



Centro de Investigación en Matemáticas, A.C.

---

# La verosimilitud perfil en la Inferencia Estadística

T E S I S

que para obtener el grado de  
Doctor en ciencias con orientación en  
Probabilidad y Estadística

**Presenta:**

**José Arturo Montoya Laos**

Directores de Tesis:

Dra. Eloísa Díaz-Francés Murguía.

Dr. David A. Sprott.

*Guanajuato, Gto. México. Junio de 2008*

Acta No.: 003

Libro No.: 002

Foja No.: 003

**CENTRO DE INVESTIGACIÓN EN MATEMÁTICAS**

En la ciudad de Guanajuato, Gto., siendo las 10:00 horas del día 9 de junio del año 2008, se reunieron los miembros del Jurado Dictaminador integrado por los señores:

<b>DR. LUIS ALBERTO ESCOBAR RESTREPO</b>	(CIMAT-LOUISIANA STATE UNIVERSITY)
<b>DR. ENRIQUE RAÚL VILLA DIHARCE</b>	(CIMAT)
<b>DR. ROMÁN VIVEROS AGUILERA</b>	(MC MASTER UNIVERSITY)
<b>DR. JOSÉ MIGUEL PONCIANO CASTELLANOS</b>	(CIMAT)
<b>DRA. ELOÍSA DÍAZ-FRANCÉS MURGUÍA</b>	(CIMAT)

bajo la presidencia del primero y con carácter de secretario el segundo, para proceder a efectuar el examen que para obtener el Grado de:

**DOCTOR EN CIENCIAS  
CON ORIENTACIÓN EN PROBABILIDAD Y ESTADÍSTICA**

sustenta:

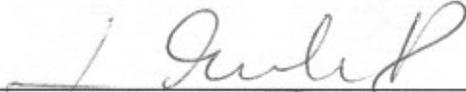
**JOSÉ ARTURO MONTOYA LAOS**

en cumplimiento con lo establecido en los reglamentos y lineamientos de estudios de posgrado del Centro de Investigación en Matemáticas, A.C., mediante la presentación del trabajo de tesis:

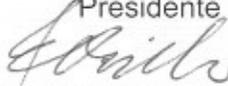
**"LA VEROSIMILITUD PERFIL EN LA INFERENCIA ESTADÍSTICA".**

Los miembros del Jurado examinaron alternadamente al (la) sustentante y después de deliberar entre sí resolvieron declararlo (a):

APROBADO

  
\_\_\_\_\_  
**DR. LUIS ALBERTO ESCOBAR RESTREPO**

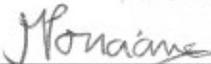
Presidente

  
\_\_\_\_\_  
**DR. ENRIQUE RAÚL VILLA DIHARCE**

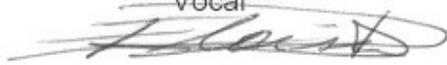
Secretario

  
\_\_\_\_\_  
**DR. ROMÁN VIVEROS AGUILERA**

Vocal

  
\_\_\_\_\_  
**DR. JOSÉ MIGUEL PONCIANO CASTELLANOS**

Vocal

  
\_\_\_\_\_  
**DRA. ELOÍSA DÍAZ-FRANCÉS MURGUÍA**

Vocal



# Agradecimientos

Por el apoyo económico que recibí para realizar mi tesis de licenciatura, para llevar a cabo mis estudios de maestría, para hacer mi tesis de maestría, para efectuar mis estudios de doctorado y para elaborar esta tesis, expreso mi eterno agradecimiento al Centro de Investigación en Matemáticas, A. C. (CIMAT). También agradezco el apoyo económico que recibí para culminar mis estudios de doctorado al Consejo Nacional de Ciencia y Tecnología (CONACYT). De igual forma agradezco al Dr. Luis Gorostiza Ortega, quien me apoyó económicamente con una beca de ayudantía de investigador en su calidad de investigador de Nivel III del Sistema Nacional de Investigadores (SNI).

Por fungir como co-directores de esta tesis, por su guía y apoyo constante, expreso mi agradecimiento a la Dra. Eloísa Díaz Francés Murguía y al Dr. David A. Sprott.

A mi esposa Mery Helen Pesantes Espinoza agradezco haberme comprendido y apoyado de manera constante e incondicional durante mi paso por las aulas del CIMAT.

A mi padre Guillermo Arturo Montoya Díaz, a mi madre Victoria Laos Conde, a mi hermana Cariana Montoya Laos, a mi hermana Lourdes Stephanny Montoya Laos y a mi sobrina Nicole agradezco la fuerza que me dieron para seguir adelante y lograr culminar mi doctorado.

A mis hermanos mexicanos Juan Manuel Mondragón Archundria, Noriki Ibarra Medina, Julio Cesar Arredondo Razo, José del Carmen Jiménez, Ivan Jiménez Robledo, Julian Sánchez y Agustín Rangel agradezco su sinceridad, lealtad y apoyo absoluto.

A mis maestros y amigos Eloísa Díaz-Francés, David A. Sprott, Miguel Nakamura Savoy, Joaquín Ortega Sánchez, Daniel Hernández Hernández, Graciela González Farías, Rogelio Ramos Quiroga, Andrés Christen Gracia, Víctor Manuel Rivero y Víctor Pérez Abreu agradezco sus enseñanzas y buenos consejos durante mi estancia en el CIMAT.

A mis colegas del doctorado Addy Bolivar, Víctor López y Antonio Murillo agradezco su compañerismo.

# Contenido

<b>Prefacio</b>	<b>5</b>
<b>1 Conceptos estadísticos básicos</b>	<b>9</b>
1.1 Función de verosimilitud . . . . .	9
1.2 Función de verosimilitud relativa . . . . .	10
1.3 Propiedades de la verosimilitud . . . . .	11
1.3.1 Las verosimilitudes no son aditivas . . . . .	11
1.3.2 Combinación de experimentos . . . . .	11
1.3.3 Invarianza funcional . . . . .	12
1.4 Aproximación continua de la función de verosimilitud y la verosimilitud exacta . . . . .	13
1.5 Intervalos de verosimilitud . . . . .	15
1.6 Intervalos de verosimilitud-confianza . . . . .	19
1.7 Estimación por separado de parámetros de interés en presencia de parámetros de estorbo. . . . .	21
1.7.1 Función de verosimilitud condicional . . . . .	22
1.7.2 Función de verosimilitud marginal . . . . .	23
1.7.3 Función de verosimilitud integrada . . . . .	24
1.7.4 Densidad marginal posterior Bayesiana . . . . .	24
1.7.5 Función de verosimilitud estimada . . . . .	25

1.7.6	Función de verosimilitud maximizada o perfil . . . . .	26
<b>2</b>	<b>Críticas a la verosimilitud perfil</b>	<b>34</b>
2.1	Críticas injustificadas a la verosimilitud perfil* . . . . .	34
2.2	Críticas justificadas a la verosimilitud perfil . . . . .	42
2.3	Discusión . . . . .	44
<b>3</b>	<b>Robusticidad de la función de verosimilitud perfil</b>	<b>46</b>
3.1	El concepto de verosimilitud perfil $\epsilon$ -perturbada . . . . .	47
3.2	Ejemplos robustos . . . . .	53
3.2.1	Modelo Binomial Negativa: Datos de conteos de garrapatas en ovejas . . . . .	54
3.2.2	Modelo Weibull para máximos: Datos de niveles máximos an- uales del mar en Port Pirie, Australia . . . . .	56
3.2.3	Modelo Binomial $(N, p)$ : Casos robustos con muestras simuladas	60
3.3	Ejemplos no robustos: Modelo Binomial $(N, p)$ . . . . .	65
3.3.1	Antecedentes . . . . .	66
3.3.2	Identificabilidad del modelo Binomial $(N, p)$ como causa princi- pal de los problemas relacionados con la estimación de $N$ cuando $p$ es desconocida . . . . .	68
3.3.3	Forma de la verosimilitud perfil $N$ . . . . .	71
3.3.4	Ejemplos de estimación de abundancia de animales con el modelo Binomial $(N, p)$ . . . . .	77
3.4	El problema ecológico de estimación de abundancia de animales usando un modelo Binomial $(N, p)$ . . . . .	83
3.4.1	Ejemplo de Impalas . . . . .	87
3.5	Discusión . . . . .	89

<b>4</b>	<b>Uso de la verosimilitud perfil para simplificar la inferencia de un parámetro de interés</b>	<b>114</b>
4.1	Introducción . . . . .	114
4.2	Caso Exponencial . . . . .	117
4.2.1	Enfoques comúnmente usados para inferencias sobre el parámetro $\theta = P(X < Y)$ . . . . .	120
4.2.2	Inferencia sobre $\theta$ a través de la verosimilitud perfil . . . . .	124
4.2.3	Ejemplo: Datos de tiempos de vida de un fluido aislante . . . . .	129
4.2.4	Simulaciones . . . . .	134
4.3	Caso Normal . . . . .	135
4.3.1	Enfoques usuales para inferencias sobre $\theta$ . . . . .	136
4.3.2	Uso de la verosimilitud perfil para inferencias sobre $\theta$ . . . . .	139
4.3.3	Ejemplos . . . . .	141
4.4	Conclusiones generales . . . . .	150
	<b>Bibliografía</b>	<b>159</b>

# Prefacio

La función de verosimilitud maximizada o perfil es un método estadístico muy simple y poderoso comúnmente usado para estimar por separado un parámetro de interés en presencia de parámetros de estorbo. Para este propósito, la función de verosimilitud maximizada o perfil es mucho más general que otras verosimilitudes tales como la condicional, la marginal o la integrada, las cuales dependen de una estructura especial del modelo de probabilidad que describe mejor al fenómeno aleatorio, y por lo tanto son más restrictivas.

En esta tesis, la verosimilitud maximizada o perfil es el tema central. Se revisarán básicamente tres aspectos de ella. Primero, se analizarán algunas críticas que ha recibido la verosimilitud perfil y se esclarecerá cuáles de ellas son justificadas y cuáles no. Otro punto importante a tratar en esta tesis es la robusticidad de la función de verosimilitud perfil frente a pequeños cambios en la estimación de los parámetros de estorbo. Para esto en este trabajo se definirá una nueva función que será llamada *función de verosimilitud perfil  $\epsilon$ -perturbada*. El último aspecto que se abordará en este trabajo es mostrar cómo la verosimilitud perfil puede ayudar a simplificar enormemente las inferencias sobre parámetros de interés en diversos modelos propuestos. Cuando sea posible, es preferible dar las inferencias sobre un parámetro de interés de la manera algebraicamente más simple y clara. A manera de ejemplo se presenta el uso de la función de verosimilitud perfil para hacer inferencias por separado sobre el parámetro de confiabilidad en los modelos tensión-fuerza. La verosimilitud perfil en este contexto

nunca antes se había utilizado.

En el Capítulo 1 se revisan algunos conceptos básicos de la teoría de inferencia estadística paramétrica que serán utilizados a lo largo de esta tesis.

En el Capítulo 2 se evalúan algunas críticas que ha recibido la verosimilitud perfil. Este capítulo se divide en dos secciones. La primera sección corresponde a las críticas injustificadas que ha recibido la función de verosimilitud perfil. Como se verá en esa sección, las críticas son injustificadas por el uso indebido de funciones de densidad que tienen singularidades para aproximar a la función de verosimilitud correcta. Así, la función de verosimilitud perfil hereda también estas singularidades y consecuentemente los problemas asociados. Al recordar que la función de verosimilitud debe ser proporcional a una probabilidad, y por tanto está acotada, se mostrará como la verosimilitud perfil se libera por completo de dichos problemas. En contraste, en la segunda sección de este capítulo se presentan algunas críticas justificadas que ha recibido la perfil. Como se verá allí, estas críticas están relacionadas con la falsa precisión que la perfil presenta en algunas situaciones. Críticas similares se pueden hacer a las verosimilitudes marginal y condicional.

El Capítulo 3 de esta tesis trata acerca de la robusticidad de la verosimilitud perfil. Para explorar la robusticidad de la función de verosimilitud perfil se define a la función de verosimilitud perfil  $\epsilon$ -perturbada. Se mostrará cómo esta función resulta ser una herramienta muy útil para evaluar si se está en una situación robusta o no. Generalmente las situaciones no robustas suelen ser problemáticas, pero a su vez ofrecen retos interesantes de estimación. Un ejemplo de este caso se tiene con la estimación del parámetro  $N$  en el modelo Binomial  $(N, p)$  bajo ciertas condiciones. Este modelo bajo otras condiciones que se discutirán en la tesis, también puede ejemplificar un caso robusto y sin problemas de estimación.

Por último, en el Capítulo 4 se considera en el caso de dos variables aleatorias de interés  $X$  y  $Y$  el problema de estimación estadística del parámetro de confiabilidad

$\theta = P(X < Y)$ . Un contexto natural de este problema son los modelos tensión-fuerza. Para ellos se muestra cómo la función de verosimilitud perfil de este parámetro resulta sumamente informativa y cómo a partir de ella se obtienen intervalos de verosimilitud-confianza con buenas coberturas para  $\theta$  a través de un procedimiento sencillo. Se aplican estas ideas a dos casos muy básicos e importantes, ampliamente estudiados en la literatura estadística. El primero es cuando  $X$  y  $Y$  son variables aleatorias independientes e idénticamente distribuidas con función de densidad exponencial. El segundo caso es cuando  $X$  y  $Y$  son normales.

Las aportaciones principales de esta tesis son:

- Recordar que las funciones de verosimilitud son proporcionales a probabilidades y por tanto deben ser acotadas.
- Presentar una versión de la función de verosimilitud perfil, la cual se identificará como *verosimilitud perfil  $\epsilon$ -perturbada*, que sirve para explorar la robusticidad de la verosimilitud perfil frente a pequeños cambios en la estimación de los parámetros de estorbo.
- Dar para el modelo Binomial  $(N, p)$  una expresión matemática cerrada y simple para aproximar el límite de la función de verosimilitud perfil relativa de  $N$  cuando este parámetro tiende a infinito. El valor de este límite determina la forma de la verosimilitud perfil relativa de  $N$ .
- Exhibir las complicaciones que surgen al estimar la abundancia  $N$  de animales usando un modelo Binomial  $(N, p)$ .
- Ejemplificar cómo hacer inferencias sobre el parámetro  $\theta = P(X < Y)$  de manera muy simple en términos de intervalos de verosimilitud-confianza, a través del uso de la verosimilitud perfil relativa de este parámetro cuando  $X$  y  $Y$  son variables aleatorias exponenciales, así como cuando ambas son normales.

- Dar una expresión matemática cerrada con una estructura algebraica muy simple y sencilla para los intervalos de verosimilitud-confianza de este parámetro  $\theta = P(X < Y)$  cuando  $X$  y  $Y$  son variables exponenciales.

Las extensiones inmediatas de este trabajo son:

- Explorar la utilidad de la verosimilitud  $\epsilon$ -perturbada como herramienta para detectar situaciones problemáticas poco estables o robustas, especialmente en situaciones donde el cálculo de la verosimilitud global se deba hacer de manera numérica.
- Explorar el uso del modelo Binomial  $(N, p)$  para estimar abundancia de animales en el caso donde se tenga información histórica abundante sobre  $p$ , con base en otros experimentos realizados en el pasado bajo condiciones similares.
- Explorar el uso de la función de verosimilitud perfil para hacer inferencias sobre el parámetro  $\theta = P(X < Y)$  cuando  $X$  y  $Y$  son variables aleatorias con distribución de probabilidad  $F$  y  $G$ , no necesariamente iguales y/o cuando se tengan observaciones censuradas. Por ejemplo, un caso de particular interés es cuando  $X$  y  $Y$  son variables aleatorias Weibull con datos censurados.

Todos las simulaciones, cálculos de funciones de verosimilitud, optimizaciones y todas las figuras mostradas en esta tesis se realizaron con el lenguaje matricial de programación Matlab 7.0. en una computadora personal Pentium IV.

# Capítulo 1

## Conceptos estadísticos básicos

En este capítulo se revisan algunos conceptos básicos de la teoría de inferencia estadística paramétrica que serán utilizados a lo largo de esta tesis. Aquí se definen conceptos generales como: la función de verosimilitud, la función de verosimilitud relativa, la aproximación continua a la función de verosimilitud y sus propiedades, intervalos de verosimilitud y verosimilitud-confianza, la función de verosimilitud condicional, marginal, integrada, la densidad marginal posterior Bayesiana y las funciones de verosimilitud esperada y maximizada o perfil que es el tema central de esta tesis.

### 1.1 Función de verosimilitud

Considérese una muestra de variables aleatorias discretas independientes e idénticamente distribuidas  $Y = (Y_1, \dots, Y_n)'$  con función de probabilidad  $P(Y = y)$  que depende de un número finito de parámetros reales desconocidos  $\theta = (\theta_1, \dots, \theta_m)' \in \Theta \subset \mathbb{R}^m$ . El espacio parametral  $\Theta$  es la región de valores posibles que puede tomar el vector de parámetros  $\theta$ . Para resaltar que la función de probabilidad depende de parámetros desconocidos  $\theta$  se escribirá  $P(Y = y; \theta)$ .

Ronald A. Fisher (1921, pág. 24) definió por primera vez a la verosimilitud en el caso

de variables aleatorias discretas como una función del parámetro  $\theta$  que es proporcional a la probabilidad de la muestra observada  $P(Y = y; \theta)$ ,

$$L(\theta; y) \propto P(Y = y; \theta), \quad (1.1)$$

donde  $y$  es un vector de observaciones, la muestra observada, y  $\theta$  puede ser un vector.

La función de verosimilitud juega un papel fundamental en la inferencia estadística. Su rol principal es inferir sobre los parámetros de la distribución que haya sido elegida para describir mejor al fenómeno aleatorio de interés a partir de una muestra observada. Nótese que esto es particularmente relevante después de un experimento, cuando ya fueron observadas las variables aleatorias. Es importante notar que la función de verosimilitud en (1.1) se define como proporcional y no igual, a la función de probabilidad  $P(Y = y; \theta)$ . Esto enfatiza que sólo el cociente de verosimilitudes,  $L(\theta'; y)/L(\theta''; y) = P(Y = y; \theta')/P(Y = y; \theta'')$ , tiene significado, y se interpreta como una medida de la plausibilidad de  $\theta'$  relativa a  $\theta''$  basada en la muestra observada  $y$ . El cociente  $L(\theta'; y)/L(\theta''; y) = k$  significa que el valor de  $\theta'$  es  $k$  veces más plausible que el valor de  $\theta''$  en el sentido de que  $\theta'$  hace a la muestra observada  $k$  veces más probable de lo que la hace  $\theta''$ . Es por ello que se sugiere usar a la función de verosimilitud relativa que definimos a continuación para hacer inferencia sobre los parámetros del modelo probabilístico  $P(Y = y; \theta)$ .

## 1.2 Función de verosimilitud relativa

Para que la función de verosimilitud tenga una única representación, que no involucre una constante arbitraria, es conveniente estandarizar a la función de verosimilitud con respecto a su máximo (Sprott 2000, pág. 9). A esta función de verosimilitud estandarizada se le llama función de verosimilitud relativa y se define como

$$R(\theta; y) = \frac{L(\theta; y)}{\max_{\theta} L(\theta; y)} = \frac{L(\theta; y)}{L(\hat{\theta}; y)}, \quad (1.2)$$

donde  $\hat{\theta} = \hat{\theta}(y)$  es el valor del parámetro que maximiza  $L(\theta; y)$  y se llama estimador de máxima verosimilitud (emv) de  $\theta$ . El emv  $\hat{\theta}$  es el valor de  $\theta$  más plausible; es decir,  $\hat{\theta}$  es el valor de  $\theta$  que hace más probable a la muestra observada. Puesto que  $P(Y = y; \theta)$  es una probabilidad, necesariamente esta acotada entre 0 y 1 por lo que el denominador en (1.2) siempre existe y es finito. Así, la función de verosimilitud relativa yace entre cero y uno,  $0 \leq R(\theta; y) \leq 1$ , para todo valor de  $\theta$  en el espacio parametral.

La función de verosimilitud relativa proporciona la plausibilidad de cualquier valor especificado de  $\theta$  relativo al máximo verosimil  $\hat{\theta}$ , basada en la muestra observada  $y$ . Valores de  $\theta$  con  $R(\theta; y)$  cercanos a uno son muy creíbles o plausibles mientras que valores cercanos a cero son poco creíbles a la luz de la muestra observada.

## 1.3 Propiedades de la verosimilitud

### 1.3.1 Las verosimilitudes no son aditivas

A diferencia de la probabilidad, la verosimilitud no es aditiva. Esta es la principal distinción entre verosimilitud y probabilidad. La verosimilitud es una función puntual cuyo dominio es el espacio parametral y su contradominio son los números reales. En cambio la probabilidad es una función de conjuntos, el dominio es una sigma-álgebra y el contradominio es el intervalo  $[0, 1]$ . La probabilidad de la unión de conjuntos ajenos  $A$  y  $B$  está bien definida,  $P(A \cup B) = P(A) + P(B)$ . Pero la verosimilitud de la unión de dos valores  $\theta_1$  y  $\theta_2$  en el espacio parametral no está definida. La unión de  $\theta_1$  y  $\theta_2$  no es un real y por tanto no puede ser argumento de una función de verosimilitud.

### 1.3.2 Combinación de experimentos

Las funciones de verosimilitud combinan datos de experimentos diferentes de manera muy simple. Como la probabilidad conjunta de eventos independientes es el producto

de sus probabilidades individuales, entonces la función de verosimilitud de  $\theta$ , definida en (1.1), basada en diferentes conjuntos de datos que provienen de eventos (experimentos) independientes es el producto de las verosimilitudes individuales basadas en cada uno de estos conjuntos de datos. Así, el logaritmo de las verosimilitudes individuales basadas en cada conjunto de datos independientes se combinan a través de su suma. En particular esto significa que la forma apropiada de combinar información de experimentos diferentes que involucran un parámetro común de interés  $\theta$  es a través de la suma de los respectivos logaritmos de las verosimilitudes individuales y esta función se maximiza para encontrar  $\hat{\theta}$  el emv común a todos los experimentos. Nótese que cada experimento podría involucrar a otros parámetros de estorbo particulares.

### 1.3.3 Invarianza funcional

La invarianza funcional es una característica muy conveniente de las verosimilitudes. Significa que cualquier declaración cuantitativa acerca de  $\theta$  implica una declaración cuantitativa correspondiente acerca de cualquier función uno a uno de  $\theta$ ,  $\delta = \delta(\theta)$ , por directa sustitución algebraica  $\theta = \theta(\delta)$ . Por ejemplo, si  $\theta > 0$  y  $\delta = \log \theta$ , entonces la verosimilitud del nuevo parámetro  $\delta$  es  $R^*(\delta; y) = R[\theta = \exp(\delta); y]$ . Como consecuencia se tiene que el emv de  $\delta$  es  $\hat{\delta} = \log \hat{\theta}$ . También  $a \leq \theta \leq b$  sí y sólo sí  $\log a \leq \delta \leq \log b$ . Estas dos declaraciones son equivalentes debido a que tienen la misma plausibilidad o incertidumbre.

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

asintóticas del emv  $\hat{\delta}$  para la muestra pequeña en cuestión, Sprott (2000, pág. 34-35).

## 1.4 Aproximación continua de la función de verosimilitud y la verosimilitud exacta

La función de verosimilitud en (1.1) se definió en términos de variables aleatorias discretas. Sin embargo, esto no involucra una pérdida de generalidad ya que en realidad los datos observados,  $y$ , siempre son discretos puesto que todo instrumento de medición tiene precisión finita, y sólo pueden registrarse mediciones con un número finito de decimales, Barnard y Sprott (1983).

Cuando  $Y_i$  es una variable aleatoria continua, la observación  $Y_i = y_i$  debe interpretarse como  $Y_i \in [y_i - \frac{1}{2}h; y_i + \frac{1}{2}h]$ , donde  $h$  es un número positivo fijo que representa la precisión del instrumento de medición. Entonces, para una muestra de variables aleatorias independientes e idénticamente distribuidas  $Y = (Y_1, \dots, Y_n)'$  con función de densidad  $f(y; \theta)$ , la función de verosimilitud de  $\theta$  es proporcional a la probabilidad conjunta de la muestra observada,

$$\begin{aligned} L(\theta; y) &\propto \prod_{i=1}^n P(y_i - \frac{1}{2}h \leq Y_i \leq y_i + \frac{1}{2}h; \theta) \\ &= \prod_{i=1}^n \int_{y_i - \frac{1}{2}h}^{y_i + \frac{1}{2}h} f(y; \theta) dy. \end{aligned} \quad (1.3)$$

En lo que sigue de esta tesis, a la función de verosimilitud de  $\theta$  definida en (1.3) se le llamará **verosimilitud exacta** de  $\theta$ .

Nótese que  $h = 0$  significa que el instrumento de medición tiene precisión infinita y que las observaciones pueden registrarse con un número infinito de decimales, lo cual es imposible en la realidad. Para una variable aleatoria continua  $Y$  se tiene que  $P(Y = y; \theta) = 0$  para todo  $y$  y  $\theta$ . Por ello, no se puede definir a la función de verosimilitud a partir de dichas probabilidades en el caso de variables continuas.

Si en contraste, se supone que la precisión del instrumento de medición es  $h > 0$  entonces bajo ciertas condiciones la función de densidad  $f(y; \theta)$  podrá usarse como una aproximación a la función de verosimilitud exacta (1.3). Sin embargo, si la función de densidad tiene una singularidad (discontinuidad infinita) en cualquier valor de  $\theta$ , entonces  $\max_{\theta} L(\theta; y)$ , el denominador en (1.2), es infinito y entonces  $R(\theta; y)$  en (1.2) no estaría definida. Este sería un caso donde la densidad  $f(y; \theta)$  no podría usarse para aproximar a la función de verosimilitud.

Sin embargo, la aproximación de la verosimilitud exacta, a través de una densidad, es justificable bajo ciertas condiciones. Por el teorema de valor medio para integrales de funciones continuas, la  $i$ -ésima integral en (1.3) es  $hf(y'; \theta)$  para algún punto intermedio  $y' \in [y_i - \frac{1}{2}h, y_i + \frac{1}{2}h]$ . Si  $f(y; \theta)$  es aproximadamente constante en este intervalo para todo valor plausible de  $\theta$ , entonces  $f(y'; \theta) \approx f(y; \theta)$  en ese mismo intervalo. Si esta aproximación es adecuada para algunos o todos los  $i \in \{1, \dots, n\}$  y si  $h$  no depende de  $\theta$  entonces las correspondientes probabilidades en (1.3) pueden reemplazarse por la función de densidad evaluada en el valor observado  $y_i$ . En la práctica, es usual que esta aproximación se use y sustituya a todas las probabilidades en (1.3). Esto es,

$$L(\theta; y) \propto \prod_{i=1}^n P(y_i - \frac{1}{2}h \leq Y_i \leq y_i + \frac{1}{2}h; \theta) \approx \prod_{i=1}^n hf(y_i; \theta) \propto \prod_{i=1}^n f(y_i; \theta).$$

Kalbfleisch (1985, Sección 9.4), Edwards (1992, pág. 6, pág. 167), Lindsey (1999), Sprott (2000, pág. 19, págs. 203-294), Lawless (2003, pág. 186), Meeker y Escobar (1998, pág. 275) discuten sobre lo necesario y lo importante que es tomar en cuenta estos argumentos. Hay que enfatizar que (1.3) no es el intento *ad hoc* para discretizar una variable aleatoria continua  $y$ , con función de densidad  $f(y; \theta)$ , la cual se considera como la base para calcular la función de verosimilitud. En realidad se trata de lo opuesto, una aproximación continua a una variable aleatoria discreta  $y$ , con función de probabilidad  $P(Y = y; \theta)$ , la cual es la base para calcular la función de verosimilitud. El propósito principal de la aproximación continua es la conveniencia matemática;

derivadas e integrales son más fáciles de calcular que diferencias y sumas finitas.

## 1.5 Intervalos de verosimilitud

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

$$IV(c) = \{\theta \mid R(\theta; y) \geq c\}, \text{ donde } 0 \leq c \leq 1.$$

Todo valor de  $\theta$  en el  $IV(c)$  tiene verosimilitud relativa igual o mayor que  $c$ , y todo valor de  $\theta$  afuera, tiene verosimilitud relativa menor. Por tanto el  $IV(c)$  separa los valores plausibles de  $\theta$  de los no plausibles a un nivel  $c$ , (Sprott, 2000, pág. 14).

Cuando  $\theta$  es unidimensional, el  $IV(c)$  se obtiene trazando una línea horizontal en la gráfica de  $R(\theta; y)$  a una distancia  $c$  paralela al eje cartesiano  $\theta$ . Además, variando  $c$  de 0 a 1 se obtiene una familia jerarquizada y anidada de intervalos de verosimilitud que converge al emv  $\hat{\theta}$  cuando  $c$  tiende a 1. El emv  $\hat{\theta}$  está contenido en todos los intervalos de verosimilitud puesto que  $R(\hat{\theta}; y) = 1$ . Así, esta familia de intervalos anidados es equivalente a la función de verosimilitud completa y reproduce la gráfica de  $R(\theta; y)$ .

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  $\theta$  adentro del intervalo. Al menos, un intervalo de verosimilitud debe estar acompañado siempre del valor del emv  $\hat{\theta}$  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  $c = .036, .15, .25$ . En lo posible se debe también graficar y analizar la función de verosimilitud relativa completa, como se hará en el siguiente ejemplo.

### Ejemplo 1.4.1

Considérese un cierto tipo de componente electrónico que falla en cualquier instante de tiempo. Sin embargo, los componentes no se deterioran con la edad, y la probabilidad de falla dentro de un periodo de tiempo dado no depende de la edad del componente. Se supone que el tiempo de vida de tales componentes sigue una distribución exponencial  $F(x; \theta)$  con función de densidad

$$f(x; \theta) = \frac{1}{\theta} \exp\left(-\frac{x}{\theta}\right), \text{ para todo } x \text{ positivo,}$$

donde  $\theta$  es el tiempo de vida esperado de los componentes. El objetivo es hacer inferencia sobre el parámetro  $\theta$  basados en observaciones de tiempo de vida de los componentes.

Supóngase que  $n$  componentes electrónicos son puestos a prueba durante periodos de tiempo fijo  $T_1, \dots, T_n$ . Considérese que  $r$  de estos componentes fallaron antes de culminar sus periodos de prueba, y que los tiempos de falla observados fueron  $t_1, \dots, t_r$ . Así, los  $(n - r)$  componentes restantes no fallaron durante sus periodos de prueba y tuvieron tiempos de vida censurados  $T_{r+1}, \dots, T_n$ . Entonces, usando la aproximación continua de la función de verosimilitud, sólo en los  $r$  componentes que fallaron, y considerando el aporte de los  $(n - r)$  tiempos de vida censurados se tiene que la función de verosimilitud de  $\theta$  es

$$\begin{aligned} L(\theta) &\propto \left[ \prod_{i=1}^r f(t_i; \theta) \right] \prod_{i=r+1}^n [1 - F(T_i; \theta)] \\ &= \left[ \prod_{i=1}^r \frac{1}{\theta} \exp\left(-\frac{t_i}{\theta}\right) \right] \prod_{i=r+1}^n \exp\left(-\frac{T_i}{\theta}\right) \\ &= \theta^{-r} \exp\left(-\frac{t}{\theta}\right), \text{ donde } t = \sum_{i=1}^r t_i + \sum_{i=r+1}^n T_i. \end{aligned}$$

Obsérvese que el emv de  $\theta$  es  $\hat{\theta} = t/r$ . Por tanto, la función de verosimilitud relativa

de  $\theta$  que resulta es

$$R(\theta) = \frac{L(\theta; y)}{L(\hat{\theta}; y)} = \left( \frac{\theta}{\hat{\theta}} \right)^{-r} \exp \left( -\frac{t}{\theta} + \frac{t}{\hat{\theta}} \right) = \left( \frac{r}{t} \theta \right)^{-r} \exp \left( -\frac{t}{\theta} + r \right).$$

Un ejemplo que aparece en la literatura es  $n = 10$ , con periodos de tiempo de prueba fijo  $T = 81, 70, 41, 31, 31, 30, 29, 72, 60, 21$  días. De estos componentes, siete fallaron antes de culminar su periodo de prueba, y los tiempos de falla observados fueron  $t = 2, 51, 33, 27, 14, 24, 4$  días. Los tres tiempos de vida censurados fueron  $T_8 = 72, T_9 = 60$  y  $T_{10} = 21$ , (Bartholomew, 1957 y Sprott, 2000, pág. 23). Así, como  $r = 7$  y  $t = 308$  entonces la función de verosimilitud relativa de  $\theta$  es

$$R(\theta) = \left( \frac{7}{308} \theta \right)^{-7} \exp \left( -\frac{308}{\theta} + 7 \right). \quad (1.4)$$

La Figura 1.1 muestra la gráfica de  $R(\theta)$  en (1.4), los intervalos de verosimilitud de nivel  $c = .036, .15$  y  $.25$  y la ubicación del emv de  $\theta$ ,  $\hat{\theta} = 252/8 = 31.5$ . Se observa claramente que la función de verosimilitud relativa de  $\theta$  es asimétrica con cola pesada a la derecha. Esto también puede verse a partir de los intervalos de verosimilitud y de la ubicación del emv  $\hat{\theta}$ . Existe una desviación evidente entre  $\hat{\theta}$  y el centro geométrico de cada intervalo, lo que da idea de la asimetría de la función de verosimilitud.

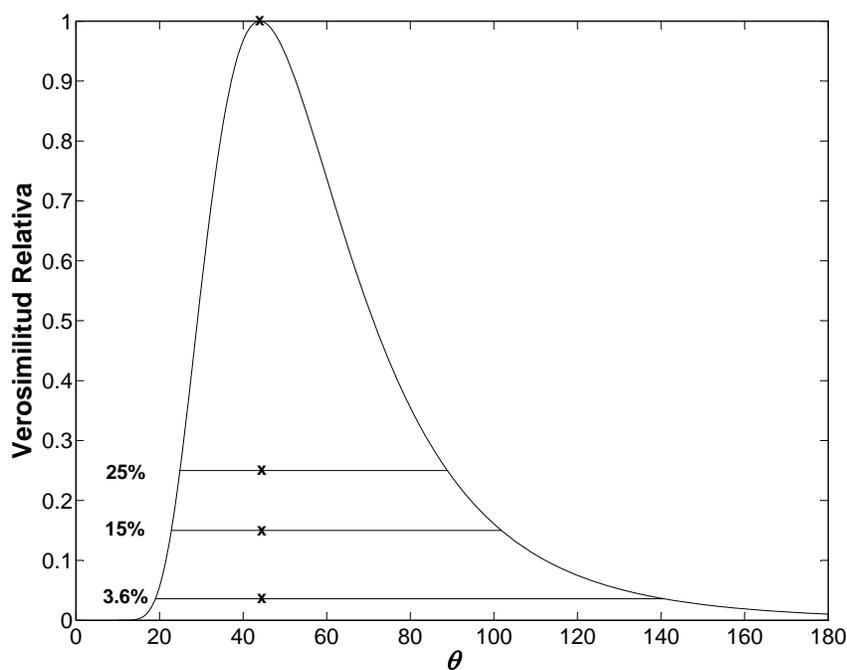


Figura 1.1: Intervalos de verosimilitud.

El resumen de las inferencias en términos de verosimilitud consiste en proporcionar el nivel de verosimilitud, el extremo inferior del intervalo,  $A$ , el env  $\hat{\theta}$  y el extremo superior  $B$  del intervalo de verosimilitud. Usualmente se dan intervalos de verosimilitud de niveles  $c = .036, .15$  y  $.25$  porque suelen estar asociados con niveles del 99, 95 y 90% de confianza, en el caso de que el parámetro sea unidimensional.

Tabla 1.1. Resumen de inferencias para  $\theta$ .

Nivel	$A$	$\hat{\theta}$	$B$
$c = .036$	19.02	44	140.64
$c = .15$	22.83	44	101.80
$c = .25$	24.89	44	88.86

Nótese que valores de  $\theta \leq 19.02$  y valores de  $\theta \geq 140.64$  tienen plausibilidad muy pequeña, menor que  $c = 0.036$ .

Antes de finalizar esta sección es importante enfatizar que la verosimilitud es una

función puntual y por lo tanto el nivel de plausibilidad  $c$  de un intervalo de verosimilitud no es una declaración de la incertidumbre del intervalo. Es una declaración acerca de la plausibilidad relativa de cualquier punto individual adentro del intervalo, cuya verosimilitud es mayor o igual a  $c$ .

## 1.6 Intervalos de verosimilitud-confianza

Supóngase que se tiene una muestra de variables aleatorias independientes e idénticamente distribuidas  $Y = (Y_1, \dots, Y_n)'$  con función de probabilidad  $P(Y = y; \theta)$ , donde  $\theta$  es un parámetro escalar desconocido. Si se tiene una muestra observada  $y$  que proviene de la distribución de  $Y$  con  $\theta$  fijo en un valor  $\theta_0$ , entonces se puede calcular, a partir de ésta muestra observada, un intervalo  $[A, B]$  para el valor verdadero  $\theta_0$ . Así, los extremos del intervalo  $[A, B]$  son variables aleatorias puesto que su valor varía cuando cambia la muestra. Obsérvese que en principio, la distribución de probabilidad de  $A$  y  $B$  se puede calcular a partir de la distribución de la variable aleatoria  $Y$  y generalmente depende de  $\theta_0$ . Además, como el intervalo  $[A, B]$  puede variar cuando varía la muestra entonces éste puede algunas veces incluir y en otras no, al valor verdadero  $\theta_0$ .

La probabilidad de cobertura de un intervalo aleatorio  $[A, B]$  es la probabilidad de que el intervalo  $[A, B]$  incluya o cubra, el verdadero valor del parámetro  $\theta_0$ ,

$$PC(\theta_0) = P(A \leq \theta_0 \leq B; \theta = \theta_0).$$

La probabilidad de cobertura  $PC(\theta_0)$  se interpreta como la fracción de veces que el intervalo  $[A, B]$  incluirá el valor verdadero  $\theta_0$  en un número muy grande de repeticiones de la muestra pero con el valor de  $\theta$  fijo en  $\theta_0$ .

Kalbfleisch (1985, pág. 113) define un intervalo de confianza para  $\theta$  de la siguiente manera. Un intervalo  $[A, B]$  se llama intervalo de confianza para  $\theta$  cuando su probabilidad de cobertura no depende de  $\theta_0$ . Es decir, cuando el valor de  $PC(\theta_0)$  es el

mismo para todo valor del parámetro  $\theta_0$ . Nosotros adoptaremos esta definición ya que es clara, adecuada y suficiente para los objetivos de este trabajo.

La probabilidad de cobertura de un  $IV(c)$  se puede calcular a través de la distribución de probabilidad de la estadística de la razón de verosimilitud para un  $\theta$  fijo en  $\theta_0$ ,  $RV \equiv -2 \log R(\theta_0)$ . Un valor particular  $\theta_0$  está en el  $IV(c)$  sí y sólo sí  $R(\theta_0) \geq c$ , o de forma equivalente,  $-2 \log R(\theta_0) \leq -2 \log(c)$ . Por lo tanto, la probabilidad de cobertura del  $IV(c)$  es

$$\begin{aligned} PC(\theta_0) &= P[\theta_0 \in IV(c); \theta = \theta_0] \\ &= P[RV \leq -2 \log(c); \theta = \theta_0]. \end{aligned} \tag{1.5}$$

Muchas veces es difícil encontrar la distribución de probabilidad exacta de  $RV$ ; sin embargo existe teoría asintótica que suele dar una aproximación buena a ésta distribución en muchos casos. Bajo algunas condiciones de regularidad, la estadística de la razón de verosimilitud  $RV \equiv -2 \log R(\theta_0)$  converge en distribución a una Ji-cuadrada con un grado de libertad para todo  $\theta_0 \in \Theta \subset \mathbb{R}$ . En este caso particular, esto equivale a que se tenga que

$$\lim_{n \rightarrow \infty} P(RV \leq x; \theta = \theta_0) = P[\chi_{(1)}^2 \leq x], \text{ para todo } x \text{ positivo.}$$

Para mayores detalles de la prueba ver Serfling (1980; pág. 155-156).

De (1.5) se tiene que  $P(RV \leq x; \theta = \theta_0)$  es la probabilidad de cobertura del  $IV(c)$ , donde  $x = -2 \log(c)$ ; esto es,  $c = \exp(-x/2)$ . Así, si se selecciona  $x = q_{(\alpha,1)}$ , donde  $q_{(\alpha,1)}$  es el cuantil  $(1 - \alpha)$  de una distribución Ji-cuadrada con un grado de libertad, entonces el  $IV(c)$  tiene una probabilidad de cobertura aproximada del  $100(1 - \alpha)\%$ , donde  $c = \exp(-q_{(\alpha,1)}/2)$ . Así, el  $IV(c)$  también es un intervalo o región de confianza para el parámetro  $\theta$  y toma consecuentemente el nombre de intervalo o región de verosimilitud-confianza para  $\theta$ .

Obsérvese en la Tabla 1.2 que 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, respectivamente.

Entonces, los  $IV(c)$  con  $c = .258, .146$  y  $.036$  tienen una probabilidad de cobertura aproximada del 90%, 95% y 99%, respectivamente.

Tabla 1.2. Confianza aproximada de intervalos de verosimilitud cuando  $\theta$  es unidimensional.

$(1 - \alpha)$	$c$	$q_{(\alpha,1)}$
.90	.258	2.706
.95	.146	3.841
.99	.036	6.635

## 1.7 Estimación por separado de parámetros de interés en presencia de parámetros de estorbo.

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.

Es importante resaltar que un parámetro puede ser considerado en algunos casos como de estorbo y en otros casos no. Por ejemplo, si se tiene una muestra de variables normales con media  $\mu$  y varianza  $\sigma^2$  y se desea estimar la media poblacional  $\mu$ , entonces  $\sigma$  es un parámetro de estorbo o no deseado. En contraste, si el parámetro de interés es  $\sigma$ , entonces ahora el parámetro de estorbo es  $\mu$ .

Supóngase que se tiene el vector de parámetros  $\theta = (\delta, \lambda) \in \mathbb{R}^m$ , donde  $\delta$  es un parámetro de interés de dimensión  $m_\delta$  y  $\lambda$  es considerado un parámetro de estorbo de dimensión  $m_\lambda$ , donde  $m = m_\delta + m_\lambda$ . A continuación se presentan diferentes métodos

para estimar por separado a  $\delta$  en presencia del parámetro  $\lambda$  de estorbo, tales como las verosimilitudes condicional, marginal, estimada e integrada. También es común usar una densidad marginal posterior Bayesiana del parámetro de interés para este fin. Finalmente se presenta el concepto central de esta tesis que es la verosimilitud maximizada o perfil.

Cabe mencionar aquí que muchas veces dos parámetros desconocidos se pueden relacionar de una manera tan estrecha y confusa que no sea posible separarlos, Sprott (2000, pág. 50). Edwards (1992, pág. 109) comenta que no hay razón lógica para suponer que siempre sea posible eliminar un parámetro de estorbo. Sin embargo, bajo el enfoque Bayesiano matemáticamente siempre es posible eliminar los parámetros de estorbo independientemente de la asociación o dependencia inextricable que guarden los parámetros. Lo mismo ocurre con la verosimilitud integrada. En realidad, esto es una gran desventaja, pues las consecuencias inmediatas de aplicar de manera automática estas herramientas son que las inferencias hechas sobre los parámetros de interés pueden estar muy alejadas de la realidad. Por ello es de vital importancia explorar siempre el grado de asociación y confusión que tengan los parámetros de interés con los de estorbo.

### 1.7.1 Función de verosimilitud condicional

Supóngase nuevamente que  $\delta$  es un parámetro de interés y que se tiene una estadística  $t$  suficiente minimal para el parámetro  $\lambda$  de estorbo y que la verosimilitud se puede factorizar de la siguiente manera

$$\begin{aligned} L(\delta, \lambda; y) &\propto P(y; \delta, \lambda) = P(y; \delta | t) P(t; \delta, \lambda) \\ &\propto L_C(\delta; y) L_{Cres}(\delta, \lambda; t). \end{aligned} \quad (1.6)$$

La función  $L_C(\delta; y)$  en (1.6) se llama función de verosimilitud condicional de  $\delta$  puesto que se basa en la distribución condicional de la muestra  $y$  dada la estadística  $t$  suficiente

minimal para  $\lambda$ . Nótese que al ser  $t$  una estadística suficiente minimal para  $\lambda$ , ocasiona que el primer factor en (1.6) no dependa de  $\lambda$  para cualquier valor especificado de  $\delta$ .

En la práctica, el uso de la función de verosimilitud condicional  $L_C(\delta; y)$  para inferencias sobre  $\delta$  depende de qué tan buena sea la factorización en (1.6) para separar la información del parámetro de interés  $\delta$  del de estorbo  $\lambda$ . Esto depende de que el factor  $L_{Cres}(\delta, \lambda; t)$  en (1.6), llamado función de verosimilitud condicional residual, contenga poca información acerca de  $\delta$  cuando se desconoce  $\lambda$ . A pesar de que esta condición es muy difícil de cuantificar, más adelante se presentará una forma gráfica de valorarla.

## 1.7.2 Función de verosimilitud marginal

Supóngase que se tiene una estadística  $a$  *ancilar* (traducción literal del inglés *ancillary*) o auxiliar para el parámetro  $\lambda$  de estorbo y que la verosimilitud se puede factorizar ahora de la siguiente manera,

$$\begin{aligned} L(\delta, \lambda; y) \propto P(y; \delta, \lambda) &= P(a; \delta) P(y; \delta, \lambda | a) \\ &\propto L_M(\delta; a) L_{Mres}(\delta, \lambda; y). \end{aligned} \quad (1.7)$$

La función  $L_M(\delta; a)$  en (1.7) se llama función de verosimilitud marginal de  $\delta$  puesto que se basa en la distribución marginal de la estadística auxiliar  $a$ . Una estadística es auxiliar para un parámetro si su distribución no depende de dicho parámetro. Obsérvese que  $a$  es una estadística auxiliar para  $\lambda$  para cualquier valor especificado de  $\delta$ , ya que su distribución en (1.7) no depende de  $\lambda$ .

En la práctica, el uso de la función de verosimilitud marginal  $L_M(\delta; y)$  para inferencias sobre  $\delta$  depende de qué tan buena sea la factorización en (1.7) para separar la información del parámetro de interés  $\delta$  del parámetro de estorbo  $\lambda$ . Esto depende de que el factor  $L_{Mres}(\delta, \lambda; y)$  en (1.7), llamado función de verosimilitud marginal residual, contenga poca información acerca de  $\delta$  cuando se desconoce  $\lambda$ . Como en la

Sección 1.6.1, a pesar de que esta situación es difícil de cuantificar, gráficamente se puede evaluar como se verá más adelante.

### 1.7.3 Función de verosimilitud integrada

Este método consiste en eliminar el parámetro de estorbo a través de integración. Si se conoce una función de densidad inicial, previa o *a priori* para el parámetro de estorbo  $\lambda$  que dependa del parámetro de interés  $\delta$  en la forma  $f(\lambda; \delta)$ , entonces esta densidad previa puede multiplicarse con la función de verosimilitud  $L(\delta, \lambda; y)$  e integrarse con respecto a  $\lambda$  para así eliminar este parámetro. Entonces se obtiene la función de verosimilitud integrada de  $\delta$ ,

$$L_I(\delta; y) \propto \int_{\Lambda} P(y; \delta | \lambda) f(\lambda; \delta) d\lambda, \quad (1.8)$$

que sólo depende del parámetro de interés  $\delta$ .

Se usará la notación  $P(z; a|b)$  cuando  $b$  tenga el nivel lógico de variable aleatoria y  $a$  sea un parámetro fijo. Entonces esto denota la distribución condicional de  $z$  dado  $b$  que depende de un parámetro fijo  $a$ .

La dificultad principal con este método para hacer inferencias sobre  $\delta$  es que depende fuertemente de información inicial muy específica, que se da a través de la densidad previa y con la que a menudo no se cuenta.

### 1.7.4 Densidad marginal posterior Bayesiana

En el enfoque estadístico Bayesiano todos los parámetros, de interés y de estorbo, son considerados variables aleatorias, a diferencia del enfoque no Bayesiano de estimación, donde se considera que los parámetros son fijos pero desconocidos.

Bajo el enfoque Bayesiano se especifica una función de densidad conjunta inicial, previa o *a priori* para  $\delta$  y  $\lambda$  en la forma  $f(\delta, \lambda)$  y entonces, a través del Teorema de

Bayes se calcula la función de densidad conjunta posterior de los parámetros  $\delta$  y  $\lambda$  dada la muestra observada  $Y = y$  de la siguiente manera:

$$f(\delta, \lambda | y) \propto P(y | \delta, \lambda) f(\delta, \lambda) \propto L(\delta, \lambda; y) f(\delta, \lambda). \quad (1.9)$$

En presencia del parámetro de estorbo  $\lambda$ , la densidad relevante para hacer inferencias sobre  $\delta$  es la densidad marginal posterior de  $\delta$  dada la muestra observada  $Y = y$ . Esta densidad marginal se calcula integrando  $f(\delta, \lambda | y)$  dada en (1.9) con respecto al parámetro de estorbo  $\lambda$ ,

$$f(\delta | y) \propto \int_{\Lambda} P(y | \delta, \lambda) f(\delta, \lambda) d\lambda,$$

y se utiliza para hacer inferencia sobre  $\delta$ .

