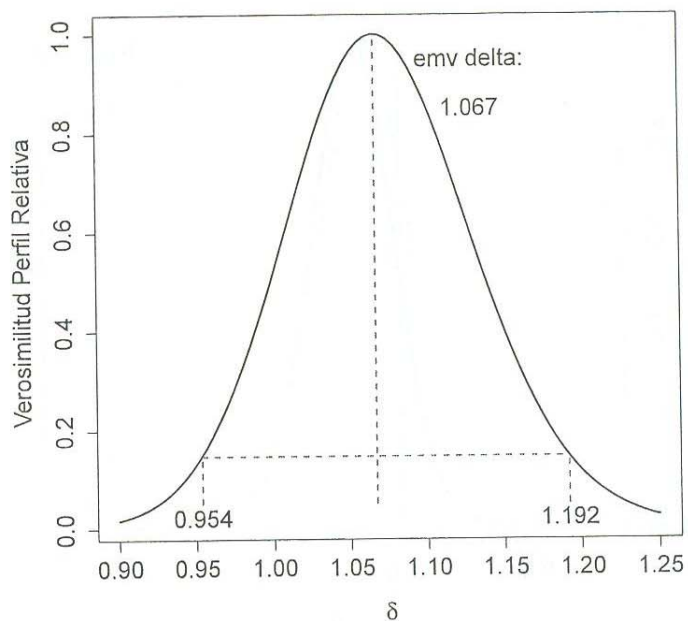


Figura 3.8: Verosimilitud perfil relativa de δ : AGPI n-6 de la especie *U. chillensis* y *U. halleri*.

fármaco y se compara con la concentración plasmática que alcanza para ese mismo preparado una inyección intravenosa.

En estudios para la evaluación de la biodisponibilidad se usan cantidades farmacocinéticas que se consideran importantes. Por ejemplo, la concentración máxima en plasma y el área bajo la curva de concentración sanguínea o plasmática versus tiempo. En estos estudios, la potencia relativa de un nuevo fármaco respecto a otro se expresa en términos del cociente de medias. Así, frecuentemente es necesario construir un intervalo de confianza para este parámetro o probar la hipótesis nula de que el cociente de medias igual a uno (no hay diferencia entre las medias de dos fármacos).

Wu, Jiang, Wong y Sun (2002) consideran el problema de comparar las medias de la concentración máxima en plasma ($C_{\text{máx}}$) de dos fármacos a través del cociente de medias. Ellos, en su Tabla 5, muestran 20 datos de $C_{\text{máx}}$ (10 para cada fármaco) y mencionan que fueron obtenidos mediante un experimento aleatorio, paralelo y controlado. Es decir, que cada paciente fue aleatoriamente asignado a un grupo y todos los pacientes en un grupo recibieron un fármaco. Wu *et al.* (2002) presentan gráficas cuantil-cuantil y pruebas de hipótesis para justificar el uso del modelo lognormal para los datos. Además, presentan inferencias sobre el cociente de medias de $C_{\text{máx}}$. En particular, prueban la hipótesis nula de que el cociente de medias igual a uno ($H_0 : \delta = 1$). Los resultados de Wu *et al.* (2002) conducen a no rechazar $H_0 : \delta = 1$ con una confianza del 95 %.

El objetivo en esta sección es mostrar que las inferencias sobre el cociente de medias de $C_{\text{máx}}$ son no *robustas* bajo el procedimiento propuesto en la Sección 3.1.

3.3.1. Análisis de robusticidad

El parámetro de interés es el cociente de medias de variables aleatorias independientes X y Y con distribución lognormal de parámetros (μ_x, σ_x) y (μ_y, σ_y) , $\delta = E(X)/E(Y)$. Los datos que se usarán para este análisis son las muestras observadas de $C_{\text{máx}}$ dadas en la Tabla 5 de Wu *et al.* (2002).

