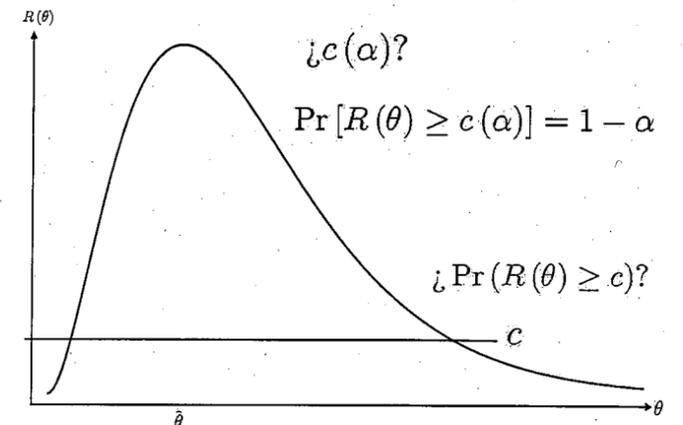


Centro de Investigación en Matemáticas, A.C.

CIMAT

## Calibración de Intervalos de Verosimilitud



**T E S I S**

Que para obtener el grado de:

**Doctor en ciencias**

Con orientación en:

**Probabilidad y Estadística**

P r e s e n t a:

**Jesús Armando Domínguez Molina**

Guanajuato, Gto. México. Febrero de 2001

## Calibración de Intervalos de Verosimilitud

*A Emilio*

*A Lupita*

*A mis padres*



CIMAT  
BIBLIOTECA

## Agradecimientos

A mi asesor el Dr. Víctor Manuel Pérez-Abreu porque gracias a él inicié y finalicé mis estudios de doctorado. Por su apoyo durante el transcurso de todos mis estudios de postgrado y por todo.

Al Dr. David Sprott por su labor como coasesor, sus fructíferos comentarios, enriquecedoras discusiones y su muy valiosa ayuda en la parte final de mi tesis.

Al Dr. Ole Barndorff-Nielsen por su coasesoría, por las discusiones y sugerencias sobre mi tesis que me resultaron de gran utilidad.

Al Dr. Miguel Nakamura porque, después de varias discusiones con él, logré iniciar la redacción final de mi tesis.

Al Dr. Jens Ledet Jensen porque las sugerencias que de él recibí me ayudaron a encontrar el camino para la solución de mi problema de tesis.

Al Dr. Adolfo Sánchez Valenzuela por el tiempo que me dedicó para explicarme algunos conceptos relacionados con geometría diferencial.

A mis sinodales, los doctores:

Rolando Biscay, Eloísa Díaz-Francés, Miguel Nakamura y Román Viveros-Aguilera. Especialmente a Román Viveros Aguilera por el reporte de la revisión de mi tesis.

C I M A T  
B I B L I O T E C A

015987

Al SNI por la beca de ayudante de SNI III que me fue asignada por el Dr. Víctor Manuel Pérez-Abreu.

Al CONACYT por mi beca No. 90238 y al CONCYTEG por su apoyo parcial.

A la Universidad de Aarhus por el hospedaje y las facilidades prestadas por la institución durante mi estancia de dos meses en el otoño de 1996.

Por supuesto, al CIMAT por toda la infraestructura que me brindó, por los apoyos económicos por medio de beca y porque me permitió trabajar un año como asistente del Dr. Miguel Nakamura.

A mis amigos por las discusiones que tuve con ellos sobre mi tesis y el interés mostrado en mi trabajo desde el principio: Dr. Enrique Villa, M.E. Román de la Vara, M.E. Netzahualcóyotl Castañeda.

Finalmente, a mi esposa Lupita por su apoyo en todos los sentidos, incluyendo el económico durante todo el doctorado y a mi hermano Ramón por su apoyo (*e.g.*, tiempo, auto e impresora) durante los últimos dos y medio años de mi doctorado.

# Contenido

<b>Introducción</b>	<i>xi</i>
<b>1 Preliminares</b>	<b>1</b>
1.1 Introducción . . . . .	1
1.1.1 Probabilidad de cobertura de regiones de verosimilitud . . . . .	3
1.2 Calibración de regiones de verosimilitud . . . . .	5
1.3 Suficiencia, auxiliaridad . . . . .	7
1.3.1 Notación . . . . .	9
1.4 La verosimilitud dirigida . . . . .	10
1.5 La verosimilitud dirigida modificada . . . . .	10
1.5.1 Sugerencias para calcular $u$ . . . . .	13
1.5.2 La verosimilitud dirigida modificada en función de $r$ . . . . .	15
<b>2 Calibración de intervalos de verosimilitud</b>	<b>19</b>
2.1 Probabilidades de cobertura más precisas . . . . .	21
2.2 Probabilidad de cobertura mínima . . . . .	22
2.3 Calibración de intervalos de verosimilitud . . . . .	24

2.3.1	Aproximación del factor de Bartlett . . . . .	8
2.4	Ejemplos sin parámetros de estorbo . . . . .	8
2.4.1	El parámetro de localización del modelo hiperbólico . . . . .	8
2.4.2	La distribución exponencial . . . . .	8
2.4.3	Distribución de extremos . . . . .	8
2.4.4	El modelo gaussiana inversa . . . . .	8
2.4.5	El modelo gaussiana inversa normal . . . . .	8
2.5	Ejemplos con Parámetros de estorbo . . . . .	6
2.5.1	El modelo normal . . . . .	6
2.5.2	El modelo Gaussiana inversa . . . . .	6
2.5.3	El modelo gama . . . . .	6
2.5.4	Regresión exponencial . . . . .	6
2.5.5	Distribución sobre una esfera . . . . .	6
2.6	Discusión . . . . .	7
<b>3</b>	<b>Justificación de resultados</b> . . . . .	<b>8</b>
3.1	Condiciones de regularidad . . . . .	8
3.2	Parámetro $\theta$ escalar . . . . .	8
3.2.1	Densidad de la verosimilitud dirigida . . . . .	8
3.2.2	Probabilidad de cobertura de un intervalo de verosimilitud . . . . .	8
3.2.3	Resultados auxiliares . . . . .	8
3.3	Verificación de cinco equivalencias . . . . .	9
3.3.1	Fórmula $p^*$ . . . . .	9
3.3.2	Factores de Bartlett . . . . .	9

3.3.3	Fórmula- $r^*$ . . . . .	95
3.3.4	Aproximación de Lugannani-Rice . . . . .	96
<b>4</b>	<b>Precisión de <math>(r^*)^2</math></b> . . . . .	<b>99</b>
4.1	$\theta$ escalar . . . . .	100
4.2	Parámetros de estorbo: $\psi$ escalar . . . . .	101
4.3	Discusión . . . . .	103
<b>A</b>	<b>Expansiones asintóticas</b> . . . . .	<b>105</b>
A.1	Álgebra de series asintóticas . . . . .	105
A.1.1	Cociente de series . . . . .	105
A.1.2	Raíz de series . . . . .	106
A.1.3	Inversa de Series. . . . .	108
A.2	Expansiones asintóticas de $r$ y $r^*$ . . . . .	108
A.2.1	Parámetro $\theta$ escalar . . . . .	108
A.2.2	Parámetro $\theta = (\psi, \chi)$ , $\psi$ escalar . . . . .	113
<b>B</b>	<b>Obtención de <math>u</math> en el modelo exponencial-<math>(m, m)</math></b> . . . . .	<b>119</b>
<b>C</b>	<b>Distribución gaussiana inversa normal</b> . . . . .	<b>123</b>
	<b>Bibliografía</b> . . . . .	<b>125</b>
	<b>Glosario</b> . . . . .	<b>131</b>

## Introducción

Esta tesis está orientada a problemas de estimación que surgen de experimentos repetibles cuyas observaciones se distribuyen de acuerdo con una función que envuelve un parámetro  $\theta$  desconocido. Con base en datos de un experimento consideramos que la manera óptima de resumir información sobre el parámetro  $\theta$  es por medio de regiones de verosimilitud complementadas con sus probabilidades de cobertura mínima. Óptima porque podemos cuantificar de dos maneras informativas y complementarias la incertidumbre sobre  $\theta$ . Específicamente tenemos dos medidas de incertidumbre: la verosimilitud para cada punto dentro de la región y la probabilidad para la región misma.

El objetivo principal de esta tesis es *calibrar* regiones de verosimilitud. Es decir, construir regiones de verosimilitud con probabilidad de cobertura mínima especificada. Esto conlleva a estudiar, como función de  $\theta$ , a la probabilidad de cobertura de las regiones de verosimilitud. Las herramientas que utilizamos para lograr nuestro objetivo son principalmente la verosimilitud dirigida,  $r$ , de Sprott (ver sec. 1.4), la verosimilitud dirigida modificada,  $r^*$ , de Barndorff-Nielsen (ver sec. 1.5) y los factores de Bartlett (ver sec. 3.3.2).

Las regiones de verosimilitud fueron introducidas por Fisher en 1956 como una forma de separar los valores plausibles de los implausibles, con base en una muestra observada

(Fisher 1993, pág. 71-78). Barnard (1965) sugiere la posibilidad de utilizar las regiones de verosimilitud como regiones de confianza para  $\theta$ . Barnard & Sprott (1983) muestran que es posible contar con regiones con nivel de verosimilitud *baja* y probabilidad de cobertura muy *baja*. A un intervalo de verosimilitud con probabilidad de cobertura constante Sprott (1984) le llama intervalo de verosimilitud-confianza. Kalbfleisch (1985, sec. 11.2) ilustra el cálculo explícito de la probabilidad de cobertura de intervalos de verosimilitud en algunos modelos estadísticos e ilustra la posible dependencia sobre  $\theta$  de esta probabilidad. Chamberlin & Sprott (1991) discuten en detalle los intervalos de verosimilitud-confianza hablando en términos de «fidelity» y «accuracy», además calculan mediante simulación algunas probabilidades de cobertura. Barnard (1996) afirma que la manera óptima de hacer inferencia sobre un parámetro es por medio de regiones de verosimilitud acompañados con una medida de confianza dado por la cota inferior de la probabilidad de cobertura, argumentando que en estos tiempos, debido a los avances en computación, ya no es impedimento calcular este valor utilizando métodos de Montecarlo.

Las regiones de verosimilitud-confianza aproximadas han sido consideradas por Sprott (1984, 1996, 2000), Viveros (1985), Viveros & Sprott (1987), Chamberlin & Sprott (1991), Díaz-Francés (1998), Meeker & Escobar (1998) y Díaz-Francés & Sprott (2000). Estos trabajos consideran principalmente regiones que son intervalos. El procedimiento que utilizan para construir intervalos de verosimilitud-confianza aproximados, consiste en aproximar mediante series de Taylor la función de verosimilitud observada a una verosimilitud más simple de trabajar (la normal, *t*-Student o Log *F*) y así obtienen un intervalo de verosimilitud aproximado con nivel de confianza dado por la verosimilitud que hayan aproximado. Cuando alguna de estas aproximaciones no ajusta bien se busca una reparametrización donde la

verosimilitud sea aproximadamente la deseada. Cabe destacar que el procedimiento que utilizan, para construir los intervalos de verosimilitud-confianza aproximados, no es invariante bajo reparametrizaciones uno a uno del parámetro  $\theta$  y no consideran que la probabilidad de cobertura puede depender de  $\theta$ .

A diferencia de los trabajos arriba mencionados, en esta tesis no consideramos aproximaciones a la función de verosimilitud. Nos enfocamos en obtener expresiones analíticas muy precisas de la probabilidad de cobertura de regiones de verosimilitud.

Utilizamos conjuntamente la verosimilitud dirigida,  $r$  (Sección 1.4), de Sprott (1973, 1975, 1980) y la verosimilitud dirigida modificada,  $r^*$  (Sección 1.5), de Barndorff-Nielsen (1986, 1991) para obtener una medida de la probabilidad de cobertura suficientemente precisa para indicarnos si existe dependencia funcional sobre  $\theta$ . Sin embargo, cabe destacar que los intervalos de verosimilitud son únicos y dependen sólo de  $r^2$ , no sobre  $r$ . Esto incrementa la precisión de sus probabilidades de cobertura, comparados con todos aquellos que se obtienen con  $r$ . Lo mismo sucede con  $r^*$  (ver Capítulo 4).