### 1.7.5 Función de verosimilitud estimada

Este método consiste en reemplazar el parámetro de estorbo  $\lambda$  por  $\tilde{\lambda}$  en la función de verosimilitud global  $L(\delta, \lambda; y)$ , donde  $\tilde{\lambda} = \tilde{\lambda}(y)$  es algún estimador “razonable” del parámetro  $\lambda$  de estorbo, Pawitan (2001, pág 292). Así, la función de verosimilitud estimada de  $\delta$  es

$$L_E(\delta; y) = L(\delta, \tilde{\lambda}; y). \quad (1.10)$$

Obsérvese que  $\tilde{\lambda}$  no es necesariamente el emv de  $\lambda$ . Además, es importante notar aquí que el valor del estimador  $\tilde{\lambda}$  no depende del parámetro  $\delta$  de interés.

La función de verosimilitud estimada de  $\delta$ ,  $L_E(\delta; y)$  en (1.10), puede interpretarse de manera geométrica como la intersección del hiperplano  $\lambda = \tilde{\lambda}$  con la superficie de verosimilitud global  $L(\delta, \lambda; y)$ . Es decir, si  $\delta$  y  $\lambda$  son parámetros unidimensionales entonces la función de verosimilitud estimada es la trayectoria sobre la función de verosimilitud  $L(\delta, \lambda; y)$  que resulta al cortar la superficie de verosimilitud con el plano vertical  $\lambda = \tilde{\lambda}$ , que es paralelo al eje  $\delta$ .

El problema fundamental con este método es que supone que no hay incertidumbre alguna en la estimación del parámetro de estorbo ya que este se reemplaza por  $\tilde{\lambda}$ , cuando en realidad se desconoce por completo a dicho parámetro. Así, la función de verosimilitud que se obtiene generalmente presenta un exceso de precisión en las inferencias que suele ser irreal.

### 1.7.6 Función de verosimilitud maximizada o perfil

La función de verosimilitud maximizada o perfil es un método estadístico muy simple y poderoso que sirve para estimar por separado un parámetro de interés en presencia de parámetros de estorbo. Aunque ya se había usado antes en la literatura estadística por Hood y Koopmans (1953, pág. 156-1957) en el contexto de modelos econométricos y denominándola ‘verosimilitud concentrada’ (véase también Seber y Wild, 2003, Sección 2.2.3) y Box y Cox (1964) la usaron y graficaron para ayuda en la selección del parámetro de forma  $\lambda$  de sus modelos de transformaciones. Sin embargo, Sprott y Kalbfleisch (1969), por vez primera le dieron el nombre de maximizada y la presentaron formalmente como un método general para eliminar parámetros de estorbo en Kalbfleisch y Sprott (1970). Para este propósito, la función de verosimilitud maximizada o perfil es mucho más general que otras verosimilitudes tales como la condicional, la marginal o la integrada, las cuales dependen de una estructura especial como en (1.6), (1.7) y (1.8), respectivamente, y por lo tanto son más restrictivas.

La función de verosimilitud maximizada o perfil del parámetro de interés  $\delta$ ,  $L_P(\delta)$ , se define como

$$L_P(\delta; y) = \max_{\lambda|\delta} L(\delta, \lambda; y) = L[\delta, \hat{\lambda}(\delta, y); y], \quad (1.11)$$

donde  $\hat{\lambda}(\delta, y)$  es el estimador de máxima verosimilitud restringido (emvr) de  $\lambda$  para un valor especificado de  $\delta$ . El emvr  $\hat{\lambda}(\delta, y)$  es el valor de  $\lambda$  que tiene mayor plausibilidad para ese valor fijo de  $\delta$  dada la muestra observada  $Y = y$ . Es decir, la verosimilitud perfil de  $\delta$ ,  $L_P(\delta; y)$ , se obtiene maximizando la función de verosimilitud  $L(\delta, \lambda; y)$

sobre  $\lambda$  pero fijando  $\delta$ . Por tanto, en especial se tiene que el emv global de  $\lambda$  coincidirá con el estimador restringido evaluado en el emv global  $\hat{\delta}$ ,  $\hat{\lambda} = \hat{\lambda}(\hat{\delta}, y)$ .

Es importante resaltar que la función de verosimilitud maximizada o perfil en (1.11) se obtiene reemplazando el parámetro de estorbo  $\lambda$  por  $\hat{\lambda}(\delta, y)$ , que depende de  $\delta$ , en la función de verosimilitud global  $L(\delta, \lambda; y)$ . En contraste, la función de verosimilitud estimada en (1.10) se obtenía reemplazando  $\lambda$  por  $\tilde{\lambda}$  fijo en la función de verosimilitud global  $L(\delta, \lambda; y)$ . Así, la función de verosimilitud perfil de  $\delta$  se calcula poniendo énfasis en una familia de estimadores puntuales de  $\lambda$ ,  $\hat{\lambda}(\delta, y)$ , indicada por  $\delta$ , mientras que la función de verosimilitud estimada sólo pone énfasis en un punto fijo  $\tilde{\lambda}$ , para todo valor de  $\delta$ . De esta manera la función de verosimilitud perfil se puede ver como una generalización de la verosimilitud estimada pero que rescata de manera adaptable información que contiene la muestra sobre el parámetro de estorbo para cada valor fijo de  $\delta$  a través del emvr  $\hat{\lambda}(\delta, y)$ .

Nótese que cuando  $\delta$  y  $\lambda$  son parámetros unidimensionales entonces la función de verosimilitud global o de ambos parámetros,  $L(\delta, \lambda; y)$ , 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  $\lambda$ , entonces la silueta o perfil que se observa de esta verosimilitud global  $L(\delta, \lambda; y)$  es justamente la función de verosimilitud maximizada de  $\delta$ . Por esta razón la verosimilitud maximizada también recibe el nombre de perfil. A manera de ilustración véase la Figura 1.2, donde  $\theta$  es el parámetro de interés y  $p$  es considerado el parámetro de estorbo.

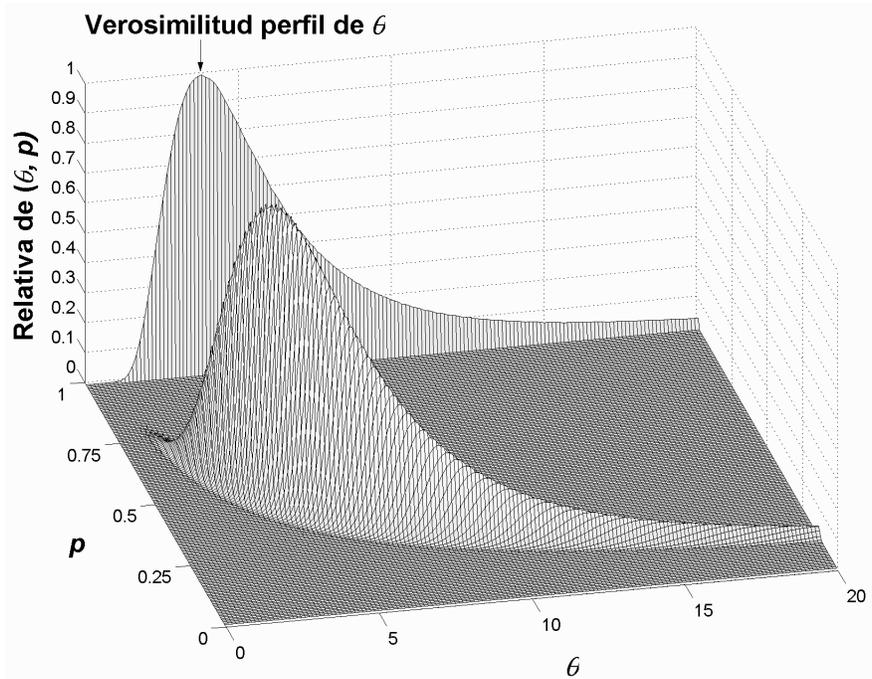


Figura 1.2: Función de verosimilitud perfil de  $\theta$ .

A continuación se presentan algunas propiedades importantes que tiene la función de verosimilitud perfil:

1. El emv perfil de  $\delta$  es igual a el emv global o no restringido  $\widehat{\delta}$ .
2. La estadística de la razón de verosimilitud perfil,

$$\text{RVP} = -2 \log[L_P(\delta_0; y) - L_P(\widehat{\delta}; y)] = -2 \log R_P(\delta_0; y),$$

converge en distribución a una Ji-cuadrada con un grado de libertad para todo  $\delta \in \Delta \subset \mathbb{R}$ . Los detalles de la prueba se pueden ver en Serfling (1980; 156-160). Es importante notar que la estadística RVP es igual a la estadística de la razón de verosimilitud generalizada para la hipótesis  $H : \delta = \delta_0$  (ver Mood, Graybill y Boes, 1985, pág. 419 ).

3. De lo anterior se tiene que un intervalo o región de verosimilitud de nivel  $c$ , obtenido a partir de la función de verosimilitud perfil de  $\delta$ ,  $\{\delta | R_P(\delta; y) \geq c\}$ , donde  $0 \leq c \leq 1$ , también es un intervalo o región de confianza aproximada para el parámetro  $\delta$  de interés. Cuando la dimensión de  $\delta$  es menor o igual que cinco,  $d_\delta \leq 5$ , entonces los IV( $c$ ) con  $c = \exp[-(d_\delta + 1)]$ ,  $\exp[-(d_\delta + 3)]$  y  $\exp[-(d_\delta + 5)]$  tienen una probabilidad de cobertura aproximada del 95%, 99% y 99.9%, respectivamente, Barndorff-Nielsen y Cox (1994, pág. 90). Nótese que si  $\delta$  es un parámetro escalar,  $d_\delta = 1$ , entonces como se vio en la Sección 1.5, los intervalos de verosimilitud perfil con  $c = .258$ ,  $.146$  y  $.036$  tienen una probabilidad de cobertura aproximada del 90%, 95% y 99%, respectivamente. Para ahondar en el tema de calibración de regiones de verosimilitud al aumentar la dimensión del parámetro de interés, véase Pawitan (2004, Sección 3.5).

4. Sea  $I_{\hat{\theta}}$  la matriz de información observada para el vector de parámetros  $\theta = (\delta, \lambda) \in \mathbb{R}^m$ . La entrada  $ij$ -ésima de esta matriz es

$$\{I_{\hat{\theta}}\}_{ij} = -\frac{\partial^2}{\partial\theta_i\partial\theta_j} \log L(\theta; y) \Big|_{\theta=\hat{\theta}}, \quad \text{para } i, j = 1, \dots, m, \quad (1.12)$$

donde  $\hat{\theta}$  es el emv de  $\theta$ . Sea  $I_{\hat{\theta}}^{-1}$  la matriz inversa de  $I_{\hat{\theta}}$  y defínase a la  $i$ -ésima entrada en la diagonal de  $I_{\hat{\theta}}^{-1}$  como  $\{I_{\hat{\theta}}^{-1}\}^{ii}$ . Considérese el caso que  $\delta$  es un parámetro unidimensional y la primera entrada del vector  $\theta$ . En este caso la información observada perfil para  $\delta$  se define como

$$I_{\hat{\delta}} = \frac{1}{\{I_{\hat{\theta}}^{-1}\}^{11}}. \quad (1.13)$$

La matriz  $I_{\hat{\theta}}^{-1}$  se suele interpretar como un estimador asintótico de la matriz de covarianza del estimador de máxima verosimilitud  $\hat{\theta}$ . Para algunas muestras pequeñas esta interpretación no es necesariamente correcta, Sprott (2000, Secciones 5.7, 9.1), Fisher (1991, pág. 161). Sin embargo, esta interpretación es

adecuada si la función de verosimilitud perfil de  $\delta$  es simétrica y acampanada como una densidad normal.

A menudo se puede tratar e interpretar a la función de verosimilitud perfil como una verosimilitud genuina unidimensional. Esto se debe a que su expansión en series de Taylor tiene la misma forma de una verosimilitud genuina de un solo parámetro,

$$\begin{aligned} \log R_P(\delta; y) &= -\frac{1}{2}(\delta - \widehat{\delta})^2 I_{\widehat{\delta}} + \sum_{i=3}^{\infty} \frac{1}{i!} (\delta - \widehat{\delta})^i \frac{\partial^i \log R_P}{\partial \widehat{\delta}^i} \\ &= -\frac{1}{2} u_{\delta}^2 + \sum_{i=3}^{\infty} \frac{(-1)^i}{i!} F_i(\widehat{\delta}; y) u_{\delta}^i, \end{aligned}$$

que es una serie de potencias en la cantidad pivotal\* lineal  $u_{\delta} = (\delta - \widehat{\delta}) \sqrt{I_{\widehat{\delta}}}$  cuyos coeficientes están determinados por las “estadísticas de forma”

$$F_i(\widehat{\theta}; y) = \frac{\partial^i \log R_P}{\partial \widehat{\delta}^i} I_{\widehat{\delta}}^{-i/2}, i = 3, \dots$$

Así, las técnicas para la construcción de intervalos de verosimilitud-confianza y aproximaciones a la función de verosimilitud se pueden aplicar, de la misma forma, a las funciones verosimilitud perfil.

A continuación se presenta un ejemplo donde se ilustra una aplicación de la verosimilitud perfil para hacer inferencia sobre un parámetro de interés en presencia de otro parámetro considerado de estorbo y se compara con la verosimilitud condicional que en este ejemplo se puede calcular para el parámetro de interés.

### Ejemplo 1.7.1

Considérese el modelo Binomial Negativo descrito por Fisher (1941),

$$P(X = i; \theta, p) = \binom{\theta + i - 1}{i} p^i (1 - p)^{\theta}, \quad i = 0, \dots,$$

---

\* Un pivotal  $u_{\theta} = u(y; \theta)$  es una función de las observaciones y del parámetro cuya distribución no depende de parámetros desconocidos.

donde  $\theta > 0$  y  $0 < p < 1$ . Supóngase que la frecuencia observada de  $X = i$  es  $f_i$ . El problema consiste en estimar solamente el parámetro  $\theta$  cuando también  $p$  es desconocida. En este caso, se considera que  $p$  es un parámetro de estorbo o de ruido.

La función de verosimilitud global del vector de parámetros  $(\theta, p)$  es

$$\begin{aligned} L[\theta, p; \{f_i\}_{i \geq 0}] &\propto \prod_{i=0}^{\infty} [P(X = i; \theta, p)]^{f_i} \\ &= \prod_{i=0}^{\infty} \left[ \binom{\theta + i - 1}{i} p^i (1-p)^\theta \right]^{f_i} \\ &= p^t (1-p)^{n\theta} \prod_{i=0}^{\infty} \binom{\theta + i - 1}{i}^{f_i}, \end{aligned} \quad (1.14)$$

donde  $t = \sum i f_i$  y  $n = \sum f_i$  es el tamaño de la muestra. Como el emv restringido de  $p$  para cada valor especificado de  $\theta$  es  $\hat{p}(\theta) = t / (t + n\theta)$ , se tiene que la función de verosimilitud perfil de  $\theta$  es

$$L_P[\theta; \{f_i\}_{i \geq 0}] \propto \hat{p}(\theta)^t [1 - \hat{p}(\theta)]^{n\theta} \prod_{i=0}^{\infty} \binom{\theta + i - 1}{i}^{f_i}. \quad (1.15)$$

En este caso, también se pueden hacer inferencias condicionales sobre el parámetro  $\theta$  debido a la estructura de la función de verosimilitud. Obsérvese que  $t$  es una estadística suficiente para el parámetro de estorbo  $p$ . Por tanto, la verosimilitud se puede factorizar como en (1.6). Como  $t$  tiene una distribución Binomial Negativa,

$$P(T = t; \theta, p) = \binom{n\theta + t - 1}{t} p^t (1-p)^{n\theta}, \quad t = 0, \dots,$$

se tiene que la función de verosimilitud condicional de  $\theta$  es

$$L_C[\theta; \{f_i\}_{i \geq 0}] \propto \prod_{i=0}^{\infty} \binom{\theta + i - 1}{i}^{f_i} / \binom{n\theta + t - 1}{t}. \quad (1.16)$$

Como se comentó en la Sección 1.6.2, el uso de la función de verosimilitud condicional  $L_C[\theta; \{f_i\}_{i \geq 0}]$  en (1.16) depende de que la función de verosimilitud condicional residual  $L_{Cres}[\theta, p; \{f_i\}_{i \geq 0}] \propto P(T = t; \theta, p)$  contenga poca información acerca de  $\theta$

cuando se desconoce  $p$ . Una forma de valorarla es a través de la función de verosimilitud condicional residual perfil de  $\theta$  que se define como,

$$L_{Cresp} [\theta; \{f_i\}_{i \geq 0}] \propto L_{Cres} [\theta, \hat{p}(\theta); \{f_i\}_{i \geq 0}] \propto \binom{n\theta + t - 1}{t} \hat{p}(\theta)^t [1 - \hat{p}(\theta)]^{n\theta}, \quad (1.17)$$

donde  $\hat{p}(\theta)$  es el emvr de  $p$  para cada valor especificado de  $\theta$ . Si esta verosimilitud fuera completamente plana, es decir  $R_{Cresp}(\theta) = 1$  para todo  $\theta$ , entonces esto indicaría que el factor residual no contiene información alguna sobre  $\theta$ . Este es el caso de factores ortogonales como en el ejemplo de la diferencia, en escala logarítmica, de las medias de dos distribuciones Poisson presentado en Sprott (2000, pág. 50, Ejemplo 4.2.1). Cualquier desviación de esto se interpreta como información residual sobre  $\theta$ .

La siguiente tabla muestra datos citados por Fisher (1941) sobre la clasificación de ovejas según el número de garrapatas encontradas en cada una de ellas.

Tabla1.3. Datos de clasificación de ovejas.

# de garrapatas	0	1	2	3	4	5	6	7	8	9	10	Total
# de ovejas $f_i$	7	9	8	13	8	5	4	3	0	1	2	60

La Figura 1.2 en la pág. 28, muestra a la función de verosimilitud global relativa,  $R(\theta, p)$ , correspondiente a (1.14). Nótese que cuando uno se posiciona en un punto muy distante sobre el eje  $p$ , entonces la silueta o perfil que se observa de la superficie de  $R(\theta, p)$  es justamente la función de verosimilitud maximizada o perfil relativa de  $\theta$ . En la Figura 1.3 se muestran juntas la función de verosimilitud perfil relativa  $R_P(\theta)$  correspondiente a (1.15), la función de verosimilitud condicional relativa  $R_C(\theta)$ , correspondiente a (1.16) y la función de verosimilitud condicional residual perfil relativa  $R_{Cresp}(\theta)$  correspondiente a (1.17). Para este ejemplo de ovejas, se tiene que la verosimilitud perfil  $R_P(\theta)$  y la condicional  $R_C(\theta)$  no difieren mucho. Además, la verosimilitud condicional residual perfil de  $\theta$ ,  $R_{Cresp}(\theta)$ , es tan solo ligeramente curva sobre el rango de valores plausibles de  $\theta$  o equivalentemente no es informativa sobre  $\theta$ .

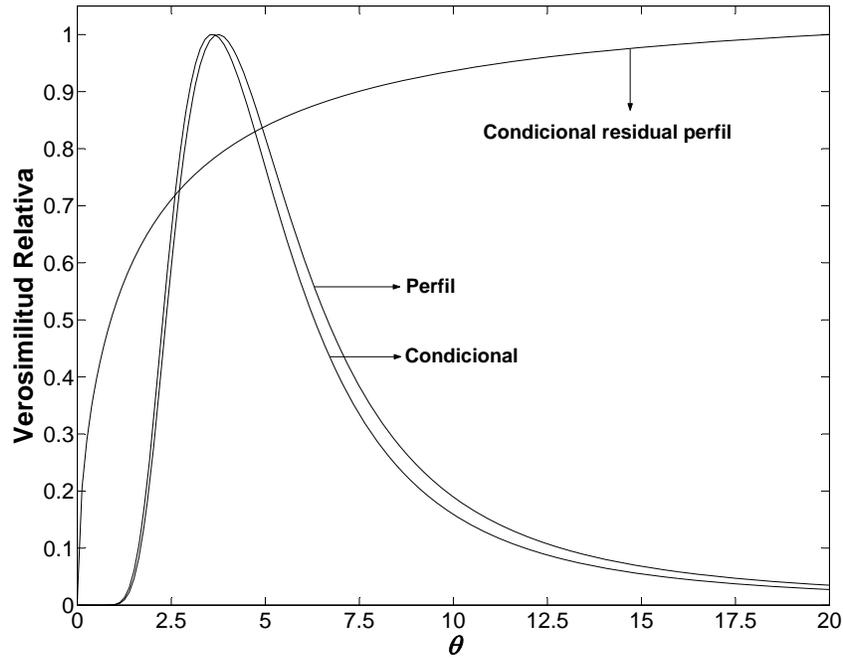


Figura 1.3: Función de verosimilitud perfil y condicional de  $\theta$

Cabe señalar aquí que en el caso de que la verosimilitud se pueda factorizar de manera marginal como en (1.7), entonces, de forma análoga al ejemplo anterior, la información marginal residual se puede valorar gráficamente a través de la función de verosimilitud marginal residual perfil.

La verosimilitud perfil es importante puesto que es un método simple más fácil de calcular que otras alternativas para estimar por separado un parámetro de interés en presencia de parámetros de estorbo. Además, para este propósito, muchas veces la verosimilitud perfil es la única alternativa puesto que otras formas de verosimilitud como la condicional, la marginal o la integrada, dependen de una estructura especial como en (1.6), (1.7) y (1.8), que no siempre se tiene. En esta tesis, la verosimilitud perfil es el tema central.

# Capítulo 2

## Críticas a la verosimilitud perfil

En este capítulo se analizarán algunas críticas que ha recibido la verosimilitud perfil y se esclarecerá cuáles de ellas son justificadas y cuáles no.

### 2.1 Críticas injustificadas a la verosimilitud perfil\*

A menudo se critica a la función de verosimilitud perfil definida en (1.11) porque arroja resultados extraños o poco intuitivos. Como se verá en los ejemplos descritos en esta sección, estos problemas generalmente se deben a que se está usando como función de verosimilitud a una función de densidad que tiene singularidades. Así, la función de verosimilitud perfil hereda estas singularidades y consecuentemente presenta problemas. Sin embargo, es importante recordar que la función de verosimilitud en (1.1) se define como proporcional a una probabilidad, de modo que no puede tener singularidades puesto que toda probabilidad está acotada por uno. Cuando este requisito se toma en cuenta, se libera por completo a la verosimilitud perfil de dichos problemas.

---

\*Montoya, Díaz-Francés y Sprott (2007)

### Ejemplo 2.1.1

Un ejemplo muy conocido es aquel en el que se considera una sola observación  $y$  de una distribución normal  $N(\lambda, \sigma^2)$ . El parámetro de interés es el de escala  $\sigma$  y el parámetro de estorbo es la media  $\lambda$ . La función de verosimilitud de  $(\sigma, \lambda)$  que se obtiene al usar la aproximación continua para  $Y = y$  es

$$L(\sigma, \lambda) \propto \frac{1}{\sigma} \exp \left[ -\frac{1}{2\sigma^2} (y - \lambda)^2 \right]. \quad (2.1)$$

Nótese que el emvr de  $\lambda$  para cada valor especificado de  $\sigma$  es  $\hat{\lambda} = \hat{\lambda}(\sigma) = y$ . Entonces, la función de verosimilitud perfil de  $\sigma$  que resulta de reemplazar  $\lambda$  por  $\hat{\lambda}(\sigma)$  en (2.1) es

$$L_P(\sigma; y) \propto \frac{1}{\sigma}.$$

Esta verosimilitud perfil es proporcional a  $\sigma^{-1}$ , y tiende a infinito cuando  $\sigma$  tiende a 0. Como la función de verosimilitud perfil relativa de  $\sigma$  es

$$R_P(\sigma; y) = \frac{L_P(\sigma; y)}{\sup_{\sigma} L_P(\sigma; y)},$$

y dado que como  $\sup_{\sigma} L_P(\sigma; y) = \infty$  se tiene que  $R_P(\sigma; y)$  no está definida.

En contraste, si consideramos que la observación  $y$  es en realidad discreta y que la precisión del instrumento de medición es  $h$ , un número positivo fijo, entonces usando (1.3) y (1.11) con  $\hat{\lambda}(\sigma) = y$ , la función de verosimilitud perfil de  $\sigma$  resulta ser

$$\begin{aligned} L_P(\sigma; y) &\propto \int_{y-\frac{1}{2}h}^{y+\frac{1}{2}h} \frac{1}{\sigma} \exp \left[ -\frac{1}{2\sigma^2} (t - y)^2 \right] dt \\ &= \int_{-\frac{2\sigma}{h}}^{\frac{2\sigma}{h}} \frac{h}{2\sigma} \exp \left( -\frac{1}{2} z^2 \right) dz. \end{aligned}$$

Ahora, por el Teorema de Convergencia Dominada de Lebesgue (véase Dudley, 1989,

pág. 101), tenemos que

$$\lim_{\sigma \rightarrow 0^+} \int_{-\frac{2\sigma}{h}}^{\frac{h}{2\sigma}} \exp\left(-\frac{1}{2}z^2\right) dz = \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}z^2\right) dz = \sqrt{2\pi}.$$

Entonces, definiendo  $L_P(\sigma = 0, y) = \sqrt{2\pi}$  se elimina la discontinuidad de  $L_P(\sigma, y)$  en cero y así  $L_P(\sigma, y)$  es ahora una función continua. Por lo tanto, la función de verosimilitud perfil relativa de  $\sigma$  es

$$R_P(\sigma; y) = \frac{1}{\sqrt{2\pi}} \int_{-\frac{df_{racch}2\sigma}{h}}^{\frac{h}{2\sigma}} \exp\left(-\frac{1}{2}z^2\right) dz, \quad (2.2)$$

que es una función de verosimilitud continua que no posee problemas de singularidad. Quizás este ejemplo parezca muy artificial, pero es similar al caso cuando se tienen mezclas no agrupadas de distribuciones normales, lo cual sí se da seguido en la práctica. Nótese que en este caso la función de densidad tiene dos singularidades en cada observación.

### Ejemplo 2.1.2

A manera de un ejemplo más reciente, Berger, Liseo y Wolpert (1999) consideran una muestra de variables aleatorias normales  $N(\theta, 1)$  independientes e idénticamente distribuidas  $X_1, \dots, X_n$  mientras que  $Y$  es otra variable aleatoria normal  $N(\lambda, \sigma_\theta^2)$  independiente, donde  $\sigma_\theta^2 = \exp(-n\theta^2) \leq 1$ . El parámetro de interés es  $\theta$  y el parámetro de estorbo es  $\lambda$ . Berger *et al.* (1999) usan este ejemplo para criticar a la función de verosimilitud perfil y sugerir que la función de verosimilitud integrada es una mejor opción para estimar por separado a  $\theta$  en presencia del parámetro de estorbo  $\lambda$ . Este ejemplo es un poco más general, aunque más artificial, que la versión del Ejemplo 2.1.1 puesto que la distribución de  $Y$  es normal.

Como en el Ejemplo 2.1.1, cuando se usa la aproximación continua tenemos que  $\hat{\lambda}(\sigma_\theta) = y$ ; así la única observación  $Y = y$  contribuye a la función de verosimilitud de

$\theta$  basada en las  $X_i$ 's con el factor  $1/\sigma_\theta$ . Entonces la función de verosimilitud “perfil” de  $\theta$  es proporcional a

$$\exp \left[ -\frac{1}{2} \sum_{i=1}^n (x_i - \theta)^2 \right] \frac{1}{\sigma_\theta} \propto \exp(n\bar{x}\theta) \rightarrow \infty, \text{ cuando } \theta \rightarrow \pm\infty,$$

dependiendo del signo de  $\bar{x} = 1/n \sum_{i=1}^n x_i$ . Berger *et al.* (1999) comentan que esta verosimilitud es una verosimilitud perfil muy extraña porque crece rápido a infinito cuando  $\theta$  tiende a  $+\infty$  o  $-\infty$ , dependiendo del signo de  $\bar{x}$  (“*Note that this is a very strange ‘likelihood,’ rapidly growing to infinity as  $\theta \rightarrow \infty$  or  $\theta \rightarrow -\infty$ , depending on the sign of  $\bar{x}$ .*”).

Sin embargo, si se considera la precisión finita del instrumento de medición, la función de verosimilitud perfil de  $\theta$  que resulta de usar (2.2) es

$$L_P(\theta; x_1, \dots, x_n, y) \propto \exp(n\bar{x}\theta)\sigma_\theta \int_{-\frac{2\sigma_\theta}{h}}^{\frac{h}{2\sigma_\theta}} \exp\left(-\frac{1}{2}z^2\right) dz, \quad (2.3)$$

donde  $\sigma_\theta = \exp(-\frac{1}{2}n\theta^2) \leq 1$ , por suposición inicial.

Por otro lado, Berger *et al.* (1999) eliminan al parámetro de estorbo  $\lambda$  integrando con respecto a  $\lambda$  la función de densidad de  $Y$ ,  $N(\lambda, \sigma_\theta^2)$ , obteniendo la función de verosimilitud integrada uniforme de  $\theta$ ,

$$L_I^U(\theta) \propto \exp \left[ -\frac{1}{2}n(\bar{x} - \theta)^2 \right] \propto \exp(n\bar{x}\theta)\sigma_\theta. \quad (2.4)$$

En esencia, esto implica suponer una distribución inicial uniforme impropia para  $\lambda$ ,  $-\infty \leq \lambda \leq \infty$ . Nótese que el mismo resultado (2.4) se obtendría si se integrara la densidad conjunta con respecto a la variable aleatoria  $Y$ , de manera que (2.4) se puede ver también como una verosimilitud marginal. Para el caso  $n = 1$ ,  $x = 1$ , y  $y = 0$ , la Figura 2.1 muestra algunos ejemplos de cómo cambios en la precisión  $h$  del instrumento de medición afectan a las funciones de verosimilitud perfil relativa de  $\theta$ .

De hecho, en este ejemplo la función de verosimilitud perfil relativa  $L_P$  en (2.3) incluye a la verosimilitud integrada en (2.4) como un caso especial cuando  $h$  es grande,

indicando la carencia de precisión en la medición de  $Y$  con respecto a la variación de  $Y$  determinada por  $\sigma_\theta \leq 1$ . La Figura 2.1 muestra como, con  $h = 5$ , la verosimilitud perfil  $L_P$  en (2.3) es prácticamente indistinguible de la verosimilitud integrada uniforme  $L_I^U$  en (2.4).

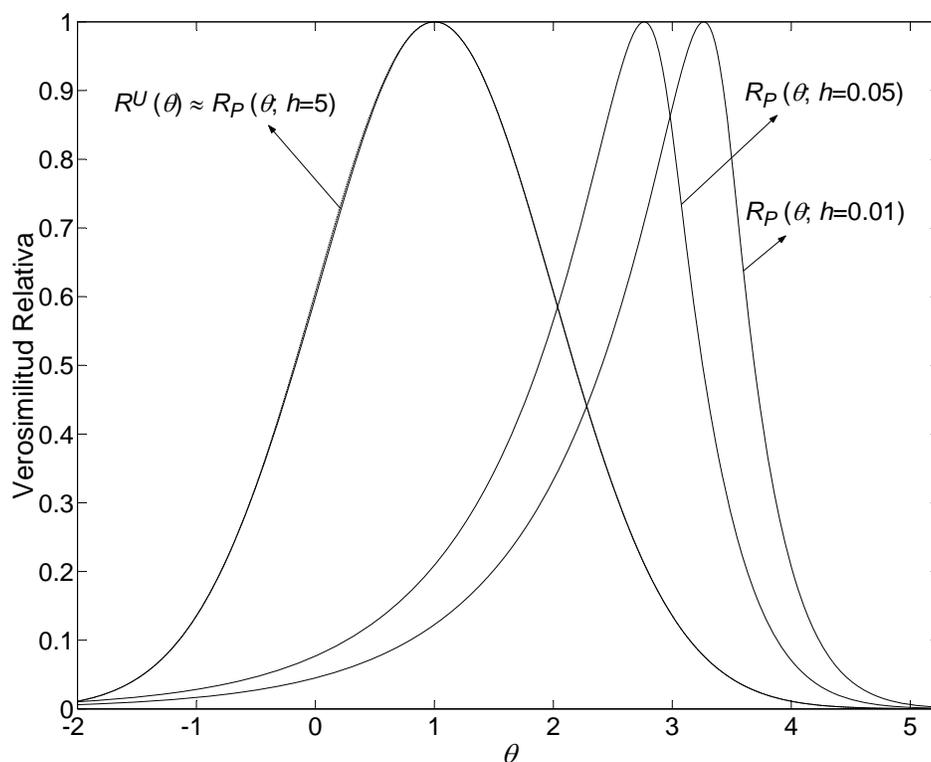


Figura 2.1: Sensibilidad de  $R_P(\theta; x, y)$  bajo cambios en la precisión del instrumento de medición.

### Ejemplo 2.1.3

En este ejemplo se considera el caso de un parámetro umbral  $\alpha$  que restringe el soporte de la variable aleatoria  $Y$ , de tal manera que  $\alpha$  puede ser una cota superior o inferior para  $Y$ . En contraste a los ejemplos previos, las observaciones determinan al rango de  $\alpha$  en la función de verosimilitud correspondiente, por lo que no se tiene un caso regular.

Esta situación ocurre con frecuencia en Confiabilidad y Valores Extremos; por

ejemplo, con las distribuciones Weibull y Fréchet de tres parámetros. Uno de dichos parámetros es una cota superior o inferior para la variable aleatoria. En ocasiones se llama vida garantizada cuando se trata de una cota inferior para la variable aleatoria. A manera de ejemplo, aquí se presenta el caso de la distribución lognormal de tres parámetros de  $Y$ ,

$$\log(Y - \alpha) = X(\alpha) \sim N(\theta, \sigma^2), \quad 0 \leq \alpha < Y.$$

La función de verosimilitud de  $(\alpha, \theta, \sigma^2)$  que se obtiene al usar la aproximación continua para una muestra observada de variables aleatorias lognormales independientes e idénticamente distribuidas  $y_1, \dots, y_n$  es

$$L(\alpha, \theta, \sigma^2; y_1, \dots, y_n) \propto \prod_{i=1}^n \frac{1}{\sigma} \frac{1}{(y_i - \alpha)} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n [x_i(\alpha) - \theta]^2 \right\}.$$

Entonces, los estimadores de máxima verosimilitud restringidos de  $\theta$  y  $\sigma^2$  para un  $\alpha$  dado son

$$\hat{\theta}(\alpha) = \bar{x}(\alpha) \quad \text{y} \quad \hat{\sigma}^2(\alpha) = \sum_{i=1}^n [x_i(\alpha) - \bar{x}(\alpha)]^2 / n.$$

Así, la función de verosimilitud perfil de  $\alpha$  que resulta es

$$\begin{aligned} L_P(\alpha; y_1, \dots, y_n) &= L(\alpha, \hat{\theta}(\alpha), \hat{\sigma}(\alpha); y_1, \dots, y_n) \\ &\propto \left[ \frac{1}{\hat{\sigma}^2(\alpha)} \right]^{df_{racn2}} \exp \left\{ -\frac{1}{2\hat{\sigma}^2(\alpha)} \sum_{i=1}^n [x_i(\alpha) - \hat{\theta}(\alpha)]^2 \right\} \prod_{i=1}^n \frac{1}{y_i - \alpha} \\ &\propto \left[ \frac{1}{\hat{\sigma}^2(\alpha)} \right]^{df_{racn2}} \prod_{i=1}^n \frac{1}{y_i - \alpha}, \quad 0 \leq \alpha < y_{(1)}, \end{aligned} \quad (2.5)$$

donde  $y_{(1)}$  es la observación más pequeña en la muestra. Obsérvese que la función de verosimilitud perfil tiene una singularidad en  $\alpha = y_{(1)}$ , heredada de la correspondiente singularidad de la función de densidad conjunta.

Como mencionan Meeker y Escobar (1998, pág. 275), existe una trayectoria en el espacio parametral para la cual la verosimilitud tiende a infinito, en particular cuando  $\sigma \rightarrow 0$  y  $\alpha \rightarrow y_{(1)}$ . Cabe resaltar que si la verosimilitud se aproxima a infinito esto no

significa necesariamente que la probabilidad de la muestra observada sea grande en esa región del espacio parametral, sino que esto puede ocurrir a causa de que la aproximación de la densidad a la función de verosimilitud sea muy mala. A menudo existe, aunque no siempre necesariamente, un máximo local para la superficie de verosimilitud correspondiente que suele coincidir con el máximo de la verosimilitud correcta que se definió en (1.3) y que se basa en la probabilidad de los datos.

Entonces, una solución posible a este problema puede ser, como en los dos ejemplos anteriores, usar las contribuciones correctas a la verosimilitud de cada observación como en (1.3), implícitas por la precisión finita del instrumento de medición. Sin embargo, en este ejemplo, se puede usar un enfoque algo diferente y más simple. Hay que notar que la precisión finita  $h$  del instrumento de medición también limita al rango del parámetro umbral tal que  $\alpha < y_{(1)} - h$ , Lawless (2003, pág. 186), Barnard (1966). Así, se tiene una manera razonable para incorporar la precisión  $h$ , dado que necesariamente la observación más pequeña pudo haber sido en realidad  $y_{(1)} - h$ , valor tal que debería ser más grande que  $\alpha$ . Es decir, considérese que  $\alpha < y_{(1)} - h$  y con esta restricción, ya se puede usar la aproximación continua a la función de verosimilitud. El valor de  $h$  depende de la precisión del instrumento de medición.

Para ilustrar este otro enfoque se simularon diez observaciones usando  $\alpha = 0$ ,  $\theta = 2$ ,  $\sigma = 1$  mostradas a continuación en orden ascendente de magnitud,

$$y = 3.33, 7.15, 7.66, 9.18, 9.25, 10.57, \\ 10.88, 14.01, 15.95, 51.40.$$

Aquí se supone que la precisión del instrumento de medición es  $h = 0.01$ . En la Figura 2.2 se muestra la función de verosimilitud perfil relativa de  $\alpha$  que se obtiene a partir de estos datos. Nótese que  $L_P(\alpha)$  no está definida en  $\alpha = y_{(1)}$  y  $L_P(\alpha)$  tiende a infinito cuando  $\alpha$  se aproxima por la izquierda a  $y_{(1)}$ . Para esta nueva verosimilitud restringida, existe un máximo en  $\hat{\alpha} = 1.65$ , el estimador de máxima verosimilitud;

nuevamente todos los otros valores de (2.5) se estandarizan para obtener la función de verosimilitud perfil relativa  $R_P(\alpha)$ .

Como se muestra en la gráfica, existe una singularidad en  $\alpha = y_{(1)} = 3.33$ , puesto que  $L_P(\alpha) \rightarrow \infty$  cuando  $\alpha \uparrow y_{(1)}$ . Sin embargo, nótese que  $R_P(\alpha = 3.325) = 0.0585$ . Incluso si uno se hubiera acercado mucho más a la singularidad en  $y_{(1)}$ , digamos a  $(y_{(1)} - 0.0005 = 3.3295)$  esto no ocasiona problemas cuando se usa la aproximación continua a la función de verosimilitud porque  $R_P(\alpha = 3.3295) = 0.0467$  es todavía una cantidad pequeña. Es decir, el valor de  $\alpha = 3.3295$  está aún muy lejano del valor  $y_{(1)}$  donde  $L_P(\alpha)$  tiene la singularidad. Por tanto, basta con acotar a  $\alpha$  por debajo de  $y_{(1)}$  por una distancia pequeña  $h$  para poder usar la aproximación continua a la verosimilitud.

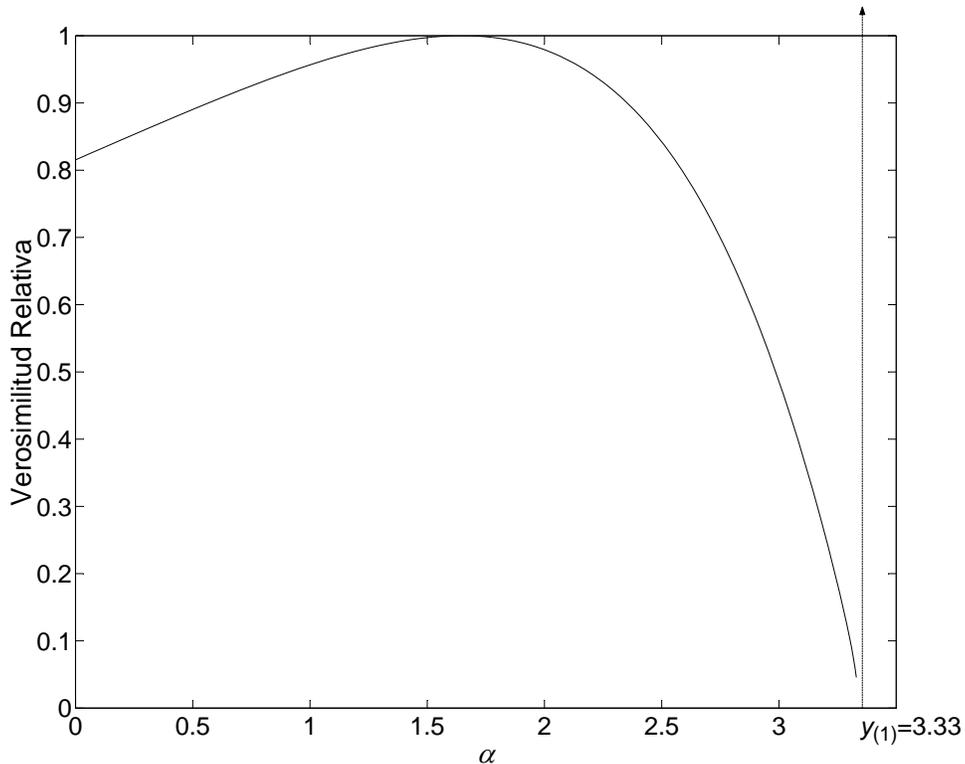


Figura 2.2: Máximo local versus global de  $R_P(\alpha; y)$ .

## 2.2 Críticas justificadas a la verosimilitud perfil

En la Sección 2.1 se mostraron tres ejemplos donde se critica a la función de verosimilitud perfil por proporcionar resultados anómalos. Sin embargo, como se vio en esa sección, dichas críticas eran injustas puesto que el problema lo ocasionaba el uso de funciones de densidad con singularidades para aproximar a la función de verosimilitud. En contraste, en esta sección se presentan ejemplos que ilustran ciertas desventajas de la función de verosimilitud perfil que han dado pie a diversas críticas, ahora sí justificadas genuinamente. A diferencia de la sección anterior, estas críticas no provienen de un error matemático en el proceso de aproximación a la verosimilitud, sino que provienen de que la función de verosimilitud perfil aparenta un exceso de precisión que no es realista para la muestra observada; esto puede ocurrir cuando el modelo estadístico involucra un número grande de parámetros de estorbo. A continuación se dan dos ejemplos de esta circunstancia.

### Ejemplo 2.2.1

Supóngase que  $X_1, \dots, X_n$  son variables aleatorias normales  $N(\mu, \sigma^2)$  independientes e idénticamente distribuidas. El parámetro de interés es el de escala  $\sigma$  y el parámetro de estorbo es la media  $\mu$ . Se conoce muy bien que el emvr de  $\mu$  para cualquier valor fijo de  $\sigma$  es la media muestral,  $\hat{\mu} = \bar{x}$ , y es importante resaltar que no depende de  $\sigma$ . Entonces usando (1.11) con  $\hat{\mu}(\sigma) = \hat{\mu} = \bar{x}$ , la función de verosimilitud perfil de  $\sigma$  resulta ser

$$L_P(\sigma; x_1, \dots, x_n) \propto \frac{1}{\sigma^n} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n [x_i - \hat{\mu}(\sigma)]^2 \right\} = \frac{1}{\sigma^n} \exp \left[ -\frac{n(n-1)s^2}{2\sigma^2} \right], \quad (2.6)$$

donde  $s^2 = \sum_{i=1}^n (x_i - \bar{x})^2 / n(n-1)$ . Obsérvese que  $\hat{\sigma}^2 = (n-1)s^2$  es el emv de  $\sigma^2$ . Como  $n(n-1)s^2/\sigma^2$  tiene una distribución Ji-cuadrada con  $(n-1)$  grados de libertad, se tiene que  $E(\hat{\sigma}^2) = (n-1)\sigma^2/n$ . Por tanto, el emv  $\hat{\sigma}^2$ , si bien es un estimador consistente para  $\sigma^2$ , presenta un sesgo al estimar  $\sigma^2$  y dicho sesgo es

$|E(\hat{\sigma}^2) - \sigma^2| = \sigma^2/n$ . Nótese que este sesgo tiende a cero al aumentar el tamaño de muestra. Así, el sesgo de  $\hat{\sigma}^2$  es de poca importancia cuando el tamaño de la muestra es moderado. Sin embargo, cuando  $n$  es pequeña este sesgo en la estimación de  $\sigma^2$  puede ser muy notorio. Por ejemplo, si  $n = 2$  entonces  $E(\hat{\sigma}^2) = \sigma^2/2$ . El problema puede convertirse en algo más serio si se combinan muchas de éstas verosimilitudes perfil, como en el problema clásico de medias distintas con varianza común que se describe en el Ejemplo 2.2.2. Allí el sesgo se acumula al punto de que el mismo estimador de  $\sigma^2$  se vuelve inconsistente.

Es importante recordar aquí que en este ejemplo, la manera adecuada para hacer inferencia sobre el parámetro de interés  $\sigma$  en presencia del parámetro de estorbo  $\mu$  es a través de la función de verosimilitud marginal de  $\sigma$  definida en (1.7). Como  $\bar{x}$  y  $s^2$  son independientes y conjuntamente suficientes para el vector de parámetros  $(\mu, \sigma^2)$ , y considerando que  $n(n-1)s^2/\sigma^2$  tiene una distribución Ji-cuadrada con  $(n-1)$  grados de libertad, la función de verosimilitud marginal de  $\sigma$  es

$$L_M(\sigma; s) \propto \frac{1}{\sigma^{n-1}} \exp\left[-\frac{n(n-1)s^2}{2\sigma^2}\right]. \quad (2.7)$$

Así, el emv marginal de  $\sigma^2$  es  $\hat{\sigma}_M^2 = n\hat{\sigma}^2/(n-1)$ ; y resulta ser un estimador consistente y además insesgado para  $\sigma^2$  puesto que  $E(\hat{\sigma}_M^2) = \sigma^2$ .

### Ejemplo 2.2.2

Supóngase que se tienen  $k$  muestras de tamaños  $n_i$ , con  $i = 1, \dots, k$ , de  $k$  distribuciones normales independientes con diferentes medias  $\mu_i$ , pero con varianza común  $\sigma^2$ . Considérese el caso más extremo que se tiene cuando  $n_i = 2$ , es decir se tienen  $k$  muestras normales  $N(\mu_i, \sigma^2)$  de observaciones apareadas  $(x_1, y_1), \dots, (x_k, y_k)$ . Como en el ejemplo anterior, el emvr de  $\mu_i$  para cualquier valor especificado de  $\sigma$  es la media muestral,  $\hat{\mu}_i = (x_i + y_i)/2$ , que no depende de  $\sigma$ . Entonces usando (1.11) con

$\hat{\mu}_i(\sigma) = \hat{\mu}_i = (x_i + y_i)/2$ , la función de verosimilitud perfil de  $\sigma$  resulta ser

$$L_P \left[ \sigma; (x_i, y_i)_{i=1}^k \right] \propto \sigma^{-2k} \exp \left[ -\frac{1}{4\sigma^2} \sum_{i=1}^k (x_i - y_i)^2 \right]. \quad (2.8)$$

Obsérvese que  $\hat{\sigma}^2 = \sum_{i=1}^k (x_i - y_i)^2 / 4k$  es el emv de  $\sigma^2$ . Además, como  $(x_i - y_i)/\sqrt{2}\sigma$  tiene una distribución normal con media 0 y varianza 1 entonces  $E(\hat{\sigma}^2) = \sigma^2/2$ . Así, cuando  $k \rightarrow \infty$ , por la Ley Fuerte de los Grandes Números tenemos que  $\hat{\sigma}^2 \rightarrow \sigma^2/2$ ; por tanto, el emv  $\hat{\sigma}^2$ , no es en este caso, un estimador consistente para  $\sigma^2$ .

Igual que en el ejemplo anterior, la manera correcta para hacer inferencia sobre el parámetro de interés  $\sigma$  en presencia de los parámetros  $\mu_i$  de estorbo es usando la verosimilitud marginal. Entonces, la verosimilitud marginal de  $\sigma$  que resulta de las  $k$  muestras apareadas  $(x_i, y_i)$ ,  $i = 1, \dots, k$ , que tienen distribución normal  $N(\mu_i, \sigma^2)$  con diferentes medias pero varianza común, es el producto de las funciones de verosimilitud marginales individuales,

$$L_M \left[ \sigma; (s_i)_{i=1}^k \right] \propto \sigma^{-k} \exp \left[ -\frac{1}{4\sigma^2} \sum_{i=1}^k (x_i - y_i)^2 \right]. \quad (2.9)$$

Así, el emv marginal de  $\sigma^2$  es  $\hat{\sigma}_M^2 = 2\hat{\sigma}^2$ , el cual resulta ser como en el ejemplo anterior, un estimador consistente y además insesgado de  $\sigma^2$  puesto que  $E(\hat{\sigma}_M^2) = \sigma^2$ .