La hipótesis de interés es $H_0 : \delta = \delta_0$, donde $\delta_0 = 1$. Se fijó un nivel de plausibilidad $c_0 = 0.15$, correspondiente a una región de verosimilitud-confianza perfil del 95 % para el parámetro de interés δ . Además, se fijó un nivel de plausibilidad $c_1 = 0.15$, correspondiente a una región de verosimilitud-confianza perfil del 95 % para el parámetro de estorbo $\rho = \sigma_x/\sigma_y$. El conjunto de valores plausibles extremos de ρ , calculado con base en la la función de verosimilitud perfil relativa de ρ dada en (3.1), fue $V_\rho(c_1) = \{0.277, 0.997\}$, véase Figura 3.9. Los valores de $R_P(\delta = \delta_0, \rho = \rho^*; x_{\text{obs}}, y_{\text{obs}})$, donde $R_P(\delta, \rho; x_{\text{obs}}, y_{\text{obs}})$ es la función de verosimilitud perfil relativa de (δ, ρ) dada en (3.3) y $\rho^* \in V_\rho(c_1)$, fueron:

$R_P(\delta = 1, \rho = 0.277; x_{obs}, y_{obs}) = 0.111$ y $R_P(\delta = 1, \rho = 0.997; x_{obs}, y_{obs}) = 0.802$. Por otro lado, el valor de $R_P(\delta = \delta_0; x_{obs}, y_{obs})$, donde $R_P(\delta; x_{obs}, y_{obs})$ es la función de verosimilitud perfil relativa de δ dada en (3.2), fue $R_P(\delta = 1; x_{obs}, y_{obs}) = 0.385$. Así,

$$\begin{aligned} A &= \min\{R_P(\delta = \delta_0; x_{obs}, y_{obs}), \min_{\rho^* \in V_\rho(c_1)} \{R_P(\delta = \delta_0, \rho = \rho^*; x_{obs}, y_{obs})\}\} \\ &= 0.111 \end{aligned}$$

y

$$\begin{aligned} B &= \max\{R_P(\delta = \delta_0; x_{obs}, y_{obs}), \max_{\rho^* \in V_\rho(c_1)} \{R_P(\delta = \delta_0, \rho = \rho^*; x_{obs}, y_{obs})\}\} \\ &= 0.802. \end{aligned}$$

Ahora, como el nivel de plausibilidad $c_0 = 0.15$ se encuentra entre A y B entonces se concluye que las inferencias para δ son no *robustas* respecto a cambios en los valores plausibles (a un nivel $c_1 = 0.15$) del parámetro de estorbo ρ . La Figura 3.10 muestra la robusticidad de las inferencias de manera gráfica. Se observa que el valor $R_P(\delta = 1, \rho = 0.277; x_{obs}, y_{obs}) = 0.111$ se encuentra por abajo del valor $c_0 = 0.15$; mientras que los valores de $R_P(\delta = 1; x_{obs}, y_{obs}) = 0.385$ y $R_P(\delta = 1, \rho = 0.997; x_{obs}, y_{obs}) = 0.802$ se encuentran por arriba del valor $c_0 = 0.15$. Es decir, se trata de un caso de inferencias no *robustas*.

Este es un caso donde se debe tener mucha cuidado con la validéz de las conclusiones derivadas de inferencias sobre δ . Se recomienda valorar y solucionar las causas de la no robusticidad antes de realizar cualquier inferencia sobre δ .

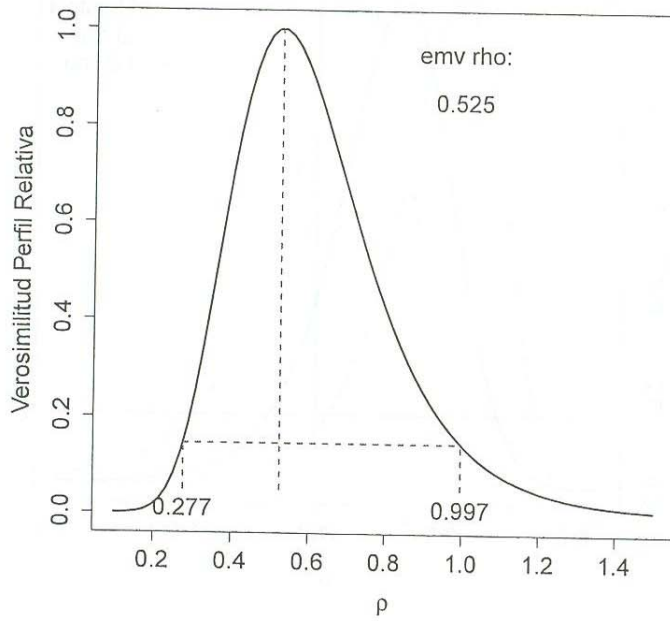


Figura 3.9: Verosimilitud perfil relativa de ρ : Datos de $C_{\text{m}\acute{\text{a}}\text{x}}$ dados en la Tabla 5 de Wu *et al.* (2002).