La verosimilitud dirigida modificada,  $r^*$ , se utiliza en el cálculo aproximado de la probabilidad de eventos cola, en la construcción de intervalos de confianza, en la construcción de *qq*-plots, en análisis de residuos y como parte de una estadística auxiliar. La idea de usar  $r^*$  para medir la probabilidad de cobertura surge debido a su relación cercana con la verosimilitud dirigida y a que tiene una distribución normal aproximada con error de orden  $O(n^{-3/2})$ . Mediante  $r$  y  $r^*$  obtenemos expresiones de la probabilidad de cobertura con error de orden  $O(n^{-2})$  y que nos permiten estudiar la dependencia funcional sobre  $\theta$  (cuando se presenta). Además, en esta tesis demostramos que las expresiones que aquí obtenemos para medir probabilidades de coberturas de intervalos de verosimilitud son equivalentes, con error

$O(n^{-2})$ , a las expresiones que obtuviéramos con otros cuatro métodos de aproximación que provienen de la teoría de aproximación por órdenes mayores.

Existen varias aproximaciones usadas comúnmente en la literatura de altos órdenes, las cuales han sido especialmente propuestas para diferentes situaciones: la verosimilitud dirigida modificada,  $r^*$ , mencionada anteriormente; la aproximación de Lugannani-Rice (Lugannani & Rice 1980) que es una aproximación de la distribución de la media de observaciones independientes y se utiliza especialmente para el cálculo aproximado de probabilidades de eventos cola; la fórmula- $p^*$  (Barndorff-Nielsen 1983) que es una aproximación a la densidad condicional del estimador de máxima verosimilitud y tiene un amplio uso en el cálculo de densidades de cantidades que involucran al estimador de máxima verosimilitud; los factores de Bartlett (Bartlett 1937, Lawley 1956 y Barndorff-Nielsen & Cox 1984) que se utilizan para construir intervalos de confianza o efectuar pruebas de hipótesis. Finalmente consideramos una fórmula que es equivalente todas las aproximaciones anteriores. En todos los casos las aproximaciones son invariantes.

Aún cuando cada aproximación citada anteriormente fue diseñada para un propósito diferente, es de interés notar que cuando se utilizan las cinco para medir coberturas de regiones de verosimilitud obtenemos aproximaciones que son equivalentes con error  $O(n^{-2})$ .

Conviene señalar que  $r^*$  y la aproximación de Lugannani-Rice están definidas sólo para  $\theta = (\psi, \chi)$ ,  $\psi$  escalar, mientras que  $p^*$  y Bartlett para  $\theta$  general. La importancia de este resultado radica en el hecho de que cuando la dimensión de  $\psi$  es mayor que uno, podemos usar  $p^*$  o Bartlett para aproximar la probabilidad de cobertura de una región de verosimilitud.

Durante el transcurso de la investigación observamos que para medir probabilidades de

cobertura de regiones de verosimilitud podemos usar al menos cinco procedimientos. Sin embargo, para calibrar las regiones de verosimilitud analíticamente es indispensable utilizar los factores de Bartlett. Lo anterior le da una nueva interpretación a los factores de Bartlett, debido a que anteriormente fueron introducidos de manera *ad hoc* como una corrección a la estadística log razón de verosimilitud para aproximarla a la distribución  $\chi^2$ . En este trabajo ilustramos que al resolver de manera analítica el problema de calibración, los factores de Bartlett emergen de manera natural.

El uso de los factores de Bartlett para calibrar regiones de verosimilitud evita la necesidad de utilizar un procedimiento numérico que utilice iterativamente la aproximación de la probabilidad de cobertura obtenida con cualquiera de los cinco procedimientos anteriores. Sin embargo los factores de Bartlett son particularmente difíciles de evaluar en presencia de parámetros de estorbo, por lo que en este trabajo proponemos aproximaciones del factor de Bartlett basadas en  $r^*$  o la aproximación de Lugannani-Rice las cuales tienen la propiedad de ser aproximaciones del factor de Bartlett con error de orden  $O(n^{-2})$ .

La organización de este trabajo es la siguiente. El Capítulo 1 contiene los conceptos claves para el desarrollo de la tesis, los cuales están estrechamente relacionados con el concepto de verosimilitud. Se muestra el procedimiento estándar para medir la probabilidad de cobertura de regiones de verosimilitud y se presenta en detalle la verosimilitud dirigida modificada  $r^*$ , así como algunas sugerencias para calcularla en algunos modelos específicos.

El Capítulo 2 se concentra en el caso  $\theta = (\psi, \chi)$  y  $\psi$  escalar. Presenta expresiones analíticas aproximadas de la probabilidad de cobertura de intervalos de verosimilitud más precisas que el procedimiento estándar del Capítulo 1. También se presentan sugerencias numéricas para calibrar la cobertura de intervalos de verosimilitud, así como varios ejemplos

analíticos y numéricos que ilustran los procedimientos propuestos en esta tesis. En particular, en la Sección 2.3.1 se proponen algunas aproximaciones numéricas del factor de Bartlett basadas en  $r^*$  o en la aproximación de Lugannani-Rice.

El Capítulo 3 contiene las demostraciones de todos los resultados de esta tesis para el caso en que  $\psi$  es escalar. El Capítulo 4 contiene la demostración, cuando  $\theta$  es escalar, de que  $(r^*)^2$  tiene distribución  $\chi_1^2$  con error de orden  $O(n^{-2})$ , en muestras repetidas ordinarias. Se ilustra numéricamente que lo anterior también se cumple en el caso en que  $\theta = (\psi, \chi)$ ,  $\psi$  escalar. Por ello obtenemos que los intervalos de confianza obtenidos con  $r^*$  tienen nivel de confianza especificado con error de orden  $O(n^{-2})$ . Cabe destacar que en el transcurso del desarrollo de este trabajo estudiamos a detalle el error de aproximación del nivel de confianza los intervalos obtenidos con  $r^*$  y se encuentra que cuando construimos intervalos de confianza con  $(r^*)^2$  el error es de orden  $O(n^{-2})$ . Esto no había sido observado en Barndorff-Nielsen (1986) donde se afirma que el error de aproximación del nivel de confianza de los intervalos construidos con  $r^*$  es de orden  $O(n^{-3/2})$ . Con  $r^*$  podemos construir intervalos de confianza con error de aproximación de orden  $O(n^{-2})$ , mediante la observación de que los intervalos construidos con  $(r^*)^2 \leq z$  son los mismos que los construidos con  $-\sqrt{z} \leq r^* \leq \sqrt{z}$ ,  $z > 0$ .

## Capítulo 1

### Preliminares

En este capítulo se presentan los conceptos claves para el desarrollo de esta tesis. La Sección 1.1 contiene los conceptos necesarios para construir regiones de verosimilitud. En la Sección 1.2 se discute el procedimiento común para medir la probabilidad de cobertura de una región de verosimilitud. La Sección 1.3 contiene los conceptos de suficiencia, auxiliaridad y derivada con respecto al estimador de máxima verosimilitud, conceptos relacionados con la función de verosimilitud vistos con el fin de definir la función de verosimilitud dirigida modificada.

#### 1.1 Introducción

Consideremos una colección  $y = (y_1, \dots, y_n)^T$  de observaciones de variables aleatorias independientes e idénticamente distribuidas (v.a.i.i.d.) con función de distribución conocida, excepto por el parámetro  $\theta \in \Omega \subseteq \mathbf{R}^d$ ,  $d \geq 1$ . La *función de verosimilitud*  $L(\theta; y)$  es una función de  $\theta$  para cada valor de  $y$ , la cual se define por

$$L(\theta; y) = K(y) \prod_{i=1}^n p_Y(y_i; \theta),$$

donde  $K(y)$  es una función positiva arbitraria que no depende de  $\theta$  y  $p_Y(y_i; \theta)$  es la función de densidad de  $y_i$ . La *función de verosimilitud relativa*  $R(\theta)$  se define como

$$R(\theta) = R(\theta; y) = \frac{L(\theta; y)}{L(\hat{\theta}; y)}, \quad (1.1)$$

donde  $\hat{\theta}$  es el *estimador de máxima verosimilitud* de  $\theta$ .

La manera primaria de resumir información sobre  $\theta$ , con base en una muestra observada, es mediante regiones de verosimilitud. Definimos una *región de verosimilitud*, con *nivel de verosimilitud*  $c$ ,  $0 < c < 1$ ; como la región que consiste de los valores  $\theta \in \Omega$  tales que  $R(\theta; y) \geq c$ . Éstas contienen los valores más plausibles de  $\theta$ , según el nivel de verosimilitud dado.

Cuando  $\dim(\theta) > 1$ , es común que se considere  $\theta = (\psi, \chi)$ , donde  $\chi$  es un *parámetro de estorbo* y al espacio paramétrico  $\Omega$  de la forma  $\Omega = \Omega_\psi \times \Omega_\chi$ . Nos interesa resumir información sobre  $\psi$  por medio de regiones similares a las de verosimilitud. Para lograr esto consideraremos la *función de verosimilitud perfil* de  $\psi$ , definida por

$$L_p(\psi) = L_p(\psi; y) = L(\psi, \hat{\chi}_\psi; y),$$

donde  $\hat{\chi}_\psi$  es el estimador de máxima verosimilitud de  $\chi$  restringido a un valor fijo de  $\psi$  (ver Capítulo 3). Existen otros conceptos de *pseudo-verosimilitud* en el caso de parámetros de estorbo; sin embargo optamos por la verosimilitud perfil porque ésta se comporta de manera más cercana a una función de verosimilitud genuina (Barndorff-Nielsen & Cox 1994, pág. 90). A su vez, también se puede disponer de las pseudo-verosimilitudes que surgen de modificaciones a la verosimilitud perfil misma (ver Barndorff-Nielsen 1985, 1990). Sprott (2000, Cap. 4) considera varios tipos de verosimilitud en el caso de parámetros de estorbo.

La *función de verosimilitud perfil relativa*  $R_p(\psi)$  se define similarmente que (1.1); es decir,  $R_p(\psi) = L_p(\psi) / L_p(\hat{\psi})$ , las *regiones de verosimilitud perfil*, con nivel de verosimilitud  $c$ , se definen mediante los valores de  $\psi \in \Omega_\psi$  tales que  $R_p(\psi) \geq c$ . Esta región contiene los valores de  $\psi$  tales que, para algún  $\chi$ , la pareja  $(\psi, \chi)$  pertenece a la región de verosimilitud  $R(\psi, \chi) \geq c$ .

En virtud de que únicamente consideraremos verosimilitud y verosimilitud perfil, a partir de ahora frecuentemente omitiremos el término perfil, quedando claro el tipo de verosimilitud cuando haya parámetros de estorbo.

### 1.1.1 Probabilidad de cobertura de regiones de verosimilitud

Estamos interesados en estudiar la *probabilidad de cobertura* de regiones de verosimilitud. Sea  $k = \dim(\psi)$ ,  $1 \leq k \leq d$  y denotemos a las regiones de verosimilitud, con nivel de verosimilitud  $c$ , por

$$LR(c; y) = \{\omega \in \Omega_\psi : R_p(\omega; y) \geq c\}. \quad (1.2)$$

Para calcular la probabilidad de cobertura de una región de verosimilitud, debemos considerar a la muestra  $Y = (Y_1, \dots, Y_n)$  como variable aleatoria y a  $\psi$  fijo en su valor verdadero. Denotaremos por  $CP(c, \theta)$  a la *probabilidad de cobertura* de una región de verosimilitud, así pues

$$CP(c, \theta) = \Pr(\psi \in LR(c; Y)). \quad (1.3)$$

Escribimos la probabilidad de cobertura (1.3) en función de  $\theta$ , porque en general el valor de  $\Pr(\psi \in LR(c; Y))$  depende de  $(\psi, \chi)$ . Dado que  $\psi \in LR(c; Y)$  si, y sólo si,  $R_p(\psi; Y) \geq c$ ,

obtenemos que

$$\text{CP}(c, \theta) = \Pr(R_p(\psi; Y) \geq c), \quad (1.4)$$

donde  $R_p(\psi; Y)$  se considera como función de  $Y$  para  $\psi$  fijo. La relación (1.4) nos muestra que con la distribución de la verosimilitud relativa se puede calcular la probabilidad de cobertura de una región de verosimilitud de manera directa. En general la distribución de  $R_p$  es complicada, por lo que conviene obtener la distribución de  $R_p$  aproximadamente. Bajo condiciones de regularidad estándar (ver Capítulo 3)  $W_p = -2 \ln R_p(\psi)$  se distribuye de acuerdo con la distribución  $\chi_k^2$  con error, en muestras repetidas ordinarias, de orden  $O(n^{-1})$ .