## 2.3 Discusión

Las funciones de verosimilitud son proporcionales a probabilidades y por tanto no pueden tener singularidades. De aquí que las críticas a la función de verosimilitud perfil que se originen de singularidades no tienen validez alguna y no pueden usarse como argumentos a favor del uso de verosimilitudes marginales e integradas. Sin embargo, hay que tener un poco de precaución al usar la función de verosimilitud perfil cuando se tengan problemas que involucren a muchos parámetros de estorbo. En estos casos, la función de verosimilitud perfil puede aparentar contar con mucha precisión en las

inferencias que no es realmente sustentada por la muestra observada, como se vio en los Ejemplos 2.2.1 y 2.2.2 de este capítulo. Es necesario mencionar que otras verosimilitudes tales como marginales, condicionales, estimadas e integradas también podrían presentar el mismo problema. Además, antes de usar verosimilitudes condicionales o marginales es necesario asegurarse de que no pierdan mucha información en su factor residual como se mencionó en las Secciones 1.7.1 y 1.7.2.

La función de verosimilitud perfil o maximizada es quizás la aplicación más común y simple en la teoría de estimación de máxima verosimilitud. A la perfil se le suele encontrar donde menos se le espera. Por ejemplo, aparece como la base de la estadística de prueba de la razón de las verosimilitudes generalizada. Como se vio en la Sección 1.7.6, la función de verosimilitud perfil se puede considerar y usar como una genuina verosimilitud que surge de modelos con un solo parámetro unidimensional.

Una ventaja adicional de la verosimilitud perfil es que, al igual que en la teoría de estimación de máxima verosimilitud, no pide ninguna estructura específica a la distribución de la variable aleatoria en consideración. De aquí que la verosimilitud perfil sea fácil de utilizar en general. Sin embargo, siempre conviene tomar ciertas precauciones como las que se describieron en este capítulo cuando se use la verosimilitud perfil.

## Capítulo 3

# Robusticidad de la función de verosimilitud perfil

El término “inferencia robusta” lo usaron por primera vez Box & Tiao (1973, pág. 152) para describir el grado en el cual las inferencias dependen de los supuestos específicos que se hagan. En esta tesis se adoptará este término con el mismo significado. Por ejemplo, Barnard (1983), Barnard y Sprott (1983), Lindsey (1999) y Sprott (2000, Secciones 7.6, 7.10) han también usado el término robusticidad en este mismo sentido.

La carencia de robusticidad sugiere inestabilidad en el sentido que cambios pequeños en las suposiciones o en los datos producen cambios grandes en las inferencias. Este concepto de robusticidad es muy general. En este capítulo se discute una situación particular que es, la robusticidad de la función de verosimilitud perfil para realizar inferencias sobre un parámetro de interés bajo cambios pequeños en los estimadores de máxima verosimilitud (emv) restringidos de los parámetros de estorbo.

Para explorar este tipo de robusticidad de la función de verosimilitud perfil se define aquí la que se llamará *función de verosimilitud perfil  $\epsilon$ -perturbada*. Se mostrará como esta función resulta ser una herramienta muy útil para evaluar si se está en el caso de situaciones poco robustas que generalmente son problemáticas pero que pueden

representar retos interesantes de estimación. Se ejemplifica el uso de esta herramienta a través de ejemplos de datos reales, robustos y no robustos.

### 3.1 El concepto de verosimilitud perfil $\epsilon$ -perturbada

En esta sección primero se definirá, para mayor claridad, a la función de verosimilitud perfil  $\epsilon$ -perturbada para el caso simple de un vector de dos parámetros escalares  $(\delta, \lambda) \in \Delta \times \Lambda \subseteq \mathbb{R}^2$ , donde  $\delta$  es el parámetro de interés y  $\lambda$  es considerado un parámetro de estorbo. Posteriormente se definirá a la función de verosimilitud perfil  $\epsilon$ -perturbada para el caso en el cual la dimensión de  $\delta$  y  $\lambda$  es mayor que uno.

La función de verosimilitud perfil de  $\delta$  pone mucho énfasis en la familia de estimadores puntuales  $\hat{\lambda}(\delta)$  indicados por  $\delta$ . Para cada valor fijo de  $\delta$  la estimación del parámetro de estorbo  $\lambda$  no toma en cuenta ningún grado de incertidumbre, sino que adopta exclusivamente el valor máximo verosimil restringido  $\hat{\lambda}(\delta)$ . Para corregir esto, aquí se propone introducir una cantidad  $\epsilon \in \mathbb{R}$ , que representa un pequeño grado de incertidumbre en la estimación del parámetro de estorbo  $\lambda$ . Esta cantidad  $\epsilon$  se mide en unidades del parámetro de estorbo. Una forma simple y natural de considerar la incertidumbre en la estimación de  $\lambda$  es reemplazar  $\hat{\lambda}(\delta)$  por  $[\hat{\lambda}(\delta) + \epsilon]$  al calcular la perfil  $L_P(\delta)$ . Así, para un valor fijo de  $\delta$ , se deben considerar los valores de esta función evaluada en  $[\delta, \hat{\lambda}(\delta) + \epsilon]$  para diversos valores pequeños de  $\epsilon$ ; cuando  $\epsilon = 0$  se tiene la verosimilitud perfil. Geométricamente esto significa que no sólo se debe considerar la cima de la función de verosimilitud conjunta, que está determinada por la verosimilitud perfil, sino que también se debe tomar en cuenta otras trayectorias de la superficie de verosimilitud, ligeramente abajo de la cima.

**Definición 1: Caso de parámetros de interés y de estorbo escalares.**

Sea  $x = (x_1, \dots, x_n)$  una muestra de variables aleatorias cuya distribución depende de un vector  $(\delta, \lambda)$  de dos parámetros escalares. El espacio parametral correspondiente es  $\Delta \times \Lambda \subseteq \mathbb{R}^2$ . El parámetro de interés es  $\delta$  y el parámetro de estorbo es  $\lambda$ . En términos de la función de verosimilitud global de  $\delta$  y  $\lambda$ ,  $L(\delta, \lambda; x)$ , la función de verosimilitud perfil  $\epsilon$ -perturbada de  $\delta$ ,  $L_{\epsilon P}(\delta; x)$ , se define como

$$L_{\epsilon P}(\delta; x) \propto L[\delta, \widehat{\lambda}(\delta) + \epsilon; x], \quad (3.1)$$

donde  $\epsilon$  es una cantidad real pequeña tal que  $[\widehat{\lambda}(\delta) + \epsilon] \in \Lambda$ .

Nótese que valores negativos de  $\epsilon$  corresponden a perturbaciones del emv restringido en dirección opuesta a la cima que la perturbación correspondiente a  $\epsilon$  positivos. Obsérvese que este procedimiento produce una familia de funciones de verosimilitud perfil  $\epsilon$ -perturbadas, donde la función de verosimilitud perfil original es simplemente el miembro de esta familia cuando  $\epsilon = 0$ . En particular, esto permite tomar en cuenta la sensibilidad o robusticidad de la función de verosimilitud perfil para cambios pequeños de  $\epsilon$ .

Cuando las inferencias obtenidas con la función de verosimilitud perfil  $\epsilon$ -perturbada con valores pequeños de  $\epsilon$  no difieren mucho de las inferencias obtenidas con la verosimilitud perfil entonces se considera que se trata de un caso robusto. Cuando esto ocurre se dice que la verosimilitud perfil es robusta o estable frente a cambios pequeños en el emv restringido del parámetro de estorbo. En contraste, cuando las inferencias obtenidas con la verosimilitud perfil  $\epsilon$ -perturbada difieren mucho de las inferencias obtenidas con la verosimilitud perfil entonces se considera que se trata de un caso no robusto. En esta situación se dice que la verosimilitud perfil es no robusta o inestable frente a cambios pequeños en el emv restringido del parámetro de estorbo.

Por el principio de parsimonia, en los casos robustos basta con considerar a la función de verosimilitud perfil para hacer inferencias sobre el parámetro de interés.

Por otro lado, cuando se tengan casos no robustos se recomienda ser cauteloso con el uso de la verosimilitud perfil para hacer inferencia sobre el parámetro de interés y de preferencia revisar de cerca el comportamiento de la superficie de verosimilitud global cerca de la cima y tratar de explicarlo.

Un problema es cómo seleccionar uno o más  $\epsilon$  de una manera objetiva. Una forma natural de conseguir información sobre  $\epsilon$  basados en una muestra observada es a través de la función verosimilitud perfil relativa de  $\epsilon$  definida como

$$R(\epsilon) = \sup_{\delta|\epsilon} \frac{L[\delta, \widehat{\lambda}(\delta) + \epsilon; x]}{L(\widehat{\delta}, \widehat{\lambda}; x)}, \quad (3.2)$$

donde los valores de  $\epsilon$  cumplen que  $[\widehat{\lambda}(\delta) + \epsilon] \in \Lambda$ . Nótese que  $(\widehat{\delta}, \widehat{\lambda})$  es el emv global de  $(\delta, \lambda)$  y que  $0 \leq R(\epsilon) \leq 1$ . Obsérvese que cuando  $\epsilon = 0$  entonces  $R(\epsilon) = 1$ ; es decir el cero es el valor más plausible para  $\epsilon$ , a la luz de la muestra observada, y la correspondiente función de verosimilitud perfil  $\epsilon$ -perturbada es, en este caso, la función de verosimilitud perfil.

La propuesta para seleccionar  $\epsilon$  consiste en considerar valores de  $\epsilon$  que tengan plausibilidad alta y que hagan a la muestra observada altamente probable. Esto equivale a moverse del borde más alto de la superficie de verosimilitud determinado por la función de verosimilitud perfil una distancia pequeña. Por ejemplo, se puede tomar  $\epsilon$  que cumpla con  $R(\epsilon) = c$ , donde  $c \in \{0.99, 0.95, 0.90\}$ . También se puede considerar a la  $\epsilon$  que corresponda a un  $c$  un poco más extremo como 0.80, 0.70, pero no menor pues implicaría un a cambio muy lejano y brusco o distante de la cima de la verosimilitud global. Esto es porque la idea de la verosimilitud perfil  $\epsilon$ -perturbada es moverse sólo un poco de la trayectoria más alta en la superficie de verosimilitud.

Se recomienda graficar la función verosimilitud perfil relativa de  $\epsilon$ ,  $R(\epsilon)$  dada en (3.2), y señalar en la misma gráfica los valores de  $\epsilon$  que cumplen con  $R(\epsilon) = c$ , donde  $c \in \{0.99, 0.95, 0.90, 0.80, 0.70\}$ . Si  $R(\epsilon)$  es simétrica alrededor del cero, entonces el tamaño de la perturbación hacia la derecha y hacia la izquierda de  $\widehat{\lambda}(\delta)$  asociado

a una misma plausibilidad será igual. En contraste, si  $R(\epsilon)$  es asimétrica entonces el tamaño de la perturbación hacia la derecha y hacia la izquierda de  $\widehat{\lambda}(\delta)$  será diferente. Si  $R(\epsilon)$  es asimétrica hacia la derecha, entonces esto se puede interpretar como que la incertidumbre en la estimación del parámetro de estorbo  $\lambda$  a través de  $\widehat{\lambda}(\delta)$  es mayor en la dirección asociada a la derecha que hacia la izquierda. De forma análoga, si  $R(\epsilon)$  es asimétrica hacia la izquierda. Como segundo paso se recomienda poner en la misma gráfica las funciones de verosimilitud perfil  $\epsilon$ -perturbadas correspondientes a todos los  $\epsilon$  elegidos a partir de la gráfica descrita en el párrafo anterior. Es decir, para valorar la sensibilidad o robusticidad de la verosimilitud perfil es recomendable poner una misma figura a todas las curvas de verosimilitud perfil  $\epsilon$ -perturbadas correspondientes a los valores de  $\epsilon$  calculados anteriormente que cumplan con que  $R(\epsilon) = c$ , donde  $c \in \{1, 0.99, 0.95, 0.90, 0.80, 0.70\}$ . Nuevamente, obsérvese que  $c = 1$  esta asociado a  $\epsilon = 0$ , y corresponde a la función de verosimilitud perfil.

Cuando no exista un cambio sustancial en las formas de estas curvas con respecto a la de la verosimilitud perfil, entonces se dirá que la función de verosimilitud perfil es robusta frente a cambios pequeños en el emv restringido del parámetro de estorbo. En este caso, es recomendable el uso de la perfil para hacer inferencia sobre el parámetro de interés. Por otro lado, si las curvas de verosimilitud perfil  $\epsilon$ -perturbadas son muy distintas con respecto a la verosimilitud perfil, entonces la función de verosimilitud perfil no es robusta y se recomienda ser cauteloso con el uso de ella para inferencias sobre el parámetro de interés.

Ahora se va presentar la definición general de la función propuesta cuando los parámetros son de dimensiones mayores.

**Definición 2: La verosimilitud perfil  $\epsilon$ -perturbada**

Sea  $x = (x_1, \dots, x_n)$  una muestra de variables aleatorias cuya distribución depende de un vector  $(\delta, \lambda)$  de parámetros. El espacio parametral correspondiente es  $\Delta \times \Lambda \subseteq \mathbb{R}^v \times \mathbb{R}^w$ .

Es decir,  $\delta = (\delta_1, \dots, \delta_v)$  es un vector  $v$ -dimensional,  $\delta \in \Delta \subseteq \mathbb{R}^v$ , y  $\lambda = (\lambda_1, \dots, \lambda_w)$  es un vector  $w$ -dimensional,  $\lambda \in \Lambda \subseteq \mathbb{R}^w$ . El parámetro de interés es  $\delta$  y el parámetro de estorbo es  $\lambda$ . En términos de la función de verosimilitud global de  $\delta$  y  $\lambda$ ,  $L(\delta, \lambda; x)$ , la función de verosimilitud perfil  $\epsilon$ -perturbada de  $\delta$ ,  $L_{\epsilon P}(\delta; x)$ , se define como

$$L_{\epsilon P}(\delta; x) \propto L[\delta, \widehat{\lambda}(\delta) + \epsilon; x], \quad (3.3)$$

donde  $\epsilon = (\epsilon_1, \dots, \epsilon_w)$  es un vector en  $\mathbb{R}^w$  tal que  $[\widehat{\lambda}(\delta) + \epsilon] \in \Lambda$ . El parámetro  $\epsilon$  seleccionado para la muestra observada como se detalla un poco más adelante se llamará *parámetro de perturbación*. Nótese que  $\widehat{\lambda}(\delta) = [\widehat{\lambda}_1(\delta), \dots, \widehat{\lambda}_w(\delta)] \in \mathbb{R}^w$  es el emv restringido del vector  $\lambda$  para cada valor fijo del vector  $\delta$ .

La selección de una colección de valores para el parámetro de perturbación  $\epsilon \in \mathbb{R}^w$  se hará a través de la función de verosimilitud perfil relativa de  $\epsilon$  definida como

$$R(\epsilon) = \sup_{\delta | \epsilon} \frac{L[\delta, \widehat{\lambda}(\delta) + \epsilon; x]}{L(\widehat{\delta}, \widehat{\lambda}; x)}, \quad (3.4)$$

donde los valores de  $\epsilon$  cumplen que  $[\widehat{\lambda}(\delta) + \epsilon] \in \Lambda$ . Nótese que  $(\widehat{\delta}, \widehat{\lambda})$  es el emv global de  $(\delta, \lambda)$  y que  $0 \leq R(\epsilon) \leq 1$ . Además, obsérvese que cuando  $\epsilon = (\epsilon_1, \dots, \epsilon_w) = \mathbf{0}$  entonces  $R(\epsilon) = 1$ ; es decir el vector cero es el valor más plausible para  $\epsilon$ , a la luz de la muestra observada, y la correspondiente función de verosimilitud perfil  $\epsilon$ -perturbada es, en este caso, la función de verosimilitud perfil.

La propuesta para seleccionar  $\epsilon = (\epsilon_1, \dots, \epsilon_w)$  consiste en considerar valores de  $\epsilon$  que asignen alta plausibilidad a la muestra observada, relativa a la  $L(\widehat{\delta}, \widehat{\lambda}; x)$ . Por ejemplo, tomar valores de  $\epsilon = (\epsilon_1, \dots, \epsilon_w)$  que cumplan con  $R(\epsilon) = c$ , donde  $c \in \{0.99, 0.95, 0.90\}$ . También se puede considerar  $\epsilon$  que corresponda a un  $c$  un poco más extremo como 0.80, 0.70, pero no menor porque se trata de lograr una pequeña perturbación de la perfil.

Para valorar la sensibilidad o robusticidad de la verosimilitud perfil se recomienda compararla con las verosimilitudes perfil  $\epsilon$ -perturbadas correspondientes a los valores

de  $\epsilon$  calculados anteriormente que cumplen con  $R(\epsilon) = c$ , donde  $c \in \{1, 0.99, 0.95, 0.90, 0.80, 0.70\}$ . Cuando no exista un cambio importante en las inferencias obtenidas con las verosimilitudes  $\epsilon$ -perturbadas con respecto a las inferencias obtenidas con la verosimilitud perfil, entonces se dirá que la función de verosimilitud perfil es robusta frente a cambios pequeños en el emv restringido del parámetro de estorbo. En caso contrario, se dirá que la perfil es no robusta. Cuando se está en un caso robusto se recomienda el uso de la verosimilitud perfil para inferencias sobre el parámetro de interés. Sin embargo, cuando se esté en un caso no robusto se recomienda ser muy cauteloso con el uso de la perfil para inferencias sobre el parámetro de interés.

La función de verosimilitud perfil  $\epsilon$ -perturbada de  $\delta = (\delta_1, \dots, \delta_v)$  dada en (3.3), se definió para cualquier  $v, w \in \mathbb{N}$ . Sin embargo, en la práctica, usualmente la dimensión del parámetro de interés  $\delta = (\delta_1, \dots, \delta_v)$  es menor o igual que dos,  $v \leq 2$ . Por tanto, es posible valorar la sensibilidad o robusticidad de la verosimilitud perfil graficamente. Por ejemplo, para el caso  $v = 1$  es recomendable poner una misma figura a todas las curvas de verosimilitud perfil  $\epsilon$ -perturbadas correspondientes a los valores de  $\epsilon$  que cumplan con que  $R(\epsilon) = c$ , donde  $c \in \{1, 0.99, 0.95, 0.90, 0.80, 0.70\}$ . Obsérvese que  $c = 1$  esta asociado a  $\epsilon = 0$ , y corresponde a la función de verosimilitud perfil. Para el caso  $v = 2$  una posibilidad es comparar los contornos de la superficie de verosimilitud correspondientes a  $L_{\epsilon P}(\delta_1, \delta_2)$ .

Nótese que cuando  $\delta$  es un parámetro escalar y  $\lambda = (\lambda_1, \lambda_2) \in \mathbb{R}^2$  entonces la función de verosimilitud perfil relativa de  $\epsilon$ ,  $R(\epsilon)$ , es una superficie de verosimilitud en  $\mathbb{R}^3$  cuyo dominio es el plano cartesiano correspondiente al espacio  $\mathbb{R} \times \mathbb{R}$ . Es decir,  $R(\epsilon)$  es una función real valuada, que está definida para cada  $\epsilon = (\epsilon_{\lambda_1}, \epsilon_{\lambda_2}) \in \mathbb{R} \times \mathbb{R}$ . En este caso es recomendable hacer una gráfica de contornos de  $R(\epsilon)$  y señalar en la misma gráfica los puntos  $(\epsilon_{\lambda_1}, \epsilon_{\lambda_2})$  que cumplan con  $R(\epsilon_{\lambda_1}, \epsilon_{\lambda_2}) = c$ , donde  $c \in \{0.99, 0.95, 0.90, 0.80, 0.70\}$ , que serán utilizados para la perturbación.

Es importante comentar aquí que en contraste a la verosimilitud perfil de un

parámetro de interés, que es invariante frente a reparametrizaciones, la función de verosimilitud perfil  $\epsilon$ -perturbada puede depender de la parametrización del modelo probabilístico supuesto para los datos. En este sentido, para evaluar la robusticidad de la verosimilitud perfil de un parámetro de interés es importante que el modelo esté parametrizado de tal manera que los parámetros tengan una interpretación natural y física en términos del problema real, Seber y Wild (2003, pág. 126). Hemos observado que en los ejemplos no robustos Binomiales mostrados en esta tesis, sí le afectaron algunas de las reparametrizaciones que experimentamos; aunque en todos estos ejemplos, la verosimilitud perfil  $\epsilon$ -perturbada de  $N$  siguió siendo no informativa y plana. Sin embargo, en los ejemplos robustos, en general la verosimilitud perfil  $\epsilon$ -perturbada es prácticamente la misma frente a reparametrizaciones distintas y fue robusta en todas ellas. Se destaca que la verosimilitud integrada, ni la posterior Bayesiana de un parámetro de interés, tampoco son invariantes frente a reparametrizaciones como las que se exploraron acá.

En la siguientes secciones se ejemplificará con datos reales y simulados el uso de la función de verosimilitud perfil  $\epsilon$ -perturbada como una nueva herramienta estadística para explorar la robusticidad de la función de verosimilitud perfil de un parámetro de interés frente a cambios pequeños en el emv restringido del parámetro de estorbo.

## 3.2 Ejemplos robustos

En esta sección se presentan tres ejemplos donde la verosimilitud perfil es robusta. En el primer ejemplo se consideran datos de conteos de garrapatas en ovejas (Ejemplo 1.7.1 del Capítulo 1) modelados con una distribución Binomial Negativa. En el segundo ejemplo se consideran datos de niveles máximos anuales del mar en Port Pirie, Australia, modelados con una distribución Weibull para máximos. Por último, se considera una muestra grande simulada de una distribución Binomial  $(N, p)$ , y otra muestra chica

simulada pero con  $p$  grande de un modelo Binomial.

### 3.2.1 Modelo Binomial Negativa: Datos de conteos de garrapatas en ovejas

La Binomial Negativa es un modelo probabilístico comúnmente usado en ecología para modelar datos de conteos con sobredispersión con respecto al modelo Poisson; es decir, cuando la varianza poblacional es más grande que la media poblacional. R. A. Fisher en 1941 derivó la distribución Binomial Negativa de la siguiente manera. Supóngase que  $X$  es una variable aleatoria Poisson con parámetro de intensidad  $\mu$  el cual se supone que es también aleatorio y que se distribuye Gamma( $\theta, p/(1-p)$ ). Entonces, la distribución marginal de  $X$  que se obtiene de la distribución conjunta de  $(X, \mu)$  es la distribución Binomial Negativa con función de probabilidad

$$P(X = i; \theta, p) = \binom{\theta + i - 1}{i} p^i (1 - p)^\theta, \quad i = 0, \dots,$$

donde  $\theta > 0$  y  $0 < p < 1$ . Cabe señalar aquí que cuando  $\theta$  tiende a  $\infty$ ,  $p$  tiende a 1, y  $\theta(1-p)$  tiende a una constante fija  $\eta$ , entonces la distribución Binomial Negativa converge a una distribución Poisson de parámetro  $\eta$ . Así, valores grandes de  $\theta$  se interpretan como un caso de poca dispersión con respecto al modelo Poisson. Por otro lado, valores pequeños de  $\theta$  se interpretan como un caso de sobredispersión con respecto a un modelo Poisson. Por esta razón al parámetro  $\theta$  se le conoce como parámetro de sobredispersión.

En el Ejemplo 1.7.1 del Capítulo 1, la Tabla 1.3 muestra datos citados por Fisher (1941) sobre la clasificación de ovejas según el número de garrapatas encontradas en cada una de ellas. Fisher (1941) utilizó un modelo Binomial Negativa  $(\theta, p)$  para ajustar estos datos. En este ejemplo se considera a  $\theta$  como el parámetro de interés y  $p$  es considerado un parámetro de estorbo.

La función de verosimilitud global de  $\theta$  y  $p$  se da en el Capítulo 1 en (1.14). Además,

el emv restringido de  $p$  para cada valor especificado de  $\theta$  es  $\widehat{p}(\theta) = t/(t + n\theta)$ , donde  $t = \sum if_i$  y  $n = \sum f_i$  es el tamaño de la muestra.

La función de verosimilitud perfil  $\epsilon$ -perturbada de  $\theta$  es en este caso, según (3.1),

$$L_{\epsilon P}(\theta; \{f_i\}_{i \geq 0}) \propto [\widehat{p}(\theta) + \epsilon]^t \{1 - [\widehat{p}(\theta) + \epsilon]\}^{n\theta} \prod_{i=0}^{\infty} \binom{\theta + i - 1}{i}^{f_i},$$

donde  $\epsilon$  es una cantidad real pequeña tal que  $0 < \widehat{p}(\theta) + \epsilon < 1$ .

En la Figura 3.1 se muestra a la función de verosimilitud perfil relativa de  $\epsilon$ ,  $R(\epsilon)$  dada en (3.4), junto con diferentes valores de  $\epsilon$  que cumplen con  $R(\epsilon) = c$ , donde  $c \in \{0.99, 0.95, 0.90, 0.80, 0.70\}$ . Nótese que esta verosimilitud es aproximadamente simétrica; es decir, en este caso las perturbaciones  $\epsilon$  negativas son aproximadamente de la misma magnitud que las positivas por lo que estarán asociadas a la misma plausibilidad  $R(\epsilon) = c$ . La Figura 3.2 presenta el comportamiento de la función de verosimilitud perfil relativa  $\epsilon$ -perturbada de  $\theta$  considerando los valores de  $\epsilon$  calculados anteriormente.

La Figura 3.2 muestra que cuando la magnitud de la perturbación  $\epsilon$  se incrementa entonces también se incrementa la diferencia entre la forma de la función de verosimilitud perfil y la forma de la función de verosimilitud perfil  $\epsilon$ -perturbada. Sin embargo, las inferencias obtenidas con las funciones de verosimilitud perfil  $\epsilon$ -perturbadas no cambian substancialmente con respecto a las inferencias obtenidas con la verosimilitud perfil. Este es un caso en el que la función verosimilitud perfil de  $\theta$  es robusta frente a cambios pequeños en el emv restringido del parámetro de estorbo  $p$ .

Cabe señalar aquí que la forma que tiene la función de verosimilitud condicional de  $\theta$  es muy parecida a la forma de la función de verosimilitud perfil (véase la Figura 1.3 en el Capítulo 1). Es decir, ambas funciones de verosimilitud producen inferencias similares sobre  $\theta$ . Sin embargo, nuestro procedimiento pone de manifiesto una ventaja, no reconocida antes, de la función de verosimilitud perfil sobre la función de verosimilitud condicional. En particular, permite tomar en cuenta la sensibilidad o robusticidad de la función de verosimilitud perfil frente a cambios pequeños en el emv restringido

del parámetro de estorbo. No existe un procedimiento similar que tome en cuenta la estabilidad o robusticidad de la función de verosimilitud condicional.

A continuación se presenta otro ejemplo con datos reales de niveles máximos anuales del mar en Port Pirie, Australia, modelados con un modelo Weibull para máximos, donde se muestra que tiene nuevamente una verosimilitud perfil robusta.

### 3.2.2 Modelo Weibull para máximos: Datos de niveles máximos anuales del mar en Port Pirie, Australia

En la siguiente tabla se presentan los niveles máximos anuales del mar en Port Pirie, al Sur de Australia, registrados desde 1923 a 1987 (obtenidos del *freeware* del libro de Coles 2001).

Tabla 3.1. Niveles máximos anuales del mar registrados en metros en Port Pirie desde 1923 a 1987.

4.03	3.96	4.06	4.11	3.91
3.83	3.85	3.71	4.24	3.72
3.65	3.93	3.96	3.96	4.00
3.88	3.75	4.06	4.21	3.66
4.01	3.63	4.55	3.74	3.62
4.08	3.57	3.79	3.85	4.33
4.18	4.25	3.89	3.88	4.55
3.80	3.97	4.11	3.66	3.75
4.36	4.05	3.85	4.11	4.08
3.96	4.24	3.86	3.71	3.90
3.98	4.22	3.86	4.18	3.88
4.69	3.73	4.21	3.90	3.94
3.85	4.37	4.01	3.78	4.33

Estos datos son típicamente modelados con alguna distribución de valores extremos (Weibull, Gumbel o Fréchet) según el Teorema de Fisher y Tippett de 1928 (véase Embrechts *et al.* 1997, pág. 121).

Aquí se empleará la distribución Weibull de tres parámetros  $(\mu, \sigma, \beta)$  desconocidos, donde  $\mu$  es un parámetro umbral,  $\mu \in \mathbb{R}$ ,  $\sigma$  es de escala,  $\sigma > 0$  y  $\beta$  es de forma,  $\beta > 0$ , para modelar estos datos. La función de densidad de esta distribución se define para  $x \leq \mu$  como

$$f(x; \mu, \sigma, \beta) = \frac{\beta}{\sigma} \left( \frac{\mu - x}{\sigma} \right)^{\beta-1} \exp \left[ - \left( \frac{\mu - x}{\sigma} \right)^\beta \right]. \quad (3.5)$$

Nótese que si  $\beta < 1$ , esta densidad tiene una singularidad en  $x = \mu$ , por lo que se tiene un caso similar al descrito en el Capítulo 2 y conviene usar la aproximación continua truncada a la verosimilitud descrita en el Ejemplo 2.1.3.

Dentro de la familia de distribuciones de valores extremos, la distribución Weibull de tres parámetros es un modelo razonable para ajustar estos datos. De hecho, la razón de verosimilitud entre el mejor modelo Weibull y el mejor modelo Fréchet es  $4.278 \times 10^{12}$ . De manera similar, al comparar el mejor modelo Weibull y el mejor modelo Gumbel se tiene que la razón de verosimilitud es 1.1291. Así, en el primer caso, el modelo Weibull es indiscutiblemente mejor que uno Fréchet. En el segundo caso, ambos modelos son igualmente válidos y siendo que el Gumbel es más sencillo, podría considerarse como un modelo a elegir. Sin embargo, como el Weibull también es razonable, lo consideramos aquí para ejemplificar un caso robusto de una verosimilitud perfil con forma plana.

La aproximación continua truncada a la función de verosimilitud global Weibull de  $(\mu, \sigma, \beta)$ , basada en una muestra observada  $x = (x_1, \dots, x_n)$  de variables aleatorias independientes e idénticamente distribuidas  $X = (X_1, \dots, X_n)$  con función de densidad

$f(x; \mu, \sigma, \beta)$  dada en (3.5), es

$$L(\mu, \sigma, \beta; x) \propto \prod_{i=1}^n \frac{\beta}{\sigma} \left( \frac{\mu - x_i}{\sigma} \right)^{\beta-1} \exp \left[ - \left( \frac{\mu - x_i}{\sigma} \right)^{\beta} \right] I \left[ \frac{h}{x_{(n)} + \frac{h}{2}, \infty \right) (\mu), \sigma > 0, \beta > 0 \quad (3.6)$$

donde  $x_{(n)} = \max \{x_1, \dots, x_n\}$  y  $h$  es la precisión del instrumento de medición. Para este ejemplo se considerará una precisión  $h = 0.001$  m. Los estimadores de máxima verosimilitud globales de  $(\mu, \sigma, \beta)$  son (7.83, 3.95, 19.96).

En este ejemplo se considera al parámetro umbral  $\mu$  como el de interés y a los parámetros  $\sigma$  y  $\beta$  como parámetros de estorbo. Nótese que no es posible dar una expresión matemática cerrada para el emv restringido de  $\sigma$  y de  $\beta$  para cada valor fijo de  $\mu$ . Sin embargo, es posible calcular la función de verosimilitud perfil de  $\mu$ ,  $L_P(\mu; x)$ , en forma numérica. Para cada valor fijo de  $\mu \geq x_{(n)} + \frac{h}{2}$  se maximiza la verosimilitud global  $L(\mu, \sigma, \beta; x)$  dada en (3.6) con respecto a los parámetros de estorbo  $(\sigma, \beta)$ ,

$$L_P(\mu; x) \propto \max_{(\sigma, \beta) | \mu} L(\mu, \sigma, \beta; x).$$

La Figura 3.3 muestra a la función de verosimilitud perfil de  $\mu$ . Se observa que esta verosimilitud crece rápidamente hasta alcanzar su máximo y luego decrece lentamente y se mantiene casi plana a un nivel de plausibilidad de  $c = 0.9$ . En este ejemplo se verá que esta verosimilitud perfil de  $\mu$  es robusta a pesar de ser plana.

Nótese que en este caso se tienen dos parámetros de estorbo,  $\sigma$  y  $\beta$ . Por tanto se tienen que considerar perturbaciones en los estimadores de máxima verosimilitud restringidos de cada uno de ellos,  $\epsilon_\sigma$  y  $\epsilon_\beta$ . Es decir,  $\epsilon = (\epsilon_\sigma, \epsilon_\beta) \in \mathbb{R}^2$ . En la Figura 3.4 se muestra la superficie de verosimilitud relativa  $R(\epsilon_\sigma, \epsilon_\beta)$  dada en (3.4). La Figura 3.5 muestra la gráfica de contornos correspondientes a  $R(\epsilon_\sigma, \epsilon_\beta)$ . En esta gráfica se indican unos puntos elegidos a manera de ejemplo  $\epsilon = (\epsilon_\sigma, \epsilon_\beta)$  que cumplen con  $R(\epsilon) = c$ , donde  $c \in \{0.99, 0.95, 0.90, 0.80, 0.70\}$ , y que serán utilizados para la perturbación. Estos puntos fueron obtenidos a partir del comando ‘CONTOUR’ del programa MATLAB

7.0 que proporciona las coordenadas de cualquier contorno de nivel  $c$  deseado. Los puntos seleccionados se presentan en la Tabla 3.2.

Tabla 3.2. Valores de  $\epsilon = (\epsilon_\sigma, \epsilon_\beta)$  seleccionados para las perturbaciones.

$R(\epsilon_\sigma, \epsilon_\beta) = 0.99$	$\epsilon_\sigma$	0	0	0.0035	-0.0035	0.0025	0.0025	-0.0025	-0.0025
	$\epsilon_\beta$	0.2740	-0.2729	0	0	0.2584	-0.1375	0.1326	-0.2583
$R(\epsilon_\sigma, \epsilon_\beta) = 0.95$	$\epsilon_\sigma$	0	0	0.0079	-0.0078	0.0025	0.0025	-0.0025	-0.0025
	$\epsilon_\beta$	0.7688	-0.7636	0	0	0.8096	-0.6207	0.6150	-0.7999
$R(\epsilon_\sigma, \epsilon_\beta) = 0.90$	$\epsilon_\sigma$	0	0	0.0114	-0.0112	0.005	0.005	-0.005	-0.005
	$\epsilon_\beta$	1.4407	-1.4257	0	0	1.4868	-1.0111	0.9654	-1.4647
$R(\epsilon_\sigma, \epsilon_\beta) = 0.80$	$\epsilon_\sigma$	0	0	0.0167	-0.0162	0.01	0.01	-0.01	-0.01
	$\epsilon_\beta$	2.3389	-2.2995	0	0	2.3934	-1.3934	1.23	-2.2716
$R(\epsilon_\sigma, \epsilon_\beta) = 0.70$	$\epsilon_\sigma$	0	0	0.0212	-0.0203	0.015	0.015	-0.015	-0.015
	$\epsilon_\beta$	3.0574	-2.9904	0	0	2.7950	-1.5150	1.1950	-2.8150

En la Figura 3.6, 3.7, 3.8, 3.9 y 3.10 se presenta el comportamiento de la función de verosimilitud perfil relativa  $\epsilon$ -perturbada de  $\mu$  considerando los puntos  $\epsilon = (\epsilon_\sigma, \epsilon_\beta)$  que cumplen con  $R(\epsilon) = c$ , donde  $c = 0.99, 0.95, 0.90, 0.80, 0.70$ , respectivamente. Estos puntos se encuentran en la Tabla 3.2. La función de verosimilitud perfil relativa de  $\mu$  se obtiene cuando  $(\epsilon_\sigma, \epsilon_\beta) = (0, 0)$ .

En estas figuras se observa que las inferencias obtenidas con las funciones de verosimilitud perfil  $\epsilon$ -perturbadas no cambian substancialmente con respecto a las inferencias obtenidas con la verosimilitud perfil. En general, casi cualquier valor de  $\mu$  tal que  $\mu > \hat{\mu} = 7.83$  es altamente plausible. En especial el hecho que valores muy grandes de  $\mu$  tengan plausibilidad muy alta es un argumento a favor de usar un modelo Gumbel puesto que la densidad Weibull converge a una Gumbel cuando  $\mu \rightarrow \infty$ .

Este es otro caso en donde la función de verosimilitud perfil es robusta frente a cambios pequeños en el emv restringido del parámetro de estorbo a pesar de tener

forma plana.

### 3.2.3 Modelo Binomial $(N, p)$ : Casos robustos con muestras simuladas

En la literatura estadística, se considera a la distribución Binomial  $(N, p)$  para modelar datos de conteos tomados con el objetivo de estimar la población total de animales  $N$  en una cierta área geográfica. En este contexto el parámetro  $p$ , que representa a la probabilidad de observar un animal, se considera un parámetro de estorbo.

En esta sección se presentan dos casos donde la verosimilitud perfil de  $N$  es robusta. En el primer caso se considera una muestra simulada muy grande de tamaño  $k$  de variables aleatorias Binomiales independientes  $X = (X_1, \dots, X_k)$  con probabilidad  $p$  pequeña. En contraste, en el segundo caso se considera una muestra simulada con  $p$  grande y un tamaño de muestra  $k$  pequeño.

Considérese una muestra  $x = (x_1, \dots, x_k)$  de tamaño  $k$  de variables aleatorias Binomiales independientes  $X = (X_1, \dots, X_k)$  con  $N$  y  $p$  desconocidos, cuya probabilidad conjunta es

$$P(X = x; N, p) = \prod_{i=1}^k P(X_i = x_i; N, p) = p^t (1 - p)^{Nk-t} \prod_{i=1}^k \binom{N}{x_i}, \quad (3.7)$$

donde  $t = \sum_{i=1}^k x_i$ . La función de verosimilitud completa o global de  $N$  y  $p$ , definida en la Sección 1.1, es proporcional a la probabilidad de observar la muestra,  $P(X = x; N, p)$  dada en (3.7),

$$L(N, p; x) \propto p^t (1 - p)^{Nk-t} \prod_{i=1}^k \binom{N}{x_i}, \quad 0 < p < 1, \quad N \geq x_{\max}. \quad (3.8)$$

Nótese que en (3.8) se condiciona a que solamente valores de  $N$  mayores que el máximo valor observado en la muestra,  $x_{\max} = \max\{x_1, \dots, x_k\}$ , tengan plausibilidad. Es absurdo estimar la población  $N$  con un valor menor a lo que ya se observó; por sentido

común uno debe condicionar las inferencias que se realicen sobre todo lo que ya se sabe o conoce para que éstas hagan sentido con lo que se desea modelar.

La función de verosimilitud perfil de  $N$  se obtiene maximizando la función de verosimilitud global de  $N$  y  $p$ ,  $L(N, p; x)$  dada en (3.8) sobre  $p$  pero fijando  $N$ . Así, para obtener esta función se reemplaza el emv restringido de  $p$  para un valor especificado de  $N$ ,  $\hat{p}(N) = t/Nk$ , por  $p$  en  $L(N, p; x)$ , esto es,

$$\begin{aligned} L_P(N; x) &= L(N, \hat{p}(N); x) \\ &\propto [\hat{p}(N)]^t [1 - \hat{p}(N)]^{nk-t} \prod_{i=1}^k \binom{N}{x_i}. \end{aligned} \quad (3.9)$$

A partir de (3.8) se tiene que la función de verosimilitud perfil  $\varepsilon$ -perturbada de  $N$ ,  $L_{\varepsilon P}(N; x)$ , para el modelo Binomial se define como

$$\begin{aligned} L_{\varepsilon P}(N; x) &= L[N, \hat{p}(N) + \varepsilon; x] \\ &= [\hat{p}(N) + \varepsilon]^t \{1 - [\hat{p}(N) + \varepsilon]\}^{Nk-t} \prod_{i=1}^k \binom{N}{x_i}, \end{aligned} \quad (3.10)$$

donde  $\varepsilon$  es una cantidad real pequeña tal que  $0 < \hat{p}(N) + \varepsilon < 1$ .

Por otro lado, nótese que bajo el modelo Binomial  $(N, p)$  se tiene que  $T = \sum_{i=1}^k X_i$  es una estadística suficiente sólo para  $p$  cuando  $N$  es conocido, con distribución Binomial  $(Nk, p)$ ,

$$P(T = t; N, p) = \binom{Nk}{t} p^t (1 - p)^{Nk-t}. \quad (3.11)$$

Así, la función de verosimilitud condicional de  $N$ , definida en la Sección 1.6.1, es proporcional a  $P(X = x; N | T = t)$ . Esta distribución condicional se obtiene a través del cociente de la probabilidad de observar la muestra,  $P(X = x; N, p)$  dada en (3.7), y la probabilidad de observar  $T = t$ ,  $P(T = t; N, p)$  dada en (3.11),

$$P(X = x; N | T = t) = \frac{P(X = x; N, p)}{P(T = t; N, p)} = \frac{\prod_{i=1}^k \binom{N}{x_i}}{\binom{Nk}{t}}. \quad (3.12)$$

Por lo que la función de verosimilitud condicional de  $N$  es entonces

$$L_C(N; x) \propto \frac{\prod_{i=1}^k \binom{N}{x_i}}{\binom{Nk}{t}}, \quad N \geq x_{\max}. \quad (3.13)$$

Nótese que la función de verosimilitud completa de  $N$  y  $p$  se factoriza en dos términos, ya que  $P(X = x; N, p) = P(X = x; N | T = t) P(T = t; N, p)$ . Así,

$$L(N, p; x) = L_C(N; x) L_{Cres}(N, p; t).$$

Luego,

$$L_{Cres}(N, p; t) \propto P(t; N, p) = \binom{Nk}{t} p^t (1-p)^{Nk-t} \quad (3.14)$$

es la función de verosimilitud condicional residual y que también podría contener información acerca de  $N$  cuando  $p$  es desconocida, dado que también es función de  $N$ . Una forma de valorar qué tan grande es la información residual es a través de la gráfica de la función de verosimilitud condicional residual perfil de  $N$ . Esta se define como

$$L_{Cresp}(N; t) \propto L_{Cres}[N, \hat{p}(N); t] \propto \binom{Nk}{t} \hat{p}(N)^t [1 - \hat{p}(N)]^{Nk-t}, \quad (3.15)$$

donde  $\hat{p}(N)$  es el emv restringido de  $p$  para cada valor especificado de  $N$ .

A continuación se presentan los dos casos, con datos simulados donde se usa a la función de verosimilitud perfil  $\varepsilon$ -perturbada, dada en (3.10), para mostrar la robusticidad de la verosimilitud perfil de  $N$  bajo cambios pequeños en  $\hat{p}(N)$ .

### **Caso 1 (Robusto): Tamaño de muestra $k$ grande y probabilidad $p$ pequeña**

Se simuló una muestra  $x = (x_1, \dots, x_k)$  de tamaño  $k = 1,000$  de variables aleatorias Binomiales independientes  $X = (X_1, \dots, X_k)$  con  $N = 100$  y  $p = 0.2$ . La función de verosimilitud global de  $N$  y  $p$  dada en (3.8) y la correspondiente gráfica de contornos se presentan en las Figuras 3.11 y 3.12, respectivamente. Se observa que la superficie de verosimilitud no presenta anomalías. Los contornos son alargados mostrando la

fuerte relación entre los parámetros  $N$  y  $p$ . Nótese que los estimadores de máxima verosimilitud de  $N$  y  $p$ ,  $(\hat{N}, \hat{p}) = (96, 0.2077)$ , están muy cerca de los verdaderos empleados para simular la muestra,  $(N, p) = (100, 0.2)$ . En general, las inferencias sobre  $N$  y  $p$  son muy razonables.

En la Figura 3.13 se muestra a la función de verosimilitud perfil relativa de  $\epsilon$ ,  $R(\epsilon)$  dada en (3.2), junto con diferentes valores de  $\epsilon > 0$  que cumplen con  $R(\epsilon) = c$ , donde  $c \in \{0.99, 0.95, 0.90, 0.80, 0.70\}$ . Nótese que esta verosimilitud está muy cerrada alrededor del cero y es aproximadamente simétrica; es decir, perturbaciones hacia la derecha o hacia la izquierda que cumplen que  $R(\epsilon) = c$  son aproximadamente de la misma magnitud en la escala de  $\epsilon$ .

La Figura 3.14 presenta el comportamiento de la función de verosimilitud perfil relativa  $\epsilon$ -perturbada de  $N$  para los valores de  $\epsilon$  calculados anteriormente. Se observa que cuando la magnitud de la perturbación  $\epsilon$  se incrementa entonces también se incrementa la diferencia entre la forma la función de verosimilitud perfil y la forma de la función de verosimilitud perfil  $\epsilon$ -perturbada. Sin embargo, las inferencias obtenidas con las funciones de verosimilitud perfil  $\epsilon$ -perturbadas de  $N$  no cambian substancialmente con respecto a las inferencias obtenidas con la verosimilitud perfil de  $N$ . Este es un caso en el que la función verosimilitud perfil de  $N$  es robusta frente a cambios pequeños en el emv restringido del parámetro de estorbo  $p$ .

En la Figura 3.15 se muestra la función de verosimilitud perfil de  $N$  dada en (3.9), la función de verosimilitud condicional dada en (3.12) y la función de verosimilitud condicional residual perfil de  $N$  dada en (3.15). Se observa que la forma de la verosimilitud perfil y la forma de la verosimilitud condicional es similar. Es decir, ambas funciones de verosimilitud producen inferencias similares sobre  $N$ . Por otro lado, la verosimilitud condicional residual perfil contiene poca información sobre  $N$  puesto que es casi plana y le da alta plausibilidad a todos los posibles valores que puede tomar  $N$  y no contradice a la verosimilitud perfil de  $N$ .

Cabe señalar aquí nuevamente que nuestro procedimiento pone de manifiesto una ventaja, no reconocida antes, de la función de verosimilitud perfil sobre la función de verosimilitud condicional. En particular, permite tomar en cuenta la sensibilidad o robusticidad de la función de verosimilitud perfil frente a cambios pequeños en el emv restringido del parámetro de estorbo  $p$ . No existe un procedimiento similar que tome en cuenta la estabilidad o robusticidad de la función de verosimilitud condicional.

### **Caso 2 (Robusto): Tamaño de muestra $k$ pequeño y probabilidad $p$ grande**

Se simuló una muestra de tamaño  $k = 5$  de variables aleatorias Binomiales independientes  $X = (X_1, \dots, X_k)$  con  $N = 100$  y  $p = 0.8$ . Los valores numéricos de esta muestra simulada fueron  $x = (71, 77, 78, 80, 81)$ . La función de verosimilitud global de  $N$  y  $p$  dada en (3.8) y la correspondiente gráfica de contornos se presentan en las Figuras 3.16 y 3.17, respectivamente. Se observa que la superficie de verosimilitud no presenta anomalías, excepto que para valores muy grandes de  $N$  y valores muy pequeños de  $p$ , la cima de la superficie de verosimilitud, que está por debajo de un nivel de plausibilidad  $c = 0.1$ , decrece muy lentamente. Nuevamente los contornos de la superficie de verosimilitud son alargados mostrando la fuerte relación entre los parámetros  $N$  y  $p$ . También se observa que los estimadores de máxima verosimilitud de  $N$  y  $p$ ,  $(\hat{N}, \hat{p}) = (89, 0.8664)$ , están muy cerca de los verdaderos valores empleados para simular la muestra,  $(N, p) = (100, 0.8)$ . En general, las inferencias sobre  $N$  y  $p$  son muy razonables.

En la Figura 3.18 se muestra a la función de verosimilitud perfil relativa de  $\epsilon$ ,  $R(\epsilon)$  dada en (3.4), junto con diferentes valores de  $\epsilon$  que cumplen con  $R(\epsilon) = c$ , donde  $c \in \{0.99, 0.95, 0.90, 0.80, 0.70\}$ . Nótese que esta verosimilitud es simétrica; es decir, valores de  $\epsilon > 0$  (perturbación hacia la derecha) o valores  $\epsilon < 0$  (perturbación hacia la izquierda) con  $R(\epsilon) = c$  son aproximadamente de la misma magnitud. En comparación con el ejemplo anterior las magnitudes de  $\epsilon$  son mucho más grandes.

La Figura 3.19 presenta el comportamiento de la función de verosimilitud perfil relativa  $\epsilon$ -perturbada de  $N$  para los valores de  $\epsilon$  calculados anteriormente. La función de verosimilitud perfil relativa de  $N$  se obtiene cuando  $\epsilon = 0$ . Nótese que las funciones de verosimilitud perfil  $\epsilon$ -perturbadas de  $N$  difieren un poco de la función de verosimilitud perfil de  $N$  en la cola derecha. Sin embargo, intervalos de verosimilitud de nivel de plausibilidad  $c \geq 0.15$  calculados con la perfil y con estas verosimilitudes  $\epsilon$ -perturbadas son casi idénticos. En este sentido, las inferencias obtenidas con las funciones de verosimilitud perfil  $\epsilon$ -perturbadas de  $N$  no cambian substancialmente con respecto a las inferencias obtenidas con la verosimilitud perfil de  $N$ . Así, este es un caso en el que la función verosimilitud perfil de  $N$  es robusta frente a cambios pequeños en el emv restringido del parámetro de estorbo  $p$ .