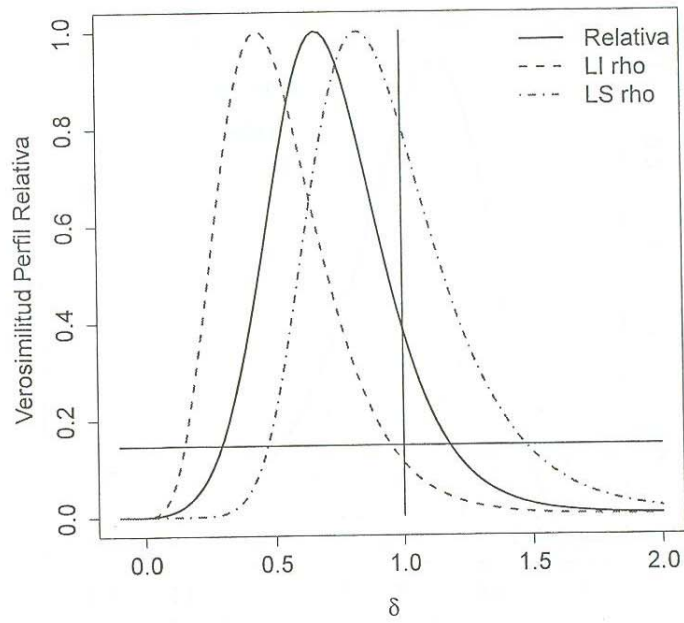


Figura 3.10: Inferencias no *robustas* para δ respecto a cambios en los valores plausibles (a un nivel $c_1 = 0.15$) del parámetro de estorbo ρ : Datos de C_{\max} dados en la Tabla 5 de Wu *et al.* (2002).

Capítulo 4

Conclusiones

En esta tesis se describen etapas fundamentales de un proceso de inferencia estadística: modelación, valoración de modelos, selección del modelo, análisis de robusticidad, inferencia sobre parámetros de interés en el contexto del problema e interpretación de los resultados en el contexto del problema. En este escenario, el tema central del trabajo es la importancia de un análisis de robusticidad en las inferencias.

En este trabajo se presenta un método novedoso para analizar, vía la función de verosimilitud perfil, la robusticidad de las inferencias sobre el cociente de medias de variables aleatorias independientes X y Y con distribución lognormal de parámetros (μ_X, σ_X) y (μ_Y, σ_Y) , $\delta = E(X)/E(Y)$, frente a cambios en los valores plausibles del parámetro de estorbo $\rho = \sigma_X/\sigma_Y$. Para mostrar la utilidad del método propuesto se consideraron dos casos de estudio con datos reales, uno robusto (comparación de medias de ácidos grasos poliinsaturados de dos especies de rayas) y otro no robusto (comparación de medias de biodisponibilidad de dos fármacos).

Con base en lo presentado en este trabajo, se concluye que es importante analizar la robusticidad de las inferencias de un parámetro de interés cuando se tiene presencia de parámetros de estorbo en el modelo usado para modelar el fenómeno aleatorio bajo estudio. Más aún cuando se sospecha que declaraciones cuantitativas acerca del parámetro

de interés pueden ser afectadas fuertemente por valores plausibles de parámetros de estorbo. Por ejemplo, es razonable sospechar que inferencias sobre la diferencia o el cociente de dos parámetros de localización pueden depender de parámetros de estorbo (en el modelo) que explican la dispersión del fenómeno aleatorio de interés. Nótese en estos casos la función de verosimilitud perfil, utiliza principalmete para hacer inferencias sobre un parámetro de interés en presencia de otros parámetros considerados de estorbo, puede ser empleada para analizar la robusticidad de las inferencias.