La aproximación es en el siguiente sentido

$$\Pr(W_p \leq w) = Q_k(w) [1 + O(n^{-1})], \quad (1.5)$$

donde  $Q_k$  es la función de distribución  $\chi_k^2$ . Mediante (1.5) obtenemos una medida de la probabilidad de cobertura de un intervalo de verosimilitud con error de orden  $O(n^{-1})$ . Para  $0 < c < 1$ , sea  $w_c = -2 \ln c$ , entonces

$$\begin{aligned} \text{CP}(c, \theta) &= \Pr(R_p \geq c) \\ &= \Pr(W_p \leq w_c) \\ &= Q_k(w_c) [1 + O(n^{-1})]. \end{aligned} \quad (1.6)$$

De (1.6) podemos decir que las regiones de verosimilitud, con nivel de verosimilitud  $c$ , tienen probabilidad de cobertura aproximada  $Q_k(w_c)$ .

## 1.2 Calibración de regiones de verosimilitud

Por *calibrar* una región de verosimilitud queremos decir que fijamos una probabilidad de cobertura  $1 - \alpha$  y calculamos el correspondiente nivel de verosimilitud que satisfaga  $\Pr(R_p \geq c) = 1 - \alpha$ . Esto es diferente a que meramente midamos la probabilidad de cobertura de una región de verosimilitud. Éste es un problema de calibración y su solución es simple si trabajamos con error de orden  $O(n^{-1})$ : por medio de (1.6) la relación entre la probabilidad de cobertura y el nivel de verosimilitud de una región de verosimilitud está dada por

$$\alpha(c) = 1 - Q_k(w_c) \quad \text{y} \quad c(\alpha) = e^{-\frac{1}{2}q_{1-\alpha}}, \quad (1.7)$$

donde  $q_{1-\alpha}$  es tal que  $Q_k(q_{1-\alpha}) = 1 - \alpha$ . Entonces el problema de calibración se resuelve al tomar  $c(\alpha) = e^{-\frac{1}{2}q_{1-\alpha}}$ .

Cuando  $\psi$  es escalar se obtiene que

$$\Pr(R_p \geq c) = [2\Phi(\sqrt{w_c}) - 1] [1 + O(n^{-1})], \quad (1.8)$$

donde  $\Phi$  es la función de distribución de la normal estándar. La relación (1.8) da lugar a las siguientes relaciones

$$\alpha(c) = 2\Phi(-\sqrt{-2 \ln c}), \quad c(\alpha) = e^{-\frac{1}{2}z_{1-\alpha/2}^2}. \quad (1.9)$$

Para ilustrar las fórmulas de (1.9) consideremos una muestra  $y = (y_1, \dots, y_n)^T$  de v.a.i.i.d. de una distribución normal  $N(\mu, 1)$ . En este caso la verosimilitud relativa es  $R(\mu) = \exp(-\frac{n}{2}(\hat{\mu} - \mu)^2)$ , de donde obtenemos que

$$W = -2 \ln R = (\hat{\mu} - \mu)^2 \stackrel{D}{=} \chi_1^2.$$

Observe que en este caso  $W$  tiene distribución exacta  $ji$ -cuadrada con un grado de libertad, en consecuencia se cumple que

$$\Pr(R_p \geq c) = 2\Phi(\sqrt{w_c}) - 1$$

y

$$\Pr\left(R_p \geq e^{-\frac{1}{2}z_1^2 - \alpha/2}\right) = 1 - \alpha,$$

donde  $z_\alpha$  es tal que  $\Phi(z_\alpha) = \alpha$ . Las gráficas en la Figura 1.1 ilustran el uso de (1.9), la Figura 1.1(a) muestra la verosimilitud normal con  $\hat{\mu} = 0$ ,  $n = 3$  y la ordenada derecha construida con la relación  $\alpha(c) = 2\Phi(-\sqrt{-2\ln c})$ . La Figura 1.1(b) es también una función de verosimilitud, pero ahora calibrada; la ordenada derecha representa la probabilidad de cobertura en escala equi-espaciada y la ordenada izquierda se construyó mediante la relación  $c(\alpha) = e^{-\frac{1}{2}z_1^2 - \alpha/2}$ .

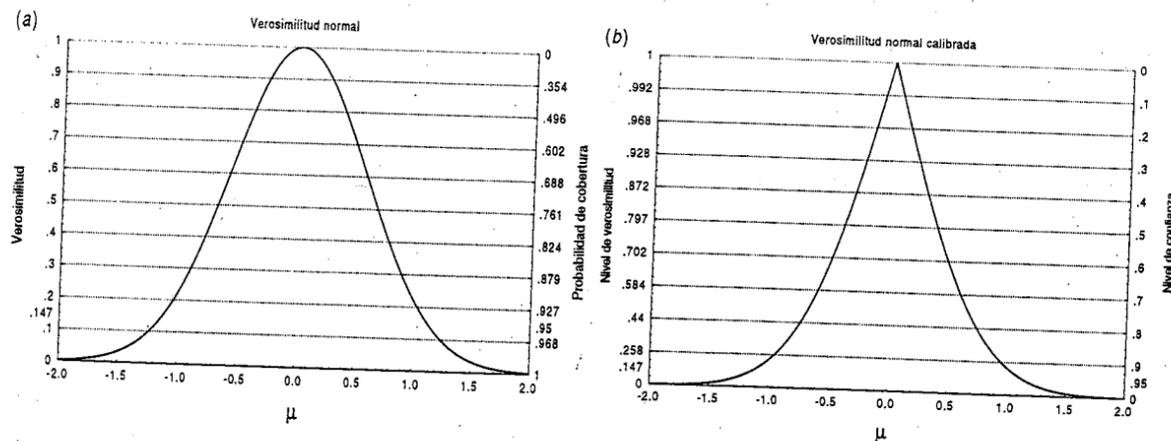


Figura 1.1. Funciones de verosimilitud relativa normal.

(a) Verosimilitud relativa normal. (b) Verosimilitud relativa normal calibrada

El procedimiento (1.6) es muy simple de implementar, únicamente se necesitan la función de verosimilitud y el estimador de máxima verosimilitud. Sin embargo presenta tres principales inconvenientes: El primero, no nos indica si existe dependencia sobre  $\theta$ . El segundo, la probabilidad de cobertura no es representativa para los datos que se hayan observado, debido a que no condiciona aunque haya estadísticas auxiliares maximales. Y el tercero, para muestras de tamaño pequeño la aproximación puede ser poco precisa. Para el caso  $\psi$  escalar, en el Capítulo 2 mostramos aproximaciones de la probabilidad de cobertura de intervalos de verosimilitud que no padecen de los problemas anteriores.

### 1.3 Suficiencia, auxiliaridad

Mediante suficiencia y auxiliaridad se puede mejorar notablemente la aproximación (1.6); se pueden obtener aproximaciones de la probabilidad de cobertura con error de orden  $O(n^{-2})$ , es decir más precisión. Además, si se presentan estadísticas auxiliares se pueden usar para calcular probabilidades de cobertura representativas para un conjunto de datos en mano. También es posible obtener explícitamente la dependencia de  $\psi$  de la probabilidad de cobertura, lo que nos ayuda a obtener probabilidades de cobertura mínima de intervalos de verosimilitud.

Una estadística  $T = T(y)$  es cualquier función de los datos y se dice que es *suficiente para  $\theta$*  si la distribución condicional de  $y$ , dado  $T$ , no depende de  $\theta$ ; es decir, dentro de cada partición de la muestra inducida por  $T$ , la distribución de  $y$  no depende de  $\theta$ . Una estadística suficiente proporciona una reducción de los datos sin ninguna pérdida de información sobre  $\theta$ . La *estadística minimal suficiente*  $S = S(y)$  es la que proporciona la mayor reducción de los datos. Cuando se tiene suficiencia observe que la función de verosimilitud es función sólo

de una estadística suficiente  $s$  y el parámetro  $\theta$ . Si la dimensión de  $s$  es la misma que la de  $\theta$  usualmente el estimador de máxima verosimilitud es función uno-a-uno de  $s$  y entonces  $\hat{\theta}$  es minimal suficiente si, y sólo si,  $s$  es minimal suficiente. De aquí que se pueda considerar a la función de verosimilitud de la forma  $L(\theta; \hat{\theta})$ , ésta es la misma que  $L(\theta; Y)$  como si los datos observados consistieran solamente de  $\hat{\theta}$ . En este caso se recomienda hacer inferencia sobre  $\theta$  con base en la distribución de  $\hat{\theta}$  o  $S$ .

En los modelos donde la estadística minimal suficiente  $S$  es de dimensión mayor que la de  $\theta$ , se recomienda construir una estadística  $a$  de manera que  $S$  sea función uno-a-uno de  $(\hat{\theta}, a)$ , para así escribir la función de verosimilitud como  $L(\theta; \hat{\theta}, a)$ . La estadística  $a$  es *auxiliar* si la distribución de  $a$  no depende de  $\theta$  y si la partición del espacio muestral generada por  $a$  es la más grande posible  $a$  es *auxiliar maximal*. En este caso se recomienda hacer inferencia sobre  $\theta$  con base en la distribución condicional de  $\hat{\theta}$  dado  $a$ . Sobre el papel de condicionar en Estadística ver Reid (1995).

Denotemos por  $\ell(\theta) = \ln L(\theta)$  al logaritmo natural de la función de verosimilitud. A  $\ell(\theta)$  se le conoce como la *función de log-verosimilitud*. Si la verosimilitud es perfil se denotará por  $\ell_p(\psi)$  a  $\ln L_p(\psi)$ . De aquí en adelante consideraremos la función de verosimilitud expresada como  $L(\theta; \hat{\theta}, a)$ , la estadística auxiliar  $a$  se mantendrá fija en todo momento. Bajo este supuesto, para valores fijos de  $\theta$  y  $a$  se puede estudiar la relación funcional de  $\ell(\theta; \hat{\theta}, a)$  como función de  $\hat{\theta}$ ; podemos hacer operaciones del cálculo infinitesimal con respecto a  $\hat{\theta}$ , por ejemplo límites y derivadas. Así, pues

$$\lim_{\hat{\theta} \rightarrow w} \ell(\theta; \hat{\theta}, a) = \lim_{x \rightarrow w} \ell(\theta; x, a) \quad \text{y} \quad \left. \frac{\partial}{\partial \hat{\theta}} \ell(\theta; \hat{\theta}, a) \right|_{\hat{\theta}=w} = \left. \frac{\partial}{\partial x} \ell(\theta; x, a) \right|_{x=w}, \quad (1.10)$$

en (1.10)  $\hat{\theta}$  juega el papel de variable muda. Considerando que  $R(\theta; Y) = R(\theta; \hat{\theta}, a)$  se

puede observar que, mediante técnicas ordinarias de cambio de variable, con la distribución de  $\hat{\theta}$  se puede calcular la distribución de  $R$ , o cualquier otra función de  $\hat{\theta}$ . En general sólo se puede conocer aproximadamente la distribución de  $\hat{\theta}$  (ver Capítulo 3).