En la Figura 3.20 se muestra la función de verosimilitud perfil de  $N$  dada en (3.9), la función de verosimilitud condicional dada en (3.13) y la función de verosimilitud condicional residual perfil de  $N$  dada en (3.15). A diferencia del caso anterior, ahora se observa que la forma de la verosimilitud perfil y la forma de la verosimilitud condicional difieren. Ambas funciones de verosimilitud producen inferencias diferentes. Nótese que la verosimilitud condicional pierde un poco de información sobre  $N$  contenida en la verosimilitud condicional residual perfil.

Nuevamente obsérvese que nuestro procedimiento pone de manifiesto la ventaja mencionada de la función de verosimilitud perfil sobre la función de verosimilitud condicional.

### 3.3 Ejemplos no robustos: Modelo Binomial $(N, p)$

Con miras a contar con un panorama general sobre los problemas inferenciales que pueden presentarse bajo el modelo Binomial  $(N, p)$  cuando se trata de estimar  $N$  cuando se desconoce  $p$ , primero se presentarán los resultados principales al respecto

en la literatura estadística. Posteriormente se presentará explícitamente el análisis de robusticidad de la verosimilitud perfil de  $N$  frente a cambios pequeños en el emv restringido de  $p$ .

### 3.3.1 Antecedentes

El problema de estimación más común asociado con el modelo Binomial  $(N, p)$  radica en estimar el parámetro  $p$  cuando  $N$  es conocido. Este problema no presenta grandes dificultades desde el punto de vista de estimación inferencial porque existe una estadística suficiente unidimensional y la verosimilitud de  $p$  es por lo general unimodal y suave a excepción de los casos extremos cuando no se observa ningún éxito, o cuando se tiene el caso contrario. Sin embargo, una situación distinta y mucho más difícil es estimar el parámetro  $N$ , el tamaño de la población de animales, cuando se desconoce también  $p$ , la probabilidad de observar un animal en una ocasión dada.

En la literatura de ecología estadística, se considera a la distribución Binomial  $(N, p)$  para modelar datos de conteos tomados con el objetivo de estimar la población total de animales  $N$  en una cierta área geográfica, Seber (1982), Borches *et al.* (2002), etc. Sin embargo, bajo este modelo, la estimación del parámetro  $N$  presenta grandes dificultades cuando el parámetro  $p$  también se desconoce. Un ejemplo de esto es que al estimar por separado al parámetro de interés  $N$  en presencia del parámetro de estorbo  $p$ , el emv de  $N$  puede ser incluso infinito cuando la media muestral es menor que la varianza muestral. En particular este estimador resulta absurdo cuando  $N$  representa el total de animales en un área, parámetro que forzosamente debe de ser finito.

Otro problema al estimar el parámetro  $N$  de la distribución Binomial  $(N, p)$  cuando  $p$  es desconocida es que los estimadores puntuales de  $N$  suelen ser muy inestables ya que la superficie de verosimilitud puede ser muy alargada y aplanada en la cima cuando el tamaño  $k$  de la muestra es chico y la probabilidad  $p$  es chica. Esta inestabilidad se manifiesta cuando un pequeño cambio de una unidad en el conteo más grande de

animales; es decir, cambiar  $x_{(n)}$  por  $x_{(n)} + 1$ , donde  $x_{(n)} = \max \{x_1, \dots, x_k\}$ , causa un cambio muy grande en el valor numérico de los estimadores puntuales de  $N$ , Olkin *et al.* (1981), Carroll y Lombard (1985), Casella (1986), Hall (1994), Gupta *et al.* (1999), DasGupta y Rubin (2004).

Aunque la motivación de Olkin *et al.* (1981) se dió a partir de conteos de crímenes, y la de Carroll y Lombard (1985) por conteos reales de animales para estimar su abundancia, estas referencias tienen como objetivo primordial proponer estimadores puntuales estables de  $N$ ; si bien Carroll y Lombard (1985) mencionan la necesidad de encontrar intervalos de estimación para  $N$ . En general, todas las referencias encontradas se han preocupado por la inestabilidad de los estimadores puntuales de  $N$ . Solamente Aitkin y Stasinopoulos (1989) consideraron la forma completa de la verosimilitud perfil de  $N$  pero llegaron a la conclusión absurda de que  $N$  podría ser infinito lo cual no es razonable en el contexto ecológico.

En la literatura estadística, varios métodos han sido utilizados para estimar por separado el parámetro de interés  $N$  en presencia del de estorbo  $p$ . Por ejemplo, la función de verosimilitud perfil de  $N$  y la función de verosimilitud condicional de  $N$  presentadas en la sección anterior en (3.9) y (3.13), respectivamente. Carroll y Lombard (1985) proponen utilizar el método de verosimilitud integrada, descrito en la Sección 1.6.3, para encontrar un estimador puntual estable de  $N$ . Ellos suponen que  $p$  tiene una densidad inicial Beta  $\pi(p; N) \propto p^a (1-p)^b$ , que no depende de  $N$ , y donde  $a$  y  $b$  son números reales conocidos, de manera tal que

$$L_I(N; x) \propto \prod_{i=1}^k \binom{N}{x_i} \int_0^1 p^{a+t} (1-p)^{b+Nk-t} dp = \binom{Nk + a + b + 1}{t + a + 1}^{-1} \prod_{i=1}^k \binom{N}{x_i}, \quad (3.16)$$

para todo  $N \geq x_{(n)}$ . Conviene resaltar aquí que Carroll y Lombard (1985) sólo se fijaron en el valor de  $N$  que maximiza la verosimilitud integrada  $L_I(N; x)$  dada en (3.16). Ellos nunca graficaron la verosimilitud integrada como función de  $N$ .

Bajo un enfoque Bayesiano, Draper y Guttman (1971) y Raftery (1988) propo-

nen utilizar la probabilidad marginal posterior Bayesiana, descrita en la Sección 1.6.4, para hacer inferencia sobre  $N$  en presencia de  $p$ . Sin embargo, en la literatura estadística se critica a este tipo de solución porque la función de densidad conjunta inicial  $\pi(N, p)$  para los parámetros  $N$  y  $p$  determina por completo la forma de la probabilidad marginal posterior  $N$  y en consecuencia las inferencias sobre  $N$ , Kahn (1987) y Aitkin y Stasinopoulos (1989). Otro problema importante relacionado con este método es la elección de la previa para  $N$  y  $p$ . En la práctica, los parámetros  $N$  y  $p$  generalmente se relacionan de una manera tan confusa y estrecha que puede ser muy complicado elicitar una función de densidad conjunta inicial para  $N$  y  $p$  apegada a la realidad.

En esta sección se dió un panorama general acerca de los problemas, descritos en la literatura estadística, relacionados con la estimación del parámetro  $N$  del modelo Binomial  $(N, p)$ . Para mayores detalles véase Montoya (2004). A continuación se describe la causa principal de todos estos problemas, la identificabilidad del modelo Binomial  $(N, p)$  cuando el tamaño de muestra  $k$  es pequeño y la probabilidad  $p$  es chica.

### **3.3.2 Identificabilidad del modelo Binomial $(N, p)$ como causa principal de los problemas relacionados con la estimación de $N$ cuando $p$ es desconocida**

Considérese una muestra de tamaño  $k$  de variables aleatorias Binomiales independientes con parámetros desconocidos  $N$  y  $p$ . Fisher (1941) comentó que cuando el tamaño de muestra  $k$  es grande no se tienen dificultades para hacer inferencia sobre el parámetro  $N$ . En particular, esto es cierto cuando la probabilidad  $p$  es chica. Esta situación se ejemplificó en la Sección 3.2.3 con una muestra simulada de tamaño  $k = 1,000$  y  $p = 0.2$ . Las inferencias sobre el parámetro  $N$ , a través de la función de verosimilitud

perfil, no presentaron grandes dificultades y fueron razonables. De hecho, para este caso se mostró que la verosimilitud perfil de  $N$  es robusta frente a cambios pequeños en el emv restringido de  $p$ .

Cuando el tamaño de muestra  $k$  es pequeño y la probabilidad  $p$  es grande, tampoco se tienen grandes dificultades para hacer inferencia sobre el parámetro  $N$ . Esta situación también se ejemplificó en la Sección 3.2.3 con una muestra simulada de tamaño  $k = 5$  y  $p = 0.8$ . Las inferencias sobre el parámetro  $N$ , a través de la función de verosimilitud perfil, tampoco presentaron grandes dificultades y fueron razonables. Para este caso, la función de verosimilitud perfil de  $N$  también fue robusta frente a cambios pequeños en el emv restringido de  $p$ . Sin embargo, en este ejemplo también se observa que la cola derecha de la perfil de  $N$  cae hasta un nivel de plausibilidad  $c = 0.1$  y luego se mantiene casi plana para valores grandes de  $N$ .

En contraste, cuando se tiene un tamaño de muestra  $k$  chico y la probabilidad  $p$  también es chica, comúnmente se tiene el problema de falta identificabilidad del modelo Binomial  $(N, p)$ . Este problema es la causa principal de los problemas relacionados con la estimación de  $N$  cuando  $p$  es desconocida, y se describirá a continuación.

### **Identificabilidad del modelo Binomial $(N, p)$ para un tamaño de muestra $k$ chico y probabilidad $p$ chica**

El problema matemático de identificabilidad del modelo Binomial consiste en que, para muestras pequeñas y probabilidad  $p$  chica, es muy difícil discernir entre dos modelos binomiales con parámetros completamente diferentes, pero que tengan la misma media  $\lambda = Np$ . A continuación se presenta un ejemplo de esta situación.

La Tabla 3.3 contiene dos muestras simuladas de tamaño  $k = 5$  de dos modelos

binomiales con parámetros muy diferentes pero con la misma media  $\lambda = 21$ .

Tabla 3.3. Muestras binomiales simuladas.

Nº	$N$	$p$	$\lambda = Np$	Muestra
1	1,000,000	$21/N = 0.000021$	21	16, 18, 22, 22, 26
2	100	$21/N = 0.21$	21	15, 21, 22, 24, 26

Obsérvese que las muestras son muy similares y que éste es un caso en el cual no es posible distinguir entre uno u otro modelo. Esto exhibe claramente la necesidad en el caso de muestras chicas de incorporar información adicional externa para así poder distinguir entre estos dos modelos por completo distintos y así poder hacer inferencias razonables sobre el total de animales  $N$  en la zona geográfica de interés.

Parece curioso que un modelo tan simple como el Binomial produzca tantos problemas a la hora de querer estimar el parámetro  $N$  en presencia del de estorbo  $p$ . La fuente principal de todos estos problemas es la falta de identificabilidad del modelo Binomial  $(N, p)$  cuando se tienen tamaños de muestra pequeños y la probabilidad  $p$  es chica. Esta situación causa que la función de verosimilitud global de  $N$  y  $p$ , que se define a partir de este modelo, herede este problema el cual se manifiesta provocando que la superficie de verosimilitud sea muy alargada y aplanada. Esto conduce a una verosimilitud perfil de  $N$  muy aplanada y a que el emv de  $N$  sea muy inestable cuando se perturba la muestra.

A continuación se presenta una cantidad utilizada, en la literatura estadística, para describir la forma de la función de verosimilitud perfil de  $N$ . Esta cantidad se llama coeficiente de estabilidad. Además, como aporte de esta tesis, se presenta el límite aproximado de la verosimilitud perfil de  $N$  cuando  $N \rightarrow \infty$ .

### 3.3.3 Forma de la verosimilitud perfil $N$

En esta sección se definen dos cantidades que dan información acerca de la forma de la verosimilitud perfil de  $N$ , el coeficiente de estabilidad y el límite aproximado de la verosimilitud perfil de  $N$  cuando  $N \rightarrow \infty$ . Ambas cantidades tienen una expresión matemática simple y se pueden calcular de una manera fácil y rápida. El límite aproximado de la verosimilitud perfil de  $N$  cuando  $N \rightarrow \infty$  no ha sido presentado antes como un indicador de la forma de la verosimilitud perfil de  $N$ .

#### Coefficiente de estabilidad

Tres variables aleatorias que sirven para describir datos de conteos son: la Binomial, la Poisson y la Binomial Negativa. Una manera de distinguir entre estos modelos probabilísticos es a través de la siguiente relación que existe entre el valor esperado  $E(X)$  y la varianza  $VAR(X)$ . En el modelo Poisson el valor esperado es igual a la varianza,  $E(X) = VAR(X)$ . En el modelo Binomial,  $E(X) > VAR(X)$ , y finalmente para la Binomial Negativa se da la relación contraria,  $E(X) < VAR(X)$ . Así, si se define al coeficiente de estabilidad ( $ce$ ) como  $ce = E(X)/VAR(X)$ , entonces lo anterior se resume de la siguiente manera:

$$ce \begin{cases} < 1, \text{ Binomial Negativa,} \\ = 1, \text{ Poisson,} \\ > 1, \text{ Binomial.} \end{cases}$$

En un caso real, la cantidad  $ce$  no se conoce. El estimador de máxima verosimilitud de  $ce$  es

$$\widehat{ce} = \frac{\widehat{\mu}}{\widehat{\sigma}^2}, \text{ donde } \widehat{\mu} = \frac{t}{k}, \text{ y } \widehat{\sigma}^2 = \frac{\sum_{i=1}^k (x_i - \widehat{\mu})^2}{k}. \quad (3.17)$$

Por tanto en ausencia de información externa adicional, una manera posible de seleccionar al mejor modelo de los tres mencionados aquí para un juego de datos observados puede ser basándose en el valor del  $\widehat{ce}$ , dependiendo de si el  $\widehat{ce}$  es menor, igual o mayor

que uno. Además, el  $\hat{c}e$  es relevante porque resume el comportamiento de la verosimilitud perfil de  $N$ ,  $L_P(N; x)$  dada en (3.9) como describen Olkin *et al.* (1981), cuando

$$ce \begin{cases} \leq 1, L_P(N; x) \text{ es una curva cóncava no decreciente que se maximiza en } N = \infty; \\ > 1, L_P(N; x) \text{ se maximiza en un valor positivo finito de } N. \end{cases}$$

Para muestras pequeñas, mientras más cerca esté el  $\hat{c}e$  de uno por arriba ( $\hat{c}e > 1$ ), entonces más plana será  $L_P(N; x)$  y más inestable será  $\hat{N}$ , el emv de  $N$ . Recuérdese que  $\hat{N}$  coincide con el emv global de  $N$ . Así, el  $\hat{c}e$  contiene información sobre la forma de  $L_P(N; x)$  y por consiguiente sobre qué tan inestable puede ser  $\hat{N}$ . Por ejemplo, la siguiente tabla presenta los coeficientes de estabilidad obtenidos con los datos simulados de la Tabla 3.3 de la sección anterior.

Tabla 3.4. Estimadores de parámetros relevantes para los datos simulados de la Tabla 3.3.

Nº	Muestra	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{c}e = \hat{\mu} / \hat{\sigma}^2$
1	16, 18, 22, 22, 26	20.8	12.16	1.71
2	15, 21, 22, 24, 26	21.6	13.84	1.56

Nótese que estos valores son cercanos a uno, lo que ocasionan que en ambos casos,  $L_P(N; x)$  sea muy aplanada. Además, como el  $\hat{c}e$  de la segunda muestra esta más cerca de uno que el de la primera muestra se tiene que  $L_P(N; x)$  con los datos de la segunda muestra es aún más plana que para la primera muestra. Así, la inestabilidad de  $\hat{N}$  con los datos de la segunda muestra es mayor.

Otro indicador de lo plano que puede ser la verosimilitud perfil de  $N$  es el límite de la cola derecha de esta verosimilitud cuando  $N$  tiende a infinito. En la literatura estadística casi nada se ha dicho sobre este límite. Quizás esto se deba a que el emv  $\hat{N}$  no tiene una fórmula matemática cerrada y esto dificulta el análisis. A continuación se presenta una aproximación del límite de la función de verosimilitud perfil relativa de  $N$  cuando  $N \rightarrow \infty$ , no antes presentada en la literatura estadística.

### Límite aproximado de la verosimilitud perfil de $N$ cuando $N \rightarrow \infty$ .

Una aproximación conveniente de  $R_P(N)$  se obtiene al usar la aproximación de Stirling

$$N! \approx \sqrt{2\pi} N^{N+0.5} \exp(-N), \quad (3.18)$$

y al reemplazar el emv  $\hat{N}$  por el estimador del método de momentos (emm)  $\bar{N}$ ; esto último siempre que los estimadores de  $\mu$  y  $\sigma^2$  cumplan que  $\hat{\mu} > \hat{\sigma}^2$ .

La media y la varianza de una variable aleatoria Binomial  $(N, p)$  es  $E(X) = Np$  y  $VAR(X) = Np(1-p)$ . Se tiene que los estimadores del método de momentos  $(\bar{N}, \bar{p})$  se pueden escribir en términos de la media y varianza muestrales dadas en (3.17)

$$\bar{N} = \frac{\hat{\mu}^2}{\hat{\mu} - \hat{\sigma}^2} \quad \text{y} \quad \bar{p} = \frac{\hat{\mu}}{\bar{N}}, \quad (3.19)$$

respectivamente. Si  $\hat{\mu} < \hat{\sigma}^2$  entonces  $\bar{N}$  sería negativo y por tanto un estimador absurdo del total de animales. Por otro lado, si  $\hat{\mu} = \hat{\sigma}^2$  entonces  $\bar{N}$  no está definido.

Así, reemplazando  $\hat{N}$  por  $\bar{N}$  dado en (3.19), siempre que  $\hat{\mu} > \hat{\sigma}^2$ , se tiene que un límite aproximado de la verosimilitud perfil relativa de  $N$  cuando  $N \rightarrow \infty$  es

$$\lim_{N \rightarrow \infty} R(N) = \lim_{N \rightarrow \infty} \frac{L_P(N; x)}{L_P(\hat{N}; x)} = \frac{\lim_{N \rightarrow \infty} L_P(N; x)}{L_P(\hat{N}; x)} \approx \frac{\lim_{N \rightarrow \infty} L_P(N; x)}{L_P(\bar{N}; x)}. \quad (3.20)$$

Nótese que la función de verosimilitud perfil de  $N$ ,  $L_P(N; x)$  dada en (3.9), es el producto de Binomiales con parámetros  $N$  y  $\hat{p}(N)$ , donde  $\lim_{N \rightarrow \infty} \hat{p}(N) = 0$  y  $N\hat{p}(N) = \hat{\mu}$  para todo  $N$ . Por tanto, cuando  $N$  tiende a infinito se tiene que cada probabilidad Binomial converge a una probabilidad Poisson con media  $\hat{\mu}$ . Así,

$$\lim_{N \rightarrow \infty} L_P(N; x) = \prod_{i=1}^k \frac{\hat{\mu}^{x_i}}{x_i!} \exp(-\hat{\mu}) = \hat{\mu}^{k\hat{\mu}} \exp(-k\hat{\mu}) \prod_{i=1}^k \frac{1}{x_i!}. \quad (3.21)$$

Reemplazando (3.21) en (3.20) se tiene que

$$\lim_{N \rightarrow \infty} R(N) \approx \frac{\hat{\mu}^{k\hat{\mu}} \exp(-k\hat{\mu}) \prod_{i=1}^k \frac{1}{x_i!}}{\left(\frac{\hat{\mu}}{\bar{N}}\right)^{k\hat{\mu}} \left(1 - \frac{\hat{\mu}}{\bar{N}}\right)^{k\bar{N} - k\hat{\mu}} \prod_{i=1}^k \binom{\bar{N}}{x_i}}. \quad (3.22)$$

Luego, desarrollando las combinaciones en el denominador de (3.22) en términos de cantidades factoriales y cancelando el término  $\prod_{i=1}^k \frac{1}{x_i!}$  se sigue que

$$\lim_{N \rightarrow \infty} R(N) \approx \frac{\widehat{\mu}^{k\widehat{\mu}} \exp(-k\widehat{\mu})}{\left(\frac{\widehat{\mu}}{\bar{N}}\right)^{k\widehat{\mu}} \left(1 - \frac{\widehat{\mu}}{\bar{N}}\right)^{k\bar{N}-k\widehat{\mu}} \prod_{i=1}^k \frac{\bar{N}!}{(\bar{N}-x_i)!}}. \quad (3.23)$$

Ahora, usando la aproximación de Stirling dada en (3.18) se tiene que

$$\begin{aligned} \prod_{i=1}^k \frac{\bar{N}!}{(\bar{N}-x_i)!} &\approx \prod_{i=1}^k \frac{\sqrt{2\pi}\bar{N}^{\bar{N}+0.5} \exp(-\bar{N})}{\sqrt{2\pi}(\bar{N}-x_i)^{\bar{N}-x_i+0.5} \exp[-(\bar{N}-x_i)]} \\ &= \prod_{i=1}^k \frac{\bar{N}^{\bar{N}+0.5}}{(\bar{N}-x_i)^{\bar{N}-x_i+0.5}} \exp(-x_i) = \\ &= \exp(-k\widehat{\mu}) \bar{N}^{k\bar{N}+0.5k} \prod_{i=1}^k \frac{1}{(\bar{N}-x_i)^{\bar{N}-x_i+0.5}}. \end{aligned} \quad (3.24)$$

Entonces, reemplazando (3.24) en (3.23) y cancelando el término  $\exp(-k\widehat{\mu})$  se tiene que

$$\lim_{N \rightarrow \infty} R(N) \approx \frac{\widehat{\mu}^{k\widehat{\mu}} \prod_{i=1}^k (\bar{N}-x_i)^{\bar{N}-x_i+0.5}}{\left(\frac{\widehat{\mu}}{\bar{N}}\right)^{k\widehat{\mu}} \left(1 - \frac{\widehat{\mu}}{\bar{N}}\right)^{k\bar{N}-k\widehat{\mu}} \bar{N}^{k\bar{N}+0.5k}}. \quad (3.25)$$

Luego, nótese que el denominador en (3.25) se puede escribir como

$$\left(\frac{\widehat{\mu}}{\bar{N}}\right)^{k\widehat{\mu}} \left(1 - \frac{\widehat{\mu}}{\bar{N}}\right)^{k\bar{N}-k\widehat{\mu}} \bar{N}^{k\bar{N}+0.5k} = \widehat{\mu}^{k\widehat{\mu}} \bar{N}^{0.5k} (\bar{N} - \widehat{\mu})^{k\bar{N}-k\widehat{\mu}}. \quad (3.26)$$

Así, reemplazando (3.26) en (3.25) y cancelando el término  $\widehat{\mu}^{k\widehat{\mu}}$  se tiene que

$$\lim_{N \rightarrow \infty} R(N) \approx \frac{\prod_{i=1}^k (\bar{N}-x_i)^{\bar{N}-x_i+0.5}}{\bar{N}^{0.5k} (\bar{N} - \widehat{\mu})^{k\bar{N}-k\widehat{\mu}}}. \quad (3.27)$$

Ahora, obsérvese que es posible escribir el denominador en (3.27) de la siguiente forma:

$$\begin{aligned}
\bar{N}^{0.5k} (\bar{N} - \hat{\mu})^{k\bar{N}-k\hat{\mu}} &= \bar{N}^{0.5k} (\bar{N} - \hat{\mu})^{-0.5k} (\bar{N} - \hat{\mu})^{k\bar{N}-k\hat{\mu}+0.5k} \\
&= \bar{N}^{0.5k} (\bar{N} - \hat{\mu})^{-0.5k} \prod_{i=1}^k (\bar{N} - \hat{\mu})^{\bar{N}-x_i+0.5} \\
&= \left( \frac{\bar{N}}{\bar{N} - \hat{\mu}} \right)^{0.5k} \prod_{i=1}^k (\bar{N} - \hat{\mu})^{\bar{N}-x_i+0.5}. \tag{3.28}
\end{aligned}$$

Entonces, reemplazando (3.28) en (3.27) se tiene la siguiente aproximación del límite de la verosimilitud relativa de  $N$ ,

$$\lim_{N \rightarrow \infty} R(N) \approx \frac{\prod_{i=1}^k (\bar{N} - x_i)^{\bar{N}-x_i+0.5}}{\left( \frac{\bar{N}}{\bar{N} - \hat{\mu}} \right)^{0.5k} \prod_{i=1}^k (\bar{N} - \hat{\mu})^{\bar{N}-x_i+0.5}}. \tag{3.29}$$

Nótese que el lado derecho de (3.29) se puede reescribir como

$$\begin{aligned}
\frac{\prod_{i=1}^k (\bar{N} - x_i)^{\bar{N}-x_i+0.5}}{\left( \frac{\bar{N}}{\bar{N} - \hat{\mu}} \right)^{0.5k} \prod_{i=1}^k (\bar{N} - \hat{\mu})^{\bar{N}-x_i+0.5}} &= \left( \frac{\bar{N}}{\bar{N} - \hat{\mu}} \right)^{-0.5k} \prod_{i=1}^k \left( \frac{\bar{N} - x_i}{\bar{N} - \hat{\mu}} \right)^{\bar{N}-x_i+0.5} \\
&= \left( \frac{\bar{N}}{\bar{N} - \hat{\mu}} \right)^{-0.5k} \prod_{i=1}^k \left( \frac{\bar{N}}{\bar{N} - \hat{\mu}} - \frac{\bar{N}}{\bar{N} - \hat{\mu}} \frac{x_i}{\bar{N}} \right)^{\bar{N}-x_i+0.5} \\
&= \left( \frac{\bar{N}}{\bar{N} - \hat{\mu}} \right)^{-0.5k} \left( \frac{\bar{N}}{\bar{N} - \hat{\mu}} \right)^{k\bar{N}-k\hat{\mu}+0.5k} \prod_{i=1}^k \left( 1 - \frac{x_i}{\bar{N}} \right)^{\bar{N}-x_i+0.5} \\
&= \left( \frac{\bar{N}}{\bar{N} - \hat{\mu}} \right)^{k(\bar{N}-\hat{\mu})} \prod_{i=1}^k \left( 1 - \frac{x_i}{\bar{N}} \right)^{\bar{N}-x_i+0.5}.
\end{aligned}$$

Entonces se sigue que

$$\lim_{N \rightarrow \infty} R(N) \approx \left( \frac{\bar{N}}{\bar{N} - \hat{\mu}} \right)^{k(\bar{N}-\hat{\mu})} \prod_{i=1}^k \left( 1 - \frac{x_i}{\bar{N}} \right)^{\bar{N}-x_i+0.5}. \tag{3.30}$$

Más aún, es posible escribir (3.30) en términos del  $\widehat{c\hat{e}}$ . De (3.17) y (3.19) se tiene que

$$\bar{N} = \widehat{\mu} \left( \frac{\widehat{c\hat{e}}}{\widehat{c\hat{e}} - 1} \right), \quad (3.31)$$

entonces reemplazando (3.31) en (3.30) se obtiene que

$$\lim_{N \rightarrow \infty} R(N) \approx (\widehat{c\hat{e}})^{k\widehat{\mu} \left( \frac{1}{\widehat{c\hat{e}} - 1} \right)} \prod_{i=1}^k \left[ 1 - \frac{x_i}{\widehat{\mu}} \left( \frac{\widehat{c\hat{e}} - 1}{\widehat{c\hat{e}}} \right) \right]^{\widehat{\mu} \left( \frac{\widehat{c\hat{e}}}{\widehat{c\hat{e}} - 1} \right) - x_i + 0.5}. \quad (3.32)$$

Este límite aproximado de la verosimilitud perfil de  $N$  cuando  $N \rightarrow \infty$  no sólo depende del  $\widehat{c\hat{e}}$  sino también de otras cantidades como la media muestral,  $\widehat{\mu}$ , y el tamaño de muestra,  $k$ . Por tanto la forma de la verosimilitud perfil depende de tres cantidades:  $\widehat{c\hat{e}}$ ,  $\widehat{widehatt\mu}$  y  $k$ .

Para la primera y segunda muestra de la Tabla 3.3 el valor aproximado de  $R_P(N)$  cuando  $N \rightarrow \infty$  es 0.7299 y 0.7835, respectivamente. Estos valores indican que la función de verosimilitud perfil de  $N$  calculada con la segunda muestra es un poco más plana que la verosimilitud perfil de  $N$  calculada con la primera muestra. Así, la inestabilidad de  $\widehat{N}$  con los datos de la segunda muestra es mayor. Véase que también se llegó a esta misma conclusión a partir del coeficiente de estabilidad de la primera muestra ( $\widehat{c\hat{e}} = 1.56$ ) y de la segunda muestra ( $\widehat{c\hat{e}} = 1.71$ ), el cual era chico para el valor de  $k = 5$ .

Para el caso robusto de la Sección 3.2.3, donde se simuló una muestra Binomial de tamaño  $k = 5$ , el valor aproximado de  $R_P(N)$  cuando  $N \rightarrow \infty$  es 0.0682. Por otra parte el  $\widehat{c\hat{e}} = 6.3235$ . En ambos casos la conclusión es la misma, que la cola derecha de la verosimilitud no es plana, por el contrario, decrece hasta alcanzar niveles de plausidad bajos. Sin embargo, para el otro caso Binomial robusto de la Sección 3.2.3, donde se simuló una muestra de tamaño  $k = 1,000$ , el valor aproximado de  $R_P(N)$  cuando  $N \rightarrow \infty$  es 0.000003 mientras que el  $\widehat{c\hat{e}} = 1.26$ . Aquí el valor del  $\widehat{c\hat{e}}$  es cercano a uno y por sí solo conduciría a concluir equivocadamente que la perfil de  $N$  es plana; véase Figura 3.14. Por el contrario, como el límite aproximado de  $R_P(N)$

cuando  $N \rightarrow \infty$  es cercano a cero, entonces la conclusión es que la perfil de  $N$  no es plana, como ocurre en realidad. Es decir hay que tener en cuenta a las tres cantidades  $(\hat{c}_e, \widehat{hatt}\mu, k)$ .

### 3.3.4 Ejemplos de estimación de abundancia de animales con el modelo Binomial $(N, p)$

A continuación se presentará explícitamente el análisis de robusticidad de la verosimilitud perfil de  $N$  frente a cambios pequeños en el emv restringido de  $p$ . A manera de ejemplo se emplearán los datos reales de conteos de manadas de impalas y conteos individuales de antílopes de agua presentados por Carroll y Lombard (1985) quienes usaron un modelo Binomial  $(N, p)$  para modelar estos datos.

*“Counts of impala herds and individual waterbucks were obtained on five successive cloudless days in a small area of the Kruger Park. Counting was done from a light aircraft by five highly trained and experienced wildlife officials”,*

manadas de impalas: 15, 20, 21, 23, 26,  $\bar{x} = 21$ ,

antílopes de agua: 53, 57, 66, 67, 72,  $\bar{x} = 63$ .

Se verá que los datos de impalas y antílopes de agua muestran que la función de verosimilitud perfil de  $N$  no es robusta frente a cambios pequeños en el emv restringido de  $p$ . Estos datos servirán también para ejemplificar una situación real con un tamaño de muestra pequeño de  $k = 5$ , donde se tienen problemas relacionados con la estimación de  $N$  cuando  $p$  es desconocida.

Los datos de impalas han sido descritos en la literatura como *inestables*, y los datos de antílopes de agua como *altamente inestables* tomando en cuenta los cambios en los estimadores puntuales de máxima verosimilitud al incrementar en una unidad la observación más grande de la muestra. A esta muestra modificada la llaman *muestra perturbada* en la literatura estadística, Olkin *et al.* (1981), Aitkin y Stasinopoulos (1989).

### Caso 1 (No robusto): Impalas

En la Figura 3.21 y 3.22 se presentan la gráfica de la función de verosimilitud global relativa de  $N$  y  $p$ ,  $R(N, p; x)$ , y la gráfica de contornos de  $R(N, p; x)$  para los datos de manadas de impalas. Estas gráficas muestran que  $R(N, p; x)$  es muy angosta y alargada, es decir decrece en  $p$  rápidamente a partir del contorno en la cima de esta superficie donde se cumple que  $Np = t/k = 21$ . La gráfica de contornos de  $R(N, p; x)$  da evidencia de la fuerte relación que existe entre los parámetros  $N$  y  $p$  mostrando contornos muy alargados e inclinados. Este es un caso donde se tiene el problema de identificabilidad de los parámetros y del modelo Binomial como se comentó en la Sección 3.3.2.

La Figura 3.23 presenta en la misma gráfica, la función de verosimilitud condicional, perfil y condicional residual perfil de  $N$ ,  $R_C(N; x)$ ,  $R_P(N; x)$  y  $R_{Cresp}(N; x)$ . Se observa que  $R_C(N; x)$  y  $R_P(N; x)$  son prácticamente planas, es decir valores muy grandes de  $N$  tienen alta plausibilidad. De hecho, el coeficiente de estabilidad es  $\hat{c}e = 1.59$  y el límite aproximado de la verosimilitud perfil de  $N$  cuando  $N \rightarrow \infty$  es 0.7767 (alto). Así, si se construyen intervalos o regiones de verosimilitud de nivel de plausibilidad  $c$  con este valor tal que  $0 < c < 0.7$ , el límite superior de estos intervalos tendría valor infinito. Esto es poco informativo y poco práctico pues el problema central de estimar la abundancia de animales  $N$  constituye el proveer límites o cotas superiores para  $N$ . La Figura 3.23 muestra además que la cola derecha de  $R_C(N; x)$  está por encima de la cola derecha de  $R_P(N; x)$  por lo que la verosimilitud condicional es incluso aún menos informativa que la perfil. Por otro lado, la función de verosimilitud condicional residual perfil relativa,  $R_{Cresp}(N; x)$ , tiene información sobre  $N$  que no está contenida en  $R_C(N; x)$ .

La Figura 3.24 muestra a la función de verosimilitud perfil relativa de  $N$ ,  $R_P(N; x)$ , junto con la ubicación del emv,  $\hat{N}$ , para el caso de la muestra sin perturbar y la muestra perturbada. Se observa que al incrementar en una unidad la observación máxima

(perturbar la muestra), el valor de  $\hat{N}$  cambia drásticamente con respecto al valor que tenía antes de perturbar la muestra. Todo esto se debe a que esta verosimilitud perfil de  $N$  es muy plana. Este es un caso donde no se puede resumir a la función de verosimilitud con un solo punto, tal como el valor donde alcanza su máximo.

Por otro lado, la Figura 3.24 exhibe que  $R_P(N; x)$  es altamente robusta frente a perturbaciones de una unidad en la observación máxima, y que además no es informativa. Es decir, es robusta porque la forma de la verosimilitud no cambia drásticamente cuando se perturba la muestra, y no es informativa porque casi todo valor de  $N$  en el rango de valores plausibles de  $N$  hacen igualmente probable a la muestra observada. De hecho  $N$  puede ser infinito con alta plausibilidad,  $R_P(N)$  es aproximadamente 0.7767 cuando  $N \rightarrow \infty$ . Sin embargo, es absurdo pensar que el total de manadas de impalas que viven en una pequeña área geográfica del Parque Kruger en Sudáfrica sea infinito. Es claro que estas verosimilitudes planas no proporcionan una solución realista al problema biológico.

La Figura 3.25 presenta en la misma gráfica a la función de verosimilitud perfil y a la función de verosimilitud integrada relativa de  $N$  de Carroll y Lombard,  $R_I(N; x)$ , con una Beta( $a = 0, b = 0$ ) (la distribución uniforme) y con una Beta( $a = 1, b = 1$ ). En esta gráfica se observa que, a diferencia de lo plana que es la verosimilitud perfil, las verosimilitudes integradas crecen rápidamente hasta alcanzar su máximo y luego decrecen hasta alcanzar niveles de plausibilidad bajos ( $c \leq 0.1$ ) para valores grandes de  $N$ . Además,  $R_I(N; x)$  con  $(a, b) = (0, 0)$  tiene cola derecha más pesada que con  $(a, b) = (1, 1)$ . En general, la cola derecha de las verosimilitudes integradas depende de manera explícita del parámetro  $a$  de la distribución inicial de  $p$  que se elija (véase Kahn, 1987 y Aitkin y Stasinopoulos, 1989). En consecuencia, la distribución inicial de  $p$  que se elija influye fuertemente en la estimación de  $N$  que se hará.

Hasta ahora se han descrito los problemas relacionados con la estimación de  $N$  cuando  $p$  es desconocida y el tamaño de muestra  $k$  es chico. A continuación se empleará

la función de verosimilitud perfil  $\epsilon$ -perturbada de  $N$  dada en (3.10) para explorar la robusticidad de la función de verosimilitud perfil  $N$  frente a cambios pequeños en el emv restringido de  $p$ .

En la Figura 3.26 se presenta a la función de verosimilitud perfil relativa de  $\epsilon$  dada en (3.2),  $R(\epsilon)$ , junto con diferentes valores de  $\epsilon$  que cumplen que  $R(\epsilon) = c$ , donde  $c \in \{0.99, 0.95, 0.90, 0.80, 0.70\}$ . Nótese que esta verosimilitud es aproximadamente simétrica; es decir, valores de  $\epsilon > 0$  (perturbación hacia la derecha) o valores  $\epsilon < 0$  (perturbación hacia la izquierda) con  $R(\epsilon) = c$  son aproximadamente de la misma magnitud.

La Figura 3.27 presenta el comportamiento de la función de verosimilitud perfil relativa  $\epsilon$ -perturbada de  $N$  para los valores de  $\epsilon$  calculados anteriormente. La función de verosimilitud perfil relativa de  $N$  se obtiene cuando  $\epsilon = 0$ . Se observa que cuando la magnitud de la perturbación  $\epsilon$  se incrementa entonces también se incrementa la diferencia entre la forma de la función de verosimilitud perfil  $\epsilon$ -perturbada y la forma de la función de verosimilitud perfil. Además, se observa que la verosimilitud perfil de  $N$  no es robusta frente a cambios pequeños en el emv restringido de  $p$  en el sentido de que cambios pequeños de  $\epsilon$  provocan grandes cambios en las inferencias sobre  $N$ . Nótese que perturbaciones hacia la derecha o hacia la izquierda, de la misma magnitud, producen verosimilitudes perfil  $\epsilon$ -perturbadas de  $N$  muy similares. Así, en este caso, se puede considerar  $\hat{p}(N) - \epsilon$ , con  $\epsilon > 0$ , para perturbaciones de la verosimilitud perfil de  $N$ .

## **Caso 2 (No robusto): Antílopes de agua**

En la Figura 3.28 y 3.29 se presentan la gráfica de la función de verosimilitud global relativa de  $N$  y  $p$ ,  $R(N, p; x)$ , y la gráfica de contornos de  $R(N, p; x)$  para los datos de antílopes de agua. De manera similar al ejemplo anterior, estas gráficas muestran que  $R(N, p; x)$  es muy estrecha o angosta. Los contornos de  $R(N, p; x)$  son muy alargados.

Este es otro caso donde se tiene nuevamente el problema de identificabilidad de los parámetros y del modelo Binomial asociado.

La Figura 3.30 presenta en la misma gráfica, la función de verosimilitud condicional, perfil y condicional residual de  $N$ ,  $R_C(N; x)$ ,  $R_P(N; x)$  y  $R_{Cresp}(N; x)$ . Nuevamente, como en el ejemplo anterior, se observa que  $R_C(N; x)$  y  $R_P(N; x)$  son muy planas. De hecho, el coeficiente de estabilidad es  $\hat{c}\epsilon = 1.30$  y el límite aproximado de la verosimilitud perfil de  $N$  cuando  $N \rightarrow \infty$  es 0.9209 (muy cercana a uno). Se observa además que la cola derecha de  $R_C(N; x)$  está por encima de la cola derecha de  $R_P(N; x)$ , sin embargo las inferencias sobre  $N$  con ambas funciones son similares. Por otro lado, la función de verosimilitud condicional residual  $R_{Cresp}(N; x)$ , tiene información sobre  $N$  que no está contenida en  $R_C(N; x)$ .

La Figura 3.31 muestra a la función de verosimilitud perfil relativa de  $N$ ,  $R_P(N; x)$ , junto el emv de  $N$ ,  $\hat{N}$ , para el caso de la muestra sin perturbar y para la muestra perturbada. Se observa que al perturbar la muestra, el valor de  $\hat{N}$  cambia muy drásticamente con respecto al valor que tenía antes de perturbar la muestra. Nuevamente, todo esto se debe a que esta verosimilitud perfil de  $N$  es muy plana. Este es otro caso donde no se puede resumir a la función de verosimilitud solamente con el emv; se requiere de la función de verosimilitud completa.

Por otro lado, la Figura 3.31 nuevamente exhibe que  $R_P(N; x)$  es altamente robusta frente a perturbaciones de la muestra, y que nuevamente resulta ser no informativa. De hecho  $N$  puede ser infinito con alta plausibilidad,  $R_P(N)$  es aproximadamente 0.9209 cuando  $N \rightarrow \infty$ . Otra vez; es absurdo pensar que el total de antílopes de agua que viven en una pequeña área geográfica del Parque Kruger en Sudáfrica sea infinito. Es claro que estas verosimilitudes no proporcionan una solución realista al problema biológico.

La Figura 3.32 presenta en la misma gráfica a la función de verosimilitud perfil y a la función de verosimilitud integrada relativa de  $N$  de Carroll y Lombard,  $R_I(N; x)$ ,

con una  $\text{Beta}(a = 0, b = 0)$  (la distribución uniforme) y con una  $\text{Beta}(a = 1, b = 1)$ . En forma similar al ejemplo anterior, en esta gráfica se observa que ambas integradas son asimétricas con cola derecha pesada que baja a niveles de plausibilidad pequeños ( $c \leq 0.1$ ). En contraste, la verosimilitud perfil de  $N$  es plana. Nuevamente  $R_I(N; x)$  con  $(a, b) = (0, 0)$  tiene cola derecha más pesada que con  $(a, b) = (1, 1)$ . Como se mencionó antes, la cola derecha de las verosimilitudes integradas depende de manera explícita del parámetro  $a$  de la distribución inicial de  $p$  que se elija.

A continuación se empleará la función de verosimilitud perfil  $\epsilon$ -perturbada de  $N$  dada en (3.10) para explorar la robusticidad de la función de verosimilitud perfil  $N$  frente a cambios pequeños en el emv restringido de  $p$ .

La Figura 3.33 presenta la función de verosimilitud perfil relativa de  $\epsilon$  dada en (3.2),  $R(\epsilon)$ , junto con diferentes valores de  $\epsilon$  asociados a una plausibilidad  $R(\epsilon) = c$ , donde  $c \in \{0.99, 0.95, 0.90, 0.80, 0.70\}$ . Nuevamente nótese que esta verosimilitud es aproximadamente simétrica; es decir, perturbaciones hacia la derecha o hacia la izquierda que cumplen que  $R(\epsilon) = c$  son aproximadamente de la misma magnitud.

La Figura 3.34 presenta el comportamiento de la función de verosimilitud perfil relativa  $\epsilon$ -perturbada de  $N$  para los valores de  $\epsilon$  calculados anteriormente. Otra vez se observa que cuando la magnitud de la perturbación  $\epsilon$  se incrementa entonces también se incrementa substancialmente la diferencia entre la forma de la función de verosimilitud perfil  $\epsilon$ -perturbada y la forma de la función de verosimilitud perfil lo cual indica que se está en un caso no robusto. Nuevamente se observa que perturbaciones hacia la derecha o hacia la izquierda, de la misma magnitud, producen verosimilitudes perfil  $\epsilon$ -perturbadas de  $N$  muy similares (pero muy diferentes a la perfil). Así, también en este caso, se puede considerar  $\hat{p}(N) - \epsilon$ , con  $\epsilon > 0$ , para perturbaciones de la verosimilitud perfil de  $N$ .

En general se concluye que las inferencias sobre  $N$  cambian dramáticamente para valores distintos de  $\epsilon$  y en particular son muy distintas de la perfil. Esto manifiesta

que **solamente con la muestra observada no se puede hacer inferencia sobre  $N$** . Falta recabar más información u observaciones.

En la siguiente sección se presentan diferentes propuestas de solución al problema ecológico de estimación de abundancia de animales bajo un modelo Binomial  $(N, p)$ .

### **3.4 El problema ecológico de estimación de abundancia de animales usando un modelo Binomial $(N, p)$**

Como se ha visto hasta ahora en este capítulo, el modelo Binomial  $(N, p)$  puede conducir a muchos problemas a la hora de querer estimar el parámetro  $N$  cuando se desconoce  $p$ . Recuérdese que  $N$  es la población total de animales en una pequeña zona y  $p$  es la probabilidad de observar un animal en una ocasión dada. De hecho, casi siempre se tienen estos problemas porque en la práctica se cuenta con pocas observaciones ( $k$  chico) y la probabilidad de observar un animal suele ser pequeña a menos que se controle y se modifique el mecanismo de observación. Sin embargo, este modelo se utiliza ampliamente para estimar abundancia de animales porque los muestreos binomiales son más baratos y rápidos que otros tipos de muestreos, como por ejemplo los de captura y recaptura. Además, otra ventaja de los muestreos binomiales es que son menos invasivos; es decir hacen menos daño a los sistemas ecológicos porque requieren una menor intervención del hombre en el hábitat del animal. Generalmente se realizan las observaciones a distancia, desde una avioneta u otro transporte.

Bajo la distribución Binomial  $(N, p)$ , una alternativa de solución al problema de estimar la población total de animales  $N$  en una zona geográfica dada es forzar a través del diseño del mecanismo de observación a que las observaciones provengan de un modelo Binomial con  $p$  grande. Así, se estaría en un caso similar al del ejemplo

Binomial robusto con  $k$  chica y  $p$  grande de la Sección 3.2. Allí se tienen inferencias razonables sobre  $N$  y además se cuenta con robusticidad de la función de verosimilitud perfil de  $N$  frente a cambios pequeños en el emv restringido de  $p$ .

Cuando el parámetro de interés  $N$  representa el tamaño de la población de animales en una cierta área geográfica, entonces  $N$  tiene un nivel lógico diferente que  $p$ . Esto es,  $N$  es en principio un parámetro observable, el número actual de animales en una cierta área geográfica. En contraste, el parámetro  $p$  considerado de estorbo es una cantidad hipotética no observable. Por tanto, resulta incuestionable que en cualquier modelo que describa a una población de animales  $N$  deba ser finito y acotado. A continuación se presenta una relación matemática entre la verosimilitud perfil  $\epsilon$ -perturbada y una cota superior  $M$  para  $N$ .

La función de verosimilitud perfil  $\epsilon$ -perturbada de  $N$ ,  $L_{\epsilon P}(N; x)$  dada en (3.10), esta condicionada a que cualquier emv restringido y perturbado de  $p$ , en particular  $\hat{p}(N) - \epsilon$ , sea mayor que cero. Así, para todo  $\epsilon > 0$ , se tiene que

$$\hat{p}(N) - \epsilon = \frac{t}{Nk} - \epsilon > 0;$$

esto implica que

$$N < \frac{t}{\epsilon k}.$$

Definamos a  $M(\epsilon) = t/\epsilon k$  como una cota superior para  $N$ , la población total de animales en una cierta área geográfica. Así, para cada valor fijo de  $\epsilon > 0$ , se tiene una función de verosimilitud perfil  $\epsilon$ -perturbada de  $N$  y una cota superior para  $N$ ,  $M(\epsilon)$ . Nótese que la verosimilitud perfil de  $N$ , que se obtiene cuando  $\epsilon = 0$ , esta asociada a  $M(\epsilon) = \infty$ . En la práctica decir cuánto vale  $M$  no es necesariamente fácil para los ecólogos. Depende de muchos factores, el animal, el medio ambiente, el alimento, entre otros, y el ecólogo debe conocerlos muy bien de cerca.

De lo anterior, se sigue que una propuesta para brindar mayor información a los ecólogos sobre la abundancia de animales  $N$  es trazar curvas de verosimilitud perfil

$\epsilon$ -perturbadas con diversas  $M$ 's asociadas a  $\epsilon$  tal que  $R(\epsilon) = c$ , donde  $c \in \{0.99, 0.95, 0.90, 0.80, 0.70\}$ . Así, ecólogos expertos pueden saber cuáles  $M$ 's y curvas asociadas son las más razonables para los datos y evaluar las consecuencias en la inferencia de  $N$  al cambiar  $M$ . Sin embargo, nótese que esto no brinda una solución al problema de estimación de abundancia  $N$  de animales en el sentido de que dada una muestra observada no se puede decidir cuál curva perfil  $\epsilon$ -perturbada es preferible. Sólo se puede decir que las inferencias dependen fuertemente de las suposiciones. Habría que recolectar más observaciones para poder concluir algo sobre  $N$ .