Apéndice A

Verosimilitud perfil de λ

Sea X una variable aleatoria con función de densidad $f(x; \mu, \sigma, \lambda)$ que pertenece a la familia de densidades de Box-Cox descrita en la Sección 1.3 del Capítulo I. Sea $x_{obs} = (x_1, \dots, x_n)$ una muestra observada de la variable aleatoria X . Entonces, la función de verosimilitud del vector de parámetros (μ, σ, λ) para el caso $\lambda \neq 0$ es (aproximación continua)

$$\begin{aligned} L(\mu, \sigma, \lambda; x_{obs}) &= c \prod_{i=1}^n \frac{x_i^{\lambda-1}}{\sqrt{2\pi}\sigma} \exp \left[-\frac{1}{2\sigma^2} (x_i^{(\lambda)} - \mu)^2 \right] \\ &= \sigma^{-n} \left(\prod_{i=1}^n x_i \right)^{\lambda-1} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i^{(\lambda)} - \mu)^2 \right], \end{aligned} \quad (\text{A.1})$$

donde por conveniencia se tomó $c = (2\pi)^{n/2}$. Por lo tanto, de (A.1) tiene que la función log-verosimilitud de (μ, σ, λ) es

$$\begin{aligned} l(\mu, \sigma, \lambda; x_{obs}) &= \ln [L(\mu, \sigma, \lambda; x_{obs})] \\ &= -n \ln(\sigma) + (\lambda - 1) \sum_{i=1}^n \ln(x_i) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i^{(\lambda)} - \mu)^2. \end{aligned} \quad (\text{A.2})$$

Cuando el parámetro λ está fijo, entonces la log-verosimilitud dada en (A.2) es una función de μ y σ . En este caso, los puntos críticos se obtienen resolviendo simultáneamente el siguiente sistema de ecuaciones:

$$0 = \frac{\partial}{\partial \mu} l(\mu, \sigma, \lambda; x_{obs}) = -\frac{1}{\sigma^2} \sum_{i=1}^n x_i^{(\lambda)} + \frac{1}{\sigma^2} n\mu, \quad (\text{A.3})$$

$$0 = \frac{\partial}{\partial \sigma} l(\mu, \sigma, \lambda; x_{obs}) = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (x_i^{(\lambda)} - \mu)^2. \quad (\text{A.4})$$

Ahora, nótese que

$$\hat{\mu}(\lambda) = \frac{1}{n} \sum_{i=1}^n x_i^{(\lambda)}. \quad (\text{A.5})$$

y

$$\hat{\sigma}^2(\lambda) = \frac{1}{n} \sum_{i=1}^n [x_i^{(\lambda)} - \hat{\mu}(\lambda)]^2 \quad (\text{A.6})$$

satisfacen (A.3) y (A.4). Además, cumplen que

$$\begin{aligned} \hat{l}_{11} &= \left. \frac{\partial^2}{\partial \mu^2} l(\mu, \sigma, \lambda; x_{obs}) \right|_{(\mu, \sigma)=[\hat{\mu}(\lambda), \hat{\sigma}(\lambda)]} = -\frac{n}{\hat{\sigma}^2(\lambda)}, \\ \hat{l}_{22} &= \left. \frac{\partial^2}{\partial \sigma^2} l(\mu, \sigma, \lambda; x_{obs}) \right|_{(\mu, \sigma)=[\hat{\mu}(\lambda), \hat{\sigma}(\lambda)]} = -\frac{2n}{\hat{\sigma}^2(\lambda)}, \\ \hat{l}_{12} &= \left. \frac{\partial^2}{\partial \mu \partial \sigma} l(\mu, \sigma, \lambda; x_{obs}) \right|_{(\mu, \sigma)=[\hat{\mu}(\lambda), \hat{\sigma}(\lambda)]} = 0, \end{aligned}$$

para cada valor fijo de λ . Así, $\hat{l}_{11} < 0$, $\hat{l}_{22} < 0$ y $\hat{l}_{11}\hat{l}_{22} - \hat{l}_{12}^2 > 0$. Por lo tanto, $\hat{\mu}(\lambda)$ y $\hat{\sigma}^2(\lambda)$, dados en (A.5) y (A.6), son los estimadores de máxima verosimilitud restringidos de μ y σ^2 para cada valor fijo de λ .