### 1.3.1 Notación

Consideremos la siguiente notación

$$\ell_{j;k}(\theta; \hat{\theta}, a) = \frac{\partial^{j+k}}{\partial \theta^j \partial \hat{\theta}^k} \ell(\theta; \hat{\theta}, a), \quad j, k \geq 0 \quad (1.11)$$

con las siguientes convenciones

$$l_i \equiv l_{i;0}, \quad l_{i;j} = l_{0;j}, \quad \hat{l}_{i;j} \equiv l_{i;j}(\hat{\theta}; \hat{\theta}, a), \quad \bar{l}_{i;j} \equiv l_{i;j}(\theta; \theta, a), \quad i, j \geq 1. \quad (1.12)$$

Cuando  $\theta = (\psi, \chi)$ ;  $\psi$  escalar, consideramos  $\ell(\psi, \chi) = \ell(\psi, \chi; \hat{\psi}, \hat{\chi}, a)$ . En este caso utilizaremos la siguiente notación:

$$\begin{aligned} l_\psi &= \frac{\partial}{\partial \psi} \ell(\psi, \chi), \quad l_{\psi\psi} = \frac{\partial^2}{\partial \psi^2} \ell(\psi, \chi), \quad \text{etc.} \\ l_\chi &= \nabla_\chi \ell(\psi, \chi) = (l_{\chi_1}, \dots, l_{\chi_{d-1}}) = \left( \frac{\partial}{\partial \chi_1} \ell(\psi, \chi), \dots, \frac{\partial}{\partial \chi_{d-1}} \ell(\psi, \chi) \right) \\ l_{\chi\chi} &= \nabla_\chi^T \nabla_\chi \ell = (l_{\chi_i \chi_j})_{i,j=1}^{d-1} = \left( \frac{\partial^2}{\partial \chi_i \partial \chi_j} \ell_{\chi_i \chi_j}(\psi, \chi) \right)_{i,j=1}^{d-1} \\ l_{\hat{\psi}} &= \frac{\partial}{\partial \hat{\psi}} \ell(\psi, \chi; \hat{\psi}, \hat{\chi}); \quad l_{\hat{\chi}} = \nabla_{\hat{\chi}} \ell(\psi, \chi; \hat{\psi}, \hat{\chi}); \quad l_{\chi; \hat{\psi}} = \nabla_\chi \ell_{\hat{\psi}}(\psi, \chi; \hat{\psi}, \hat{\chi}); \\ l_{\chi; \hat{\chi}} &= \nabla_\chi^T \nabla_{\hat{\chi}} \ell(\psi, \chi; \hat{\psi}, \hat{\chi}), \quad l_{\psi \chi; \hat{\chi}} = \nabla_\chi^T \nabla_{\hat{\chi}} \ell_\psi(\psi, \chi; \hat{\psi}, \hat{\chi}), \quad \text{etc.} \\ l_{\theta\theta} &= \nabla_\theta^T \nabla_\theta \ell = (l_{\theta_i \theta_j})_{i,j=1}^d = \left( \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ell_{\theta_i \theta_j}(\psi, \chi) \right)_{i,j=1}^d \end{aligned}$$

Se usará, sobre una función de  $(\psi, \chi)$ , una tilde “~” o un circunflejo “^” de acuerdo a si se evalúa en  $\hat{\chi}_\psi$  o en  $\hat{\theta}$ . Así  $\tilde{\ell} = \ell(\psi, \hat{\chi}_\psi; \hat{\theta}, a)$  y  $\hat{\ell} = \ell(\hat{\theta}; \hat{\theta}, a) = \ell(\hat{\psi}, \hat{\chi}; \hat{\psi}, \hat{\chi}, a)$ . Además, cuando consideramos cantidades que se consideran función de  $\hat{\psi}$ , utilizaremos una barra “-” arriba de la cantidad para indicar evaluación en  $\hat{\psi} = \psi$ . Por ejemplo,  $\bar{\ell} = \ell(\psi, \chi; \hat{\chi}, \psi)$ .

Los conceptos que se presentan en las siguientes secciones son para el caso  $\psi$  escalar.

### 1.4 La verosimilitud dirigida

Sprott (1973, 1975, 1980) inició el estudio de la *estadística log razón de verosimilitud* para un parámetro de interés escalar. Por simplicidad nos referiremos a ella como la *verosimilitud dirigida*, la cual está dada por

$$r = \text{sgn}(\hat{\psi} - \psi) \left\{ 2 \left[ l_p(\hat{\psi}) - l_p(\psi) \right] \right\}^{\frac{1}{2}}. \quad (1.13)$$

La verosimilitud dirigida sigue aproximadamente la distribución de una variable aleatoria con distribución normal estándar con error, en muestras repetidas ordinarias, de orden  $O(n^{-1/2})$ . La aproximación es en el siguiente sentido

$$\Pr(r \leq x) = \Phi(x) [1 + O(n^{-1/2})], \quad (1.14)$$

además, dado que  $r^2 = W_p$ , por (1.5) se tiene

$$\Pr(r^2 \leq x^2) = Q_1(x^2) [1 + O(n^{-1})], \quad (1.15)$$

donde  $Q_1$  es la distribución de la  $\chi^2$ -cuadrada con un grado de libertad. Con (1.15) obtenemos la probabilidad de cobertura aproximada para intervalos de verosimilitud.

Debido a que  $R_p(\psi) = e^{-\frac{1}{2}r^2}$ , se puede verificar fácilmente que los valores de  $\psi$  que cumplen  $R_p(\psi) \geq c$  son los mismos que los que satisfacen  $-\sqrt{2 \ln c} \leq r \leq \sqrt{2 \ln c}$ .

### 1.5 La verosimilitud dirigida modificada

Con el fin de mejorar la aproximación normal de la verosimilitud dirigida, ésta ha sufrido varias modificaciones; una de ellas es la *verosimilitud dirigida modificada definida* por Barndorff-

Nielsen (1986) como

$$r^* = r + \frac{1}{r} \log(u/r), \quad (1.16)$$

aquí  $r$  es la verosimilitud dirigida (1.13) y

$$u = \frac{|\hat{\ell}_{;\hat{\theta}} - \bar{\ell}_{;\hat{\theta}} \quad \bar{\ell}_{x;\hat{\theta}}|}{|\tilde{J}_{xx}|^{1/2} |\hat{j}|^{1/2}}, \quad (1.17)$$

donde  $[\hat{\ell}_{;\hat{\theta}} - \bar{\ell}_{;\hat{\theta}} \quad \bar{\ell}_{x;\hat{\theta}}]$  es una matriz  $d \times d$  cuyo  $i$ -ésimo renglón es  $(\hat{\ell}_{;\hat{\theta}_i} - \bar{\ell}_{;\hat{\theta}_i}, \bar{\ell}_{x_1;\hat{\theta}_i}, \dots, \bar{\ell}_{x_{d-1};\hat{\theta}_i})$ ,  $|\cdot|$  es la función determinante y  $\hat{j} = -\hat{\ell}_{\theta\theta}$ ,  $\tilde{J}_{xx} = -\bar{\ell}_{xx}$  (ver Sección 1.3.1).

El signo de  $u$  debe ser el mismo que el de  $r$ , si no lo es, se debe multiplicar  $u$  por  $-1$ .

Otra expresión de  $u$  es

$$u = C \tilde{u}, \quad (1.18)$$

donde

$$\tilde{u} = \hat{j}_p^{-1/2} \bar{\ell}_{p/\hat{\psi}}, \quad C = C(\psi) = |\bar{\ell}_{x;\hat{x}}| \{|\tilde{J}_{xx}| |\hat{J}_{xx}|\}^{-1/2}$$

y

$$\hat{j}_p = \frac{\partial^2}{\partial \psi^2} l_p(\psi) \Big|_{\psi=\hat{\psi}}, \quad \bar{\ell}_p = \hat{\ell} - l_p, \quad \bar{\ell}_{p/\hat{\psi}} = \frac{\partial}{\partial \hat{\psi}} \bar{\ell}_p = \hat{\ell}_{;\hat{\psi}} - \bar{\ell}_{;\hat{\psi}} + \hat{\chi}_{/\hat{\psi}} (\hat{\ell}_{;\hat{x}} - \bar{\ell}_{;\hat{x}}).$$

La derivada del  $\hat{\chi}$  con respecto a  $\hat{\psi}$ ,  $\hat{\chi}_{/\hat{\psi}} = \frac{d}{d\hat{\psi}} \hat{\chi}$  se obtiene de la relación  $l_x(\psi, \hat{\chi}_\psi; \hat{\psi}, \hat{\chi}) = 0$ , para  $\hat{\chi}_\psi$  fijo. Así, pues

$$\hat{\chi}_{/\hat{\psi}} = -\bar{\ell}_{x;\hat{\psi}} (\bar{\ell}_{x;\hat{x}})^{-1}.$$

Cuando no hay parámetros de estorbo  $C = 1$ . Ver Barndorff-Nielsen (1986,1991), Barndorff-Nielsen y Cox (1994), Reid (1993) y Reid & Fraser (2000).

La verosimilitud dirigida modificada es una cantidad invariante bajo transformaciones 1-1 del parámetro de interés. Bajo muestras repetidas ordinarias  $r^*$  sigue la distribución normal estándar con aproximación de orden  $O(n^{-3/2})$ , i.e.

$$\Pr(r^* \leq x|a) = \Phi(x) [1 + O(n^{-3/2})]. \quad (1.19)$$

En virtud de que el lado derecho de (1.19) no depende de  $a$  el resultado es válido marginalmente.

Una aportación de esta tesis es probar que, en ausencia de parámetros de estorbo,  $r^*$  cumple

$$\Pr((r^*)^2 \leq x^2) = 2\Phi(x) - 1 + O(n^{-2}), \quad (1.20)$$

lo cual muestra que los intervalos de confianza construidos con  $(r^*)^2$  tienen probabilidad de cobertura de orden  $O(n^{-2})$  en muestras repetidas ordinarias. La relación (1.20) se demuestra en la Sección 4.1. En presencia de parámetros de estorbo, no se ha verificado analíticamente que se cumpla (1.20), sin embargo la evidencia numérica sugiere que en efecto lo cumple (ver Capítulo 4). Podemos utilizar la propiedad (1.20) para construir intervalo de confianza con nivel de confianza aproximado con error de orden  $O(n^{-2})$ ; sin embargo, a diferencia de Sección 1.4, estos intervalos en general no son intervalos de verosimilitud.

En el Capítulo 2 se muestran otras nuevas propiedades de  $r^*$ , relacionadas con la utilidad de  $r^*$  para medir la probabilidad de cobertura de los intervalos de verosimilitud y calibrar intervalos de verosimilitud.

### 1.5.1 Sugerencias para calcular $u$

El artificio principal para calcular  $u$  de (1.17) o (1.18) consiste en escribir la verosimilitud en función de  $\hat{\theta}$  y  $a$ . Esto es simple en la familia exponencial, ya que en esta familia la función de log-verosimilitud es de la forma

$$\ell(\theta, Y) = \ell(\theta, S) = \theta^T S - k(\theta),$$

por lo tanto se cumple que

$$\nabla k(\theta)|_{\theta=\hat{\theta}} = S,$$

en consecuencia la función de log-verosimilitud expresada en términos de  $\hat{\theta}$  y  $\theta$  es

$$\ell(\theta; \hat{\theta}) = \theta^T \nabla k(\hat{\theta}) - k(\theta), \quad (1.21)$$

en este caso no se necesita condicionar, ya que  $\hat{\theta}$  es minimal suficiente.

En el modelo de localización escala es simple también puesto que, con  $a_i = (y_i - \hat{\mu})/\hat{\sigma}$ , la función de log-verosimilitud se expresa como

$$\begin{aligned} \ell(\theta, Y) &= \sum_{i=1}^n \ln f\left(\frac{y_i - \mu}{\sigma}\right) \\ &= \sum_{i=1}^n \ln f\left[\frac{\hat{\sigma}}{\sigma} \left(a_i + \frac{\hat{\mu} - \mu}{\hat{\sigma}}\right)\right]. \end{aligned} \quad (1.22)$$

En particular si el modelo es de localización, se obtiene que

$$\ell(\mu; Y) = \sum_{i=1}^n \ln f(a_i + \hat{\mu} - \mu), \quad (1.23)$$

donde  $a_i = y_i - \hat{\mu}$ . Y si el modelo es de escala, con  $a_i = y_i/\hat{\sigma}$  tenemos

$$\ell(\sigma; Y) = \sum_{i=1}^n \ln f\left(a_i \frac{\hat{\sigma}}{\sigma}\right). \quad (1.24)$$

En otro caso hay que tener bien definida la estadística auxiliar y hacer el cambio de variable  $Y \longleftrightarrow (\hat{\theta}, a)$ .

Después de escribir la función de verosimilitud en términos de  $\hat{\theta}$  y  $a$ , el resto es utilizar las relaciones (1.17) o (1.18) como recetas matemáticas.

Cuando no hay parámetros de estorbo  $u$  se simplifica a

$$u = \hat{j}^{-1/2} (\hat{\ell}_{;\hat{\theta}} - \ell_{;\hat{\theta}}). \quad (1.25)$$

En particular en el modelo exponencial se obtiene que

$$u = [k''(\hat{\theta})]^{1/2} (\hat{\theta} - \theta), \quad (1.26)$$

y en el modelo de localización  $u$  se expresa como

$$u = \left[ \sum_{i=1}^n g''(a_i) \right]^{-1/2} \sum_{i=1}^n g'(a_i + \hat{\mu} - \mu). \quad (1.27)$$

En presencia de parámetros de estorbo ocurre un caso importante cuando  $\hat{\chi}_\psi = \hat{\chi}$ . Si  $\hat{\chi}_\psi = \hat{\chi}$  se cumple lo siguiente

$$\hat{j}_p = \hat{j}_{\psi\psi}, \quad \bar{\ell}_{p/\psi} = \hat{\ell}_{;\hat{\psi}} - \bar{\ell}_{;\hat{\psi}}, \quad C = \hat{j}_{XX}^{-1/2} \bar{j}_{XX}^{1/2},$$

de donde

$$u = \hat{j}_{XX}^{-1/2} \bar{j}_{XX}^{1/2} \hat{j}_{\psi\psi}^{-1/2} (\hat{\ell}_{;\hat{\psi}} - \bar{\ell}_{;\hat{\psi}}), \quad (1.28)$$

La expresión de  $u$  dada en (1.28) se cumple también cuando  $\psi$  es ortogonal al parámetro de estorbo, en cuyo caso  $\hat{\chi}_\psi = \hat{\chi} + O(n^{-1})$  (ver Barndorff-Nielsen 1991). Otros casos especiales para el modelo exponencial se pueden ver en el Apéndice B.