En ocasiones el ecólogo puede contar con más información sobre  $p$  pues ésta depende del diseño de muestreo y del procedimiento para observar a los animales. En ese caso se sugiere comparar gráficamente la verosimilitud perfil  $\epsilon$ -perturbada,  $L_{\epsilon P}(N; x)$ , con la verosimilitud integrada de Carroll y Lombard (1985),  $L_I(N; x)$  dada en (3.16), para varias previas  $\pi(p)$  que sean razonables y que reflejen el conocimiento que se tenga de  $p$ . Aunque no hay una correspondencia uno a uno entre estos dos tipos de curva, se recomienda lo siguiente: dada una distribución previa  $\pi(p)$ , calcular la  $L_I(N; x)$  y luego seleccionar la  $L_{\epsilon P}(N; x)$  que ajuste mejor a esta verosimilitud integrada. Es decir, seleccionar el valor de  $\epsilon$  que haga que la forma de la función  $L_{\epsilon P}(N; x)$  sea similar a la forma de la función  $L_I(N; x)$ , en especial en la parte de la cola derecha de las curvas. Posteriormente, si el valor seleccionado de  $\epsilon$  es sustentado por los datos (por ejemplo que  $\epsilon$  sea tal que  $R(\epsilon) \geq 0.15$ ), y la correspondiente cota superior  $M(\epsilon)$  para  $N$  es razonable para los ecólogos, entonces se pueden hacer inferencias sobre  $N$  a través de la función de verosimilitud integrada o de la función de verosimilitud  $\epsilon$ -perturbada de  $N$ ; ambas serán muy parecidas.

Ahora, si el valor seleccionado de  $\epsilon$  es sustentado por los datos; pero la cota superior  $M(\epsilon)$  para  $N$  asociada a esta  $\epsilon$  no es razonable para los ecólogos, entonces esto se puede interpretar como una discrepancia entre la información previa de  $p$  y la información previa sobre la cota superior  $M$  para  $N$ . En ese caso, antes de usar alguna

de estas verosimilitudes para hacer inferencias sobre  $N$  es recomendable revisar ambos supuestos, tanto de la previa como de  $M$ .

Por otro lado, si el valor seleccionado de  $\epsilon$  que ajuste mejor la  $L_{\epsilon P}(N; x)$  a la  $L_I(N; x)$  no fuera sustentado por los datos (por ejemplo que  $\epsilon$  sea tal que  $R(\epsilon) \leq 0.15$ ), entonces esto se puede interpretar como una discrepancia fuerte entre la información previa de  $p$  y la información paramétrica sobre  $N$  y  $p$  contenida en la muestra observada. Recuérdese que la propuesta para seleccionar  $\epsilon$  consiste en considerar valores de  $\epsilon$  que asignen plausibilidad alta a la muestra observada, relativa a la  $L(\hat{N}, \hat{p}; x)$ , y que correspondan a moverse un poco del borde más alto de la superficie de verosimilitud determinado por la función de verosimilitud perfil. Así, si el valor seleccionado de  $\epsilon$  no está sustentado por los datos, esto equivale a haberse alejado demasiado de la perfil de  $N$ . En ese caso, es recomendable revisar los supuestos sobre la previa de  $p$ .

También se puede actuar de manera inversa; es decir, dada una cota superior  $M$  para  $N$  buscar una distribución de probabilidad previa  $\pi(p)$  para  $p$  que haga ahora que la forma de la función  $L_I(N; x)$  sea similar a la forma de la función  $L_{\epsilon P}(N; x)$ . Bajo este procedimiento, si el valor de  $\epsilon$  asociado a esta cota superior  $M$  para  $N$  está sustentado por los datos (por ejemplo que  $\epsilon$  sea tal que  $R(\epsilon) \geq 0.15$ ), y la previa  $\pi(p)$  que haga que la forma de la función  $L_I(N; x)$  sea similar a la forma de la función  $L_{\epsilon P}(N; x)$  es razonable también para los ecólogos, entonces se puede dar información sobre  $N$  a través de la función de verosimilitud integrada o de la función de verosimilitud  $\epsilon$ -perturbada de  $N$ , ya que ambas serán muy parecidas. En caso contrario, habrá que revisar los supuestos.

A continuación se usarán los datos de impalas para ejemplificar los procedimientos anteriores. Se emplearán estos datos, y no los de antílopes de agua, porque al menos para ellos se tiene un poco de información externa, de un ecólogo experto en mamíferos, sobre una cota superior  $M$  para  $N$ .

### 3.4.1 Ejemplo de Impalas

La Figura 3.35 presenta el comportamiento de la función de verosimilitud perfil relativa  $\epsilon$ -perturbada de  $N$  para diferentes valores de  $\epsilon$  asociados a una plausibilidad  $R(\epsilon) = c$ , donde  $c \in \{0.99, 0.95, 0.90, 0.80, 0.70\}$ . Además se presenta la cota superior  $M(\epsilon)$  para  $N$  asociada con cada valor de  $\epsilon$ . La función de verosimilitud perfil relativa de  $N$  se obtiene cuando  $R(\epsilon) = 1$ ; es decir  $\epsilon = 0$  y en este caso la cota superior para  $N$  es infinito,  $N \leq M(\epsilon) = \infty$ .

Las curvas  $\epsilon$ -perturbadas en la Figura 3.35 dan información muy distinta unas de otras. Esto evidencia la dependencia fuerte que tienen en los supuestos sobre  $M$ . Por ello, a menos que se cuente con mucha información sobre  $M$  certera, no se puede estimar  $N$  a través de estas curvas.

En esta ocasión no se tiene información previa sobre  $p$  en forma de una distribución de probabilidad  $\pi(p)$  dada por algún ecólogo. Sin embargo, con el objetivo de explicar nuestro procedimiento descrito anteriormente, se empleará como una previa razonable para  $p$  la densidad Beta ( $a = 1, b = 1$ ),  $\pi(p) \propto p(1-p)$ , propuesta por Carroll y Lombard (1985), quienes presentaron estos datos. Ellos proponen utilizar el método de verosimilitud integrada para encontrar un estimador puntual estable de  $N$ . Sólo se fijaron en el valor de  $N$  que maximiza la verosimilitud integrada  $L_I(N; x)$  y no presentan gráfica alguna de esta verosimilitud.

En la Figura 3.36 se presenta, al lado izquierdo, dos densidades Beta( $\alpha, \beta$ ) previas de  $p$ , y al lado derecho de cada densidad previa de  $p$ , en la misma gráfica, a la correspondiente función de verosimilitud integrada relativa de  $N$  junto con una verosimilitud perfil relativa  $\epsilon$ -perturbada de  $N$ . En el primer caso, se fijó la densidad previa Beta( $a = 1, b = 1$ ) para  $p$  de Carroll y Lombard (1985). Luego se calculó la verosimilitud integrada correspondiente y se le ajustó una verosimilitud perfil  $\epsilon$ -perturbada de  $N$  de manera aproximada. Esta verosimilitud perfil  $\epsilon$ -perturbada esta asociada a una cota superior de  $M = 840$ . En el segundo caso se procedió de manera inversa.

Se fija la verosimilitud perfil  $\epsilon$ -perturbada de  $N$  correspondiente a  $\epsilon = 0.0142$ , donde  $R(\epsilon = 0.0142) = 0.90$ , y luego se busca una previa que haga que la integrada se ajustara a la  $\epsilon$ -perturbada de manera aproximada. La previa asociada con esta integrada es la densidad  $\text{Beta}(a = 1.5, b = 2.5)$ .

Esta previa da mayor probabilidad a valores pequeños de  $p$  y menor probabilidad a valores muy cercanos a uno. La Figura 3.36 da a los ecólogos expertos información acerca del impacto que tiene la distribución previa de  $p$  en las inferencias sobre  $N$ . En particular, muestra el impacto que tiene sobre la cota superior  $M$  para  $N$ .

Obsérvese que cuando la distribución previa de  $p$  asigne mayor probabilidad a valores pequeños de  $p$  entonces la cola de la verosimilitud integrada se volverá cada vez más aplanada. Así, se requerirán valores de  $\epsilon$  cada vez más pequeños para obtener curvas de verosimilitud  $\epsilon$ -perturbadas que ajusten bien de manera aproximada a estas verosimilitudes integradas. En consecuencia, las cotas superiores  $M = t/k\epsilon$  asociadas con estos valores de  $\epsilon$  serán cada vez más grandes. Además, cuando la previa de  $p$  es proporcional  $1/p$  se tiene que la forma de la integrada es idéntica a la forma de la verosimilitud condicional, que es muy aplanada en este ejemplo; Aitkin y Stasinopoulos (1989). En resumen, en los casos no robustos, la inferencia sobre  $N$  depende fuertemente de los supuestos que se hagan, ya sea sobre  $p$ , a través de su previa, o sobre la cota superior  $M$  (y consecuentemente sobre  $\epsilon$ ).

Si bien, la verosimilitud integrada (con Beta adecuada) y la  $\epsilon$ -perturbada (con  $\epsilon > 0$ ) pueden arrojar inferencias tales que  $N$  sea finita y que sean razonables, estas curvas no constituyen en sí mismas una solución al problema de estimación de abundancia de animales. Solamente permiten comparar curvas muy distintas entre sí, asociadas a supuestos diferentes. Habría que recabar más observaciones o incluso cambiar el método de observación para obtener un caso robusto para así poder realizar inferencias sobre  $N$  que sean confiables y que se basen en los datos observados.

## 3.5 Discusión

La función de verosimilitud perfil  $\epsilon$ -perturbada permite explorar la robusticidad de la función de verosimilitud perfil frente a cambios pequeños en la estimación de los parámetros de estorbo. En este capítulo se ha ejemplificado este uso y se ha mostrado la utilidad de dicha función para reconocer situaciones donde la verosimilitud perfil del parámetro de interés no sea confiable. La función de verosimilitud perfil resulta ser un miembro particular de la familia de funciones de verosimilitud perfil  $\epsilon$ -perturbada, que se obtiene cuando  $\epsilon = 0$ .

El modelo Binomial  $(N, p)$  permite ejemplificar situaciones robustas y no robustas en la verosimilitud perfil de  $N$ . Como se vió en la Sección 3.3, el modelo Binomial  $(N, p)$  conduce a muchos problemas a la hora de querer estimar el parámetro  $N$  cuando también  $p$  es desconocida. Estos problemas son una consecuencia de tener superficies de verosimilitud de  $N$  y  $p$  alargadas y aplanadas. La causa principal de este comportamiento peculiar de la verosimilitud radica en que el modelo Binomial  $(N, p)$  es no identificable en el límite. Cuando  $N \rightarrow \infty$ ,  $p \rightarrow 0$  y tal que  $Np = \lambda$  es fijo, entonces se tiene una aproximación a la distribución Poisson con media  $\lambda$ .

El límite aproximado de la verosimilitud perfil de  $N$  cuando éste parámetro tiende a infinito, es un buen indicador de la forma de la perfil de  $N$ . Otro indicador es la terna de valores  $(\hat{c}\epsilon, \hat{\mu}, k)$  como ya se mencionó.

La verosimilitud perfil  $\epsilon$ -perturbada pone en evidencia una ventaja, no reconocida antes, de la función de verosimilitud perfil sobre la función de verosimilitud condicional. En particular, permite tomar en cuenta la sensibilidad o robusticidad de la función de verosimilitud perfil frente a cambios pequeños de  $\epsilon$ . No existe un procedimiento similar que permita tomar en cuenta la estabilidad o robusticidad de la función de verosimilitud condicional.

Tanto la estimación puntual de  $N$  como los intervalos de verosimilitud perfil de  $N$  pueden resultar absurdos cuando la verosimilitud perfil de  $N$  es no robusta. Si

se desea hacer inferencias razonables sobre  $N$  es necesario aumentar el tamaño de la muestra recolectando más observaciones bajo circunstancias similares y esto puede no ser práctico o posible.

En casos no robustos, se tienen cuatro alternativas para incorporar información adicional a los datos observados que al menos garantizan intervalos de estimación para  $N$  finitos que no sean absurdos. Estas son:

- (a) el enfoque Bayesiano, que agrega información adicional sobre  $N$  y  $p$  a través de una distribución de probabilidad conjunta previa para  $N$  y  $p$ ,
- (b) la verosimilitud integrada de  $N$ , que incorpora información adicional sólo sobre  $p$  en forma de una densidad previa para  $p$ ,
- (c) el uso de una cota superior  $M$  para  $N$  que se incorpora a la verosimilitud perfil  $\epsilon$ -perturbada de  $N$ ,
- (d) la selección de una o varias  $\epsilon$  plausibles a través de  $R(\epsilon)$  para obtener una o varias curvas  $\epsilon$ -perfil perturbadas.

Sin embargo, con todas estas opciones las inferencias sobre  $N$  dependen fuertemente de los supuestos que se hagan y generalmente son muy distintas e incluso contradictorias entre sí.

El modelo Binomial se usa ampliamente en ecología porque los muestreos binomiales son menos invasivos; es decir hacen menos daño a los sistemas ecológicos porque requieren una menor intervención del hombre en el hábitat del animal. Además su implementación es de costos menores que otras alternativas. A la luz de esto, una propuesta de solución al problema ecológico de estimación de abundancia de animales usando un modelo Binomial  $(N, p)$  es lograr que con alta probabilidad el mecanismo de observación sea tal que sea razonable suponer que el parámetro  $p$  sea grande. Así, se estaría en un caso similar al del ejemplo Binomial robusto con  $k$  chica y  $p$  grande

de la Sección 3.2, donde se pueden realizar inferencias razonables sobre  $N$  y además se cuenta con robusticidad de la función de verosimilitud perfil de  $N$  frente a cambios pequeños en el emv restringido de  $p$ .

Una segunda propuesta es cambiar el diseño de muestreo para poder usar un modelo distinto y así tener un caso robusto con alta probabilidad. Así se evitan los problemas del modelo Binomial para estimar  $N$ . Un ejemplo de modelos no problemáticos son los probabilísticos considerados para captura y recaptura (Moran, 1951). Sin embargo, esto implica cambiar el diseño del experimento y esto queda fuera del contexto de esta tesis.

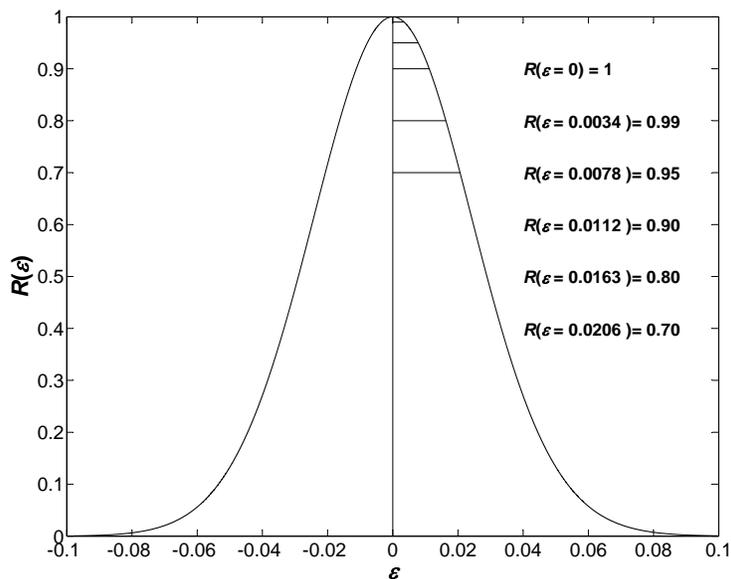


Figura 3.1: Verosimilitud perfil relativa de  $\epsilon$  para los datos de garrapatas en ovejas. Valores de  $\epsilon$  con alta plausibilidad  $c$ .

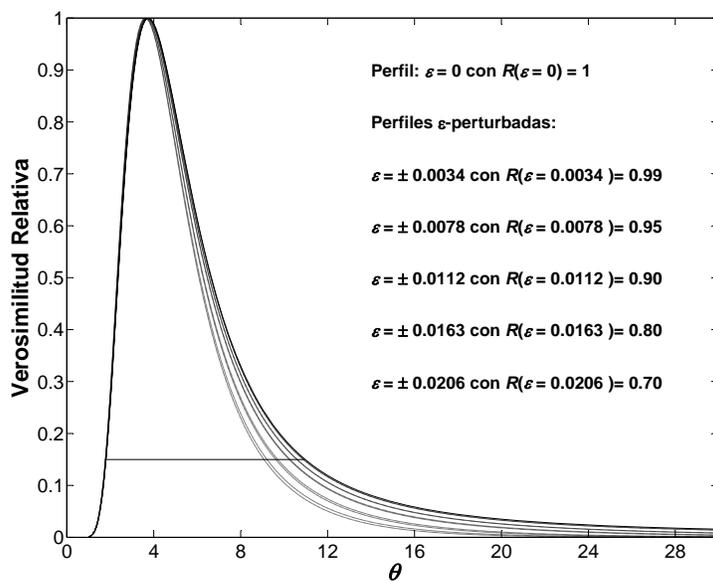


Figura 3.2: Verosimilitudes perfil  $\epsilon$ -perturbadas de  $\theta$  para los datos de garrapatas en ovejas e intervalos del 15% de verosimilitud asociados.

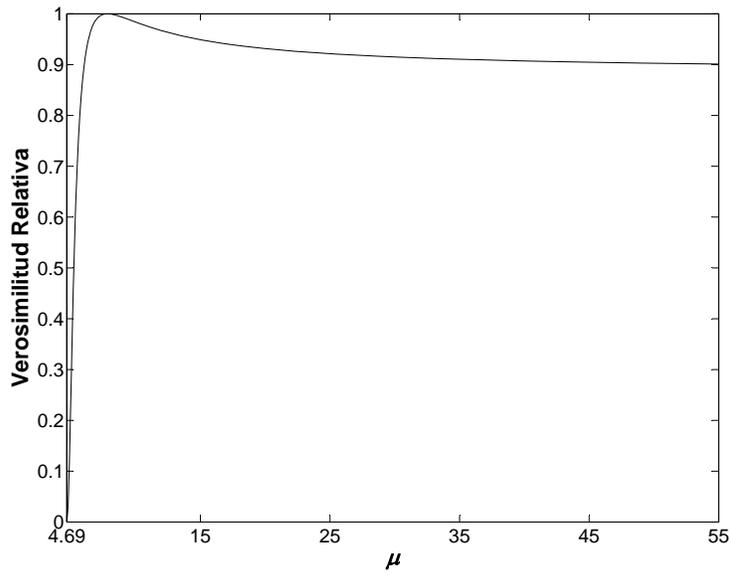


Figura 3.3: Función de verosimilitud perfil de  $\mu$  para los datos de niveles máximos anuales del mar en Port Pirie.

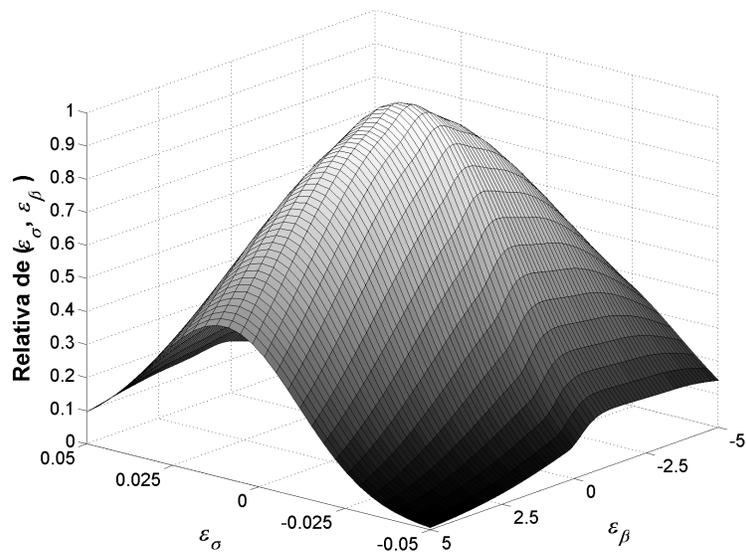


Figura 3.4: Superficie de verosimilitud relativa  $R(\epsilon_\sigma, \epsilon_\beta)$  para los datos de niveles máximos anuales del mar en Port Pirie.

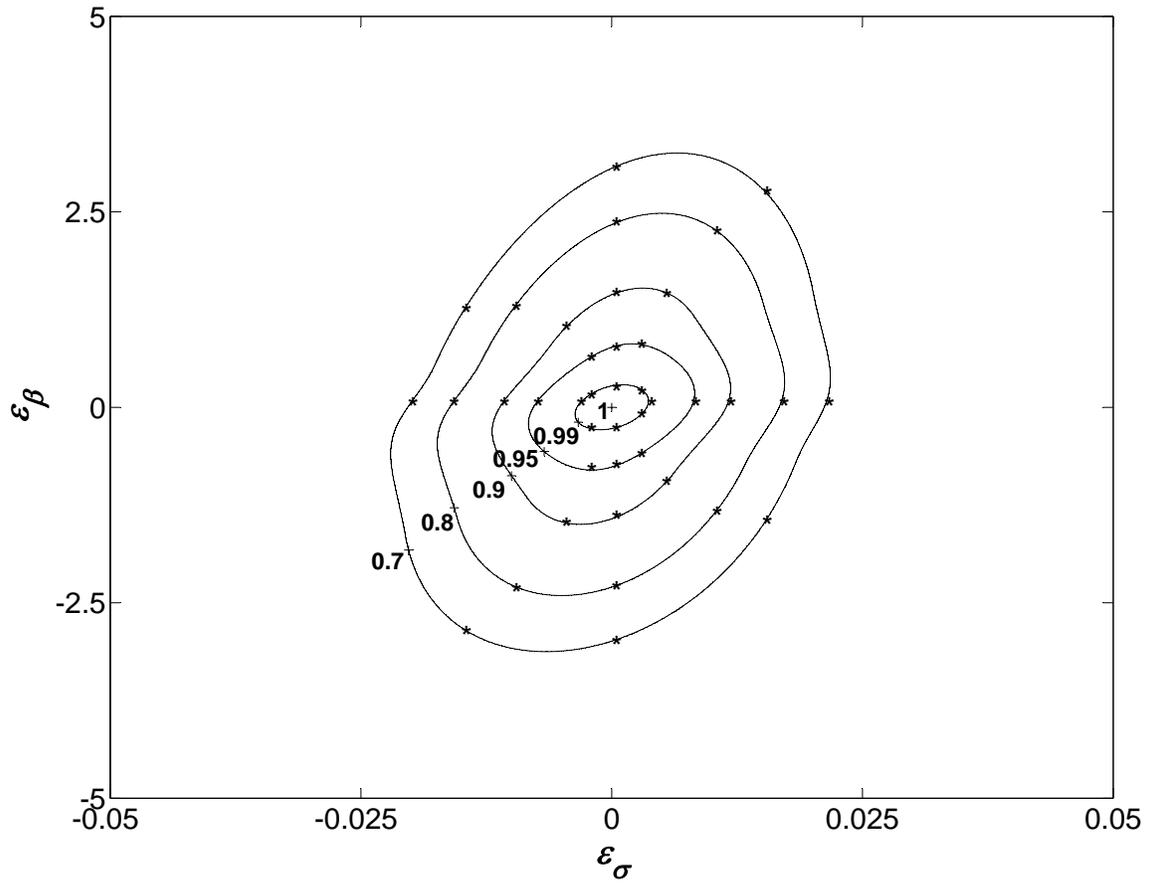


Figura 3.5: Contornos de  $R(\epsilon_\sigma, \epsilon_\beta)$  para los datos de niveles máximos anuales del mar en Port Pirie. Se marcan (\*) los puntos  $(\epsilon_\sigma, \epsilon_\beta)$  que serán utilizados para la perturbación.

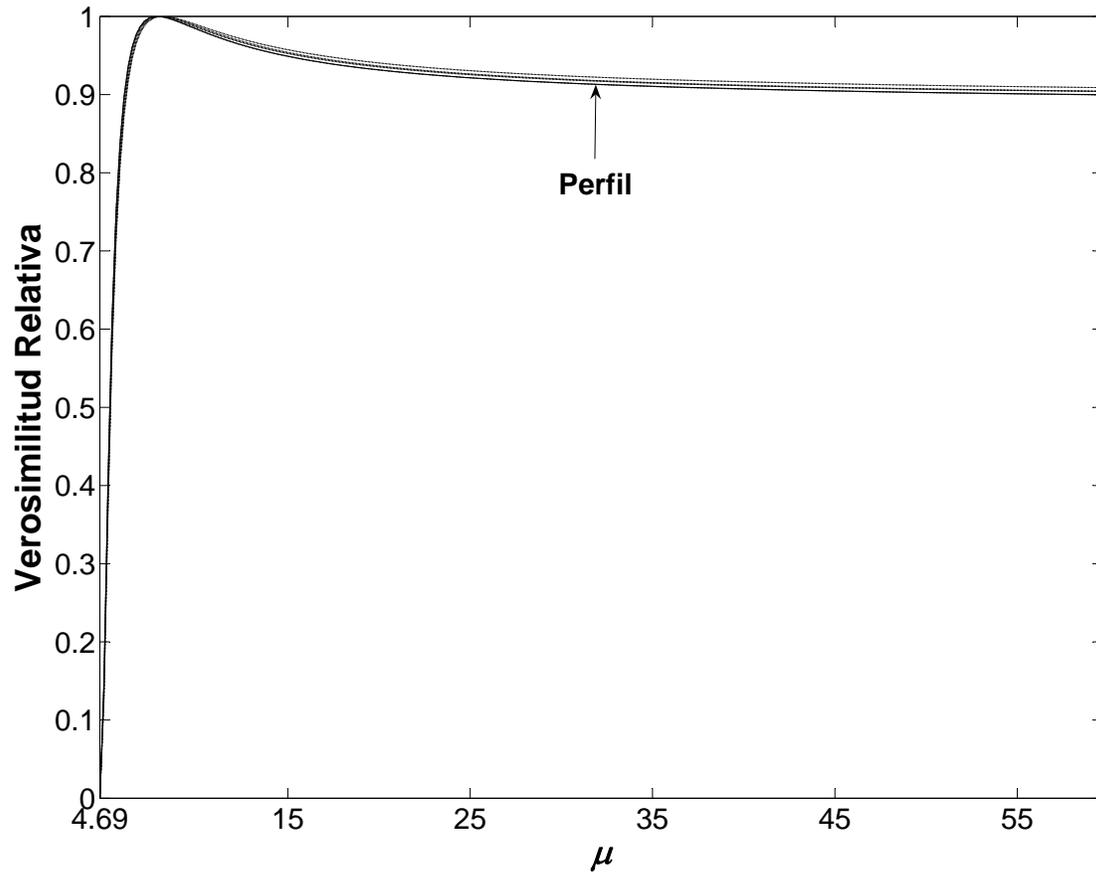


Figura 3.6: Port Pirie: Verosimilitud perfil y ocho verosimilitudes perfiles  $\epsilon$ -perturbadas de  $\mu$  correspondientes a los contornos de  $R(\epsilon_\sigma, \epsilon_\beta) = 0.99$  mostrados en la Figura 3.5.

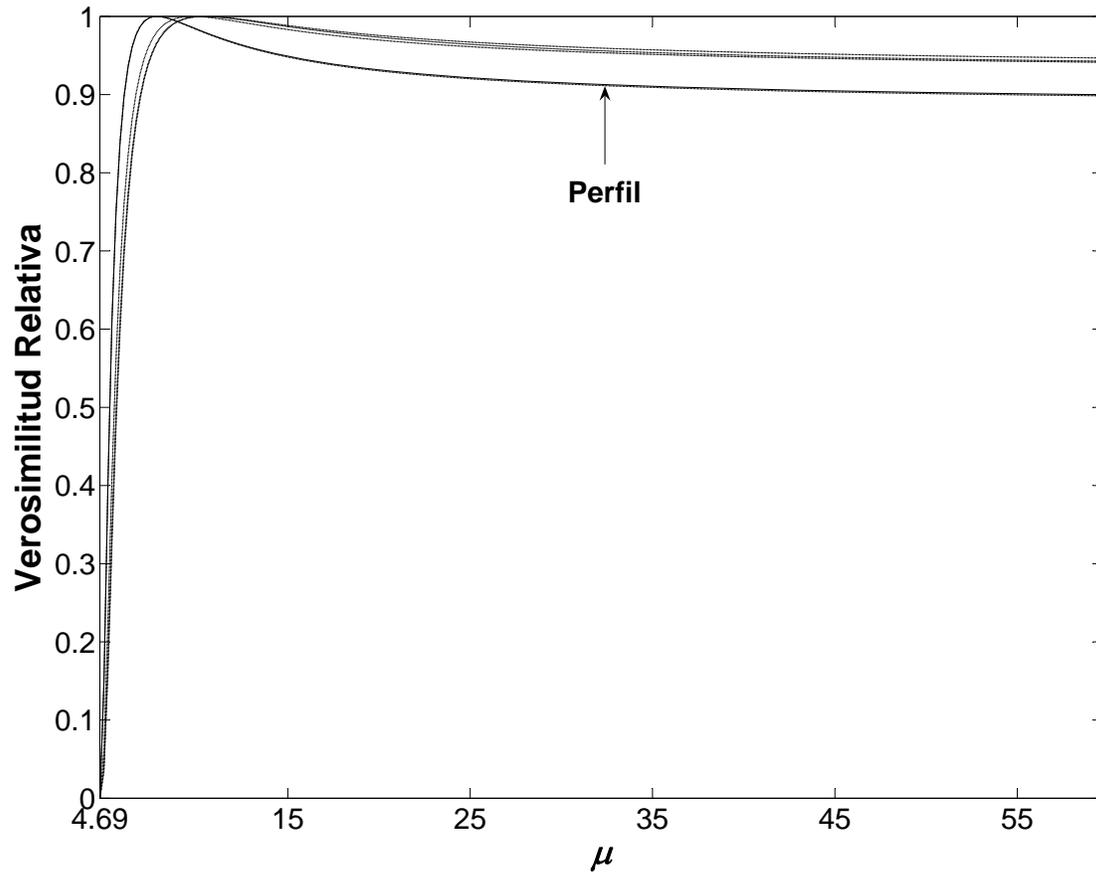


Figura 3.7: Port Pirie: Verosimilitud perfil y ocho verosimilitudes perfiles  $\epsilon$ -perturbadas de  $\mu$  correspondientes a los contornos de  $R(\epsilon_\sigma, \epsilon_\beta) = 0.95$  mostrados en la Figura 3.5.

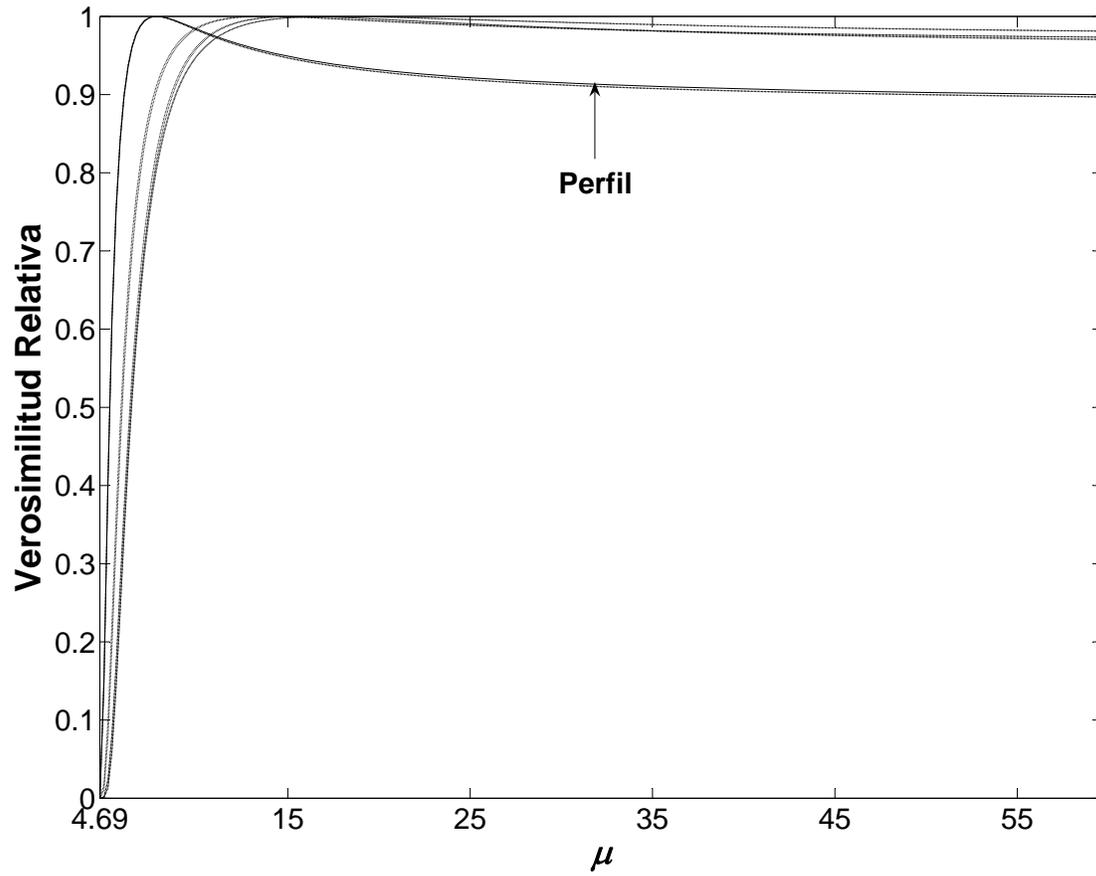


Figura 3.8: Port Pirie: Verosimilitud perfil y ocho verosimilitudes perfiles  $\epsilon$ -perturbadas de  $\mu$  correspondientes a los contornos de  $R(\epsilon_\sigma, \epsilon_\beta) = 0.90$  mostrados en la Figura 3.5.

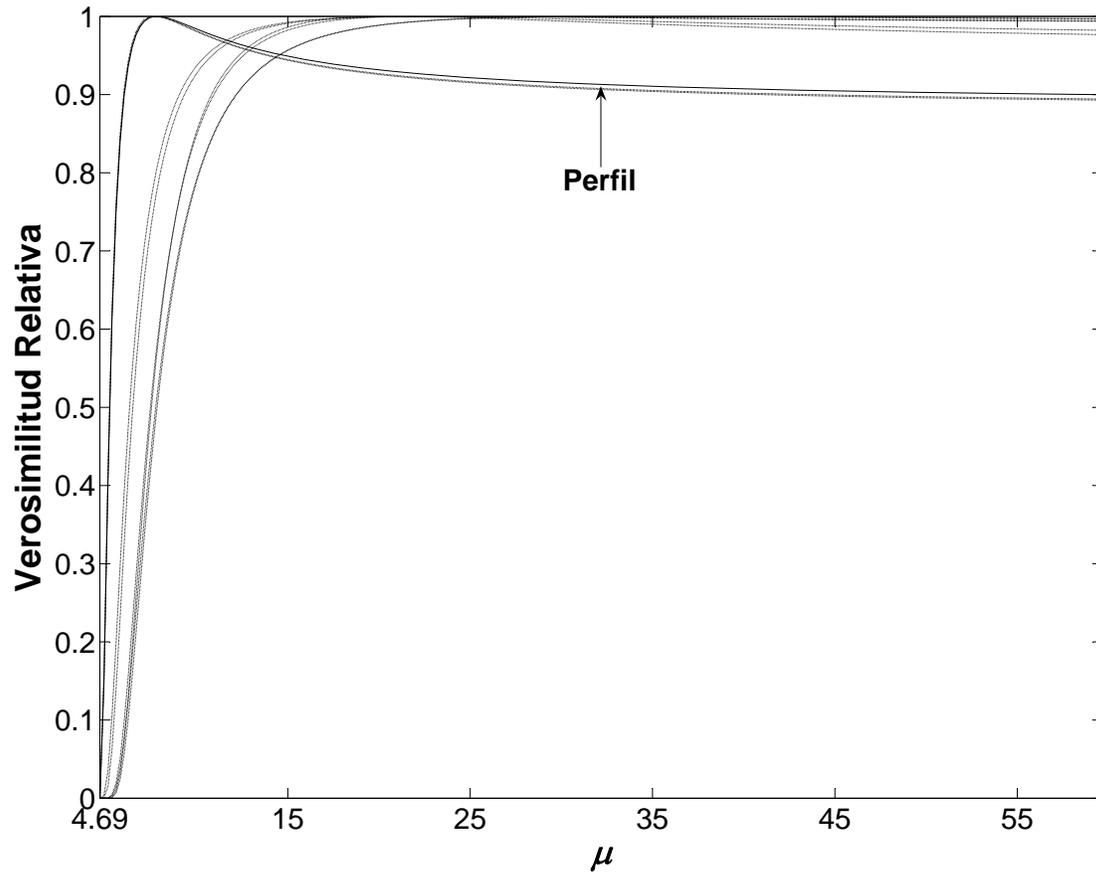


Figura 3.9: Port Pirie: Verosimilitud perfil y ocho verosimilitudes perfiles  $\epsilon$ -perturbadas de  $\mu$  correspondientes a los contornos de  $R(\epsilon_\sigma, \epsilon_\beta) = 0.80$  mostrados en la Figura 3.5.

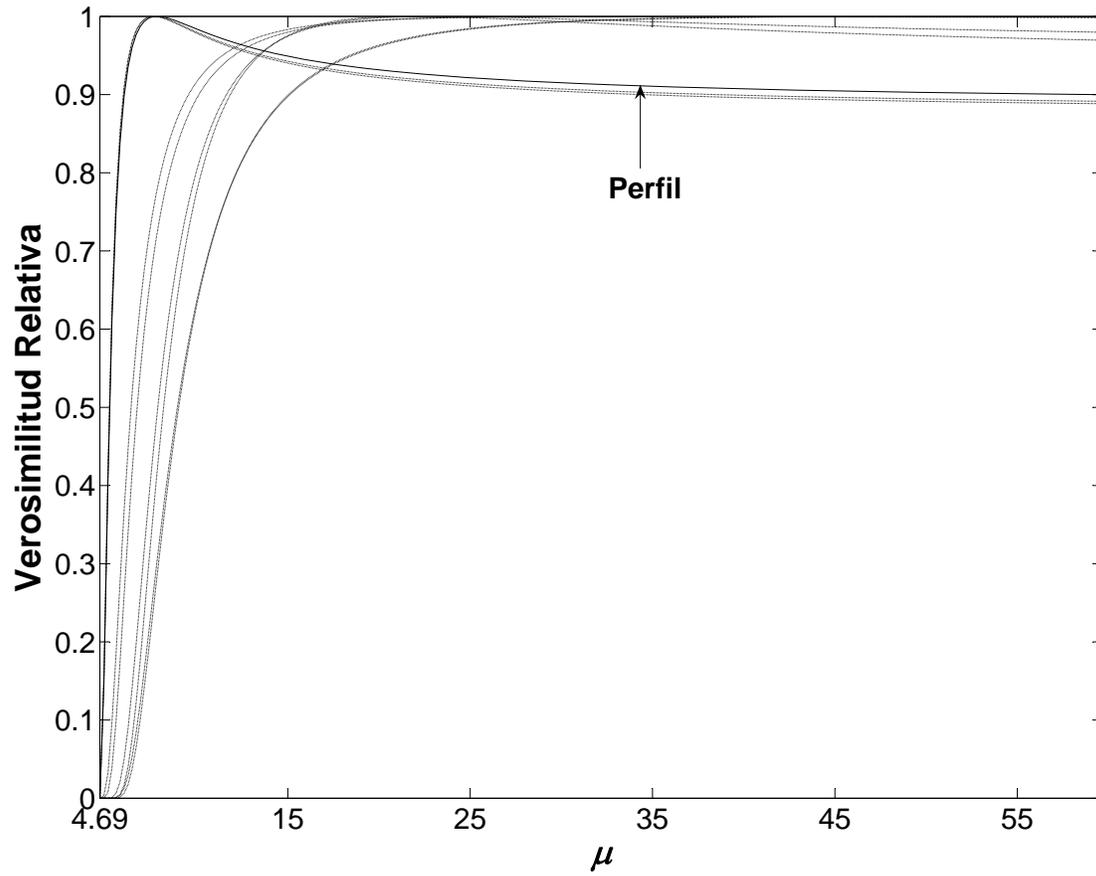


Figura 3.10: Port Pirie: Verosimilitud perfil y ocho verosimilitudes perfiles  $\epsilon$ -perturbadas de  $\mu$  correspondientes a los contornos de  $R(\epsilon_\sigma, \epsilon_\beta) = 0.70$  mostrados en la Figura 3.5.

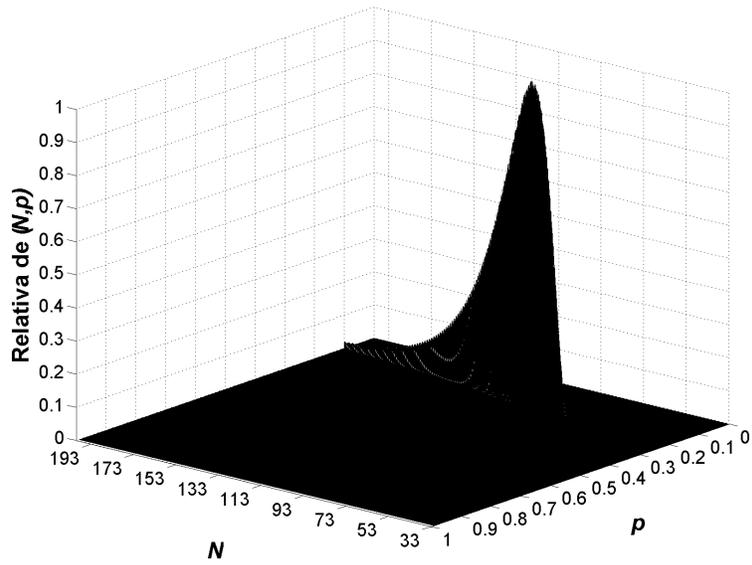


Figura 3.11: Función de verosimilitud global de  $N$  y  $p$  para los datos simulados con  $N = 100$ ,  $k = 1,000$  y  $p = 0.2$ .

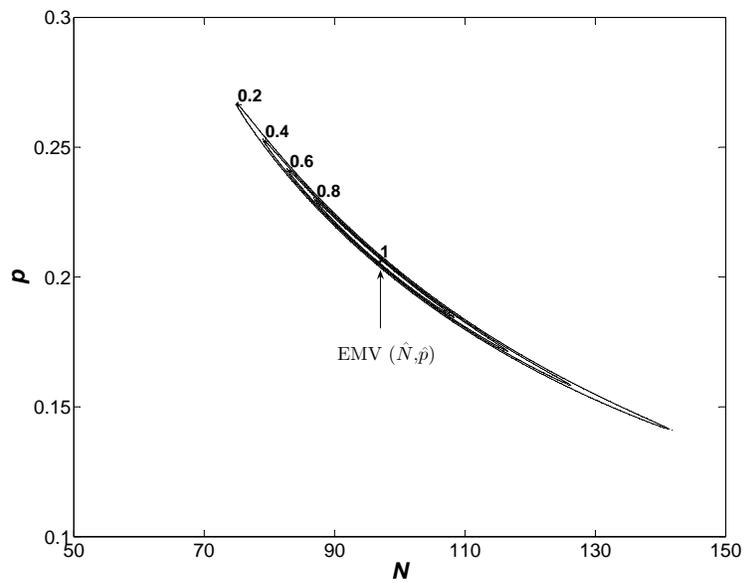


Figura 3.12: Contornos de la función de verosimilitud global de  $N$  y  $p$  para los datos simulados con  $N = 100$ ,  $k = 1,000$  y  $p = 0.2$ .

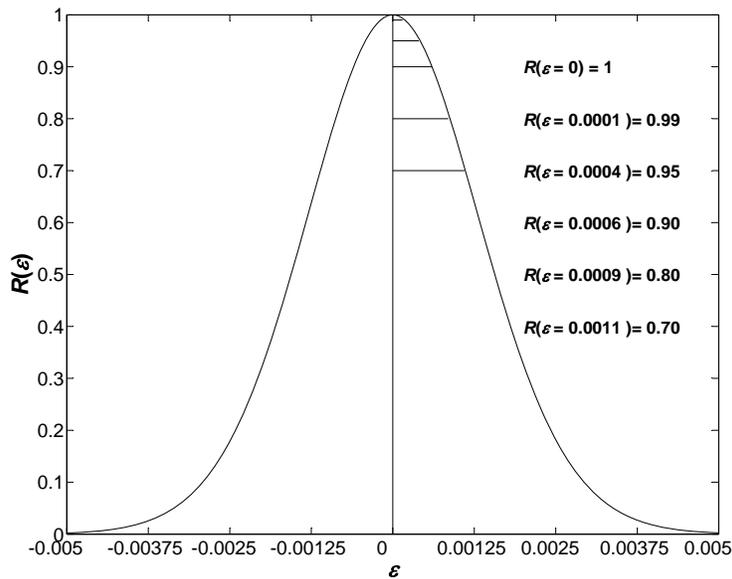


Figura 3.13: Verosimilitud perfil relativa de  $\epsilon$  para los datos simulados con  $N = 100$ ,  $k = 1,000$  y  $p = 0.2$ . Valores de  $\epsilon$  con alta plausibilidad  $c$ .

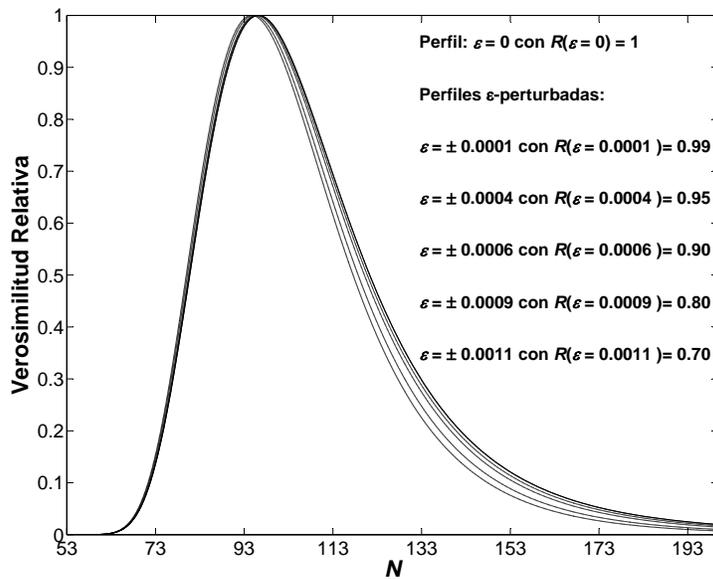


Figura 3.14: Verosimilitudes perfil  $\epsilon$ -perturbadas de  $N$  para los datos simulados con  $N = 100$ ,  $k = 1,000$  y  $p = 0.2$ . Valores de  $\epsilon$  con alta plausibilidad  $c$ .

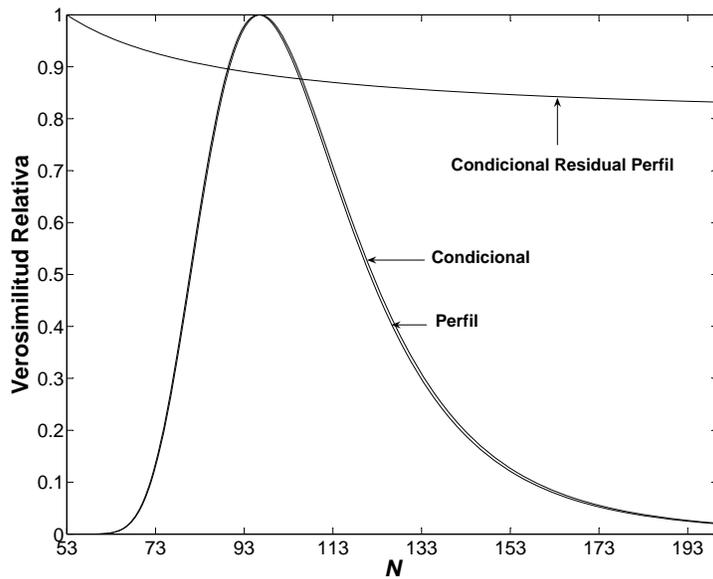


Figura 3.15: Función de verosimilitud perfil, condicional y condicional residual perfil de  $N$  para los datos simulados con  $N = 100$ ,  $k = 1,000$  y  $p = 0.2$ .

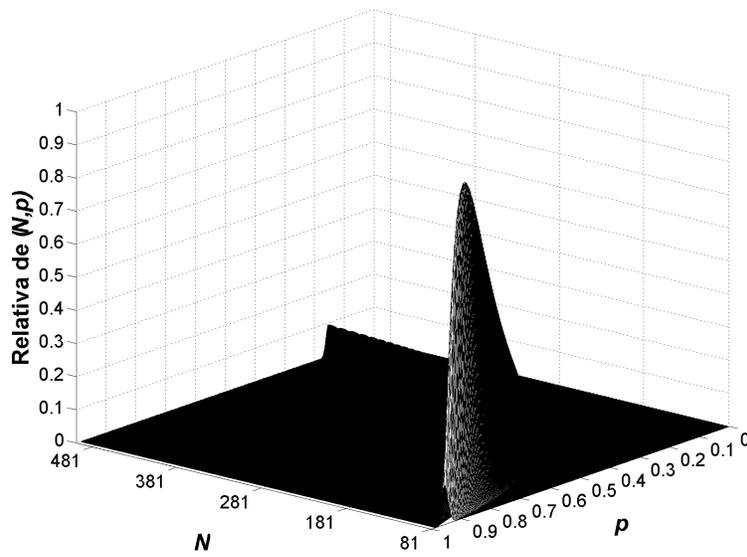


Figura 3.16: Función de verosimilitud global de  $N$  y  $p$  para los datos simulados con  $N = 100$ ,  $k = 5$  y  $p = 0.8$ .