Sustituyendo (A.5) y (A.6) en (A.1) se tiene que la verosimilitud perfil de λ , para $\lambda \neq 0$ es

$$\begin{aligned} L(\lambda; x_{obs}) &= [\hat{\sigma}^2(\lambda)]^{-n/2} \left(\prod_{i=1}^n x_i \right)^{\lambda-1} \exp \left\{ -\frac{1}{2\hat{\sigma}^2(\lambda)} \sum_{i=1}^n [x_i^{(\lambda)} - \hat{\mu}(\lambda)]^2 \right\} \\ &= [\hat{\sigma}^2(\lambda)]^{-n/2} \left(\prod_{i=1}^n x_i \right)^{\lambda-1} \exp \left(-\frac{n}{2} \right). \end{aligned}$$

En el caso $\lambda = 0$, el valor de la verosimilitud perfil es

$$\begin{aligned} L(\lambda; x_{obs}) &= [\hat{\sigma}^2(\lambda)]^{-n/2} \exp \left\{ -\frac{1}{2\hat{\sigma}^2(\lambda)} \sum_{i=1}^n [\ln(x_i) - \hat{\mu}(\lambda)]^2 \right\} \\ &= [\hat{\sigma}^2(\lambda)]^{-n/2} \exp \left(-\frac{n}{2} \right), \end{aligned}$$

donde $\hat{\mu}(\lambda) = (1/n) \sum_{i=1}^n \ln(x_i)$ y $\hat{\sigma}^2(\lambda) = (1/n) \sum_{i=1}^n [\ln(x_i) - \hat{\mu}]^2$ son los e.m.v. restringidos de μ y σ que resultan del modelo lognormal.

Bibliografía

- BOX, G. E. P. y COX, D. R.: «An Analysis of Transformations». *Journal of the Royal Statistical Society. Series B (Methodological)*, 1964, **26**, pp. 211–252.
- CASELLA, G. y BERGER, R. L.: *Statistical Inference*. Duxbury, 2002.
- EVANS, M.; HASTINGS, N. y B. PEACOCK: *Statistical Distributions*. John Wiley, 2000.
- FISHER, R. A.: «On the “probable error” of a coefficient of correlation deduced from a small sample». *Metron*, 1921, **1**, pp. 3–32.
- GALTON, F.: «The Geometric Mean, in Vital and Social Statistics». *The Royal Society*, 1879, **29**, pp. 365–367.
- HUBER, P. J. y RONCHETTI, E.: *Robust Statistics*. John Wiley, 2009.
- KOTZ, S. y JOHNSON, N.L.: *Encyclopedia of Statistical Science*. John Wiley, 1982.
- LIMPERT, E.; STAHEL, W. A. y ABBT, M.: «Log-Normal Distributions across the Sciences: Keys and Clues». *BioScience, American Institute of Biological Sciences*, 2001, **51(5)**, pp. 341–352.
- MCALISTER, D.: «The Law of the Geometric Mean». *The Royal Society*, 1879, **29**, pp. 367–376.

- MESA, M. D.; AGUILERA, C. M. y GIL, A.: «Importancia de los lípidos en el tratamiento nutricional de las patologías de base inflamatoria». *Nutrición Hospitalaria*, 2006, 21, pp. 30–43.
- MONTOYA, J.: *La verosimilitud perfil en la Inferencia Estadística*. Tesis doctoral, Centro de Investigación en Matemáticas A. C., 2008.
- NASIFF, A. y MERIÑO, E.: «Ácidos grasos omega-3: pescados de carne azul y concentrados de aceites de pescado. Lo bueno y lo malo». *Revista Cubana de Medicina*, 2003, 42(2), p. 7.
- SERFLING, R.: *Approximation Theorems of Mathematical Statistics*. John Wiley, 1980.
- SPROTT, D.A.: *Statistical Inference in Science*. Springer, 2000.
- R Development Core Team: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2011. ISBN 3-900051-07-0. <http://www.R-project.org/>
- WU, J.; JIANG, G.; WONG, A. C. M. y SUN, X.: «Likelihood Analysis for the Ratio of Means of Two Independent Log-Normal Distributions». *The College Mathematics Journal*, 2002, 58, pp. 463–469.