### 1.5.2 La verosimilitud dirigida modificada en función de $r$

La verosimilitud dirigida modificada ha mostrado su utilidad en el cálculo de probabilidades de eventos cola, en el análisis de residuos, en la construcción de intervalos de confianza y como parte de una estadística auxiliar. En esta tesis probamos que también es útil para calibrar intervalos de verosimilitud. Para este fin en el Capítulo 2 veremos que es necesario expresar  $r^*$  en función de  $r$ , para valores fijos de  $\psi$ ,  $\hat{\chi}_\psi$  y  $a$ .

Supondremos que  $r^*$  y  $r$  son funciones uno-a-uno de  $\hat{\psi}$ , para valores fijos de  $\psi$ ,  $\hat{\chi}_\psi$ ,  $a$ . Para expresar la verosimilitud dirigida modificada como función de  $r$  es suficiente con expresar  $u$  en función de  $r$ . Así tendremos que

$$r^*(\psi; r) = r + \frac{1}{r} \ln \left( \frac{u(\psi; r)}{r} \right), \quad (1.29)$$

los Ejemplos 2.4.4, 2.4.5 y 2.5.5 presentan la relación exacta. Pero en general dicha relación sólo se puede expresar aproximadamente. De la Sección A.2 obtenemos que la expresión de  $u$  en función de  $r$  es de la forma

$$u(\psi; r) = r + \bar{t}_3 r^2 + (\bar{t}_3^2 - \frac{1}{2}b) r^3 + O_p(n^{-3/2}), \quad (1.30)$$

donde  $\bar{t}_3$ ,  $b$  son funciones de  $\psi$ ,  $\hat{\chi}_\psi$  y  $a$ . Usando (1.30) y (1.29) junto con la conocida expansión  $\ln(1+x) = x - \frac{1}{2}x^2 + \dots$  aplicada con  $x = \bar{t}_3 r + (\bar{t}_3^2 - \frac{1}{2}b) r^2$  se llega a que

$$r^* = \bar{t}_3 + (1 + \frac{1}{2}\bar{t}_3^2 - \frac{1}{2}b) r + O_p(n^{-3/2})$$

La expresiones de  $\bar{t}_3$  y  $b$  en el caso de  $\theta$  escalar son

$$\begin{aligned} \bar{t}_3 &= \frac{1}{6} (\bar{\ell}_3 + 3\bar{\ell}_{2;1}) \bar{j}^{-3/2} \\ b(\theta, a, n) &= \frac{1}{24n} \{ 5\bar{\ell}_3^2 + 24\bar{\ell}_{2;1} (\bar{\ell}_3 + \bar{\ell}_{2;1}) + [3\bar{\ell}_4 + 12(\bar{\ell}_{3;1} + \bar{\ell}_{2;2})] \bar{j} \} \bar{j}^{-3}. \end{aligned} \quad (1.31)$$

Para el caso con parámetros de estorbo en la Sección A.2.2 se ilustra cómo obtener  $b$ . La cantidad  $\bar{t}_3$  no es de utilidad para calibrar intervalos de verosimilitud, por lo que no es necesario obtenerla explícitamente. En el Capítulo 2 presentamos una aproximación de  $b$  en términos de  $r^*$  la cual es útil principalmente en presencia de parámetros de estorbo.

En los modelos exponencial y de localización el factor  $b$  está dado por

$$b(\theta, a, n) = \frac{1}{24} (5\bar{\ell}_3^2 + 3\bar{\ell}_4\bar{j}) \bar{j}^{-3}, \quad (1.32)$$

en el modelo exponencial  $b$  depende de  $\theta$ , i.e.,  $b = b(\theta, n)$  y en el de localización  $b$  depende de  $a$ , pero no de  $\theta$ , i.e.,  $b = b(a, n)$ .

Otra forma de evaluar  $r^*$ , como función de  $r$ , es resolviendo con respecto a  $\hat{\psi}$  la ecuación

$$r(\psi; \hat{\psi}) = r_0. \quad (1.33)$$

Los valores de  $\hat{\psi}$  que resuelven (1.33) se obtienen de manera simple sólo cuando  $\theta$  es escalar o  $\chi$  es ortogonal a  $\psi$ . Estos valores son

$$\hat{\psi} = \psi + \bar{j}_p^{-1/2} r_0 + \frac{1}{6} (2\bar{\ell}_{p3} + 3\bar{\ell}_{p2;1}) \bar{j}_p^{-2} r_0^2 \pm \frac{1}{72} (5(2\bar{\ell}_{p3} + 3\bar{\ell}_{p2;1})^2 + 3(3\bar{\ell}_{p4} + 8\bar{\ell}_{p3;1} + 6\bar{\ell}_{p2;2}) \bar{j}_p) \bar{j}_p^{-7/2} r_0^3 + O(n^{-2}), \quad (1.34)$$

donde

$$\ell_{pj;k}(\psi; \hat{\psi}, a) = \frac{\partial^{j+k}}{\partial \psi^j \partial \hat{\psi}^k} \ell(\psi, \hat{\chi}; \hat{\psi}, \hat{\chi}, a), \quad j, k \geq 0,$$

y se cumple que  $\frac{d}{d\hat{\psi}} \hat{\chi} = 0$ .

Sin embargo si  $\psi$  no es ortogonal a  $\chi$ , puede ser más conveniente resolver (1.33) numéricamente.

Antes de resolver (1.33) numérica o analíticamente, conviene ver con más detalle la relación funcional entre  $r$  y  $\hat{\psi}$ . Por el momento usaremos en la notación todas las variables involucradas. Entonces, la verosimilitud dirigida se escribe como

$$r(\psi; \hat{\psi}) = r(\psi, \hat{\chi}_\psi; \hat{\psi}, \hat{\chi}, a).$$

Observe que el lado izquierdo sólo indica dependencia con respecto a  $\psi$  y  $\hat{\psi}$ . Esto es porque para considerar a  $r$  como función de  $\hat{\psi}$ , se deben mantener fijos los valores de  $\psi$ ,  $a$  y  $\hat{\chi}_\psi$ . Donde  $\hat{\chi}_\psi$  se considera parte de la estadística auxiliar, supondremos que  $\hat{\chi}_\psi$  es suficiente para  $\chi$  y auxiliar para  $\psi$ . El estimador de máxima verosimilitud  $\hat{\chi}$  de  $\chi$  no aparece en la parte izquierda porque al mantener fijos  $\psi$ ,  $\hat{\chi}_\psi$  y  $a$  y exigir que se siga cumpliendo  $\ell_\chi(\psi, \hat{\chi}_\psi; \hat{\psi}, \hat{\chi}, a) = 0$  obliga a  $\hat{\chi}$  a ser función de  $\hat{\psi}$ , entonces debemos resolver

$$r(\psi, \hat{\chi}_\psi; \hat{\psi}, \hat{\chi}(\hat{\psi}; \psi, \hat{\chi}_\psi, a), a) = r_0, \quad (1.35)$$

en general resolver (1.35) con respecto a  $\hat{\psi}$  no es más complicado que resolver (1.33) con respecto a  $\psi$ , para valores fijos de  $\hat{\psi}$  y  $\hat{\chi}$ . En los ejemplos con parámetros de estorbo del Capítulo 2 se presentan expresiones explícitas de  $\hat{\chi}$  en función de  $\hat{\psi}$ .

Si  $\hat{\psi}_0$  es solución de (1.35) entonces

$$r^*(\psi; r_0) = r^*(\psi; \hat{\psi}_0). \quad (1.36)$$

## Capítulo 2

### Calibración de intervalos de verosimilitud

En este capítulo consideramos que el parámetro  $\theta$  se descompone de la forma  $\theta = (\psi, \chi)$ ,  $\psi$  escalar. Suponemos que las regiones de verosimilitud  $LR(c, Y)$  definidas en (1.2) son intervalos de verosimilitud y los denotaremos por  $LI(c, Y)$ , es decir

$$LI(c, Y) = \{\psi \in \Omega_\psi : R_p(\psi; Y) \geq c\}.$$

El procedimiento para calcular la probabilidad de cobertura de un intervalo de verosimilitud discutido en los preliminares se reduce a

$$\begin{aligned} CP(c; \theta) &= \Pr(\psi \in LI(c, Y)) \\ &= \Pr(R_p(\psi; Y) \geq c) \\ &= 2\Phi(v_c) - 1 + O(n^{-1}), \end{aligned} \tag{2.1}$$

donde  $v_c = \sqrt{-2 \ln c}$ . La aproximación (2.1) puede ser poco precisa en muestras de tamaño pequeño, sin embargo un problema más serio que presenta tal aproximación es la falta de

información: no informa si la probabilidad de cobertura depende de  $\psi$ , ni condiciona en estadísticas auxiliares, cuando fuera necesario.

En este capítulo mostramos que trabajando conjuntamente con la verosimilitud dirigida,  $r$ , y la verosimilitud dirigida modificada,  $r^*$ , se pueden obtener expresiones para la probabilidad de cobertura que presentan las siguientes ventajas:

1. Más informativas: Se tiene la forma explícita de la dependencia de la probabilidad de cobertura sobre  $\psi$ . Esto es útil para calcular las cotas mínimas de la probabilidad de cobertura de intervalos de verosimilitud que sugiere Barnard (1996).
2. Más representativas: condiciona en estadísticas auxiliares, si se requiere.
3. Más precisas, ahora el error es de orden  $O(n^{-2})$  en muestras repetidas ordinarias.

Las principales aportaciones de esta tesis están incluidas en las ecuaciones (2.3), (2.4), (2.8), (2.10), (2.12), (2.13), (3.36) y (3.41), los resultados se presentan de manera expositiva y se ilustran con varios ejemplos, en el Capítulo 3 se dan las condiciones de regularidad y las demostraciones de todos los resultados aquí presentados. Los ejemplos discutidos en este capítulo se presentan en dos secciones: los ejemplos para modelos sin parámetros de estorbo se encuentran en la Sección 2.4 y con parámetros de estorbo en la Sección 2.5.

## 2.1 Probabilidades de cobertura más precisas

Una aportación de esta tesis es, encontrar y demostrar las siguientes equivalencias, con error de orden  $O(n^{-2})$ , para la probabilidad de cobertura de un intervalo de verosimilitud

$$CP(c, \psi) = \Pr(R_p(\psi; \hat{\psi}, a) \geq c|a) \quad (2.2)$$

$$= \Phi(r^*(\psi; v_c)) - \Phi(r^*(\psi; -v_c)) + O(n^{-2}) \quad (2.3)$$

$$= 2\Phi(v_c) - 1 - bv_c\phi(v_c) + O(n^{-2}) \quad (2.4)$$

donde  $v_c = \sqrt{-2 \ln c}$ ,  $r^*(v_c; \psi)$  denota a la verosimilitud dirigida modificada expresada como función de  $r = v_c$  para un valor fijo de  $\psi$ , la relación funcional es de la forma

$$r^*(\psi; r) = r + \frac{1}{r} \ln \left( \frac{u(\psi; r)}{r} \right), \quad (2.5)$$

donde  $u(\psi; r)$  está dada en (1.30) y  $b = b(\psi, \hat{\chi}_\psi, a, n)$  es un factor de tipo Bartlett, el cual es de orden  $O(n^{-1})$ . Si  $\theta$  es escalar  $b$  está dado en (1.31) (ver Sección 1.4.1). Si  $\theta = (\psi, \chi)$ ,  $\psi$  escalar, no damos una expresión general para  $b$ . El factor de Bartlett con el que trabajamos en esta tesis lo definimos como dos veces el coeficiente que acompaña a  $r^2$  en la expansión de  $r/u$  alrededor de  $r = 0$  (ver Sección A.2).

Usaremos la siguiente notación:

$$\begin{aligned} \alpha(c, b) &= 1 - 2\Phi(v_c) + 1 + bv_c\phi(v_c) \\ &= 2\Phi(-v_c) + bv_c\phi(v_c). \end{aligned} \quad (2.6)$$

Entonces

$$CP(c, \psi) = 1 - \alpha(c, b) + O(n^{-2}),$$

si  $\alpha(c, b)$  es menor que cero o mayor que uno, se recomienda evaluar (2.2) con la aproximación (2.3).