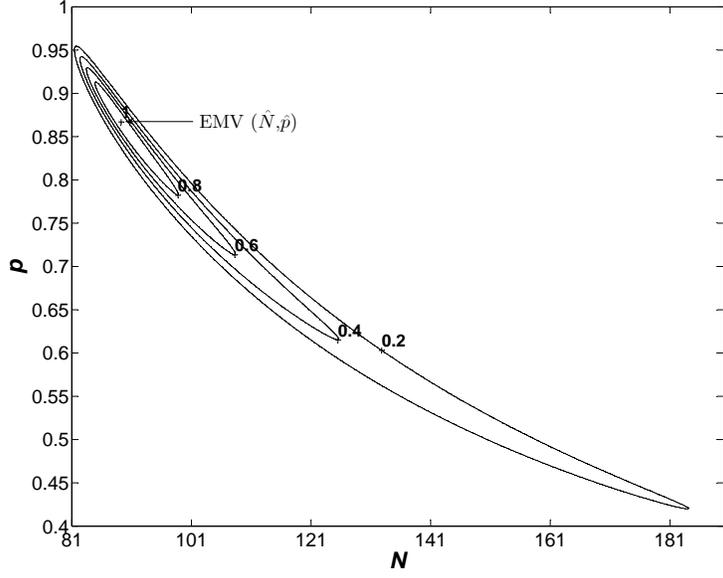


Figura 3.17: Contornos de la función de verosimilitud global de  $N$  y  $p$  para los datos simulados con  $N = 100$ ,  $k = 5$  y  $p = 0.8$ .

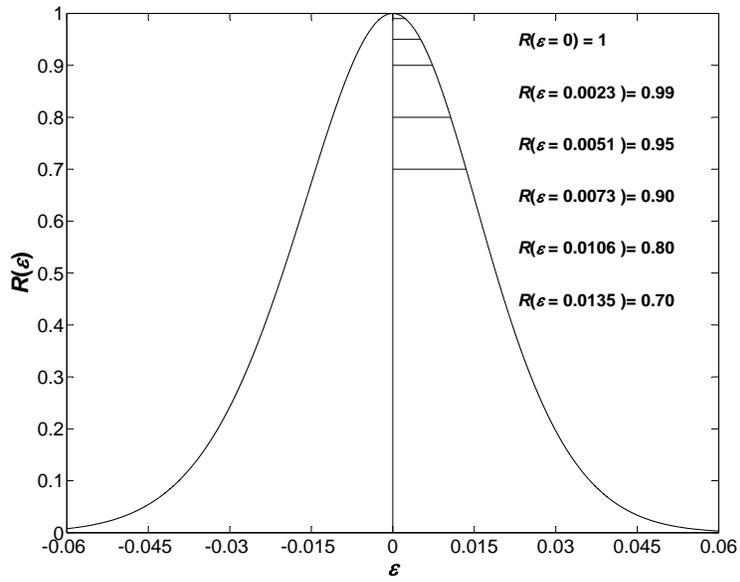


Figura 3.18: Verosimilitud perfil relativa de  $\epsilon$  para los datos simulados con  $N = 100$ ,  $k = 5$  y  $p = 0.8$ . Valores de  $\epsilon$  con alta plausibilidad  $c$ .

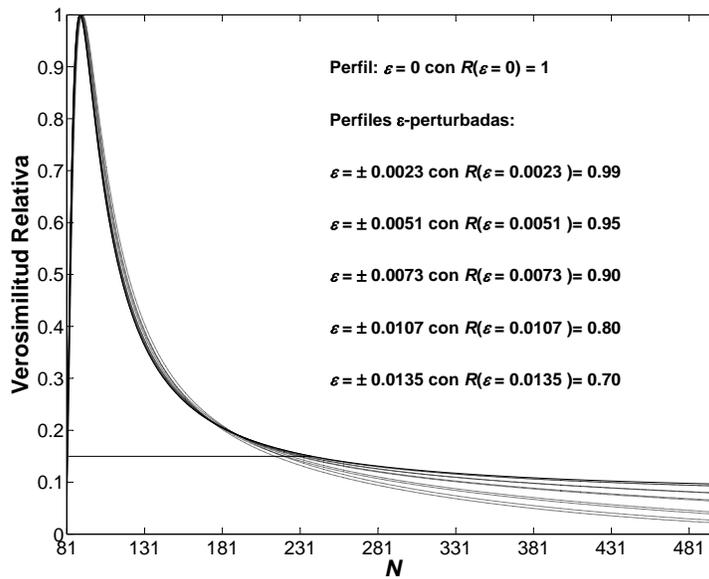


Figura 3.19: Verosimilitudes perfil  $\epsilon$ -perturbadas de  $N$  para los datos simulados con  $N = 100$ ,  $k = 5$  y  $p = 0.8$  e intervalos del 15% de verosimilitud asociados.

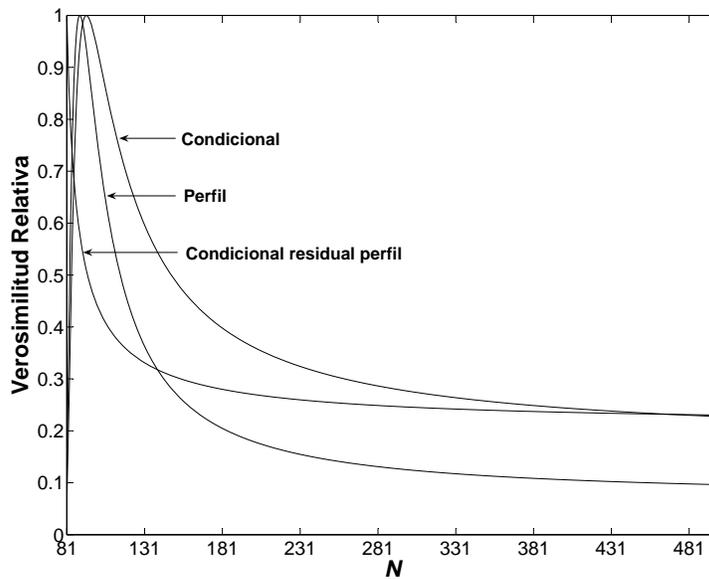


Figura 3.20: Función de verosimilitud perfil, condicional y condicional residual perfil de  $N$  para los datos simulados con  $N = 100$ ,  $k = 5$  y  $p = 0.8$ .

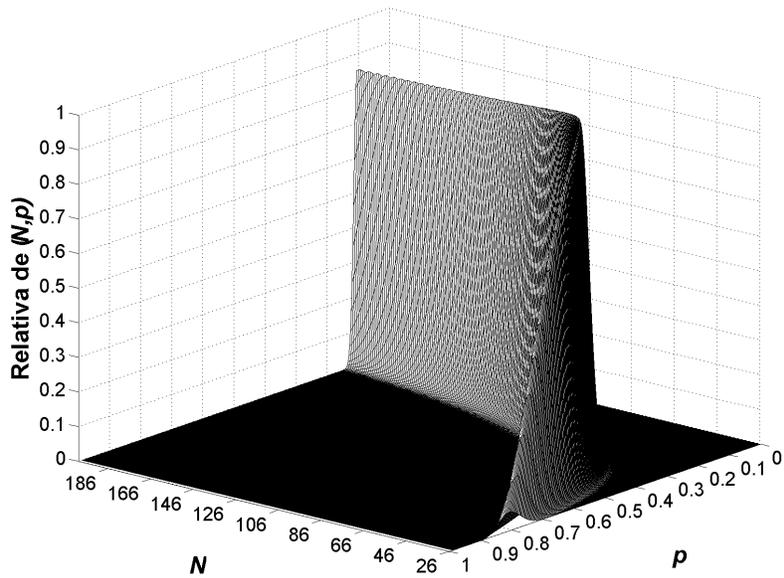


Figura 3.21: Función de verosimilitud global de  $N$  y  $p$  para los datos de impalas.

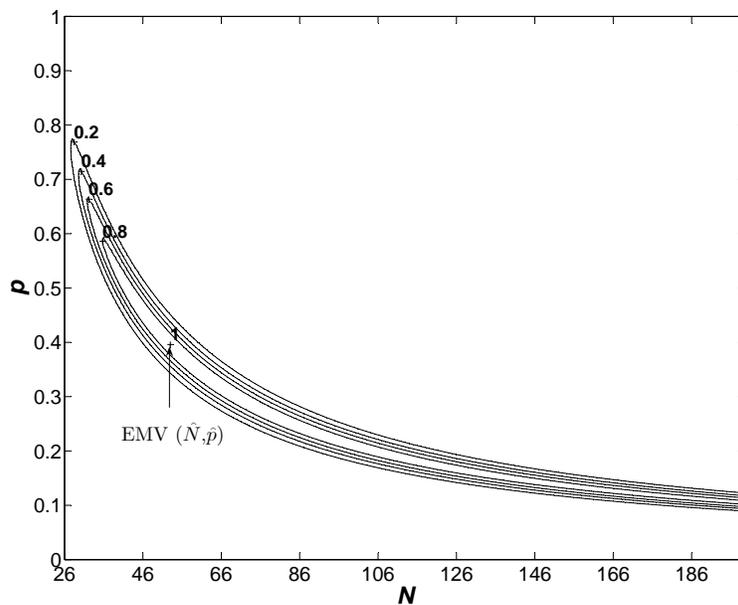


Figura 3.22: Contornos de la función de verosimilitud global de  $N$  y  $p$  para los datos de impalas.

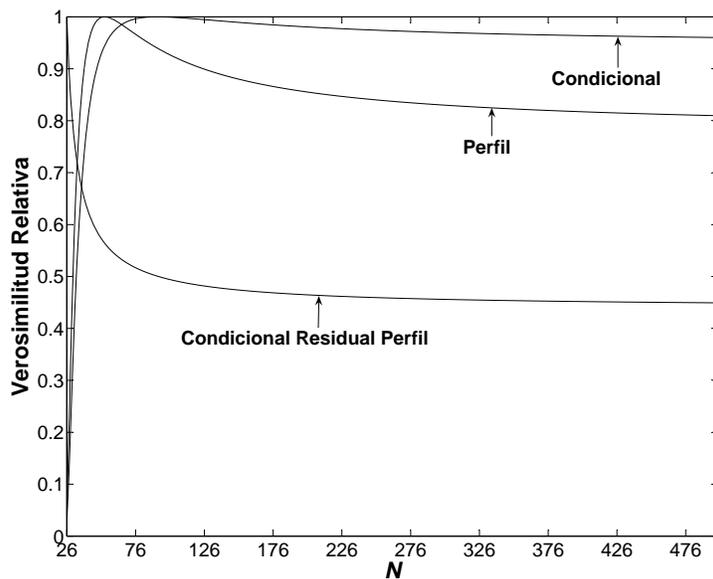


Figura 3.23: Función de verosimilitud perfil, condicional y condicional residual perfil de  $N$  para los datos de impalas.

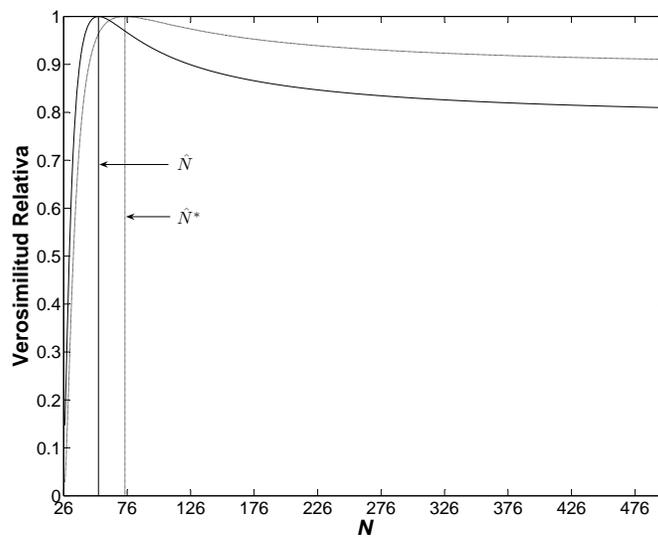


Figura 3.24: Funciones de verosimilitud perfil de  $N$  al perturbar la muestra,  $x_{(5)} = 26$  (—) a  $x_{(5)} = 27$  (- - -), para los datos de impalas.

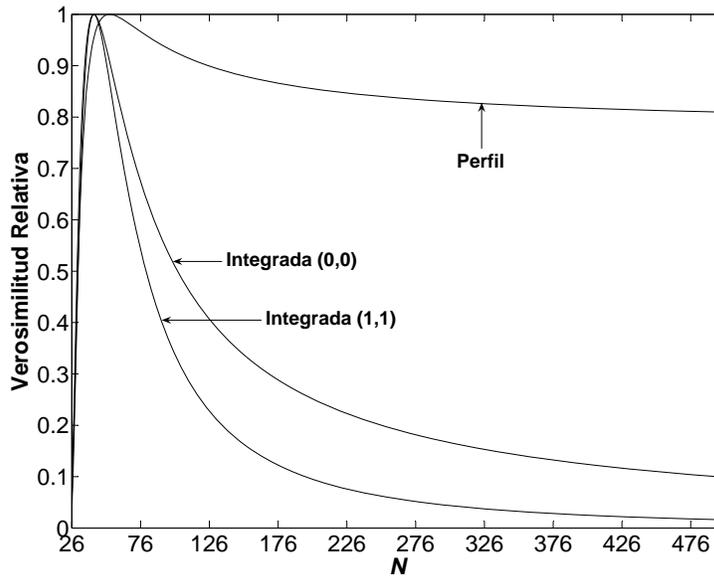


Figura 3.25: Función de verosimilitud perfil e integrada de  $N$  para los datos de impalas.

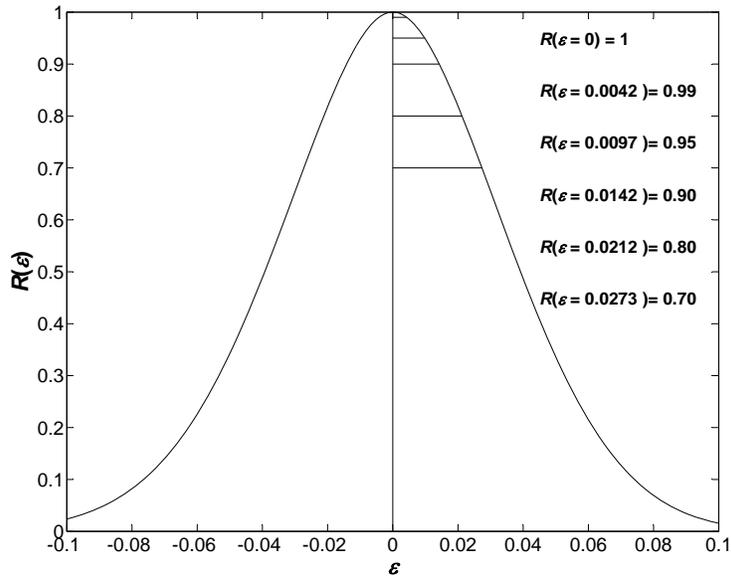


Figura 3.26: Verosimilitud perfil relativa de  $\epsilon$  para los datos de impalas. Valores de  $\epsilon$  con alta plausibilidad  $c$ .

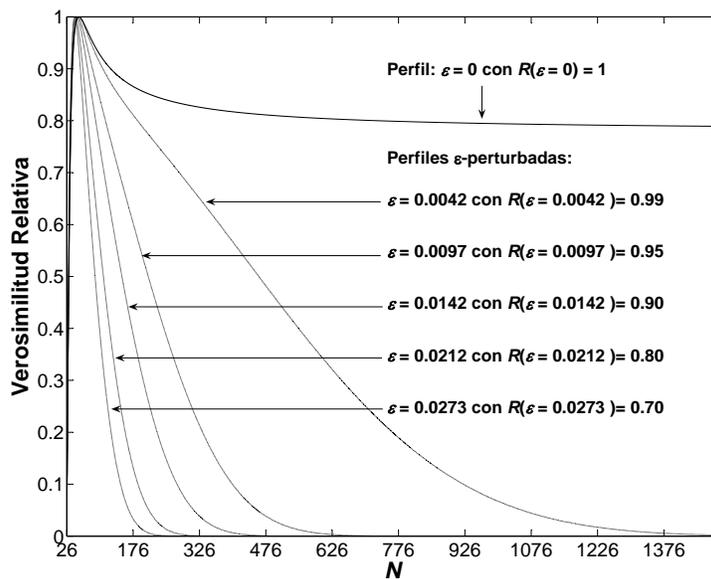


Figura 3.27: Verosimilitudes perfil  $\epsilon$ -perturbadas de  $N$  para los datos de impalas. Valores de  $\epsilon$  con alta plausibilidad  $c$ .

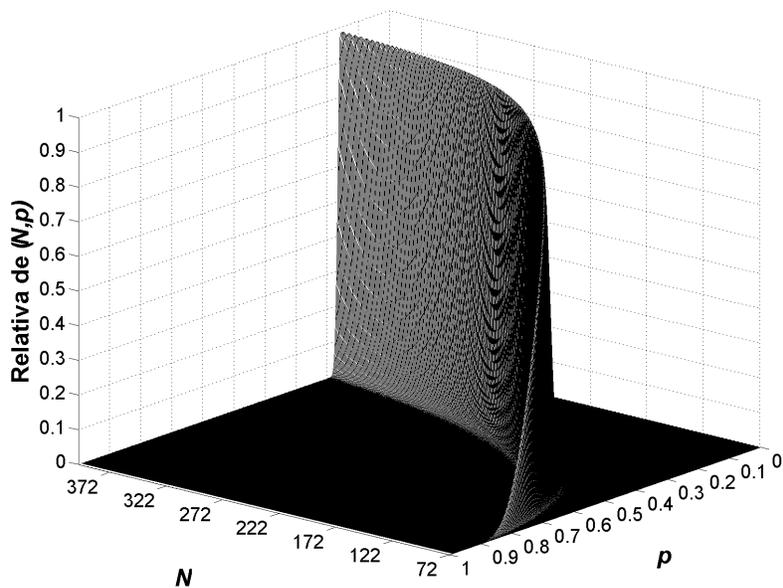


Figura 3.28: Función de verosimilitud global de  $N$  y  $p$  para los datos de antílopes.

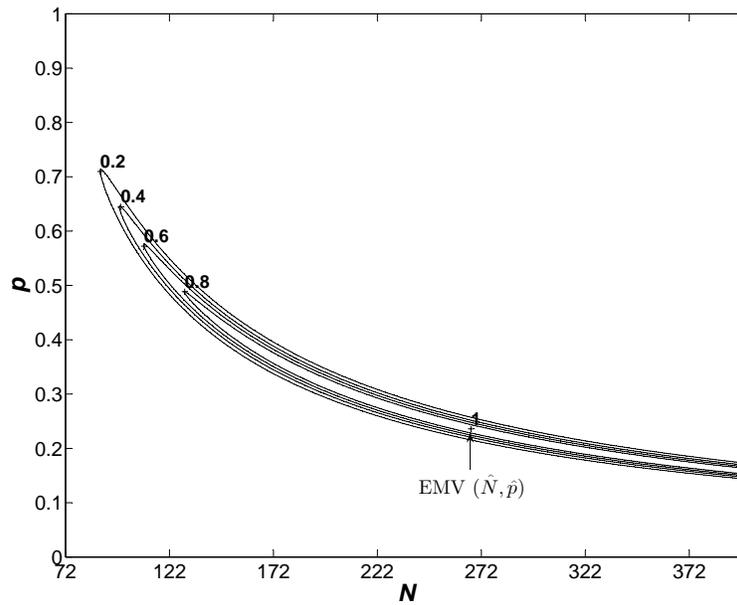


Figura 3.29: Contornos de la función de verosimilitud global de  $N$  y  $p$  para los datos de antílopes.

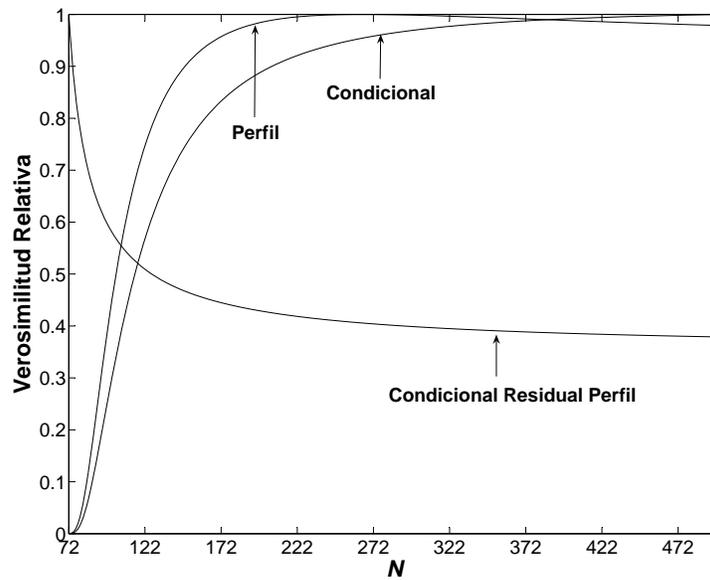


Figura 3.30: Función de verosimilitud perfil, condicional y condicional residual perfil de  $N$  para los datos de antílopes.

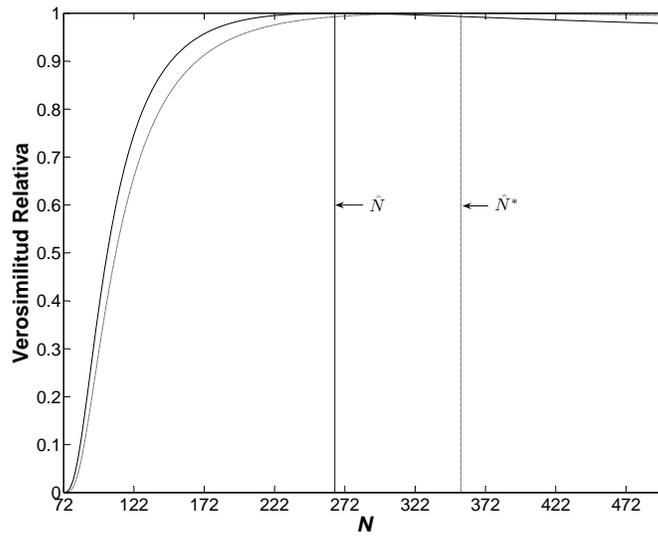


Figura 3.31: Funciones de verosimilitud perfil de  $N$  al perturbar la muestra,  $x_{(5)} = 72$  (—) a  $x_{(5)} = 73$  (- - -), para los datos de antílopes.

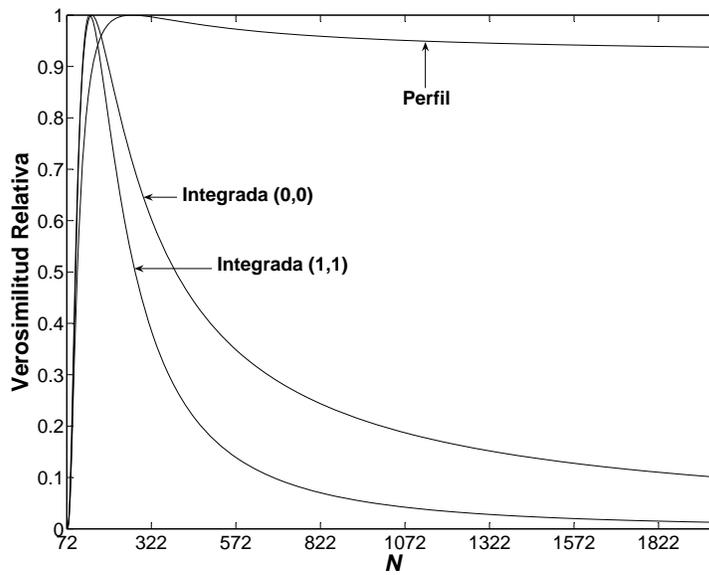


Figura 3.32: Función de verosimilitud perfil e integrada de  $N$  para los datos de antílopes.

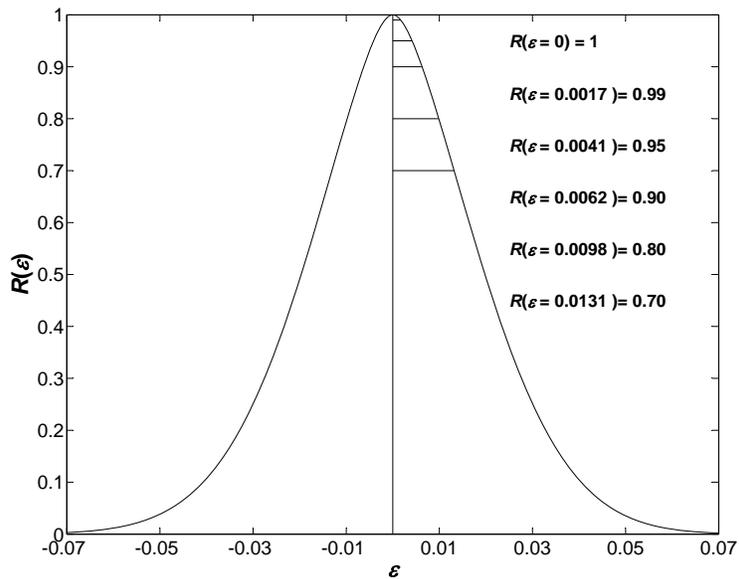


Figura 3.33: Verosimilitud perfil relativa de  $\epsilon$  para los datos de antílopes. Valores de  $\epsilon$  con alta plausibilidad  $c$ .

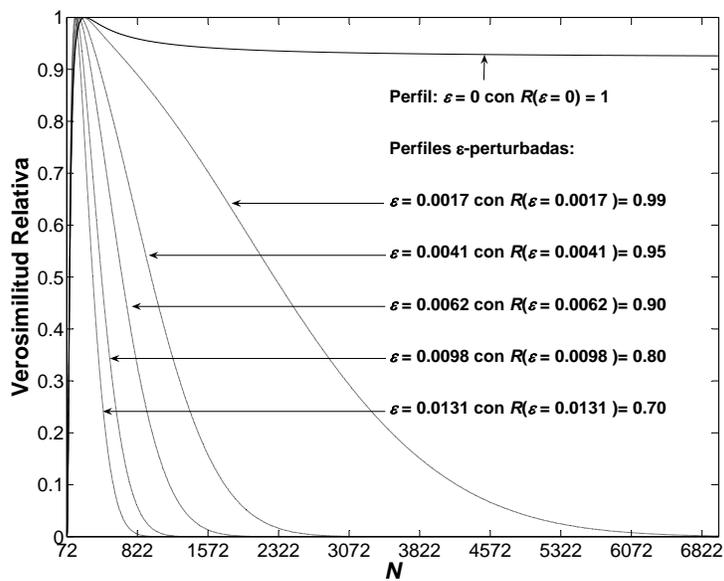


Figura 3.34: Verosimilitudes perfil  $\epsilon$ -perturbadas de  $N$  para los datos de antílopes. Valores de  $\epsilon$  con alta plausibilidad  $c$ .

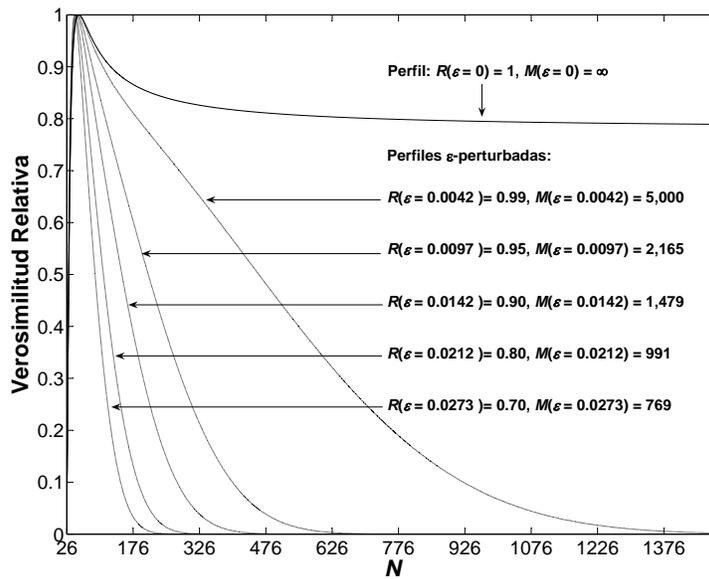


Figura 3.35: Verosimilitudes perfil  $\epsilon$ -perturbadas de  $N$  correspondientes a distintos valores de  $\epsilon$  con plausibilidad  $c$  y asociados a una cota superior  $M$  para  $N$  para los datos de impalas.

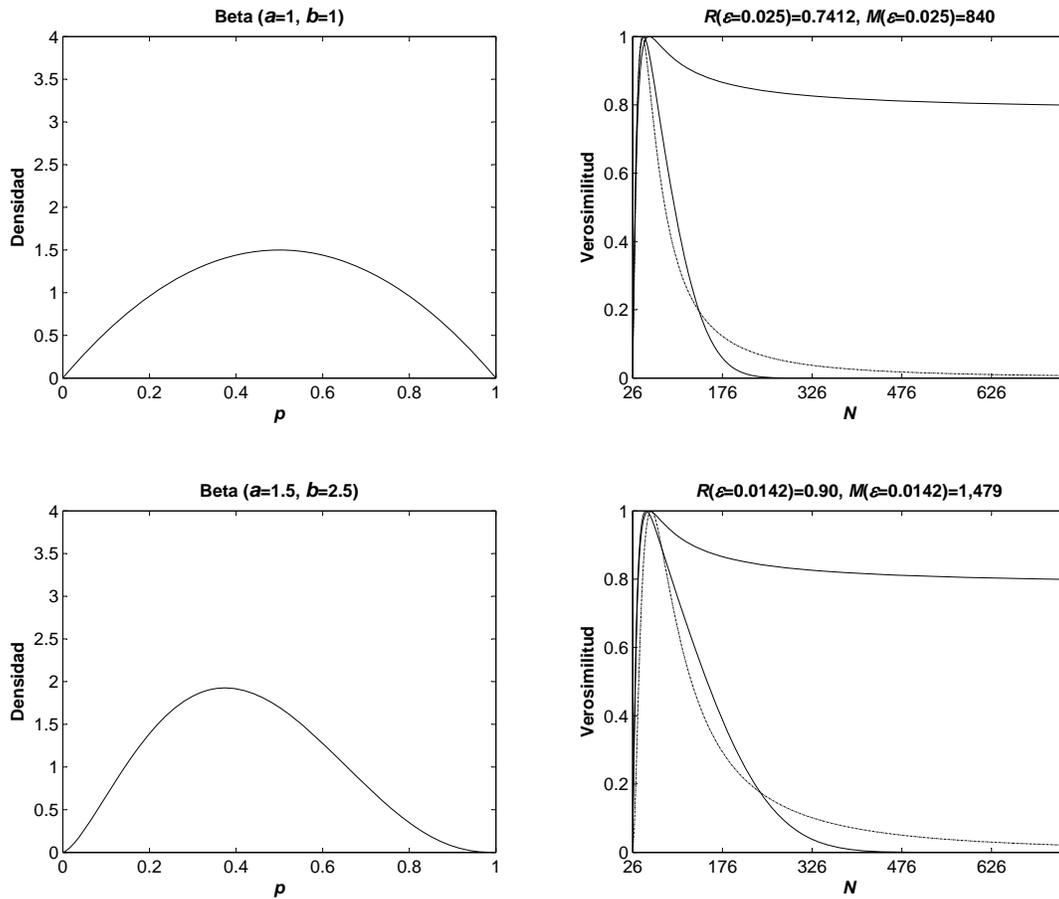


Figura 3.36: A la izquierda se presenta la densidad previa Beta de  $p$ . A la derecha se muestra, en la misma gráfica, la correspondiente verosimilitud integrada (---) y la verosimilitud perfil  $\epsilon$ -perturbada de  $N$  (—) para los datos de impalas.

# Capítulo 4

## Uso de la verosimilitud perfil para simplificar la inferencia de un parámetro de interés

### 4.1 Introducción

En este capítulo se mostrará como la verosimilitud perfil puede ayudar a simplificar enormemente las inferencias sobre un parámetro de interés. En general, es recomendable dar dichas inferencias de la manera algebraicamente más simple y clara. Aquí se considerará, a manera de ejemplo, el problema de hacer inferencia sobre el parámetro de confiabilidad en el modelo tensión-fuerza a través de la verosimilitud perfil.

En general, la confiabilidad de un componente electrónico es la capacidad que tiene para realizar y mantener sus funciones en circunstancias normales de trabajo, así como en circunstancias hostiles o inesperadas. Con frecuencia se reporta en términos de la probabilidad de que el componente no falle.

El modelo tensión-fuerza (*'stress-strength'*) es una abstracción matemática de la confiabilidad de un componente. En este modelo la confiabilidad de un componente

se puede evaluar en términos de las variables aleatorias  $X$  y  $Y$ , donde  $X$  representa la tensión que se ejerce sobre el componente y  $Y$  representa la fuerza que tiene el componente para soportar la tensión. Si la tensión excede o sobrepasa a la fuerza del componente ( $X > Y$ ), entonces el componente fallará. En caso contrario, el componente no falla. Así, la confiabilidad del componente es la probabilidad de que no falle, la cual se representa a través del parámetro fijo y desconocido  $\theta = P(X < Y)$  al cual se le llama parámetro de confiabilidad.

Por definición,  $\theta = P(X < Y)$  toma valores en el intervalo  $[0, 1]$  y valores grandes de éste parámetro (cercaos a uno) son buenos porque significan que con alta probabilidad el componente soporta la tensión. En contraste, valores pequeños de  $\theta$  (cercaos a cero) son muy malos. En términos de las densidades de  $X$  y de  $Y$  valores grandes o pequeños de  $\theta$  significan que éstas densidades se intersecan muy poco. Por otro lado, cuando  $\theta$  toma valores cercaos a 0.5 entonces las densidades de  $X$  y de  $Y$  se intersecan casi por completo.

Birnbaum (1956) presentó esta idea por vez primera y posteriormente la desarrollaron en Birnbaum y McCarty (1958) la desarrollaron. Sin embargo el término tensión-fuerza aparece por primera vez en Church y Harris (1970). El desarrollo de la probabilidad y de la estadística para los modelos tensión-fuerza, durante las últimas cuatro décadas, se resume de manera muy completa en el libro de Kotz *et al.* (2003). En este libro se presentan muchos resultados teóricos sobre el modelo tensión-fuerza; así como algunas de sus aplicaciones en la industria, la economía y la medicina.

Aunque nuestra motivación para hacer inferencia sobre el parámetro  $\theta = P(X < Y)$  es el modelo tensión-fuerza, existen otros escenarios o contextos donde este parámetro resulta ser de interés. Por ejemplo, cuando  $X$  y  $Y$  son variables aleatorias que representan el Índice de Masa Corporal (IMC) de mujeres y hombres en una región específica, o cuando representan el tiempo de vida de un componente electrónico fabricado por dos proveedores distintos, etc. Por esta razón, en lo que sigue de este capítulo se con-

siderará el problema general de hacer inferencia sobre el parámetro  $\theta = P(X < Y)$  basados en dos muestras independientes observadas  $x = (x_1, \dots, x_n)$  y  $y = (y_1, \dots, y_m)$ , con  $n$  y  $m$  no necesariamente iguales, a través de la función de verosimilitud perfil, sin estar forzosamente bajo un modelo de tensión-fuerza. Cuando los tamaños de muestras sean iguales tampoco se supondrá que se tienen datos pareados.

El problema de hacer inferencia sobre el parámetro de confiabilidad  $\theta = P(X < Y)$ , cuando  $X$  y  $Y$  son variables aleatorias con función de densidad conjunta de probabilidad  $f(\cdot; \varphi)$ , donde  $\varphi \in \mathbb{R}^k$  es un vector de parámetros desconocidos, ha sido ampliamente discutido en la literatura estadística, Tong (1974, 1975), Chao (1982), Reiser y Guttman (1986), Kundu y Gupta (2005), Adimari y Chiogna (2005), Gupta y Kundu (2007), etc. El enfoque usual para abordar este problema es calcular estimadores puntuales eficientes para  $\theta$ , por ejemplo estimadores de máxima verosimilitud, estimadores insesgados de mínima varianza, estimadores de Bayes, etc. Además, se calculan intervalos de confianza a través de cantidades pivotaes, cuando es posible, o intervalos de confianza asintóticos. Poco o casi nada se ha hecho desde el enfoque de verosimilitud. Es decir, no se ha considerado a la función de verosimilitud perfil de  $\theta$  completa, sino que sólo el valor donde alcanza su máximo, el estimador de máxima verosimilitud (emv), para hacer inferencia sobre el parámetro de confiabilidad  $\theta$ .

En este capítulo se considera el problema de estimación estadística del parámetro  $\theta = P(X < Y)$  a través de la función de verosimilitud perfil de este parámetro de interés para dos casos muy básicos, ampliamente estudiados e importantes. El primero es cuando  $X$  y  $Y$  son variables aleatorias independientes con función de densidad exponencial; el segundo caso es cuando  $X$  y  $Y$  son normales.

## 4.2 Caso Exponencial

Supóngase que  $X$  y  $Y$  son dos variables aleatorias independientes con función de densidad  $f_X(x; \alpha) = (1/\alpha) \exp(-x/\alpha)$  y  $f_Y(y; \beta) = (1/\beta) \exp(-y/\beta)$ , donde  $\alpha, \beta \in \mathbb{R}^+$  y son parámetros fijos y desconocidos. Se tiene interés en hacer inferencias sobre el parámetro  $\theta = P(X < Y; \alpha, \beta)$  basados en dos muestras independientes observadas  $x = (x_1, \dots, x_n)$  y  $y = (y_1, \dots, y_m)$  provenientes de  $f_X(x; \alpha)$  y  $f_Y(y; \alpha)$ , respectivamente.

Por definición  $\theta = P(X < Y; \alpha, \beta)$  se puede calcular como

$$\theta = P(X < Y; \alpha, \beta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y; \alpha, \beta) I(x < y) dx dy, \quad (4.1)$$

donde  $f_{X,Y}(x, y; \alpha, \beta)$  es la densidad conjunta del vector aleatorio  $(X, Y)$  y  $I(\cdot)$  es la función indicadora. Como  $X$  y  $Y$  son variables exponenciales independientes, entonces son no negativas y su distribución conjunta es el producto de sus distribuciones,  $f_{X,Y}(x, y; \alpha, \beta) = f_X(x; \alpha) f_Y(y; \beta)$ . Así, el parámetro de interés  $\theta$  dado en (4.1) se puede escribir como

$$\begin{aligned} \theta &= \int_0^{\infty} \int_0^y f_X(x; \alpha) f_Y(y; \beta) dx dy \\ &= \int_0^{\infty} \left[ \int_0^y f_X(x; \alpha) dx \right] f_Y(y; \beta) dy \\ &= \int_0^{\infty} F_X(y; \alpha) f_Y(y; \beta) dy, \end{aligned}$$

donde  $F_X(\cdot; \alpha)$  es la función de distribución de  $X$ . Entonces, el parámetro de interés es

$$\theta = \int_0^{\infty} \left[ 1 - \exp\left(-\frac{y}{\alpha}\right) \right] \frac{1}{\beta} \exp\left(-\frac{y}{\beta}\right) dy = \frac{\beta}{\beta + \alpha}. \quad (4.2)$$

Para interpretar el parámetro de interés  $\theta = P(X < Y; \alpha, \beta)$  en términos de los tiempos medios de vida  $\alpha$  y  $\beta$  de las variables  $X$  y  $Y$ , es conveniente escribir (4.2) de la siguiente forma:

$$\theta = \frac{1}{1 + \frac{\alpha}{\beta}}.$$

Nótese que el cociente  $\alpha/\beta$  es siempre mayor que cero porque  $\alpha$  y  $\beta$  son mayores que cero. Así, valores grandes de  $\theta$  (cercaos a uno) se conseguirán cuando el cociente  $\alpha/\beta$  sea cercano a cero. Esto ocurrirá cuando el tiempo medio de vida de  $Y$ ,  $E[Y] = \beta$ , sea grande en comparación con el tiempo medio de vida de  $X$ ,  $E[X] = \alpha$ . En contraste, valores pequeños de  $\theta$  (cercaos a cero), se conseguirán cuando el cociente  $\alpha/\beta$  sea grande. Esto ocurrirá cuando  $\beta$ , sea pequeño en comparación con  $\alpha$ . Por otro lado,  $\theta$  tomará valores cercaos a 0.5 cuando el cociente  $\alpha/\beta$  sea cercano a uno. Esto ocurrirá cuando  $\alpha$  y  $\beta$  tengan valores muy parecidos. La siguiente tabla ejemplifica estas situaciones.

Tabla 4.1. Comportamiento del parámetro  $\theta$  para diferentes valores del cociente  $\alpha/\beta$ .

$\alpha = E[X]$	$\beta = E[Y]$	$\alpha/\beta$	$\theta = P(X < Y)$
10	990	0.0101	0.99
10	190	0.0526	0.95
10	90	0.1111	0.90
10	56.67	0.1765	0.85
10	40	0.25	0.80
10	30	0.3333	0.75
10	23.33	0.4286	0.70
10	15	0.6667	0.60
10	10	1	0.50

En la Figura 4.1 se presentan las gráficas de las densidades de  $X$  y de  $Y$  correspondientes a valores de  $\theta = 0.95, 0.90, 0.75, 0.60$  dados en la Tabla 4.1. Esto con el objetivo de mostrar la información que proporciona el parámetro  $\theta = P(X < Y)$  en términos de las densidades de  $X$  y  $Y$ .

Valores de  $\theta$  cercaos a uno significan que la densidad de  $Y$  tiene una cola derecha

más pesada que la densidad de  $X$ , y que el traslape de ambas densidades es pequeño. De manera análoga, se sigue que valores pequeños de  $\theta$  significan que ahora la densidad de  $X$  tiene una cola derecha más pesada que la densidad de  $Y$ , y que nuevamente el traslape de ambas densidades es pequeño. Por otro lado, valores de  $\theta$  cercanos a 0.5 significan que ambas densidades se traslapan mucho. De hecho,  $\theta = 0.5$  se obtiene cuando  $\alpha = \beta$ . Es decir, cuando  $X$  y  $Y$  tienen la misma densidad.

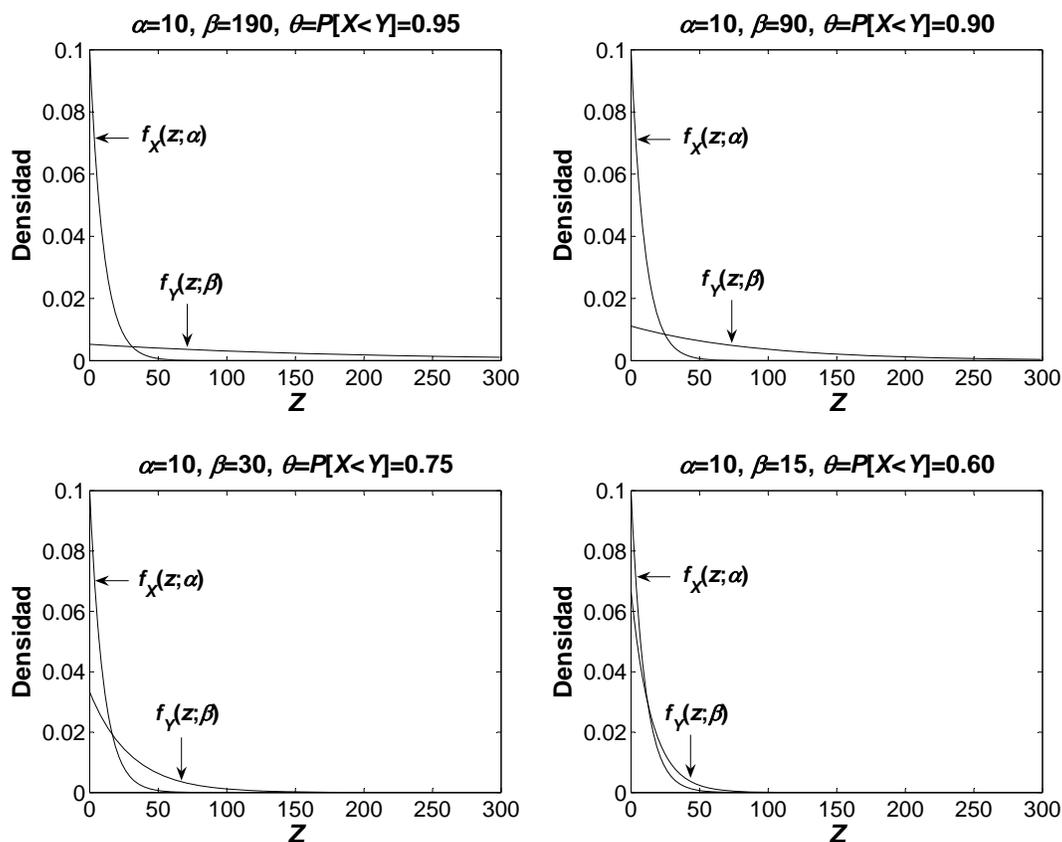


Figura 4.1: Funciones de densidad de  $X$  y de  $Y$  correspondientes a valores de  $\theta = 0.95, 0.90, 0.75, 0.60$  dados en la Tabla 4.1.

En la siguiente sección se presenta una breve revisión de diferentes enfoques estadísticos empleados para hacer inferencia sobre el parámetro de interés  $\theta$  dado en

(4.2).

### 4.2.1 Enfoques comúnmente usados para inferencias sobre el parámetro $\theta = P(X < Y)$

#### Estimación puntual

En la literatura estadística varios autores centran su atención en construir estimadores insesgados de mínima varianza (iminvar) para  $\theta$ , en aproximar el error cuadrático medio del emv de  $\theta$ ,  $\text{ecm}(\hat{\theta}) = E(\hat{\theta} - \theta)^2$  o en construir cotas superiores para éste, Tong (1974, 1975), Kelley *et al.* (1976), Sathe y Shah (1981), Chao (1982), Jana (1997), Kotz *et al.* (2003). Por ejemplo, Tong (1974) presentó el emv de  $\theta$ ,

$$\hat{\theta} = \frac{\bar{y}}{\bar{x} + \bar{y}}, \quad (4.3)$$

donde

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{y} \quad \bar{y} = \frac{1}{m} \sum_{i=1}^m y_i,$$

y el estimador iminvar para  $\theta$ ,

$$\tilde{\theta} = \begin{cases} Q_1(n, m, n\bar{x}, m\bar{y}), & \text{si } m\bar{y} \leq n\bar{x}, \\ Q_2(n, m, n\bar{x}, m\bar{y}), & \text{si } m\bar{y} > n\bar{x}, \end{cases} \quad (4.4)$$

donde

$$Q_1(a, b, c, d) = \sum_{i=0}^{a-2} (-1)^i \frac{\Gamma(a) \Gamma(b)}{\Gamma(a-i-1) \Gamma(b+i+1)} \left(\frac{d}{c}\right)^{i+1},$$

$$Q_2(a, b, c, d) = \sum_{i=0}^{a-2} (-1)^i \frac{\Gamma(a) \Gamma(b)}{\Gamma(a+i) \Gamma(b-i)} \left(\frac{c}{d}\right)^i.$$

Chao (1982), bajo la reparametrización del modelo exponencial correspondiente a  $E[X] = 1/\alpha'$  y  $E[Y] = 1/\beta'$ , utilizó la expansión en series de Taylor de  $(\hat{\theta} - \theta)$  alrededor de  $(\bar{x} - 1/\alpha')$  y  $(\bar{y} - 1/\beta')$  para calcular, solamente cuando  $n$  es igual a  $m$ ,

la siguiente expresión asintótica para el  $\text{ecm}(\widehat{\theta})$ ,

$$\text{ecm}(\widehat{\theta}) = 2\tau^2 (1 + \tau)^{-4} n^{-1} + 4\tau^2 (2\tau - 1) (\tau - 2) (1 + \tau)^{-6} n^{-2} + o(n^{-2}), \quad (4.5)$$

donde  $\tau = \alpha' / \beta'$ . Nótese que  $\widehat{\alpha}' = 1/\bar{x}$  y  $\widehat{\beta}' = 1/\bar{y}$  son los estimadores de máxima verosimilitud de  $\alpha'$  y  $\beta'$ , respectivamente. Entonces, estimando  $\tau$  con  $\widehat{\tau} = \widehat{\alpha}' / \widehat{\beta}' = \bar{y} / \bar{x}$  y reemplazando  $\widehat{\tau}$  en (4.5) se obtiene el siguiente estimador para el  $\text{ecm}(\widehat{\theta})$  dado en (4.5),

$$\widehat{\sigma}^2 = \widehat{\text{ecm}(\widehat{\theta})} = 2\widehat{\tau}^2 (1 + \widehat{\tau})^{-4} n^{-1} + 4\widehat{\tau}^2 (2\widehat{\tau} - 1) (\widehat{\tau} - 2) (1 + \widehat{\tau})^{-6} n^{-2}. \quad (4.6)$$

Recientemente, Kotz *et al.* (2003, pág. 20-22) calculó el estimador iminvar de la varianza de  $\tilde{\theta}$  dado en (4.4),

$$\tilde{\sigma}^2 = \widehat{\text{Var}(\tilde{\theta})} = \tilde{\theta}^2 - \frac{(n-1)(n-2)(m-1)(m-2)}{n^2 m^2 \bar{x}^{n-1} \bar{y}^{m-1}} H(n, m, \bar{x}, \bar{y}), \quad (4.7)$$

donde  $H(n, m, \bar{x}, \bar{y})$  esta dada por la siguiente integral múltiple

$$H(n, m, \bar{x}, \bar{y}) = \iiint\limits_W \left( \bar{x} - \frac{w_1 + w_2}{n} \right)^{n-3} \left( \bar{y} - \frac{w_3 + w_4}{m} \right)^{m-3} dw_1 dw_2 dw_3 dw_4, \quad (4.8)$$

donde

$$W = \{(w_1, w_2, w_3, w_4) : w_1 + w_2 < n\bar{x}, w_3 + w_4 < m\bar{y}, \\ 0 < w_1 < w_3, 0 < w_2 < w_4\}.$$

En general, la integral en (4.8) no tiene una forma matemática cerrada y se tiene que calcular de forma numérica. Nótese que ésta no es una tarea fácil y puede demandar mucho tiempo y esfuerzo. Hay que integrar numéricamente un polinomio de grado  $(n-3)(m-3)$ , que puede ser muy alto, sobre una región restringida  $W \subset R^4$ .