El factor de Bartlett  $b \equiv b(\psi, \hat{\chi}_\psi, a, n)$  nos muestra la manera en que la probabilidad de cobertura de un intervalo de verosimilitud depende de  $\psi, \hat{\chi}_\psi, a$  y  $n$ , lo que no se tiene en la aproximación dada en (2.1).

La verosimilitud dirigida y la verosimilitud dirigida modificada se mantienen en el centro de esta tesis porque la definición que usamos del factor de Bartlett depende fuertemente de estas cantidades. Incluso la expresión (1.31) se obtiene de esa manera. Conviene señalar que en general no es necesario calcular el factor de Bartlett para evaluar (2.3) como lo ilustran los ejemplos 2.4.4, 2.4.5 y 2.5.5.

**Observación:** Si se utiliza cualquiera de las tres siguientes cantidades: *fórmula-p\** (Sec. 3.3.1), aproximación de Lugannani-Rice (Sec. 3.3.4) o factores de Bartlett (Sec. 3.3.2), también se obtiene (2.4). La relación (2.4) se puede encontrar en Efron & Hinkley (1978) y DiCiccio (1986) en el contexto de modelos de localización y en Jensen (1988) en un contexto general.

## 2.2 Probabilidad de cobertura mínima

Cuando la probabilidad de cobertura de un intervalo de verosimilitud depende funcionalmente del valor verdadero de  $\psi$ , no es posible dar un valor numérico de ella. Sin embargo, con ayuda de la expresión analítica (2.4) de  $CP(c, \psi)$  es posible conocer las cotas de la probabilidad de cobertura de un intervalo de verosimilitud, con error de orden  $O(n^{-2})$ . Para cada nivel de verosimilitud,  $c$ , las cotas de  $CP(c, \psi)$  proporcionan un intervalo conocido y

fijo en donde se encuentra el verdadero valor de la probabilidad de cobertura del intervalo  $LI(c, Y)$ . Estas cotas se obtienen con ayuda de las cotas del factor de Bartlett. Sean

$$b^-(a, \hat{\chi}_\psi, n) = \inf_{\omega \in \Omega_\psi} b(\omega, a, \hat{\chi}_\psi, n) \quad \text{y} \quad b^+(a, \hat{\chi}_\psi, n) = \sup_{\omega \in \Omega_\psi} b(\omega, a, \hat{\chi}_\psi, n), \quad (2.7)$$

donde  $\Omega_\psi$  es el espacio paramétrico de  $\psi$ . Se requiere que  $b^-$  y  $b^+$  sean acotados y que el ínfimo y el supremo se alcancen en el interior de  $\Omega_\psi$ . Así, con error de orden  $O(n^{-2})$ , se cumple lo siguiente

$$1 - \alpha(c; b^+) \leq \Pr(R_p(\psi; \hat{\psi}, a) \geq c|a) \leq 1 - \alpha(c; b^-). \quad (2.8)$$

A  $1 - \alpha(c; b^+)$  le llamaremos *probabilidad de cobertura mínima* del intervalo con nivel de verosimilitud  $c$ . Similarmente a  $1 - \alpha(c; b^+)$  se le llamará *probabilidad de cobertura máxima*. Si  $b$  no depende de  $\psi$ ,  $b^- = b^+$  y diremos que el intervalo de verosimilitud tiene *probabilidad de cobertura estándar (marginal o condicional)*.

Se puede dar el caso en que alguno de los valores de (2.7) no sea acotado. Esto sucede generalmente en la frontera de  $\Omega_\psi$ , donde pueden fallar las aproximaciones (2.2) y (2.4), debido a que una condición de regularidad exige que  $\psi$  sea un punto interior de  $\Omega_\psi$  (ver Sección 3.1). Las desigualdades (2.8) se siguen cumpliendo si  $b^-$  y  $b^+$  son acotados y el ínfimo o supremo se alcanza suavemente en un punto frontera  $\psi^\#$ . Por suavemente queremos decir que  $\lim_{\psi \rightarrow \psi^\#} b(\psi)$  existe y es finito. El punto frontera puede ser  $\pm\infty$ . Ver ejemplos 2.5.3 y 2.5.4.

En la práctica, si  $\psi \in (-\infty, +\infty)$  se recomienda calcular el mínimo en el intervalo  $[-M, M]$ , donde  $M$  es tan grande como se desee; si  $\psi \in (0, \infty)$  el mínimo se puede tomar en  $[\varepsilon, M]$ , donde  $\varepsilon$  es tan pequeño como se desee y si es un intervalo acotado  $[a, b]$  se recomienda tomar las cotas en  $[a + \varepsilon, b - \varepsilon]$ , etc.

## 2.3 Calibración de intervalos de verosimilitud

En esta tesis consideramos que la manera primaria de resumir información sobre un parámetro de interés  $\psi$ , con base en una muestra observada, es por medio de intervalos de verosimilitud construidos con nivel de verosimilitud  $c$ . La elección de  $c$  es de acuerdo a la plausibilidad que se desee. El objetivo principal de esta tesis es elegir un valor de  $c$ , basado en probabilidad de cobertura. Nos interesa fijar un valor de la probabilidad de cobertura (mínima o estándar), digamos  $1 - \alpha$ , y encontrar un nivel de verosimilitud  $c(\alpha)$  de tal manera que  $CP(c(\alpha); Y) \cong 1 - \alpha$ . Este es un problema de calibración.

Para obtener el nivel de verosimilitud correspondiente a  $\alpha$ , se procede a despejar  $c$  de (2.6), es decir  $c$  debe satisfacer

$$2\Phi(v_c) - 1 - bv_c\phi(v_c) = 1 - \alpha, \quad (2.9)$$

donde  $v_c = \sqrt{-2 \ln c}$ . Mediante técnicas de inversión de series (ver Sección A.1.4) se llega a que

$$c(\alpha; b) = \exp \left[ -\frac{1}{2} \left( 1 + \frac{1}{2}b \right)^2 z_{1-\alpha/2}^2 \right], \quad (2.10)$$

satisface (2.9) con error de orden  $O(n^{-2})$ . A  $c(\alpha, b)$  le llamaremos el *nivel de verosimilitud calibrado*. Cualquier otro valor que difiera de (2.10) por un término de orden  $O(n^{-2})$ , por ejemplo  $c'(\alpha; b) = e^{-\frac{1}{2}z_{1-\alpha/2}^2} \left( 1 - \frac{1}{2}z_{1-\alpha/2}^2 b \right)$ , satisface

$$\Pr \left( R_p(\psi; \hat{\psi}, a) \geq c(\alpha; b) | a \right) = 1 - \alpha + O(n^{-2}). \quad (2.11)$$

De (2.11) se ve que el nivel de verosimilitud depende del verdadero  $\psi$ , por lo que es necesario calcular cotas del nivel de verosimilitud que se puedan evaluar numéricamente.

En este caso sólo daremos el *nivel de verosimilitud mínimo* aproximado,  $c(\alpha; b^+)$ , el cual satisface que

$$\inf_{\psi \in \Omega_\psi} \Pr \left( R_p(\psi; \hat{\psi}, a) \geq c(\alpha; b^+) | a \right) = 1 - \alpha + O(n^{-2}),$$

así, con error de orden  $O(n^{-2})$ , se cumple la siguiente desigualdad.

$$1 - \alpha \leq \Pr \left( R_p(\psi; \hat{\psi}, a) \geq c(\alpha; b^+) | a \right) \leq 1 - \alpha(c(\alpha; b^+), b^-). \quad (2.12)$$

El intervalo de nivel de verosimilitud  $c(\alpha; b^+)$  tiene por lo menos probabilidad de cobertura  $1 - \alpha$ . El *nivel de verosimilitud máximo*,  $c(\alpha, b^-)$ , da un intervalo de verosimilitud que a lo más cubre a  $\psi$  el  $100(1 - \alpha)\%$  de las veces, en muestras repetidas ordinarias. No se discutirá más sobre  $c(\alpha, b^-)$ .

### 2.3.1 Aproximación del factor de Bartlett

En esta sección proponemos una aproximación numérica del factor de Bartlett basada en  $r^*$ .

El factor de Bartlett involucra derivadas de hasta cuarto orden; es decir, derivadas de la forma  $\ell_{j;k}$ , donde  $2 \leq j + k \leq 4$ ,  $j \geq 0$ ,  $k \geq 0$ .

En general la expresión analítica de  $b$  puede ser inmanejable, especialmente si hay más de dos parámetros de estorbo o si  $\chi$  no es ortogonal a  $\psi$ . En esta situación puede ser más práctico y más simple obtener  $b$  numéricamente.

A continuación proponemos una fórmula aproximada para  $b$  y damos indicaciones para evaluarla.

De las aproximaciones (2.3) y (2.4) obtenemos la siguiente aproximación para el factor

de Bartlett

$$b^*(\psi, \hat{\chi}_\psi, a, n; c) = \frac{\Phi(r^*(\psi; v_c)) - \Phi(r^*(\psi; -v_c)) - 2\Phi(v_c) + 1}{2v_c\phi(v_c)} + O(n^{-2}). \quad (2.13)$$

Para evaluar (2.13) es necesario especificar valores de  $c, \psi, a$  y  $\hat{\chi}_\psi$ . La aproximación (2.13) depende levemente de  $c$ , sin embargo recomendamos que  $c$  esté cerca de 1, puesto que

$$\lim_{c \rightarrow 1^-} b^*(\psi, \hat{\chi}_\psi, a, n; c) = b(\psi, \hat{\chi}_\psi, a, n).$$

Para evaluar  $r^*(\psi; \pm v_c)$  se procede como en la Sección 1.4.1. Si  $b$  no depende de  $\psi$  la parte derecha de (2.13) no dependerá de  $\psi$ . El ejemplo 2.5.2 utiliza  $b^*$  y da resultados muy precisos.

Por medio de las aproximaciones (2.3), (3.41), (2.4) y (3.36) se pueden obtener otras aproximaciones para el factor de Bartlett. Y dadas las siguientes relaciones

$$\sqrt{1+b} = 1 + \frac{1}{2}b + O(n^{-2}) = \frac{1}{1 - \frac{1}{2}b} + O(n^{-2}) = \frac{1}{\sqrt{1-b}} + O(n^{-2}),$$

podemos obtener más aproximaciones del factor de Bartlett mediante combinaciones de (2.3) y (3.40) con (2.4) y (3.36). Todas las aproximaciones del factor de Bartlett que se obtengan con estos procedimientos son equivalentes, con error de orden  $O(n^{-2})$ .

## 2.4 Ejemplos sin parámetros de estorbo

### 2.4.1 El parámetro de localización del modelo hiperbólico

Barndorff-Nielsen (1977) introdujo la *distribución hiperbólica* con función de densidad

$$p_Y(y; \alpha, \beta, \delta, \mu) = \frac{\sqrt{\alpha^2 - \beta^2}}{2\alpha K_1(\delta\sqrt{\alpha^2 - \beta^2})} \exp\left\{-\alpha\sqrt{\delta^2 + (y - \mu)^2} + \beta(y - \mu)\right\}, \quad (2.14)$$

donde  $y \in \mathbf{R}$ ,  $\mu \in \mathbf{R}$ ,  $\delta > 0$ ,  $0 \leq |\beta| \leq \alpha$ , y  $K_1$  es la función de Bessel modificada del tercer tipo. Si el parámetro de interés es  $\mu$  y tomando  $\delta = 1$ ,  $\alpha = 1$ ,  $\beta = \frac{1}{2}$  la función de densidad (2.14) se puede expresar de la forma

$$p_Y(y; \mu) = \frac{\sqrt{3}}{4K_1(\sqrt{3}/2)} e^{-\sqrt{1+(y-\mu)^2} + \frac{1}{2}(y-\mu)}, \quad (2.15)$$

donde  $K_1(\sqrt{3}/2) \cong 0.761932$ .

La función de log-verosimilitud de  $\mu$  está dada por

$$\ell(\mu; y) = -\sum_{i=1}^n \sqrt{1+(y_i - \mu)^2} + \frac{1}{2} \sum_{i=1}^n (y_i - \mu).$$

De aquí que la correspondiente función de log-verosimilitud expresada en términos de  $\mu, \hat{\mu}$  y  $a$  es

$$\ell(\mu; \hat{\mu}, a) = -\sum_{i=1}^n \sqrt{1+(a_i + \hat{\mu} - \mu)^2} + \frac{1}{2} \sum_{i=1}^n (a_i + \hat{\mu} - \mu). \quad (2.16)$$

donde  $a_i = y_i - \hat{\mu}$  satisface

$$\sum_{i=1}^n \frac{a_i}{\sqrt{1+a_i^2}} = \frac{n}{2},$$

debido a que la derivada de  $\ell(\mu; y)$  es

$$\ell_\mu = \sum_{i=1}^n \frac{y_i - \mu}{\sqrt{1+(y_i - \mu)^2}} - \frac{1}{2}n.$$

De (2.16) obtenemos que

$$\ell_{;\hat{\mu}}(\mu; \hat{\mu}, a) = -\sum_{i=1}^n \frac{a_i + \hat{\mu} - \mu}{\sqrt{1+(a_i + \hat{\mu} - \mu)^2}} + \frac{1}{2}n$$

además

$$\ell_{\mu\mu}(\mu; \hat{\mu}, a) = -\sum_{i=1}^n [1+(a_i + \hat{\mu} - \mu)^2]^{-3/2}$$

entonces

$$\hat{j} = \ell_{\mu\mu}(\hat{\mu}; \hat{\mu}, a) = - \sum_{i=1}^n (1 + a_i^2)^{-3/2}$$

de donde se obtiene fácilmente  $u$ , al evaluar  $u = \hat{j}^{-1/2} (\hat{\ell}_{;\hat{\mu}} - \ell_{;\hat{\mu}})$ .

Las derivadas de la función de log-verosimilitud necesarias para evaluar el factor de Bartlett son

$$\bar{\ell}_2 = \bar{\ell}_{\mu\mu}, \quad \bar{\ell}_3 = -3 \sum_{i=1}^n a_i (1 + a_i^2)^{-5/2}, \quad \bar{\ell}_4 = 3 \sum_{i=1}^n (1 - 4a_i^2) (1 + a_i^2)^{-7/2}.$$

Por ejemplo, de Barndorff-Nielsen (1982) obtenemos una muestra de cinco observaciones del modelo hiperbólico, los cuales son  $y = \{-0.26, 3.92, 1.16, -0.41, -0.6\}$ . Con estas observaciones obtenemos que  $\hat{\mu} = -0.6504$  y en consecuencia  $a = \{.3904, 4.5704, 1.8104, .2404, .0504\}$  y  $\hat{j} = 2.8465$ ,  $\bar{\ell}_3 = -1.7481$ ,  $\bar{\ell}_4 = 5.3213$ . De (1.32) se tiene que el factor de Bartlett para los datos anteriores es  $b(a) = 0.1097$ . La Figura 2.1. muestra las gráficas de verosimilitud relativa correspondiente a la densidad (2.15). En la Figura 2.1-(a) el eje derecho contiene las probabilidades de cobertura de acuerdo con su nivel de confianza. La Figura 2.1-(b) es la verosimilitud relativa de  $\mu$  con el eje derecho equi-espaciado.

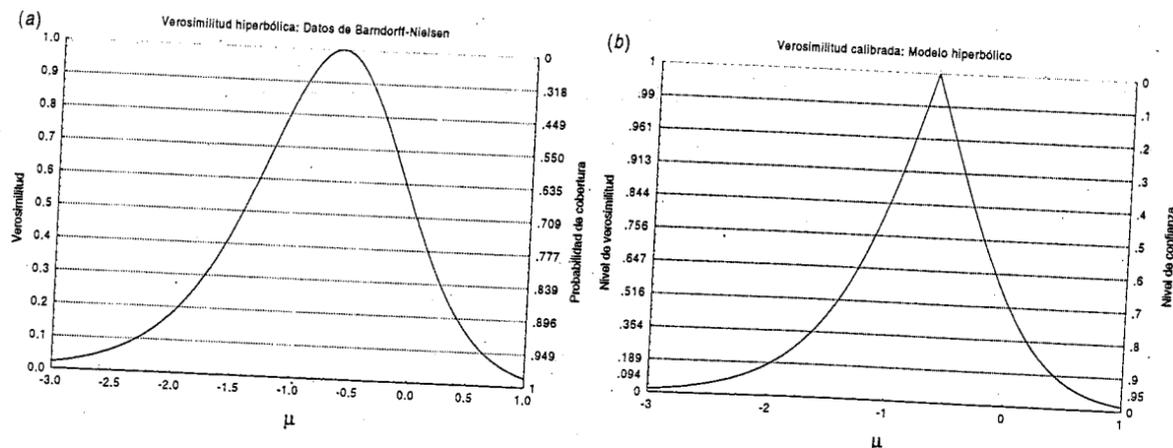


Figura 2.1. Funciones de verosimilitud relativa hiperbólica.

(a) Verosimilitud relativa hiperbólica. (b) Verosimilitud relativa hiperbólica calibrada

### 2.4.2 La distribución exponencial

Sea  $y = (y_1, \dots, y_n)^T$  una muestra de v.a.i.i.d. con función de densidad

$$f(y, \lambda, \mu) = \frac{\lambda^\lambda}{\Gamma(\lambda)} \frac{1}{\mu} \left(\frac{y}{\mu}\right)^{\lambda-1} e^{-\lambda \frac{y}{\mu}}. \quad (2.17)$$

Cuando  $\lambda$  es conocida (2.17) es un miembro del modelo de localización, esto se hace aparente si tomamos  $v = \log y$  y  $\phi = \log(\mu)$  en cuyo caso

$$p(v; \phi, \lambda) = \frac{\lambda^\lambda}{\Gamma(\lambda)} e^{\lambda(v-\phi-e^{v-\phi})},$$

lo que da un modelo de localización. Aún así, consideraremos el modelo como uno de escala y tomaremos  $\mu$  como el parámetro de interés con  $\lambda$  conocido, entonces la función log verosimilitud expresada en términos de  $\hat{\mu} = \bar{y}$  es

$$\ell(\mu; \hat{\mu}) = -n\lambda \log \mu - n\lambda \hat{\mu} / \mu.$$

Las derivadas correspondientes para calcular el factor de Bartlett están dadas por

$$\bar{\ell}_2 = -\frac{n}{\mu^2}, \quad \bar{\ell}_3 = 4\frac{n}{\mu^3}, \quad \bar{\ell}_4 = -18\frac{n}{\mu^4}, \quad \bar{\ell}_{2,1} = -2\frac{n}{\mu^3}, \quad \bar{\ell}_{2,2} = 0, \quad \bar{\ell}_{3,1} = 6\frac{n}{\mu^4}$$

y de acuerdo con (1.31) el ajuste de Bartlett es

$$b = \frac{1}{6n}.$$

La Tabla 2.1 muestra los niveles de verosimilitud calibrados de acuerdo con la probabilidad de cobertura especificada. Observe que los valores son prácticamente exactos para cualquier tamaño de muestra.

Tabla 2.1. Niveles de verosimilitud calibrados obtenidos de la fórmula (2.10).

$n$	$c(.9, \frac{1}{6n})$	pce	$c(.95, \frac{1}{6n})$	pce	$c(.99, \frac{1}{6n})$	pce
1	.204	.903	.105	.953	.0203	.991
2	.23	.9006	.124	.951	.0273	.9903
5	.247	.90001	.137	.9501	.0324	.99005
10	.253	.900003	.142	.950002	.0343	.9900009
pce = probabilidad de cobertura exacta						

La Tabla 2.2 muestra la probabilidad de cobertura obtenida por los cinco procedimientos equivalentes. La segunda columna indica la probabilidad de cobertura evaluada con (3.36),  $\alpha(c, b)$  está dado en (2.6), la cuarta columna se obtiene con (3.40), la quinta con (2.3) y la última con (3.33) la cual es igual a la exacta.

Tabla 2.2. Probabilidad de cobertura obtenida mediante cinco procedimientos de altos órdenes.

$n, c = .05$	Bartlett	$\alpha(.05, \frac{1}{6n})$	L-R	$r^*$	pce= $p^*$
1	.976560	.977487	.977390	.977691	.959298
2	.981313	.981556	.981465	.981543	.981638
5	.983957	.983997	.983975	.983988	.983995
10	.984801	.984811	.984805	.984808	.984809
pce = probabilidad de cobertura exacta					

### 2.4.3 Distribución de extremos

La función de densidad de extremos se expresa mediante

$$f(y; \sigma) = \frac{1}{\sigma} e^{\frac{y}{\sigma} - e^{\frac{y}{\sigma}}}$$

Para una muestra de  $n$  observaciones de esta distribución la función de log-verosimilitud en función de  $\hat{\sigma}$  y  $a$  es

$$\ell(\sigma; \hat{\sigma}, a) = -n \ln \sigma + \frac{\hat{\sigma}}{\sigma} \sum_{i=1}^n a_i - \sum_{i=1}^n e^{\frac{\hat{\sigma}}{\sigma} a_i} \quad (2.18)$$

donde  $a_i = \frac{y_i}{\hat{\sigma}}$ . Para  $n = 1$  la relación exacta entre  $y$  y  $\hat{\sigma}$  es

$$\hat{\sigma}(y) = \begin{cases} -0.741y & \text{si } y \leq 0 \\ 1.24y & \text{si } y > 0 \end{cases}$$

Observe que  $\hat{\sigma}$  no es función uno-a-uno de  $y$ . De aquí tenemos dos estadísticas auxiliares completamente conocidas

$$a(y) = \begin{cases} -1.35 & \text{si } y \leq 0 \\ 0.8065 & \text{si } y > 0 \end{cases}$$

En este caso la densidad condicional exacta de  $\hat{\sigma}$  está dada por

$$f(\hat{\sigma}, \sigma | a) = \frac{|a|}{\sigma} e^{\frac{\hat{\sigma}}{\sigma} a - e^{\frac{\hat{\sigma}}{\sigma} a}} \Rightarrow \begin{aligned} f(\hat{\sigma}, \sigma | a = -1.35) &= \frac{1.35e}{\sigma} e^{-1.35 \frac{\hat{\sigma}}{\sigma} - e^{-1.35 \frac{\hat{\sigma}}{\sigma}}} \\ f(\hat{\sigma}, \sigma | a = .8065) &= \frac{.8065e}{e-1} e^{.8065 \frac{\hat{\sigma}}{\sigma} - e^{.8065 \frac{\hat{\sigma}}{\sigma}}} \end{aligned}$$

Para  $n > 1$ , la densidad condicional exacta de  $\hat{\sigma}$  es

$$p(\hat{\sigma}; \sigma | a) = k(a, n) \frac{1}{\sigma} \left(\frac{\hat{\sigma}}{\sigma}\right)^{n-1} \exp\left(\frac{\hat{\sigma}}{\sigma} \sum_{i=1}^n a_i - \sum_{i=1}^n e^{\frac{\hat{\sigma}}{\sigma} a_i}\right), \quad (2.19)$$

donde  $k(a, n)$  es la constante de integración.

La primer derivada de la función de log-verosimilitud es

$$\ell_1(\sigma; \hat{\sigma}, a) = -\frac{n}{\sigma} - \frac{\hat{\sigma}}{\sigma^2} \sum_{i=1}^n a_i + \frac{\hat{\sigma}}{\sigma^2} \sum_{i=1}^n a_i e^{\frac{\hat{\sigma}}{\sigma} a_i},$$

debido a que  $\hat{\ell}_1 = 0$ , la ecuación que define a  $a$  es

$$\sum_{i=1}^n a_i (e^{a_i} - 1) = n$$

Las cantidades para evaluar (1.31) son

$$l_2 = \frac{n}{\sigma^2} + 2 \frac{\hat{\sigma}}{\sigma^3} \sum_{i=1}^n a_i - 2 \frac{\hat{\sigma}}{\sigma^3} \sum_{i=1}^n a_i e^{\frac{\hat{\sigma}}{\sigma} a_i} - \frac{\hat{\sigma}^2}{\sigma^4} \sum_{i=1}^n a_i^2 e^{\frac{\hat{\sigma}}{\sigma} a_i},$$

$$l_3 = -2 \frac{n}{\sigma^3} - 6 \frac{\hat{\sigma}}{\sigma^4} \sum_{i=1}^n a_i + 6 \frac{\hat{\sigma}}{\sigma^4} \sum_{i=1}^n a_i e^{\frac{\hat{\sigma}}{\sigma} a_i} + 6 \frac{\hat{\sigma}^2}{\sigma^5} \sum_{i=1}^n a_i^2 e^{\frac{\hat{\sigma}}{\sigma} a_i} + \frac{\hat{\sigma}^3}{\sigma^6} \sum_{i=1}^n a_i^3 e^{\frac{\hat{\sigma}}{\sigma} a_i},$$