Los resultados mostrados en esta sección son de interés porque, como se verá más adelante en esta misma sección, dan la base para construir intervalos de confianza asintóticos para  $\theta$ .

## Intervalos de confianza a través de una cantidad pivotal

La distribución exponencial es uno de los casos excepcionales donde se pueden construir intervalos de confianza para  $\theta$  a partir del pivotal

$$\varsigma = \left[ 1 + \frac{m\hat{\theta}(1-\theta)}{n\theta(1-\hat{\theta})} \right]^{-1},$$

que, bajo la reparametrización del modelo exponencial correspondiente a  $E[X] = 1/\alpha'$  y  $E[Y] = 1/\beta'$ , tiene una distribución Beta con parámetros conocidos  $n$  y  $m$ . Par ver que  $\varsigma$  tiene dicha distribución, basta notar que se puede escribir como  $\varsigma = \alpha'n\bar{x}/(\alpha'n\bar{x} + \beta'm\bar{y})$ , donde  $\alpha'n\bar{x}$  y  $\beta'm\bar{y}$  tienen distribución Gamma con parámetros  $(n, 1)$  y  $(m, 1)$ , respectivamente. Así, a partir de esta cantidad pivotal, un intervalo de confianza del  $100(1 - \gamma)\%$  para  $\theta$  es

$$P \left[ \frac{m\hat{\theta}q_1}{n(1-\hat{\theta})(1-q_1) + m\hat{\theta}q_1} < \theta < \frac{m\hat{\theta}q_2}{n(1-\hat{\theta})(1-q_2) + m\hat{\theta}q_2} \right] = 1 - \gamma, \quad (4.9)$$

donde  $q_1$  y  $q_2$  satisfacen que

$$I_{q_1}(n, m) - I_{q_2}(n, m) = 1 - \gamma, \quad (4.10)$$

y donde

$$I_z(n, m) = \frac{\Gamma(n+m)}{\Gamma(n)\Gamma(m)} \int_0^z w^{n-1} (1-w)^{m-1} dw.$$

Nótese que el intervalo de confianza para  $\theta$  dado en (4.9) es válido para cualquier  $n$  y  $m$ . Por otro lado, para calcularlo se necesita resolver la ecuación (4.10) para  $q_1$  y  $q_2$ . Sin embargo, esta ecuación no tiene una solución única. La gente que usa este enfoque busca el intervalo de confianza para  $\theta$  de nivel  $\gamma$  que tenga mínima longitud; es decir, se tiene que resolver la ecuación (4.10) restringiéndose a que la distancia entre  $q_1$  y  $q_2$  sea mínima. Kotz *et al.* (2003, pág. 37) comentan que este problema de optimización numérica frecuentemente no es trivial y proponen reemplazar estos intervalos de confianza por intervalos de confianza asintóticos.

## Intervalos de confianza asintóticos

Kotz *et al.* (2003) presentan dos intervalos de confianza asintóticos para  $\theta$ . Ambos intervalos son simétricos, uno alrededor de  $\tilde{\theta}$  y el otro alrededor de  $\hat{\theta}$ , y tienen probabilidad de cobertura aproximada de  $1 - \gamma$ . El primero es

$$\theta \in \tilde{\theta} \pm z_{\gamma/2} \tilde{\sigma}, \quad (4.11)$$

donde  $z_{\gamma/2}$  es el cuantil  $(1 - \gamma/2)$  de una distribución normal estándar,  $\tilde{\theta}$  es el estimador ininvar de  $\theta$  dado en (4.4), y  $\tilde{\sigma}^2$  es el estimador ininvar de la varianza de  $\tilde{\theta}$  dado en (4.7). Kotz *et al.* (2003), proponen estos intervalos en (4.11) y dan una nota de advertencia sobre el esfuerzo numérico substancial que se requiere para calcularlos, incluso con las facilidades computacionales modernas.

El segundo intervalo propuesto por Kotz *et al.* (2003) es para el caso de tamaño de muestras iguales  $n = m$ . Este intervalo es

$$\theta \in \hat{\theta} \pm z_{\gamma/2} \hat{\sigma}, \quad (4.12)$$

donde  $\hat{\theta}$  es el emv de  $\theta$  dado en (4.3) y  $\hat{\sigma}^2$  es un estimador del ecm( $\hat{\theta}$ ) dado en (4.6).

Aquí se desea resaltar que en general, los intervalos de confianza simétricos pueden ser muy engañosos ya que pueden sub o sobre estimar al parámetro de interés si la función de verosimilitud presenta una fuerte asimetría. La varianza de un estimador por sí sola no incorpora información acerca de la asimetría de la verosimilitud o de su forma.

A continuación se presenta la sección central de este capítulo donde se presenta la propuesta de inferencia del parámetro de interés  $\theta$  a través de su función de verosimilitud perfil completa.

## 4.2.2 Inferencia sobre $\theta$ a través de la verosimilitud perfil

La función de verosimilitud global de  $\alpha$  y  $\beta$  es

$$\begin{aligned} L(\alpha, \beta; x, y) &\propto \prod_{i=1}^n f_X(x_i; \alpha) \prod_{j=1}^m f_Y(y_j; \beta) \\ &= \alpha^{-n} \beta^{-m} \exp \left[ - \left( \frac{n\bar{x}}{\alpha} + \frac{m\bar{y}}{\beta} \right) \right]. \end{aligned} \quad (4.13)$$

Como el parámetro  $\theta$  es el parámetro de interés, conviene reparametrizar la verosimilitud (4.13) en términos de este parámetro. Para hacerlo, se despeja  $\alpha$  de (4.2) en términos de  $\theta$  y  $\beta$ . Esto es,

$$\alpha = \alpha(\theta, \beta) = \frac{(1 - \theta)\beta}{\theta}. \quad (4.14)$$

Entonces, la función de verosimilitud global de  $\theta$  y  $\beta$  se obtiene reemplazando  $\alpha$  en (4.13) por  $\alpha(\theta, \beta)$  dado en (4.14). Así, la verosimilitud queda en función de los parámetros  $(\theta, \beta)$ .

$$\begin{aligned} L(\theta, \beta; x, y) &\propto \left[ \frac{(1 - \theta)\beta}{\theta} \right]^{-n} \beta^{-m} \exp \left\{ - \left[ \frac{n\bar{x}\theta}{(1 - \theta)\beta} + \frac{m\bar{y}}{\beta} \right] \right\} \\ &= \left( \frac{\theta}{1 - \theta} \right)^n \beta^{-(n+m)} \exp \left\{ - \frac{1}{\beta} \left[ \frac{m\bar{y} + \theta(n\bar{x} - m\bar{y})}{1 - \theta} \right] \right\}. \end{aligned} \quad (4.15)$$

En este caso, el emv restringido de  $\beta$  para cada valor especificado de  $\theta$  es

$$\widehat{\beta}(\theta) = \frac{m\bar{y} + \theta(n\bar{x} - m\bar{y})}{(n + m)(1 - \theta)}. \quad (4.16)$$

Por tanto, la función de verosimilitud perfil de  $\theta$  se obtiene reemplazando  $\beta$  por  $\widehat{\beta}(\theta)$  en (4.15)

$$\begin{aligned} L_P(\theta; x, y) &\propto L[\theta, \widehat{\beta}(\theta); x, y] \\ &= \left( \frac{\theta}{1 - \theta} \right)^n \left[ \frac{m\bar{y} + \theta(n\bar{x} - m\bar{y})}{(n + m)(1 - \theta)} \right]^{-(n+m)} \exp[-(n + m)] \\ &\propto \theta^n (1 - \theta)^m [m\bar{y} + \theta(n\bar{x} - m\bar{y})]^{-(n+m)}. \end{aligned} \quad (4.17)$$

Obsérvese que si (4.13) se hubiera reparametrizado en términos de  $(\theta, \alpha)$ , y se hubiera calculado la función de verosimilitud perfil de  $\theta$ , entonces se hubiera llegado a la misma expresión matemática para la perfil de  $\theta$  dada en (4.17) porque la verosimilitud perfil de  $\theta$  es invariante frente a la reparametrización que se elija para los parámetros de estorbo.

Para hacer inferencias sobre el parámetro de interés  $\theta$  conviene graficar la función de verosimilitud perfil relativa de  $\theta$  para evaluar su asimetría y su localización. También se recomienda marcar el emv y los intervalos de verosimilitud de nivel  $c = .036, .15$  y  $.25$ . Como se vió en la Sección 1.5, estos intervalos de verosimilitud tienen una probabilidad de cobertura aproximada del 90%, 95% y 99%, respectivamente. Se resalta aquí que el cálculo computacional de la perfil de  $\theta$ , del emv y de los intervalos de verosimilitud-confianza es muy sencillo, fácil y rápido hacer.

### Asimetría de la verosimilitud perfil

Como se verá más adelante, en la sección de ejemplos, la verosimilitud perfil de  $\theta$  suele ser muy asimétrica para muestras pequeñas. Las siguientes cantidades, que se pueden calcular una vez que se observa la muestra, proporcionan una medida de la asimetría y lo grueso de las colas de una función de verosimilitud perfil de  $\theta$ ,

$$F_3(\hat{\theta}) = \left[ \frac{\partial^3}{\partial \theta^3} \log L_P(\theta) \Big|_{\theta=\hat{\theta}} \right] I_{\hat{\theta}}^{-\frac{3}{2}} \quad (4.18)$$

y

$$F_4(\hat{\theta}) = \left[ \frac{\partial^4}{\partial \theta^4} \log L_P(\theta) \Big|_{\theta=\hat{\theta}} \right] I_{\hat{\theta}}^{-2}, \quad (4.19)$$

donde  $I_{\hat{\theta}}$  es la información observada de Fisher (véase Sprott 2000, pág. 165). Si la función de verosimilitud de  $\theta$  es simétrica y acampanada entonces  $F_3$  y  $F_4$  serán cercanas a cero.

Para el caso exponencial, estas cantidades  $F_3$  y  $F_4$  son

$$F_3(\hat{\theta}) = \frac{2 \left[ n + (n + m) (1 - 3\hat{\theta}) \right]}{\sqrt{nm(n + m)}} \quad (4.20)$$

y

$$F_4(\hat{\theta}) = \frac{-6}{nm(n+m)} \left[ c_1 \hat{\theta}^2 - c_2 \hat{\theta} (1 - \hat{\theta}) + c_3 (1 - \hat{\theta})^2 \right], \quad (4.21)$$

donde  $c_1 = 3m^2 + 3nm + n^2$ ,  $c_2 = 2m^2 + 6nm + 2n^2$  y  $c_3 = m + 3nm + 3n^2$ .

Nótese que para muestras pequeñas, como ocurre en la práctica,  $F_3$  dado en (4.20) y  $F_4$  dado en (4.21) pueden tomar valores grandes dando evidencia de una fuerte asimetría de la función de verosimilitud perfil de  $\theta$ ,  $L_P(\theta; x, y)$  dada en (4.17). Por ejemplo,

para $\hat{\theta} = 0.99$ :	(a) Si $n = m = 8$ entonces $F_3(\hat{\theta}) = -1.47$ y $F_4(\hat{\theta}) = -2.54$ .
	(b) Si $n = 8$ y $m = 4$ entonces $F_3(\hat{\theta}) = -1.60$ y $F_4(\hat{\theta}) = -3.13$ .

Aquí,  $L_P(\theta; x, y)$  es altamente asimétrica con respecto al emv  $\hat{\theta}$ . Además,  $L_P(\theta; x, y)$  correspondiente al caso (b) es más asimétrica que la perfil de  $\theta$  correspondiente al caso (a). Al parecer tener tamaños de muestra diferentes ocasionan una mayor asimetría en la verosimilitud perfil de  $\theta$ .

Como  $L_P(\theta; x, y)$  suele ser muy asimétrica para muestras pequeñas, entonces intervalos de confianza simétricos basados en  $\hat{\theta}$  resultan engañosos porque valores implausibles de  $\theta$  pueden estar incluidos en el intervalo, así como también se pueden excluir otros valores plausibles de  $\theta$ . En la siguiente sección se muestra el uso de reparametrizaciones uno a uno  $\delta = \delta(\theta)$  para simetrizar la verosimilitud de  $\delta$  y así poder usar resultados asintóticos en términos de este parámetro que sí sean razonables. Todo esto con el fin de usar la eficiencia asintótica del emv  $\delta$  y la propiedad de invarianza de la verosimilitud,  $\theta = \delta^{-1}(\cdot)$ , para construir intervalos de verosimilitud-confianza para  $\theta$  que tengan una forma matemática cerrada y que también sean fáciles de calcular.

### Estimación de máxima verosimilitud

La cantidad  $u_\theta = (\hat{\theta} - \theta) \sqrt{I_{\hat{\theta}}}$  surge de manera natural al desarrollar en series de Taylor el logaritmo de la verosimilitud alrededor del emv como se mostró en el Capítulo

1, Sección 1.7.6. Resulta ser una cantidad aproximadamente pivotal que es lineal en el parámetro y que converge en distribución a una normal estándar. Cuando la verosimilitud es simétrica alrededor del emv, este resultado asintótico puede usarse para la muestra finita en cuestión como una buena aproximación. Como se ejemplificó anteriormente en (a) y (b), la verosimilitud puede ser muy asimétrica; sin embargo es posible que una transformación  $\delta = \delta(\theta)$  pueda simetrizarla y con ello resulta entonces creíble y razonable decir que  $u_\delta$  es aproximadamente normal estándar para las muestras observadas. De hecho se mostrará que la transformación

$$\delta = \begin{cases} \log\left(\frac{\theta}{1-\theta}\right), & \text{si } n = m \\ \left(\frac{\theta}{1-\theta}\right)^{\frac{1}{r}}, & \text{si } n \neq m, \end{cases} \quad (4.22)$$

donde  $r = \frac{-3(n+m)}{(n-m)}$ , ayuda a esta misión. Entonces, la cantidad  $u_\delta = (\hat{\delta} - \delta)\sqrt{I_\delta}$ , la cual es lineal en  $\delta$ , le puede heredar a los intervalos de verosimilitud de  $\delta$  una confianza aproximada a través de su distribución normal. Además, por la propiedad de invarianza de la verosimilitud se consiguen inmediatamente los intervalos de verosimilitud-confianza para  $\theta$  reparametrizando de regreso a través de (4.22).

**Diseño balanceado del experimento** ( $n = m$ ): Al considerar la transformación  $\delta = \log(\theta/(1-\theta))$  se obtiene tras realizar algunos cálculos que

$$F_3(\hat{\delta}) = \frac{(n-m)}{\sqrt{nn(n+m)}} = 0 \quad \text{y} \quad F_4(\hat{\delta}) = \frac{1}{n}. \quad (4.23)$$

Entonces la función de verosimilitud de  $\delta$  es simétrica alrededor de  $\hat{\delta}$ , aproximadamente normal y en consecuencia es razonable suponer que la función lineal  $u_\delta = (\hat{\delta} - \delta)\sqrt{I_\delta}$  sigue una distribución normal estándar. Nótese que para el ejemplo (a):

Sin transformación	Con transformación
$F_3(\hat{\theta}) = -1.47, \quad F_4(\hat{\theta}) = -2.54$	$F_3(\hat{\delta}) = 0, \quad F_4(\hat{\delta}) = .125.$

Ahora, como  $\hat{\delta} = \log(\bar{y}/\bar{x})$  y  $I_{\hat{\delta}} = n/2$  entonces los intervalos de verosimilitud-confianza aproximados de nivel de verosimilitud  $c$  y confianza aproximada del  $100(1 - \gamma)\%$  son

$$\theta \in \left[ 1 + \frac{1}{\frac{\bar{y}}{\bar{x}} \exp\left(\pm z_{\gamma/2} \sqrt{\frac{2}{n}}\right)} \right]^{-1}, \quad (4.24)$$

donde  $z_{\gamma/2}$  es el cuantil  $\gamma/2$  de una distribución normal estándar.

Es claro que si  $n \neq m$  entonces  $F_3(\hat{\delta})$  en (4.23) es diferente de cero y esta transformación ya no simetrizaría a la verosimilitud. A continuación, se muestra cual sería la reparametrización apropiada de  $\theta$  para el caso  $n \neq m$ .

**Diseño no balanceado del experimento ( $n \neq m$ ):** Si se considera la transformación  $\delta = (\theta/(1 - \theta))^{1/r}$ , donde  $r = \frac{-3(n + m)}{(n - m)}$ , entonces se obtiene tras realizar algunos cálculos que

$$F_3(\hat{\delta}) = 0 \quad \text{y} \quad F_4(\hat{\theta}) = -\frac{2}{9} \left( \frac{1}{n} + \frac{1}{m} + \frac{13}{n + m} \right).$$

Igual que antes, la función de verosimilitud de  $\delta$  es aproximadamente normal y en consecuencia la función lineal  $u_{\delta} = (\hat{\delta} - \delta)\sqrt{I_{\hat{\delta}}}$  es aproximadamente una variable aleatoria normal estándar. Nótese que para el ejemplo (b):

Sin transformación	Con transformación
$F_3(\hat{\theta}) = -1.60, \quad F_4(\hat{\theta}) = -3.13$	$F_3(\hat{\delta}) = 0, \quad F_4(\hat{\delta}) = -.157.$

Ahora, como  $\hat{\delta} = (\bar{y}/\bar{x})^{1/r}$  y  $I_{\hat{\delta}} = nmr^2\hat{\delta}^{-2}/(n + m)$  entonces los intervalos de verosimilitud-confianza aproximados de nivel de verosimilitud  $c$  y confianza aproximada del  $100(1 - \gamma)\%$  son

$$\theta \in \left[ 1 + \frac{1}{\frac{\bar{y}}{\bar{x}} \left( 1 \pm z_{\gamma/2} \frac{1}{r} \sqrt{\frac{(n + m)}{nm}} \right)^r} \right]^{-1}, \quad (4.25)$$

donde  $z_{\gamma/2}$  es el cuantil  $\gamma/2$  de una distribución normal estándar.

En la Sección 4.2.4 se efectúa un estudio de simulación para verificar que los intervalos de verosimilitud-confianza aproximados para  $\theta$  dados en (4.24) y en (4.25), tienen una cobertura aproximada del  $100(1 - \gamma)\%$ . Por otro lado, éstos intervalos tienen una estructura algebraica muy simple y son fáciles de calcular. Sobre todo, en contraste con los que se proponen en la literatura estadística en la actualidad que requieren un esfuerzo computacional muy grande (cuando  $n$  es diferente de  $m$ ).

A continuación se presenta un ejemplo con datos reales donde se calculan estos intervalos de verosimilitud-confianza aproximados para  $\theta$  y se comparan con los intervalos exactos obtenidos a partir de la gráfica de la verosimilitud perfil de  $\theta$ .

### 4.2.3 Ejemplo: Datos de tiempos de vida de un fluido aislante

Los datos en la Tabla 4.2 fueron tomados de Nelson (1990, pág.129) y serán utilizados para ejemplificar el uso de la función de verosimilitud perfil para hacer inferencias sobre el parámetro de interés  $\theta = P(X < Y)$ . Estos datos son tiempos de vida (en minutos) de un fluido aislante trabajando a dos niveles distintos de voltaje, 36 kV y 30 kV. Nelson (1990) considera una distribución Weibull para modelar los tiempos de vida del fluido aislante,

$$f(z; \lambda, \xi) = \frac{\xi}{\lambda} z^{\xi-1} \exp\left(-\frac{1}{\lambda} z^\xi\right), \quad z > 0, \quad \lambda > 0, \quad \xi > 0.$$

Sin embargo, comenta que no es posible rechazar la hipótesis de que el parámetro de forma  $\xi$  de la distribución Weibull sea igual a uno. Nótese que la distribución exponencial se obtiene como caso particular de la distribución Weibull cuando  $\xi = 1$ . De hecho, para los datos de tiempos de vida correspondientes a 36 kV ( $X$ ), la razón de verosimilitud entre el mejor modelo Weibull (cuyos emv son  $\hat{\lambda}_X = 3.6519$ ,  $\hat{\xi}_X = 0.8892$ ) y el mejor modelo Exponencial (cuyo emv es  $\hat{\alpha}_X = 4.6060$ ) es 1.24. Para el caso de 30 kV ( $Y$ ), la razón de verosimilitud entre el mejor modelo Weibull (cuyos emv son  $\hat{\lambda}_Y =$

100.2067,  $\hat{\xi}_Y = 1.0588$ ) y el mejor modelo Exponencial (cuyo emv es  $\hat{\beta}_Y = 75.7818$ ) es 1.03. Además, la gráfica cuantil-cuantil presentada en la Figura 4.2 sustenta que una distribución exponencial es razonable para estos datos. Es importante comentar aquí que se simularon muestras de tamaño  $n = 15$  y  $m = 11$  provenientes de una distribución Exponencial de parámetros  $\hat{\alpha}_X = 4.6060$  y  $\hat{\beta}_Y = 75.7818$ , respectivamente. Se observó que las gráficas cuantil-cuantil obtenidas a partir de los datos simulados fueron muy parecidas a las mostradas en la Figura 4.2. Estas gráficas no son presentadas aquí. Por todo esto, en este ejemplo se considerará a la distribución exponencial como un modelo parsimonioso para modelar el tiempo de vida del fluido aislante en ambos niveles.

Tabla 4.2. Tiempos de falla de un fluido aislante.

	$X$ (36kV)	$Y$ (30kV)
1	0.35	7.74
2	0.59	17.05
3	0.96	20.46
4	0.99	21.02
5	1.69	22.66
6	1.97	43.40
7	2.07	47.30
8	2.58	139.07
9	2.71	144.12
10	2.90	175.88
11	3.67	194.90
12	3.99	
13	5.35	
14	13.77	
15	25.50	

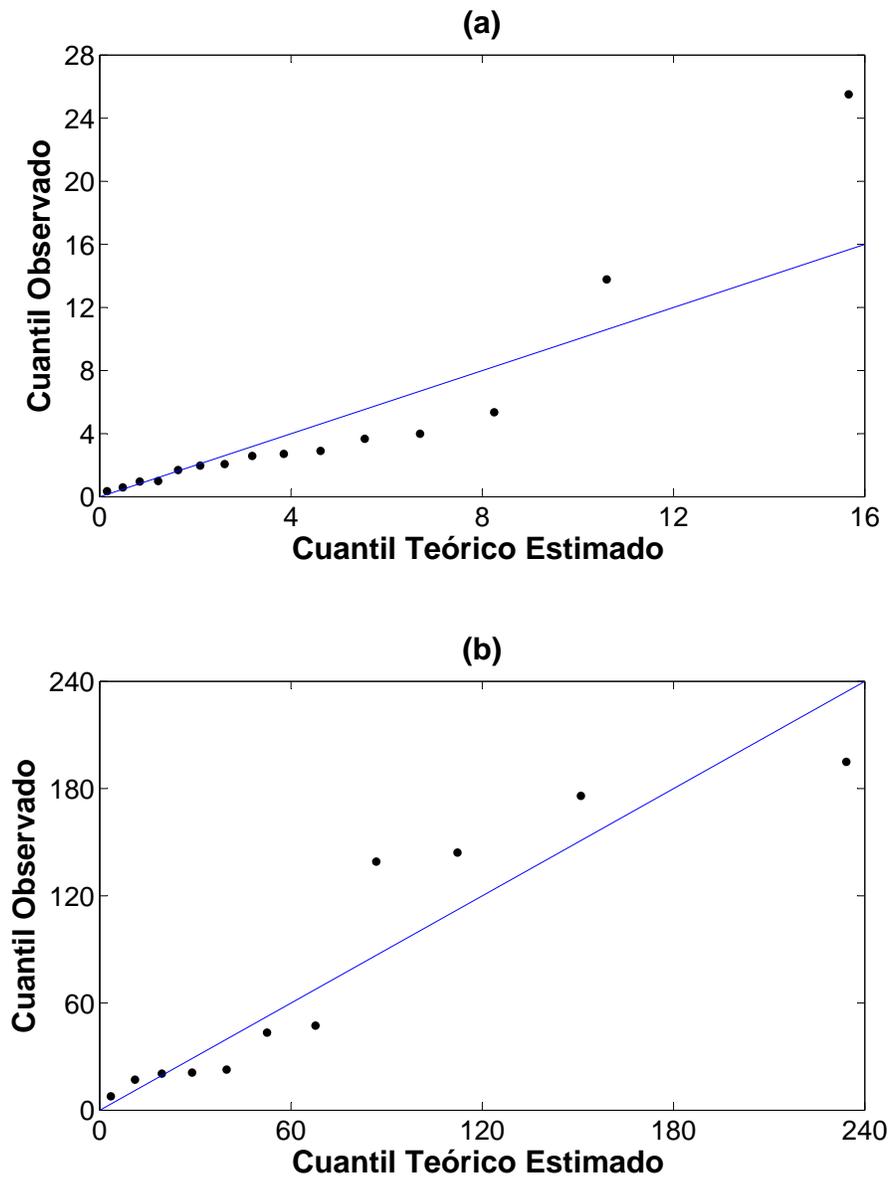


Figura 4.2: En (a) y en (b) se presentan las gráficas cuantil-cuantil para los datos de tiempos de falla de un fluido aislante trabajando a 36kV y 30kV, respectivamente.

La Figura 4.3 muestra la gráfica de la función de verosimilitud perfil relativa de  $\theta$  correspondiente a (4.17),  $R_P(\theta)$ , los intervalos de verosimilitud de nivel  $c = 0.036, 0.15$  y  $0.25$  y la ubicación del emv de  $\theta$ ,  $\hat{\theta} = 0.9427$ . Se observa claramente que la función de verosimilitud relativa de  $\theta$  es asimétrica con cola pesada a la izquierda. Además, valores del parámetro  $\theta$  menores a  $0.8$  y mayores que  $0.99$  son implausibles.

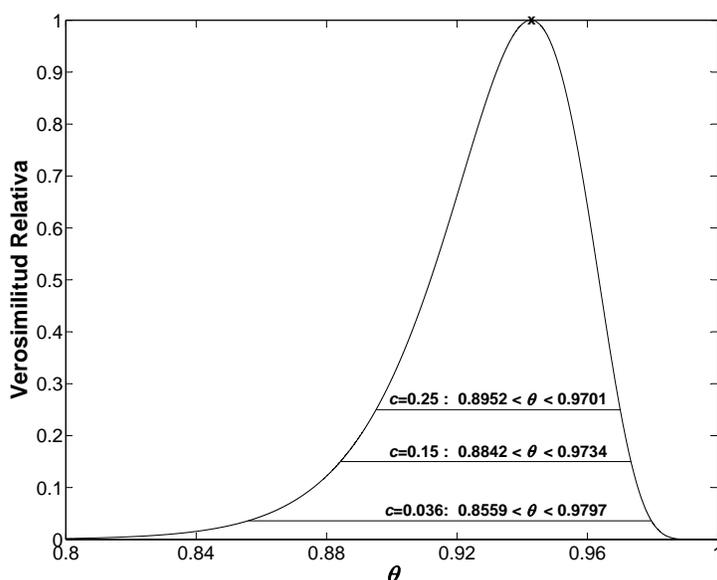


Figura 4.3: Verosimilitud perfil relativa de  $\theta$  para los datos de tiempos de falla de un fluido aislante de la Tabla 4.2.

En la Tabla 4.3 se presentan los intervalos de verosimilitud-confianza para  $\theta$  obtenidos a partir de la gráfica de verosimilitud de  $\theta$  (trazando una línea horizontal en la gráfica de  $R_P(\theta; x, y)$  a una distancia  $c$  paralela al eje  $\theta$ ). También se muestran los intervalos aproximados de verosimilitud-confianza para  $\theta$  obtenidos a través de la estimación de máxima verosimilitud dados en (4.25). Se observa que los límites

superiores e inferiores de ambos intervalos son casi iguales.

Tabla 4.3. Estimación por intervalos para  $\theta$ .

Verosim. relativa	Confianza aprox.	Int. verosimilitud	Int. aprox. normal
$c = 0.25$	0.90	(0.8952, 0.9701)	(0.8964, 0.9697)
$c = 0.15$	0.95	(0.8842, 0.9734)	(0.8847, 0.9733)
$c = 0.036$	0.99	(0.8559, 0.9797)	(0.8586, 0.9792)

Obsérvese que el emv de  $\theta$  es cercano a uno,  $\hat{\theta} = 0.9427$ . Esto significa que la densidad de  $Y$  (30kV) tiene una cola derecha más pesada que la densidad de  $X$  (36kV), y que el traslape de ambas densidades es pequeño (véase la Figura 4.4).

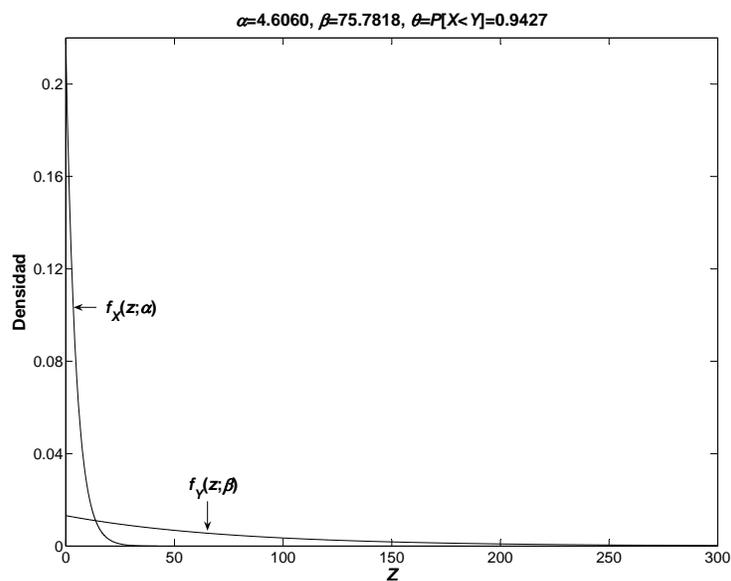


Figura 4.4: Función de densidad estimada de  $X$  y de  $Y$  con los datos de tiempos de falla de un fluido aislante de la Tabla 4.2.

#### 4.2.4 Simulaciones

Se llevó a cabo un estudio de simulación para verificar las probabilidades de cobertura de los intervalos de verosimilitud-confianza aproximados para  $\theta$  dados en (4.24) y en (4.25). Adicionalmente, para el caso de tamaños de muestras iguales, se compararon las probabilidades de cobertura de (4.24) con la del intervalo propuesto por Kotz *et al.* (2003) dado en (4.11). Sólo se comparan las probabilidades de cobertura de estos intervalos porque ambos están disponibles para tamaños de muestras iguales, y tienen en común una forma matemática cerrada y simple.

El estudio se realizó de la siguiente manera. Para valores fijos de  $\alpha$  y  $\beta$  se simularon muestras exponenciales de tamaño  $n$  y  $m$ , respectivamente. Luego, según el caso ( $n = m$  o  $n \neq m$ ), se calcularon los intervalos dados en (4.11), (4.24) y (4.25). Después se revisó que los intervalos calculados incluyeran o no el verdadero valor del parámetro  $\theta$  que depende de  $\alpha$  y de  $\beta$  y se muestra en (4.2). Las probabilidades de cobertura de estos intervalos fueron estimadas con las coberturas empíricas observadas en 10,000 iteraciones del proceso anterior.

Las Tablas 4.4 a 4.11, que se presentan al final de este capítulo, muestran las probabilidades de cobertura de los intervalos de verosimilitud-confianza aproximados y del intervalo asintótico de Kotz *et al.* (2003) para los niveles de confianza del 90%, 95% y 99% ( $\gamma = .1, .05, .01$ ), y para diferentes combinaciones de  $\alpha$  y  $\beta$  que producen valores de  $\theta$  iguales a .70, .90, .95 y .99. En la Tabla 4.4 y 4.5 se reportan las probabilidades de cobertura obtenidas para tamaños de muestra grandes e iguales  $(n, m) = (50, 50), (100, 100)$ , respectivamente. En la Tabla 4.6 y 4.7 se reportan las probabilidades de cobertura obtenidas para tamaños de muestra grandes pero diferentes  $(n, m) = (50, 100), (100, 50)$ , respectivamente. En la Tabla 4.8 y 4.9 se presentan las probabilidades de cobertura obtenidas para tamaños de muestra pequeños e iguales  $(n, m) = (4, 4), (8, 8)$ , respectivamente. Por último, en la Tabla 4.10 y 4.11 se muestran las probabilidades de cobertura para tamaños de muestra pequeños pero diferentes

$(n, m) = (4, 8), (8, 4)$ , respectivamente.

Los resultados de la simulación muestran que para tamaños de muestra grandes, iguales o diferentes, y para diferentes combinaciones de  $\alpha$  y  $\beta$ , las probabilidades de cobertura de ambos intervalos fueron un poco mayores o iguales que el nivel de confianza esperado, salvo unas pocas excepciones donde fueron un poco menores pero aún muy aceptables.

Para tamaños de muestra pequeños, iguales o diferentes, y para diferentes combinaciones de  $\alpha$  y  $\beta$ , las probabilidades de cobertura de los intervalos de verosimilitud-confianza aproximados de  $\theta$  estuvieron un poco por debajo de la probabilidad de cobertura esperada pero con niveles también muy aceptables. Sin embargo, en todos los casos donde  $n$  fue igual a  $m$  y los valores de  $\theta$  fueron iguales a 0.5 y 0.7, los intervalos de verosimilitud-confianza aproximados tuvieron una probabilidad de cobertura superior a la del intervalo asintótico de Kotz *et al* (2003). También se observó que en todos los casos donde  $n$  fue igual a  $m$  y los niveles de confianza esperados fueron de 95% y 99%, los intervalos de verosimilitud-confianza aproximados tuvieron una probabilidad de cobertura superior o igual a la del intervalo asintótico de Kotz *et al.* (2003).

### 4.3 Caso Normal

En la sección anterior se consideró el problema de estimación estadística del parámetro  $\theta = P(X < Y)$  a través de la función de verosimilitud perfil de este parámetro de interés cuando  $X$  y  $Y$  son variables aleatorias independientes con función de densidad exponencial. En esta sección, se considerará el mismo problema pero cuando  $X$  y  $Y$  son normales.

Supóngase que  $X$  y  $Y$  dos variables aleatorias normales e independientes  $N(\mu_x, \sigma_x^2)$  y  $N(\mu_y, \sigma_y^2)$ , donde  $(\mu_x, \sigma_x, \mu_y, \sigma_y)$  es un vector de parámetros desconocidos. Se supone además que las medias y varianzas poblacionales no son necesariamente iguales. Se

tiene interés en hacer inferencias sobre el parámetro  $\theta = P(X < Y)$  basados en dos muestras independientes observadas  $x = (x_1, \dots, x_n)$  y  $y = (y_1, \dots, y_m)$  de  $N(\mu_x, \sigma_x^2)$  y  $N(\mu_y, \sigma_y^2)$ , respectivamente. En este caso,  $n$  y  $m$  tampoco tienen que ser necesariamente iguales.

Como  $X$  y  $Y$  son variables aleatorias normales e independientes entonces el parámetro de interés  $\theta = P(X < Y)$  se puede expresar como

$$\begin{aligned} \theta &= P(X < Y) \\ &= P(X - Y < 0) \\ &= P\left[\frac{(X - Y) - (\mu_x - \mu_y)}{\sqrt{\sigma_x^2 + \sigma_y^2}} < \frac{\mu_y - \mu_x}{\sqrt{\sigma_x^2 + \sigma_y^2}}\right] \\ &= \Phi(\eta), \quad \text{donde } \eta = \frac{\mu_y - \mu_x}{\sqrt{\sigma_x^2 + \sigma_y^2}} \end{aligned} \quad (4.26)$$

y  $\Phi$  es la función de distribución acumulada normal estándar.

Obsérvese que por la propiedad de invarianza de la verosimilitud, el emv de  $\theta$  es

$$\hat{\theta} = \Phi(\hat{\eta}), \quad \text{donde } \hat{\eta} = \frac{\hat{\mu}_y - \hat{\mu}_x}{\sqrt{\hat{\sigma}_x^2 + \hat{\sigma}_y^2}},$$

y  $(\hat{\mu}_x, \hat{\sigma}_x, \hat{\mu}_y, \hat{\sigma}_y)$  son los estimadores de máxima verosimilitud de  $(\mu_x, \sigma_x, \mu_y, \sigma_y)$ . Nótese que  $\hat{\mu}_x = \bar{x}$  y  $\hat{\mu}_y = \bar{y}$  son las medias muestrales y

$$\hat{\sigma}_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad \text{y} \quad \hat{\sigma}_y^2 = \frac{1}{m} \sum_{i=1}^m (y_i - \bar{y})^2.$$

### 4.3.1 Enfoques usuales para inferencias sobre $\theta$

De manera similar al caso exponencial presentado en la sección anterior, en la literatura estadística muchos autores han centrado su atención en construir estimadores invarianza para  $\theta$  para diversas situaciones, como por ejemplo: a) cuando todos los parámetros en el modelo,  $(\mu_x, \sigma_x, \mu_y, \sigma_y)$ , son desconocidos, Downton (1973), b) cuando  $\mu_x$  y  $\sigma_x$  son parámetros conocidos, Mazundar (1970), Woodward y Kelley (1977), c) o cuando  $\mu_x$  y  $\mu_y$  o  $\sigma_x$  y  $\sigma_y$  son conocidos, Ivshin y Lumelskii (1995).

Desde el enfoque de estimación por intervalos, Church y Harris (1970) calcularon límites inferiores de confianza aproximada para  $\theta$  para el caso particular de  $\mu_x$  y  $\sigma_x$  conocidos. Ellos muestran que la estadística  $T = (\bar{y} - \mu_x) / \sqrt{\sigma_x^2 + \hat{\sigma}_y^2}$  se distribuye asintóticamente normal con media  $\eta$  dada en (4.26), y varianza  $\sigma_T^2$ , que depende de los parámetros desconocidos  $\mu_y$  y  $\sigma_y$ . Entonces, obtienen que la  $P[\theta > \Phi(t - z_\gamma \sigma_T)] = 1 - \gamma$ , donde  $z_\gamma$  es el cuantil  $\gamma$  de una distribución normal estándar. Para calcular el límite inferior de confianza para  $\theta$  reemplazan  $\sigma_T$  por un estimador  $\hat{\sigma}_T$ .

Reiser y Guttman (1986) también calcularon límites inferiores de confianza aproximada para  $\theta$ ; pero consideraron el caso general donde todos los parámetros ( $\mu_x$ ,  $\sigma_x$ ,  $\mu_y$ ,  $\sigma_y$ ) son desconocidos. Ellos muestran que la estadística  $W = \hat{\zeta} \sqrt{\widehat{M}}$ , donde

$$\hat{\zeta} = \frac{(\bar{y} - \bar{x})}{\sqrt{\frac{n}{n-1} \hat{\sigma}_x^2 + \frac{m}{m-1} \hat{\sigma}_y^2}} \text{ y } \widehat{M} = \frac{\frac{n}{n-1} \hat{\sigma}_x^2 + \frac{m}{m-1} \hat{\sigma}_y^2}{\frac{1}{n-1} \hat{\sigma}_x^2 + \frac{1}{m-1} \hat{\sigma}_y^2},$$

se distribuye aproximadamente como una variable aleatoria  $t$  de Student con

$$\hat{v} = \frac{\left( \frac{n}{n-1} \hat{\sigma}_x^2 + \frac{m}{m-1} \hat{\sigma}_y^2 \right)^2}{\frac{n^2}{(n-1)^3} \hat{\sigma}_x^4 + \frac{m^2}{(m-1)^3} \hat{\sigma}_y^4}$$

grados de libertad y con parámetro de centralidad  $\eta \sqrt{\widehat{M}}$ . Así, el límite inferior de confianza para  $\eta$  es el valor  $\eta^*$  que satisface la siguiente ecuación:

$$P \left[ t_{\hat{v}} \left( \eta^* \sqrt{\widehat{M}} \right) < W \right] = 1 - \gamma.$$

En general, esta ecuación sólo se puede resolver numéricamente. Entonces el límite inferior del 100  $(1 - \gamma)$  % para  $\theta$  es  $\Phi(\eta^*)$ . Es decir;  $P[\theta > \Phi(\eta^*)] = 1 - \gamma$ .

Weerahandi y Johnson (1992) consideran el problema de prueba de hipótesis de una cola,  $H_0 : \theta \leq \theta_0$  versus  $H_1 : \theta > \theta_0$ . Ellos utilizan una prueba de p-valor generalizado definida por Tsui y Weerahandi (1989). En general, esta prueba está definida en términos de una variable de prueba y una región crítica generalizada. La variable de

prueba es de la forma  $T(Z; z, \lambda)$ , y es una función de una variable aleatoria  $Z$ , del valor observado  $Z = z$  y de un parámetro de estorbo  $\lambda$ . Además, esta variable debe de cumplir que: a)  $T(Z = z; z, \lambda)$  no dependa de  $\lambda$ , b) la distribución de  $T(Z; z, \lambda)$  tampoco dependa de  $\lambda$ , c) la  $P[T(Z; z, \lambda) \geq t; \theta]$  sea no decreciente como función del parámetro de interés  $\theta$  para todo valor de  $z$  y  $\lambda$  fijo. La región crítica es de la forma  $C_z(\lambda) = \{Z : T(Z; z, \lambda) \geq 0\}$ , y es una función del valor observado  $z$  y del parámetro de estorbo  $\lambda$ . Así, dadas las condiciones anteriores, el p-valor de esta prueba es  $\rho = P[Z \in C_z(\lambda); \theta = \theta_0]$ , y se puede calcular puesto que no depende del parámetro de estorbo  $\lambda$ .

Weerahandi y Johnson (1992) basan su prueba en el siguiente p-valor generalizado:

$$\rho = 1 - E \left\{ G_{n+m-2} \left[ \frac{(\bar{y} - \bar{x}) \sqrt{n+m-2}}{\sqrt{\frac{\hat{\sigma}_x^2}{B} + \frac{\hat{\sigma}_y^2}{1-B}}} \right] \right\},$$

donde  $G$  es la función de distribución acumulada  $t$  de Student con  $v = n + m - 2$  grados de libertad y parámetro de no centralidad

$$\omega_{\theta_0}(B) = \theta_0 \sqrt{\frac{m(1-B)\hat{\sigma}_x^2 + nB\hat{\sigma}_y^2}{(1-B)\hat{\sigma}_x^2 + B\hat{\sigma}_y^2}}.$$

La esperanza se toma con respecto a  $B$  que se distribuye como una Beta  $[(m-1)/2, (n-1)/2]$ . Además, comentan que es posible utilizar este p-valor para construir intervalos de confianza para  $\theta$ .

Este es un caso donde la verosimilitud perfil no ha sido considerada para hacer inferencias sobre el parámetro de interés  $\theta$ . A continuación se usará la función de verosimilitud perfil de  $\theta$  completa para hacer inferencia sobre este parámetro.

### 4.3.2 Uso de la verosimilitud perfil para inferencias sobre $\theta$

La función de verosimilitud global de  $(\mu_x, \sigma_x, \mu_y, \sigma_y)$  es

$$\begin{aligned} L(\mu_x, \sigma_x, \mu_y, \sigma_y; x, y) &\propto \prod_{i=1}^n \frac{1}{\sigma_x} \exp\left[-\frac{1}{2} \frac{(x_i - \mu_x)^2}{\sigma_x^2}\right] \prod_{j=1}^m \frac{1}{\sigma_y} \exp\left[-\frac{1}{2} \frac{(y_j - \mu_y)^2}{\sigma_y^2}\right] \\ &\propto \sigma_x^{-n} \sigma_y^{-m} \exp\left[-\frac{1}{2\sigma_x^2} \sum_{i=1}^n (x_i - \mu_x)^2 + \right. \\ &\quad \left. -\frac{1}{2\sigma_y^2} \sum_{i=1}^m (y_i - \mu_y)^2\right]. \end{aligned} \quad (4.27)$$

Como el parámetro  $\theta$  es el parámetro de interés, conviene reparametrizar la verosimilitud (4.27) en términos de este parámetro. Para hacerlo, se despeja  $\mu_x$  de (4.26) en términos de  $\theta$  y  $(\sigma_x, \mu_y, \sigma_y)$ . Esto es,

$$\mu_x = \mu_x(\theta, \sigma_x, \mu_y, \sigma_y) = \mu_y - \Phi^{-1}(\theta) \sqrt{\sigma_x^2 + \sigma_y^2}, \quad (4.28)$$

donde  $\Phi^{-1}$  es la inversa de la función de distribución acumulada normal estándar.

Entonces, la función de verosimilitud global de  $\theta$  y  $(\sigma_x, \mu_y, \sigma_y)$  se obtiene reemplazando  $\mu_x$  en (4.27) por  $\mu_x(\theta, \sigma_x, \mu_y, \sigma_y)$  dado en (4.28). Así, la verosimilitud queda en función de los parámetros  $(\theta, \sigma_x, \mu_y, \sigma_y)$ .

$$\begin{aligned} L(\theta, \sigma_x, \mu_y, \sigma_y; x, y) &\propto \sigma_x^{-n} \sigma_y^{-m} \exp\left\{-\frac{1}{2\sigma_x^2} \sum_{i=1}^n \left[x_i - \mu_y + \Phi^{-1}(\theta) \sqrt{\sigma_x^2 + \sigma_y^2}\right]^2 + \right. \\ &\quad \left. -\frac{1}{2\sigma_y^2} \sum_{i=1}^m (y_i - \mu_y)^2\right\}. \end{aligned} \quad (4.29)$$

En este caso, la función de verosimilitud perfil de  $\theta$  se puede obtener maximizando numéricamente (4.29) sobre  $(\sigma_x, \mu_y, \sigma_y)$  para cada  $\theta$  fija,

$$L_P(\theta; x, y) \propto \max_{\mu_y, \sigma_x, \sigma_y | \theta} L(\theta, \sigma_x, \mu_y, \sigma_y; x, y). \quad (4.30)$$

Sin embargo, nótese que esta expresión general de la perfil de  $\theta$  se puede simplificar de la siguiente forma. El emv restringido de  $\mu_y$  para cada valor fijo de los otros parámetros

$(\theta, \sigma_x, \sigma_y)$  tiene una fórmula matemática cerrada dada por

$$\hat{\mu}_y(\theta, \sigma_x, \sigma_y) = \frac{m\bar{y}\sigma_x^2 + n\bar{x}\sigma_y^2 + n\Phi^{-1}(\theta)\sigma_y^2\sqrt{\sigma_x^2 + \sigma_y^2}}{m\sigma_x^2 + n\sigma_y^2}. \quad (4.31)$$

Por tanto, otra expresión equivalente pero más simple para calcular la función de verosimilitud perfil de  $\theta$  dada en (4.30), se obtiene reemplazando  $\mu_y$  en (4.29) por  $\hat{\mu}_y(\theta, \sigma_x, \sigma_y)$  dado en (4.31), y luego maximizando numéricamente sobre  $(\sigma_x, \sigma_y)$  para cada  $\theta$  fija,

$$L_P(\theta; x, y) \propto \max_{\sigma_x, \sigma_y | \theta} L[\theta, \sigma_x, \hat{\mu}_y(\theta, \sigma_x, \sigma_y), \sigma_y; x, y]. \quad (4.32)$$

Cabe señalar aquí que en la literatura estadística, en algunas situaciones se plantean problemas inferenciales en términos de una función uno a uno de  $\theta$ ,  $\psi = \psi(\theta)$ , y no directamente en términos de  $\theta$ . Por ejemplo,  $\psi(\theta) = \Phi^{-1}(\theta) = \eta$ . En estos casos,  $\psi$  es el nuevo parámetro de interés y la función de verosimilitud perfil de este parámetro se obtiene, por la propiedad de invarianza de la verosimilitud frente a reparametrizaciones uno a uno, sustituyendo simplemente  $\theta = \theta(\psi)$  en la verosimilitud global  $L(\theta, \sigma_x, \mu_y, \sigma_y; x, y)$  dada en (4.29), y maximizando numéricamente sobre  $(\sigma_x, \mu_y, \sigma_y)$  para cada  $\psi$  fija. Nótese, que nuevamente por la propiedad de invarianza de la verosimilitud, el emv restringido de  $\hat{\mu}_y$  para cada valor fijo de  $(\psi, \sigma_x, \sigma_y)$  se obtiene reemplazando  $\theta = \theta(\psi)$  en  $\hat{\mu}_y(\theta, \sigma_x, \sigma_y)$  dado en (4.31),  $\hat{\mu}_y(\psi, \sigma_x, \sigma_y) = \hat{\mu}_y[\theta = \theta(\psi), \sigma_x, \sigma_y]$ . Por tanto, la función de verosimilitud perfil de  $\psi$  se puede expresar como

$$L_P(\psi; x, y) \propto \max_{\sigma_x, \sigma_y | \psi} L[\theta = \theta(\psi), \sigma_x, \hat{\mu}_y(\psi, \sigma_x, \sigma_y), \sigma_y; x, y]. \quad (4.33)$$

Como en la sección anterior, para hacer inferencias sobre el parámetro de interés  $\theta$  se recomienda graficar la función de verosimilitud perfil relativa de este parámetro y marcar el emv y los intervalos de verosimilitud de nivel  $c = .036, .15$  y  $.25$  que están asociados a una probabilidad de cobertura aproximada del 90%, 95% y 99%, respectivamente. Otra vez se observa aquí que el cálculo computacional de la perfil,

del emv y de los intervalos de verosimilitud-confianza es muy sencillo, fácil y rápido hacer si se adopta a la verosimilitud perfil como herramienta de trabajo.

### 4.3.3 Ejemplos

#### **Ejemplo 1: Datos experimentales de motores de cohetes.**

La idea básica tras un cohete de combustible líquido es bastante sencilla. Un combustible y un oxidante, ambos en estado líquido, son introducidos en una cámara de combustión y encendidos. Por ejemplo, el Transbordador Espacial usa hidrógeno líquido como su combustible y oxígeno líquido como el oxidante. Los gases calientes producidos por la combustión escapan rápidamente a través de la tobera cónica, produciendo así el impulso. Uno de los principales factores para el buen funcionamiento del motor de un cohete es la presión en la cámara de combustión que este motor genera. Así, si  $X$  es una variable aleatoria que representa la presión dentro de la cámara de combustión de un motor de cohete (*'stress'*) y  $Y$  es otra variable aleatoria que representa la fuerza que tiene la cámara para soportar presión (*'strength'*). Entonces, una medida de la confiabilidad del motor del cohete es  $\theta = P(X < Y)$ .

En la Tabla 4.3 se presentan  $n = 24$  observaciones de la presión dentro de la cámara de combustión de un motor de cohete ( $X$ ) trabajando a una alta temperatura, 59 grados centígrados, junto con  $m = 17$  observaciones de la fuerza que tiene la cámara para soportar presión ( $Y$ ). Estos datos fueron reportados por Guttman, Johnson, Bhattacharyya, y Reiser (1988).

Tabla 4.3. Datos experimentales de motores de cohetes.

	Presión al operar ( $X$ )	Fuerza para resistir explosiones internas ( $Y$ )
1	7.74010	15.30
2	7.77490	17.10
3	7.72270	16.30
4	7.77925	16.05
5	7.96195	16.75
6	7.44720	16.60
7	8.07070	17.10
8	7.89525	17.50
9	8.07360	16.10
10	7.49650	16.10
11	7.57190	16.00
12	7.79810	16.75
13	7.87640	17.50
14	8.19250	16.50
15	8.01705	16.40
16	7.94310	16.00
17	7.71835	16.20
18	7.87785	
19	7.29040	
20	7.75750	
21	7.31960	
22	7.63570	
23	8.06055	
24	7.91120	

Guttman *et al.* (1988) suponen que ambas variables,  $X$  y  $Y$ , son normales e independientes. Ellos comentan que por confidencialidad, no es posible proporcionar más detalles acerca de estos datos y del experimento. La gráfica cuantil-cuantil presentada en la Figura 4.5 sustenta la suposición de normalidad como razonable. Los emv de los parámetros de ambos modelos son:  $\hat{\mu}_x = 7.7888$ ,  $\hat{\sigma}_x = 0.2327$ ,  $\hat{\mu}_y = 16.4853$  y  $\hat{\sigma}_y = 0.5664$ .

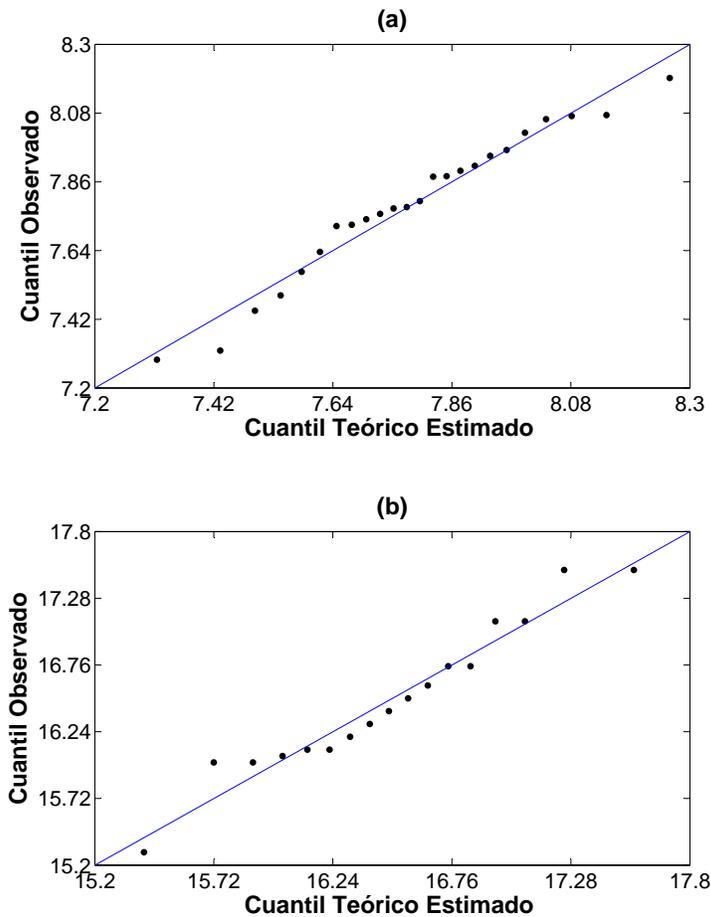


Figura 4.5: En (a) y en (b) se presentan las gráficas cuantil-cuantil para los datos de motores de cohetes  $X$  y  $Y$ , respectivamente.

Cabe señalar aquí que Weerahandi y Johnson (1992) consideran estos datos y plantean la hipótesis

$$H_0 : \theta \leq 0.999999 \text{ versus } H_1 : \theta > 0.999999,$$

o equivalentemente

$$H_0 : \eta \leq 4.75059 \text{ versus } H_1 : \eta > 4.75059,$$

donde  $\eta = \Phi^{-1}(\theta)$ . Ellos rechazan esta hipótesis con un p-valor de 0.0000042. Además, proporcionan un intervalo aproximado del 95% de confianza para  $\eta$ , [9.52, 17.67].

En este ejemplo se considerará que el parámetro de interés es  $\eta$ . Esto con el objetivo de comparar los resultados obtenidos por Weerahandi y Johnson (1992) con los que se obtendrán a través de la verosimilitud perfil. La Figura 4.6 muestra la gráfica de la función de verosimilitud perfil relativa de  $\eta$ ,  $R_P(\eta; x, y)$ , correspondiente a (4.33), donde  $\psi(\theta) = \Phi^{-1}(\theta) = \eta$ . Además, se presentan los intervalos de verosimilitud de nivel  $c = 0.036, 0.15$  y  $0.25$  y la ubicación del emv de  $\eta$ ,  $\hat{\eta} = 14.2021$ . Se observa claramente que la función de verosimilitud relativa de  $\eta$  es simétrica con respecto al emv  $\hat{\eta}$ . Además, valores del parámetro  $\eta$  menores a 7 y mayores que 21 son altamente implausibles.

En la Tabla 4.4 se presentan los intervalos de verosimilitud-confianza para  $\eta$  obtenidos a partir de la verosimilitud perfil relativa de  $\eta$ ,  $R_P(\eta; x, y)$ , junto con los niveles de confianza correspondientes.

Tabla 4.4. Intervalos de verosimilitud-confianza para  $\eta$ .

Verosim. relativa	Confianza aprox.	Int. verosim. exacto
$c = 0.25$	0.90	(10.6984, 17.7196)
$c = 0.15$	0.95	(10.1164, 18.3016)
$c = 0.036$	0.99	(8.8586, 19.6345)

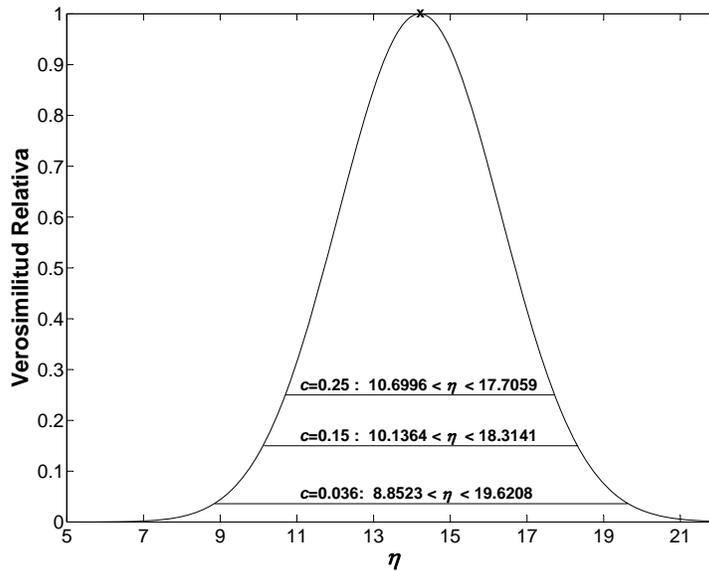


Figura 4.6: Verosimilitud perfil relativa de  $\eta$  para los datos de motores de cohetes de la Tabla 4.3.

Nótese que los límites del intervalo de verosimilitud-confianza para  $\eta$  correspondiente a una cobertura aproximada del 95%, obtenidos con la verosimilitud perfil de  $\eta$ , y los límites del intervalo dado por Weerahandi y Johnson (1992), [10.0955, 18.3266] y [10.1164, 18.3016], son similares. Por otro lado, en la Figura 4.6 se observa que valores de  $\eta$  menores que 4.75059 son extremadamente implausibles a la luz de la muestra observada. Así, se tiene una fuerte evidencia en contra de la hipótesis considerada por Weerahandi y Johnson (1992). Al parecer, el esfuerzo computacional que demanda el cálculo del intervalo de Weerahandi y Johnson (1992), basado en método de p-valor generalizado, es mayor que realizar un simple corte de la función de verosimilitud perfil relativa  $R_P(\eta; x, y)$ .

Por otro lado, como  $\hat{\theta} = \Phi(\hat{\eta} = 14.2021)$  es un valor muy cercano a uno, esto significa que las densidades estimadas correspondientes a  $X$  y a  $Y$  se encuentran muy separadas (se traslapan muy poco). Además, la densidad de  $X$  se encuentra localizada hacia la derecha de la densidad de  $Y$  (véase la Figura 4.7).

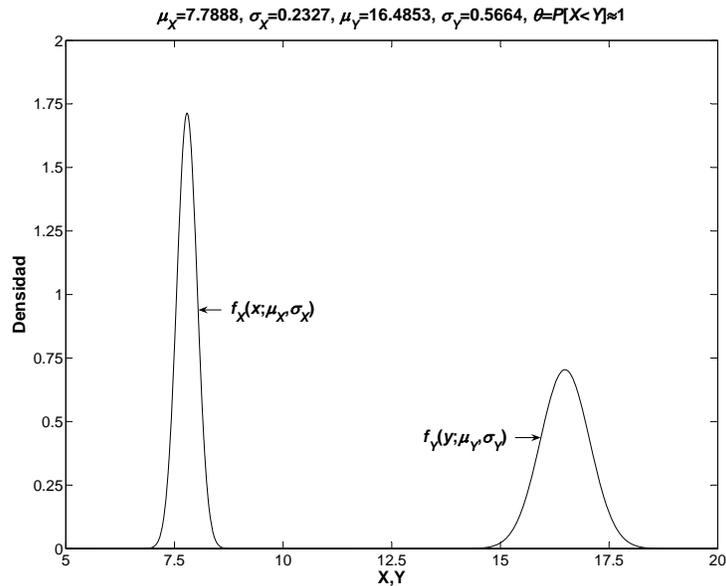


Figura 4.7: Función de densidad estimada de  $X$  y de  $Y$  con los datos de motores de cohetes de la Tabla 4.3.

### Ejemplo 2: Datos del Índice de Masa Corporal (IMC) en México

El índice de masa corporal (IMC), se calcula fácilmente a partir de la estatura  $h$  en metros y el peso  $w$  de una persona,  $IMC=w/h^2$ . A nivel mundial, el IMC es uno de los estándares de mayor uso entre los médicos para definir a la obesidad y al sobrepeso, ya que es barato, fácil de obtener y de interpretar. Debido a la existencia de un dimorfismo sexual en las mediciones del cuerpo y de los niveles de distribución de grasa en mamíferos, comúnmente se estudian las poblaciones de hombres y mujeres por separado.

Supóngase que se desea comparar el IMC de los hombres con el de las mujeres en una población determinada. Se tiene interés en conocer cuál es la probabilidad de que el IMC de las mujeres sea menor que el de los hombres. Así, si  $X$  y  $Y$  son variables aleatorias que representan el IMC de las mujeres y de los hombres en la población de

estudio, respectivamente, entonces el parámetro de interés es  $\theta = P(X < Y)$ .

En este ejemplo se considera como población de estudio a las personas mayores de 18 años, hombres y mujeres, que se atienden en el Hospital Aranda de la Parra en León, Guanajuato, México con la médico internista Dra. Yolanda Méndez Romero. Se cuenta con 285 observaciones del IMC de mujeres ( $X$ ) y 605 observaciones del IMC de hombres ( $Y$ ). Se desea hacer inferencias sobre el parámetro  $\theta = P(X < Y)$  basadas en estos datos. Cabe señalar aquí que análisis estadísticos preliminares de ellos indicaron que el supuesto de normalidad tanto para  $X$  como para  $Y$  es poco razonable puesto que siguen distribuciones asimétricas. Debido a que para poblaciones afro-americanas con distribución con forma similar del IMC se empleó la familia de densidades de transformaciones de Box y Cox (1964) exitosamente, se decidió también considerar a esta familia para describir a los datos de mexicanos (véase López, 2004).

Si  $W$  es el IMC de un grupo poblacional, se supondrá que su densidad es

$$f(w; \lambda, \mu, \sigma) = \frac{w^{\lambda-1}}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2} [w^{(\lambda)} - \mu]^2\right\} I_{(0,\infty)}(w),$$

donde

$$w^{(\lambda)} = \begin{cases} \frac{w^\lambda - 1}{\lambda} & (\lambda \neq 0), \\ \ln w & (\lambda = 0), \end{cases}$$

y  $W^{(\lambda)}$  sigue una distribución aproximadamente normal con media  $\mu$  y varianza  $\sigma^2$ . Obsérvese que como las transformaciones asociadas son monótonas y tanto hombres como mujeres se transformaron con la misma  $\lambda$ , entonces el parámetro de interés  $\theta$  no sufre ningún cambio,  $\theta = P(X < Y) = P[X^{(\lambda)} < Y^{(\lambda)}]$ . Es decir, se seleccionó **la misma transformación**  $\lambda$  tanto para  $X$  como para  $Y$  y se vió que esto fuese sensato para los datos. Esta selección se hizo a través de la verosimilitud perfil de  $\lambda$ , que se muestra en la Figura 4.8. Se eligió el valor  $\lambda = -0.4$  porque tiene una plausibilidad razonablemente alta en ambos casos,  $R(\lambda = -0.4; x) = 0.20$  para mujeres y  $R(\lambda = -0.4; y) = 0.28$  para hombres. La gráfica cuantil-cuantil del IMC transformado

con  $\lambda = -0.4$  para hombres y para mujeres se muestra en la Figura 4.9. Se observa que el supuesto de normalidad es razonable en ambos casos.

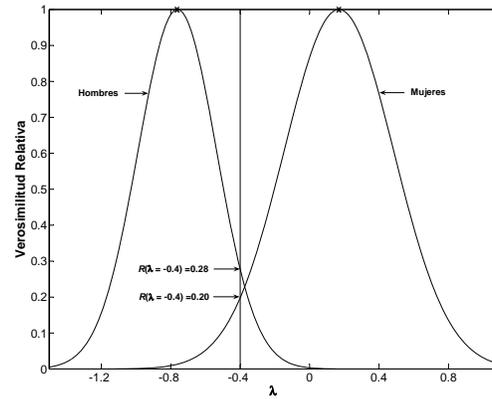


Figura 4.8: Verosimilitud perfil relativa de  $\lambda$  para los datos del IMC de mujeres y hombres de León, Guanajuato.

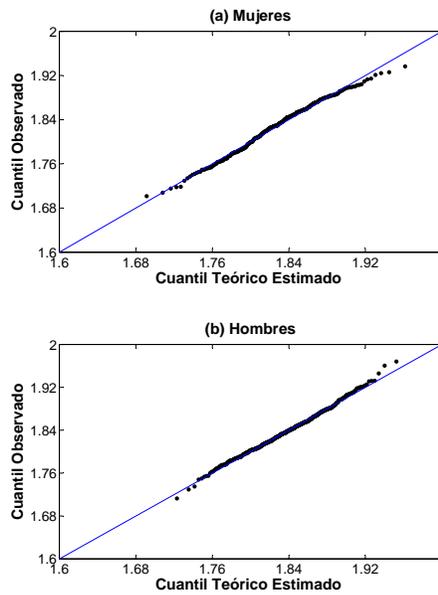


Figura 4.9: En (a) y en (b) se presentan las gráficas cuantil-cuantil para los datos del IMC transformado con  $\lambda = -0.4$  para mujeres y hombres, respectivamente.

En la Figura 4.10 se presentan las funciones de densidad estimadas para el IMC de las poblaciones de hombres y mujeres en León, Guanajuato. Se observa que la densidad estimada de la población de hombres se encuentra desplazada un poco hacia la derecha con respecto a la densidad estimada de la población de mujeres; sin embargo las colas derechas de ambas densidades casi coinciden.

La Figura 4.11 muestra la gráfica de la función de verosimilitud perfil relativa de  $\theta$ , correspondiente a (4.32). Además, se presentan los intervalos de verosimilitud de nivel  $c = 0.036, 0.15$  y  $0.25$  y la ubicación del emv de  $\theta$ ,  $\hat{\theta} = 0.5756$ . Se observa que la función de verosimilitud relativa de  $\theta$  es simétrica con respecto al emv  $\hat{\theta}$ . También, que valores del parámetro  $\theta$  menores a  $0.5215$  y mayores que  $0.6282$  son altamente implausibles.

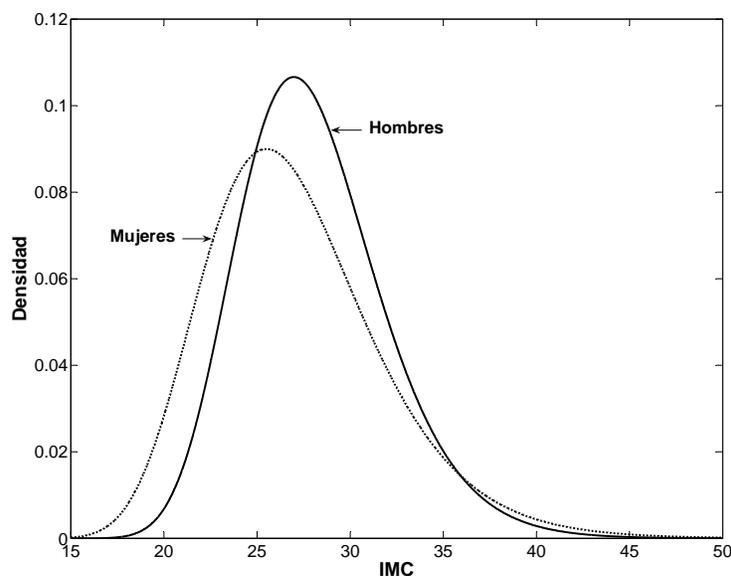


Figura 4.10: Función de densidad estimada del IMC para la población de mujeres y hombres en León, Guanajuato.

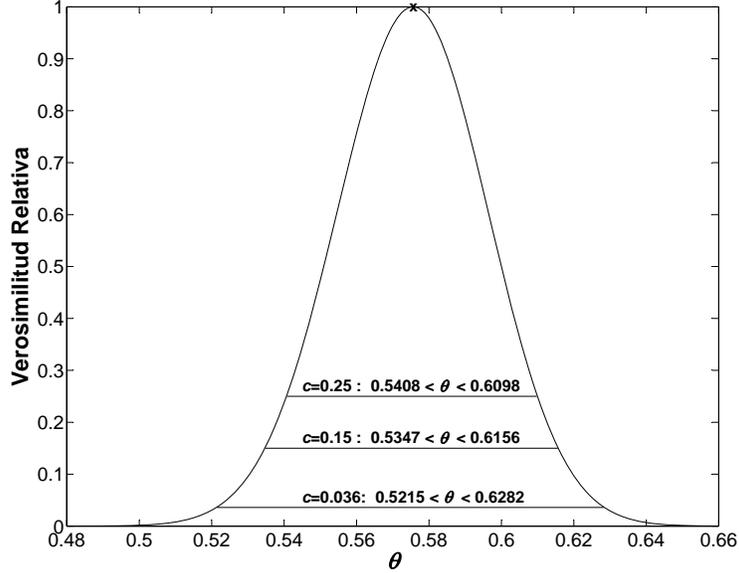


Figura 4.11: Verosimilitud perfil relativa de  $\theta$  para los datos del IMC de personas mayores de 18 años, mujeres y hombres, que se atienden en el Hospital Aranda de la Parra en León, Guanajuato, México.

## 4.4 Conclusiones generales

Resulta sorprendente que la verosimilitud perfil, un método estadístico simple y comúnmente usado en la estimación por separado de un parámetro de interés en presencia de otros de estorbo, no haya sido considerada con anterioridad para realizar inferencias sobre el parámetro de interés  $\theta = P(X < Y)$  descrito en este capítulo.

Las inferencias sobre  $\theta$  vía la función de verosimilitud perfil son muy simples y pertinentes incluso cuando se tienen muestras pequeñas. Más aún, los intervalos aproximados de verosimilitud-confianza propuestos en esta tesis para  $\theta$ , obtenidos a partir del uso de reparametrizaciones normalizadoras de la función de verosimilitud perfil, tienen una estructura algebraica muy simple, tienen buenas coberturas y son mucho más fáciles de calcular que las alternativas existentes en la literatura.

La función de verosimilitud perfil es muy flexible en el sentido que si se desea hacer inferencias sobre otro parámetro que se relaciona a través de una función uno a uno con  $\theta$ , sólo se requiere una reparametrización de la verosimilitud en términos de este nuevo parámetro de interés.

Estos resultados muestran que la verosimilitud perfil es una alternativa inferencial muy prometedora y eficiente para estimar el parámetro  $\theta$  en forma muy simple en el contexto descrito.

Tabla 4.4. Probabilidades de cobertura: Tamaño de muestra grande e igual.

$M=10000$ $n=m=50$			Intervalos					
			Verosimilitud-Confianza			Kotz et al. (2003)		
$(1-\gamma)$			0.90	0.95	0.99	0.90	0.95	0.99
$\alpha$	$\beta$	$\theta$						
0.10	0.10	0.50	0.8952	0.9486	0.9906	0.8872	0.9421	0.9866
	0.23	0.70	0.8979	0.9501	0.9893	0.8917	0.9423	0.9861
	1.90	0.95	0.9005	0.9501	0.9887	0.9063	0.9481	0.9844
	9.90	0.99	0.8985	0.9469	0.9875	0.9105	0.9472	0.9806
0.50	0.50	0.50	0.8969	0.9481	0.9889	0.8882	0.9403	0.9850
	1.17	0.70	0.8964	0.9483	0.9899	0.8914	0.9445	0.9854
	9.50	0.95	0.8975	0.9487	0.9892	0.9044	0.9480	0.9834
	49.50	0.99	0.9016	0.9493	0.9924	0.9101	0.9527	0.9853
1.00	1.00	0.50	0.8994	0.9501	0.9867	0.8924	0.9438	0.9829
	2.33	0.70	0.8970	0.9480	0.9893	0.8911	0.9406	0.9848
	19.00	0.95	0.8965	0.9465	0.9884	0.9035	0.9470	0.9837
	99.00	0.99	0.8945	0.9475	0.9877	0.9068	0.9488	0.9819
5.00	5.00	0.50	0.8968	0.9472	0.9900	0.8911	0.9404	0.9859
	11.67	0.70	0.8958	0.9464	0.9886	0.8893	0.9405	0.9841
	95.00	0.95	0.9013	0.9492	0.9881	0.9061	0.9508	0.9830
	495.00	0.99	0.8955	0.9454	0.9889	0.9088	0.9517	0.9829
20.00	20.00	0.50	0.9030	0.9484	0.9891	0.8954	0.9422	0.9851
	46.70	0.70	0.8965	0.9467	0.9893	0.8917	0.9433	0.9853
	380.00	0.95	0.9048	0.9512	0.9889	0.9105	0.9516	0.9838
	1980.00	0.99	0.9020	0.9505	0.9897	0.9138	0.9534	0.9861

Tabla 4.5. Probabilidades de cobertura: Tamaño de muestra grande e igual.

$M=10000$ $n=m=100$			Intervalos						
			Verosimilitud-Confianza			Kotz et al. (2003)			
$(1-\gamma)$			0.90	0.95	0.99	0.90	0.95	0.99	
$\alpha$	$\beta$	$\theta$							
0.10	0.10	0.50	0.8995	0.9501	0.9908	0.8963	0.9463	0.9893	
		0.23	0.70	0.8959	0.9471	0.9919	0.8943	0.9458	0.9892
		1.90	0.95	0.8956	0.9498	0.9885	0.9025	0.9493	0.9871
		9.90	0.99	0.9058	0.9518	0.9882	0.9105	0.9522	0.9859
0.50	0.50	0.50	0.9003	0.9516	0.9907	0.8958	0.9491	0.9888	
		1.17	0.70	0.9016	0.9498	0.9891	0.8982	0.9472	0.9877
		9.50	0.95	0.9035	0.9484	0.9881	0.9056	0.9472	0.9853
		49.50	0.99	0.8988	0.9490	0.9893	0.9037	0.9490	0.9847
1.00	1.00	0.50	0.9005	0.9489	0.9887	0.8963	0.9465	0.9872	
		2.33	0.70	0.8979	0.9493	0.9895	0.8954	0.9462	0.9866
		19.00	0.95	0.8988	0.9480	0.9894	0.9046	0.9496	0.9855
		99.00	0.99	0.8974	0.9492	0.9905	0.9065	0.9504	0.9863
5.00	5.00	0.50	0.8978	0.9507	0.9908	0.8939	0.9472	0.9894	
		11.67	0.70	0.8969	0.9477	0.9867	0.8959	0.9449	0.9859
		95.00	0.95	0.8993	0.9509	0.9917	0.9051	0.9522	0.9881
		495.00	0.99	0.9035	0.9522	0.9903	0.9082	0.9506	0.9868
20.00	20.00	0.50	0.8992	0.9482	0.9892	0.8964	0.9450	0.9873	
		46.70	0.70	0.8968	0.9484	0.9901	0.8948	0.9462	0.9870
		380.00	0.95	0.8961	0.9459	0.9896	0.9032	0.9477	0.9866
		1980.00	0.99	0.8991	0.9500	0.9878	0.9074	0.9507	0.9858

Tabla 4.6. Probabilidades de cobertura: Tamaño de muestra grande y diferente.

$M=10000$ $n=50$ y $m=100$			Intervalo de verosimilitud-confianza		
$(1-\gamma)$			0.90	0.95	0.99
$\alpha$	$\beta$	$\theta$			
0.10	0.10	0.50	0.8979	0.9460	0.9885
	0.23	0.70	0.8972	0.9472	0.9904
	1.90	0.95	0.8959	0.9474	0.9879
0.50	9.90	0.99	0.8982	0.9462	0.9901
	0.50	0.50	0.8952	0.9489	0.9897
	1.17	0.70	0.9023	0.9499	0.9894
1.00	9.50	0.95	0.8947	0.9488	0.9901
	49.50	0.99	0.9000	0.9482	0.9876
	1.00	0.50	0.8970	0.9487	0.9889
5.00	2.33	0.70	0.9005	0.9488	0.9883
	19.00	0.95	0.8958	0.9460	0.9867
	99.00	0.99	0.8992	0.9465	0.9897
20.00	5.00	0.50	0.8982	0.9484	0.9887
	11.67	0.70	0.8999	0.9490	0.9878
	95.00	0.95	0.8990	0.9484	0.9901
	495.00	0.99	0.9037	0.9497	0.9901
	20.00	0.50	0.9007	0.9489	0.9895
	46.70	0.70	0.9006	0.9531	0.9908
	380.00	0.95	0.9046	0.9513	0.9905
	1980.00	0.99	0.8970	0.9491	0.9885

Tabla 4.7. Probabilidades de cobertura: Tamaño de muestra grande y diferente.

$M=10000$ $n=100$ y $m=50$			Intervalo de verosimilitud-confianza		
$(1-\gamma)$			0.90	0.95	0.99
$\alpha$	$\beta$	$\theta$			
0.10	0.10	0.50	0.8972	0.9497	0.9916
	0.23	0.70	0.8986	0.9479	0.9898
	1.90	0.95	0.9009	0.9514	0.9912
	9.90	0.99	0.8982	0.9504	0.9908
0.50	0.50	0.50	0.8984	0.9500	0.9903
	1.17	0.70	0.9021	0.9503	0.9881
	9.50	0.95	0.8972	0.9485	0.9885
	49.50	0.99	0.9023	0.9494	0.9897
1.00	1.00	0.50	0.9016	0.9529	0.9902
	2.33	0.70	0.8994	0.9470	0.9889
	19.00	0.95	0.9041	0.9492	0.9914
	99.00	0.99	0.9016	0.9522	0.9917
5.00	5.00	0.50	0.9010	0.9527	0.9919
	11.67	0.70	0.8965	0.9491	0.9901
	95.00	0.95	0.9006	0.9513	0.9907
	495.00	0.99	0.9010	0.9462	0.9887
20.00	20.00	0.50	0.8940	0.9483	0.9893
	46.70	0.70	0.9022	0.9528	0.9888
	380.00	0.95	0.9028	0.9509	0.9897
	1980.00	0.99	0.8961	0.9495	0.9908

Tabla 4.8. Probabilidades de cobertura: Tamaño de muestra pequeño e igual.

$M=10000$ $n=m=4$			Intervalos						
			Verosimilitud-Confianza			Kotz et al. (2003)			
$(1-\gamma)$			0.90	0.95	0.99	0.90	0.95	0.99	
$\alpha$	$\beta$	$\theta$							
0.10	0.10	0.50	0.8823	0.9355	0.9828	0.8397	0.8943	0.9511	
		0.23	0.70	0.8789	0.9326	0.9790	0.8442	0.8941	0.9507
		1.90	0.95	0.8763	0.9318	0.9804	0.8981	0.9235	0.9550
		9.90	0.99	0.8821	0.9350	0.9823	0.9052	0.9291	0.9554
0.50	0.50	0.50	0.8774	0.9335	0.9824	0.8390	0.8894	0.9494	
		1.17	0.70	0.8759	0.9313	0.9809	0.8433	0.8938	0.9510
		9.50	0.95	0.8790	0.9327	0.9809	0.9047	0.9289	0.9561
		49.50	0.99	0.8784	0.9333	0.9816	0.9031	0.9257	0.9540
1.00	1.00	0.50	0.8820	0.9337	0.9809	0.8378	0.8928	0.9480	
		2.33	0.70	0.8809	0.9364	0.9820	0.8489	0.9004	0.9532
		19.00	0.95	0.8844	0.9361	0.9836	0.9017	0.9302	0.9578
		99.00	0.99	0.8842	0.9384	0.9840	0.9091	0.9335	0.9602
5.00	5.00	0.50	0.8877	0.9368	0.9809	0.8419	0.8995	0.9522	
		11.67	0.70	0.8880	0.9364	0.9832	0.8542	0.9036	0.9537
		95.00	0.95	0.8822	0.9320	0.9807	0.9026	0.9258	0.9550
		495.00	0.99	0.8891	0.9399	0.9832	0.9104	0.9298	0.9572
20.00	20.00	0.50	0.8824	0.9362	0.9817	0.8379	0.8933	0.9521	
		46.70	0.70	0.8747	0.9289	0.9799	0.8431	0.8920	0.9487
		380.00	0.95	0.8803	0.9297	0.9809	0.9087	0.9305	0.9567
		1980.00	0.99	0.8849	0.9348	0.9819	0.9085	0.9295	0.9569

Tabla 4.9. Probabilidades de cobertura: Tamaño de muestra pequeño e igual.

$M=10000$ $n=m=8$			Intervalos						
			Verosimilitud-Confianza			Kotz et al. (2003)			
$(1-\gamma)$			0.90	0.95	0.99	0.90	0.95	0.99	
$\alpha$	$\beta$	$\theta$							
0.10	0.10	0.50	0.8900	0.9387	0.9858	0.8584	0.9118	0.9658	
		0.23	0.70	0.8903	0.9415	0.9851	0.8677	0.9205	0.9673
		1.90	0.95	0.8888	0.9426	0.9864	0.9127	0.9398	0.9706
		9.90	0.99	0.8901	0.9439	0.9848	0.9099	0.9373	0.9668
0.50	0.50	0.50	0.8839	0.9367	0.9847	0.8547	0.9062	0.9622	
		1.17	0.70	0.8861	0.9383	0.9866	0.8650	0.9150	0.9660
		9.50	0.95	0.8868	0.9386	0.9856	0.9104	0.9373	0.9675
		49.50	0.99	0.8875	0.9400	0.9848	0.9120	0.9376	0.9663
1.00	1.00	0.50	0.8930	0.9429	0.9850	0.8598	0.9139	0.9663	
		2.33	0.70	0.8975	0.9443	0.9862	0.8787	0.9212	0.9671
		19.00	0.95	0.8900	0.9408	0.9847	0.9095	0.9365	0.9654
		99.00	0.99	0.8875	0.9424	0.9873	0.9098	0.9343	0.9664
5.00	5.00	0.50	0.8891	0.9439	0.9864	0.8562	0.9127	0.9672	
		11.67	0.70	0.8917	0.9447	0.9866	0.8734	0.9223	0.9692
		95.00	0.95	0.8855	0.9382	0.9857	0.9071	0.9348	0.9641
		495.00	0.99	0.8911	0.9399	0.9854	0.9124	0.9360	0.9656
20.00	20.00	0.50	0.8873	0.9403	0.9851	0.8561	0.9106	0.9636	
		46.70	0.70	0.8924	0.9415	0.9842	0.8680	0.9205	0.9666
		380.00	0.95	0.8890	0.9409	0.9873	0.9078	0.9358	0.9668
		1980.00	0.99	0.8892	0.9402	0.9870	0.9147	0.9383	0.9650

Tabla 4.10. Probabilidades de cobertura: Tamaño de muestra pequeño y diferente.

$M=10000$ $n=4$ y $m=8$			Intervalo de verosimilitud-confianza		
$(1-\gamma)$			0.90	0.95	0.99
$\alpha$	$\beta$	$\theta$			
0.10	0.10	0.50	0.8885	0.9410	0.9851
	0.23	0.70	0.8831	0.9385	0.9858
	1.90	0.95	0.8823	0.9366	0.9864
	9.90	0.99	0.8905	0.9401	0.9833
0.50	0.50	0.50	0.8828	0.9357	0.9834
	1.17	0.70	0.8880	0.9405	0.9854
	9.50	0.95	0.8851	0.9363	0.9856
	49.50	0.99	0.8893	0.9398	0.9853
1.00	1.00	0.50	0.8799	0.9370	0.9866
	2.33	0.70	0.8874	0.9406	0.9847
	19.00	0.95	0.8888	0.9359	0.9829
	99.00	0.99	0.8830	0.9371	0.9846
5.00	5.00	0.50	0.8869	0.9392	0.9838
	11.67	0.70	0.8850	0.9366	0.9837
	95.00	0.95	0.8893	0.9402	0.9856
	495.00	0.99	0.8832	0.9380	0.9836
20.00	20.00	0.50	0.8849	0.9377	0.9822
	46.70	0.70	0.8832	0.9367	0.9829
	380.00	0.95	0.8794	0.9348	0.9837
	1980.00	0.99	0.8867	0.9378	0.9836

Tabla 4.11. Probabilidades de cobertura: Tamaño de muestra pequeño y diferente.

$M=10000$ $n=8$ y $m=4$			Intervalo de verosimilitud-confianza			
$(1-\gamma)$			0.90	0.95	0.99	
$\alpha$	$\beta$	$\theta$				
0.10	0.10	0.50	0.8847	0.9396	0.9837	
		0.23	0.70	0.8886	0.9390	0.9848
		1.90	0.95	0.8840	0.9395	0.9860
		9.90	0.99	0.8884	0.9387	0.9850
0.50	0.50	0.50	0.8871	0.9391	0.9859	
		1.17	0.70	0.8894	0.9402	0.9858
		9.50	0.95	0.8896	0.9401	0.9845
		49.50	0.99	0.8887	0.9395	0.9847
1.00	1.00	0.50	0.8814	0.9337	0.9832	
		2.33	0.70	0.8853	0.9392	0.9836
		19.00	0.95	0.8832	0.9375	0.9845
		99.00	0.99	0.8867	0.9393	0.9833
5.00	5.00	0.50	0.8858	0.9425	0.9860	
		11.67	0.70	0.8861	0.9364	0.9837
		95.00	0.95	0.8831	0.9379	0.9830
		495.00	0.99	0.8852	0.9354	0.9834
20.00	20.00	0.50	0.8851	0.9404	0.9867	
		46.70	0.70	0.8861	0.9395	0.9839
		380.00	0.95	0.8860	0.9351	0.9830
		1980.00	0.99	0.8924	0.9442	0.9862

# Bibliografía

- [1] Adimari, G. y Chiogna, M. (2005). Partially Parametric Interval Estimation of  $\Pr\{Y > X\}$ . *Computational Statistics & Data Analysis*, V. 51, pág. 1875 – 1891.
- [2] Aitkin, M. y Stasinopoulos, M. (1989). Likelihood Analysis of a Binomial Sample Size Problem. In *Contributions to Probability and Statistics*, eds. L. J. Gleser, M. D. Perlman, S. J. Press and A. R. Sampson. Nueva York: Springer, pág. 339-411.
- [3] Barnard, G. A. (1966). The Use of the Likelihood Function in Statistical Practice. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, V. 1, pág. 27 – 40.
- [4] Barnard, G. A. (1983). Pivotal Inference and the Conditional View of Robustness. *Statistical Science, Data Analysis, and Robustness*, Academic Press Inc.
- [5] Barnard, G. A. y Sprott, D. A. (1983). Likelihood. *Encyclopedia of Statistical Science*, V. 4, pág. 639 – 644. Editado por Kotz, S. y Johnson, N. L.: Nueva York: John Wiley & Sons.
- [6] Barndorff-Nielsen, O. y Cox, D. R. (1994). *Inference and Asymptotics*. Londres: Chapman y Hall.
- [7] Bartholomew, D. J. (1957). A Problem in Life Testing . *Journal of the American Statistical Association*, V. 52, pág. 350 – 355.

- [8] Berger, J. O., Liseo, B. y Wolpert, R. L. (1999). Integrated Likelihood Methods for Eliminating Nuisance Parameters. *Statistical Science*, V. 14, pág. 1 – 28.
- [9] Birnbaum, Z. W. (1956). On a Use of the Mann–Whitney Statistic. *Proceedings of the Third Berkeley Symposium on Mathematics Statistics and Probability*, V. 1, pág. 13 – 17.
- [10] Birnbaum, Z. W. y McCarthy, R. C. (1958). A Distribution Free Upper Confidence Bound for  $P(Y < X)$  Based on Independent Samples of  $X$  and  $Y$ . *Annals of Mathematical Statistics*, V. 29, pág. 558 – 562.
- [11] Borches, D. L, Buckland, S. T. y Zucchini, W. (2002). *Estimation Animal Abundance: Closed Populations*. Nueva York: Springer-Verlag.
- [12] Box, G. E. P. y Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical, Series B*, V. 26, pág. 211 – 252.
- [13] Box, G. E. P. y Tiao, G. C. (1973). *Bayesian Statistical Inference*. Addison Wesley: Reading, Mass.
- [14] Carroll, R. J. y Lombard, F. (1985). A Note on  $n$  Estimators for the Binomial Distribution. *Journal of the American Statistical Association*, V. 80, pág. 423–426.
- [15] Casella, G. (1986). Stabilizing Binomial  $n$  Estimators. *Journal of the American Statistical Association*, V. 81, pág. 172 – 175.
- [16] Chao, A. (1982). On Comparing Estimators of  $P(Y < X)$  in the Exponential Case. *IEEE Transactions on Reliability*, V. 31, pág. 389 – 392.
- [17] Church, J. D. y Harris, B. (1970). The Estimation of Reliability from Stress–Strength Relationship. *Technometrics*, V. 12 , pág. 49 – 54.

- [18] Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Londres: Springer-Verlag.
- [19] DasGupta, A. y Rubin, H. (2004). Estimation of Binomial Parameters When Both  $n, p$  are Unknown. *Journal of Statistical Planning and Inference*, V. 130, pág. 391 – 404.
- [20] Downton, F. (1973). The Estimation of  $\Pr(Y < X)$  in the Normal Case. *Technometrics*, V. 15, pág. 551 – 558.
- [21] Draper, N. y Guttman, I. (1971). Bayesian Estimation of the Binomial Parameter. *Technometrics*, V. 13, pág. 667 – 673.
- [22] Dudley, R. M. (1989). *Real Analysis and Probability*. California: Wadsworth.
- [23] Edwards, A. W. F. (1992). *Likelihood*. Baltimore: The Johns Hopkins University Press.
- [24] Embrechts, P., Klüppelberg, C. y Mikosch, T. (1997). *Modelling Extremal Events*. Berlin: Springer-Verlag.
- [25] Fisher, R. A. (1921). On the “Probable Error” of a Coefficient of Correlation Deduced from a Small Sample. *Metron*, V. 1, pág. 3 – 32.
- [26] Fisher, R. A. (1941). The Negative Binomial Distribution. *Annals of Eugenics*, V. 6, pág. 391 – 398.
- [27] Fisher, R. A. (1991). *Statistical Methods and Scientific Inference* (Tercera Edición). Oxford: Oxford University Press.
- [28] Gupta, A. K., Nguyen, T. T. y Wang, Y. (1999). On Maximum Likelihood Estimation of the Binomial Parameter  $n$ . *The Canadian Journal of Statistics*, V. 27, pág. 599 – 606.

- [29] Gupta, R. D. y Kundu, D. (2007). Generalized Exponential Distribution: Existing Results and Some Recent Developments. *Journal of Statistical Planning and Inference*, V. 137, pág. 3537 – 3547.
- [30] Guttman, I., Johnson, R. A., Bhattacharyya, G. K. y Reiser, B. (1988). Confidence Limits for Stress-Strength Models with Explanatory Variables. *Technometrics*, V. 30, pág. 161 – 168.
- [31] Hall, P. (1994). On the Erratic Behavior of Estimators of  $N$  in the Binomial  $N, p$  Distribution. *Journal of the American Statistical Association*, V. 89, pág. 344 – 351.
- [32] Hood, W. C. y Koopmans, T. C. (Editores) (1953). *Studies in Econometric Method*. Nueva York: John Wiley.
- [33] Ivshin V. V. y Lumelskii Y. P. (1995). Unbiased Estimators for Density Functions and Probabilities of Linear Inequalities in the Multivariate Normal Case, *Proceedings of the Fifteenth Seminar-Stability Problems for Stochastic models*, Perm, Russia, pág. 71-80.
- [34] Jana, P. K. (1997). Comparison of Some Stress-Strength Reliability Estimators. *Calcutta Statist. Assoc. Bull.*, V. 47, pág. 239 – 247.
- [35] Kalbfleisch, J. G. (1985). *Probability and Statistical Inference. Vol. 2: Statistical Inference* (Segunda Edición). Nueva York: Springer-Verlag.
- [36] Kalbfleisch, J. G. y Sprott, D. A. (1970). Application of Likelihood Methods to Models Involving Large Numbers of Parameters (con discusión). *Journal of the Royal Statistical Society*, V. B 32, pág. 175 – 208.
- [37] Kahn, W. D. (1987). A Cautionary Note for Bayesian Estimation of the Binomial Parameter  $n$ . *The American Statistician*, V. 41, pág. 38 – 40.

- [38] Kelley, G. D., Kelley, J. A. y Schucany, W. R. (1976). Efficient Estimation of  $P(Y < X)$  in the Exponential Case. *Technometrics*, V. 18, No. 3, pág. 359 – 360.
- [39] Kotz, S., Lumelskii, Y. y Pensky, M. (2003). *The Stress-Strength Model and its Generalizations: Theory and Applications*. Singapore: World Scientific.
- [40] Kundu, D. y Gupta, R. D. (2005). Estimation of  $P[Y < X]$  for Generalized Exponential Distribution. *Metrika*, V. 61, pág. 291 – 308.
- [41] Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*. Nueva York: John Wiley & Sons.
- [42] Lindsey, J. K. (1999). Some Statistical Heresies (con discusión). *The Statistician*, V. 48, pág. 1 – 40.
- [43] Mazundar, M. (1970). Some Estimating of Reliability Using Inference Theory. *Naval Research Logistics Quarterly*, V. 17, pág. 159 – 165.
- [44] Meeker, W. Q. y Escobar, L. A. (1998). *Statistical Methods for Reliability Data*. Nueva York: John Wiley & Sons.
- [45] Montoya, J. A. (2004). *El Modelo Binomial  $(n, p)$  para Estimar la Abundancia de Animales*. Tesis de Maestría en Ciencias con Especialidad en Probabilidad y Estadística, CIMAT, Guanajuato, Gto., México.
- [46] Montoya, J. A., Díaz-Francés, E. y Sprott, D. A. (2007). On a Criticism of the Profile Likelihood Function. *Statistical Papers*. Disponible en Internet a partir de marzo 2007.
- [47] Mood, A. M., Graybill, F. A. y Boes, D. C. (1985). *Introduction to the Theory of Statistics*. Singapore: McGraw-Hill.

- [48] Moran, P. A. P. (1951). A Mathematical Theory of Animal Trapping. *Biometrika*, V. 38, pág. 307 – 311.
- [49] Nelson, W. (1990). *Accelerated Testing: Statistical Models, Test Plans and Data Analyses*. Nueva York: John Wiley & Sons.
- [50] Olkin, I., Petkau, A. J. y Zidek, J. V. (1981). A Comparison of  $n$  Estimators for the Binomial Distribution. *Journal of the American Statistical Association*, V. 76, pág. 637 – 642.
- [51] Pawitan, Y. (2004). *In all Likelihood: Statistical Modelling and Inference using Likelihood*. Oxford: Clarendon Press.
- [52] Raftery, A., E. (1988). Inference for the Binomial  $n$  Parameter: A hierarchical Bayes Approach. *Biometrika*, V. 75, pág. 223 – 228.
- [53] Reiser, B. y Guttman, I. (1986). Statistical Inference for  $\Pr(Y < X)$  : The Normal Case. *Technometrics*, V. 28, pág. 253 – 257.
- [54] Sathe, Y. S. y Shah, S. P. (1981). On Estimating  $P(X < Y)$  for the Exponential Distribution. *Communications in Statistics-Theory and Methods*, V. A10, pág. 39 – 47.
- [55] Seber, G. A. F. (1982). *The Estimation of Animal Abundance and Related Parameters*. Londres: Charles Griffin.
- [56] Seber, G. A. F. y Wild, C. J. (2003). *Non Linear Regression*. Nueva Jersey: John Wiley & Sons.
- [57] Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Nueva York: John Wiley & Sons.
- [58] Sprott, D. A. (2000). *Statistical Inference in Science*. Nueva York: Springer-Verlag.

- [59] Sprott, D. A. y J. D. Kalbfleisch (1969). Examples of Likelihoods and Comparison with Point Estimates and Large Sample Approximations. *Journal of the American Statistical Association*, V. 64, pág. 468 – 484.
- [60] Tong, H. (1974). A Note on the Estimation of  $\Pr(Y < X)$  in the Exponential Case. *Technometrics*, V. 16, pág. 625.
- [61] Tong, H. (1975). Errata: A Note on the Estimation of  $\Pr(Y < X)$  in the Exponential Case. *Technometrics*, V. 17, pág. 395.
- [62] Tsui, K. y Weerahandi, S. (1989). Generalized p-Values in Significance Testing of Hypotheses in Presence of Nuisance Parameters. *Journal of the American Statistical Association*, V. 84, pág. 602 – 607.
- [63] Weerahandi, S. y Johnson, R. A. (1992). Testing Reliability in a Stress-Strength Model When  $X$  and  $Y$  are Normally Distributed. *Technometrics*, V. 34, pág. 83 – 91.
- [64] Woodward, W. A. y Kelley, G. D. (1977). Minimum Variance Unbiased Estimation of  $P[Y < X]$  in the Normal Case. *Technometrics*, V. 19, No. 1, pág. 95 – 98.