$$l_4 = 6 \frac{n}{\sigma^4} + 24 \frac{\hat{\sigma}}{\sigma^5} \sum_{i=1}^n a_i - 24 \frac{\hat{\sigma}}{\sigma^5} \sum_{i=1}^n a_i e^{\frac{\hat{\sigma}}{\sigma} a_i} - 36 \frac{\hat{\sigma}^2}{\sigma^6} \sum_{i=1}^n a_i^2 e^{\frac{\hat{\sigma}}{\sigma} a_i} - 12 \frac{\hat{\sigma}^3}{\sigma^7} \sum_{i=1}^n a_i^3 e^{\frac{\hat{\sigma}}{\sigma} a_i} - \frac{\hat{\sigma}^4}{\sigma^8} \sum_{i=1}^n a_i^4 e^{\frac{\hat{\sigma}}{\sigma} a_i},$$

$$l_{3;1} = -\frac{6}{\sigma^4} \sum_{i=1}^n a_i + \frac{6}{\sigma^4} \sum_{i=1}^n a_i e^{\frac{\hat{\sigma}}{\sigma} a_i} + 18 \frac{\hat{\sigma}}{\sigma^5} \sum_{i=1}^n a_i^2 e^{\frac{\hat{\sigma}}{\sigma} a_i} + 9 \frac{\hat{\sigma}^2}{\sigma^6} \sum_{i=1}^n a_i^3 e^{\frac{\hat{\sigma}}{\sigma} a_i} + \frac{\hat{\sigma}^3}{\sigma^7} \sum_{i=1}^n a_i^4 e^{\frac{\hat{\sigma}}{\sigma} a_i},$$

$$l_{2;1} = \frac{2}{\sigma^3} \sum_{i=1}^n a_i - \frac{2}{\sigma^3} \sum_{i=1}^n a_i e^{\frac{\hat{\sigma}}{\sigma} a_i} - 4 \frac{\hat{\sigma}}{\sigma^4} \sum_{i=1}^n a_i^2 e^{\frac{\hat{\sigma}}{\sigma} a_i} - \frac{\hat{\sigma}^2}{\sigma^5} \sum_{i=1}^n a_i^3 e^{\frac{\hat{\sigma}}{\sigma} a_i},$$

$$l_{2;2} = -\frac{6}{\sigma^4} \sum_{i=1}^n a_i^2 e^{\frac{\hat{\sigma}}{\sigma} a_i} - 6 \frac{\hat{\sigma}}{\sigma^5} \sum_{i=1}^n a_i^3 e^{\frac{\hat{\sigma}}{\sigma} a_i} - \frac{\hat{\sigma}^2}{\sigma^6} \sum_{i=1}^n a_i^4 e^{\frac{\hat{\sigma}}{\sigma} a_i}.$$

Sea  $A_k = \sum_{i=1}^n a_i^k e^{a_i}$ , al evaluar en  $\hat{\theta} = \theta$  las expresiones anteriores, éstas se reducen a

$$\bar{l}_2 = \sigma^{-2} \left( n + 2 \sum_{i=1}^n a_i - 2A_1 - A_2 \right), \quad \bar{l}_3 = \sigma^{-3} \left( -2n - 6 \sum_{i=1}^n a_i + 6A_1 + 6A_2 + A_3 \right),$$

$$\bar{l}_4 = \sigma^{-4} \left( 6n + 24 \sum_{i=1}^n a_i - 24A_1 - 36A_2 - 12A_3 - A_4 \right),$$

$$\bar{l}_{3;1} = \sigma^{-4} \left( -6 \sum_{i=1}^n a_i + 6A_1 + 18A_2 + 9A_3 + A_4 \right),$$

$$\bar{l}_{2;1} = \sigma^{-3} \left( 2 \sum_{i=1}^n a_i - 2A_1 - 4A_2 - A_3 \right), \quad \bar{l}_{2;2} = \sigma^{-4} (-6A_2 - 6A_3 - A_4)$$

De la ecuación (1.31) se obtiene el factor de Bartlett y además se demuestra que  $b(\sigma, a, n) = b(a, n)$ ; es decir,  $b$  no depende de  $\sigma$ .

La Tabla 2.3 presenta las coberturas condicionales exactas y aproximadas de los intervalos con nivel de verosimilitud .147. Las estadísticas auxiliares se obtuvieron al simular muestras de tamaño  $n = 1, 2, 3$  y  $7$ . Observe que las aproximaciones son muy precisas para todos los tamaños de muestra presentados. Los valores aproximados se calcularon con la ecuación (2.4) y los exactos con (2.19).

Tabla 2.3. Valores exactos y aproximados de la probabilidad de cobertura de los intervalos de verosimilitud de la distribución de extremos.

c = .147			
n	$\alpha(c, b)$	pce	diferencia
1	.9171	.9241	-.007
1'	.9084	.916	-.0076
2	.9292	.9295	-.0003
3	.9367	.9347	.002
7	.9447	.9462	-.0015
pce = probabilidad de cobertura exacta			

$$n = 1 : a = -1.35, \quad n = 1' : a = .80647, \quad n = 2 : a = (.7096528, .887066)$$

$$n = 3 : a = (-.2361641, .4723283, 1.180821)$$

$$n = 7 : a = (1.0089642, -.27296148, -.71629832, .42605554, .6753304, -1.5236152, -2.9002252).$$

### 2.4.4 El modelo gaussiana inversa

La densidad de la distribución gaussiana inversa  $IG(\lambda, \psi)$  está dada por

$$p_Y(y; \lambda, \psi) = \frac{\sqrt{\lambda} e^{\sqrt{\lambda\psi}}}{\sqrt{2\pi}} y^{-\frac{3}{2}} e^{-\frac{1}{2}(\lambda y^{-1} + \psi y)}, \quad y, \psi, \lambda > 0. \quad (2.20)$$

**Caso 1. Parámetro de interés  $\lambda$  y  $\psi$  conocido.**

Si  $y_1, \dots, y_n$  se distribuyen de acuerdo con (2.20) con  $\psi$  conocido, la función de log verosimilitud es

$$\ell(\lambda; \psi, y) = \frac{n}{2} \ln \lambda + n\sqrt{\lambda\psi} - \frac{1}{2}\lambda \sum_{i=1}^n y_i^{-1},$$

el estimador de máxima verosimilitud de  $\lambda$ , para el valor fijo de  $\psi$  es

$$\hat{\lambda}_\psi = 4 \left( -\sqrt{\psi} + \sqrt{\psi + 4\bar{y}} \right)^{-2} \quad (2.21)$$

donde  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i^{-1}$ . Dado que  $\bar{y}$  es suficiente y  $\hat{\lambda}_\psi$  es función uno-a-uno de  $\bar{y}$ , entonces  $\hat{\lambda}_\psi$  es suficiente. En este caso no se necesita especificar ninguna estadística auxiliar. La expresión de la función de log-verosimilitud se obtiene al despejar  $\bar{y}$  de (2.21) el cual es

$$\bar{y} = \hat{\lambda}_\psi^{-1} + \sqrt{\psi \hat{\lambda}_\psi^{-1/2}},$$

así, pues la función de log-verosimilitud de  $\lambda$ , como función de  $\lambda$  y  $\hat{\lambda}_\psi$  está dada por

$$\ell(\lambda; \hat{\lambda}_\psi) = \frac{n}{2} \ln \lambda + n\sqrt{\lambda\psi} - \frac{n}{2}\lambda \left( \hat{\lambda}_\psi^{-1} + \sqrt{\psi \hat{\lambda}_\psi^{-1/2}} \right)$$

y la verosimilitud dirigida se obtiene de

$$\frac{1}{n} r^2 = \ln \frac{\hat{\lambda}_\psi}{\lambda} + \sqrt{\psi \hat{\lambda}_\psi} - 2\sqrt{\lambda\psi} - 1 + \lambda \left( \hat{\lambda}_\psi^{-1} + \sqrt{\psi \hat{\lambda}_\psi^{-1/2}} \right).$$

Los cálculos necesarios para obtener  $r^*$  son

$$\ell_{;1} = \frac{n}{4} \frac{\lambda}{\hat{\lambda}_\psi^2} \left( 2 + \sqrt{\psi \hat{\lambda}_\psi} \right), \quad j = \frac{1}{4} \frac{n}{\lambda^2} \left( 2 + \sqrt{\lambda\psi} \right),$$

de aquí se obtiene que la estadística  $u$  es

$$u = \frac{\sqrt{n}}{\sqrt{2}} \left( \hat{\lambda}_\psi^{-1} + \frac{1}{2} \sqrt{\psi \hat{\lambda}_\psi^{-3/2}} \right)^{1/2} \left( \hat{\lambda}_\psi - \lambda \right).$$

Observe que

$$\lim_{\lambda \rightarrow 0} \frac{u}{r} = +\infty \quad \text{y} \quad \lim_{\lambda \rightarrow \infty} \frac{u}{r} = \frac{1}{2}.$$

Las derivadas necesarias para obtener el factor de Bartlett son

$$\bar{\ell}_2 = -j, \quad \bar{\ell}_3 = \frac{1}{8} \frac{n}{\lambda^3} \left( 8 + 3\sqrt{\lambda\psi} \right), \quad \bar{\ell}_4 = -\frac{3}{16} \frac{n}{\lambda^4} \left( 16 + 5\sqrt{\lambda\psi} \right).$$

Entonces de (1.31) obtenemos que

$$b(\lambda, n) = \frac{1}{12n} \frac{16 + 3\sqrt{\lambda\psi}}{(2 + \sqrt{\lambda\psi})^3}. \quad (2.22)$$

De donde la cobertura aproximada como función de  $\lambda$  está dada por

$$\Pr(R(\lambda) \geq c) = 2\Phi(v_c) - 1 - \frac{1}{6n} \frac{16 + 3\sqrt{\lambda\psi}}{(2 + \sqrt{\lambda\psi})^3} v_c \phi(v_c) + O(n^{-2}),$$

donde  $v_c = \sqrt{-2 \ln c}$ .

De (2.22) obtenemos que

$$b^-(n) = \lim_{\lambda \rightarrow \infty} b(\lambda, n) = 0 \quad \text{y} \quad b^+(n) = \lim_{\lambda \rightarrow 0} b(\lambda, n) = \frac{1}{6n}, \quad (2.23)$$

los cuales se alcanzan en las fronteras de  $\Omega_\lambda = (0, \infty)$ . Sin embargo en virtud de la suavidad de  $b(\lambda, n)$  como función de  $\lambda$ , la relación (2.8) se cumple con las cotas del factor de Bartlett dadas en (2.23).

La Tabla 2.4 contiene los valores exactos y aproximados de la probabilidad de cobertura mínima de los intervalos de verosimilitud para  $\lambda$ . Los valores aproximados se calcularon con la ecuación (2.4) y los valores exactos con ayuda de las relaciones (2.24) y (2.25) con  $\lambda = 10^{-5}$  y  $\psi = 1$ .

**Tabla 2.4.** Valores exactos y aproximados de la probabilidad de cobertura mínima de los intervalos de verosimilitud para  $\lambda$ . Los valores aproximados se calcularon con la ecuación (2.4).

	Probabilidad de cobertura	$c = .258$	$c = .147$	$c = .036$
$n = 1$	Mínima aproximada	.8440	.9117	.9778
	Mínima exacta	.8536	.9196	.9809
$n = 2$	Mínima aproximada	.8721	.9307	.9839
	Mínima exacta	.8736	.9323	.9844

**Densidades exactas de  $\hat{\lambda}_\psi$  para  $n = 1$  y  $2$**

Sea  $\tilde{y} = \frac{1}{n} \sum y_i^{-1}$ , la relación entre  $\tilde{y}$  y  $\hat{\lambda}_\psi$  está dada por  $\hat{\lambda}_\psi = 4(-\sqrt{\psi} + \sqrt{\psi + 4\tilde{y}})^{-2}$  y la relación inversa por  $\tilde{y} = \hat{\lambda}_\psi^{-1} + \sqrt{\psi} \hat{\lambda}_\psi^{-1/2}$ . Entonces, dado que  $\hat{\lambda}_\psi$  es función uno-a-uno de  $\tilde{y}$ , si conocemos la función de densidad de  $\tilde{y}$  se obtiene directamente la de  $\hat{\lambda}_\psi$ . Por lo tanto únicamente escribiremos las expresiones para la densidad de  $\tilde{y}$ .

La función de distribución de la gaussiana inversa se expresa como

$$F(y; \psi, \lambda) = \int_0^y p_x(t; \lambda, \psi) dt = \Phi\left(\sqrt{\frac{\lambda}{y}}\left(\sqrt{\frac{\psi}{\lambda}}y - 1\right)\right) + e^{2\sqrt{\lambda\psi}}\Phi\left(-\sqrt{\frac{\lambda}{y}}\left(1 + \sqrt{\frac{\psi}{\lambda}}y\right)\right),$$

donde  $p_x(y; \lambda, \psi)$  está dada en la ecuación (2.20).

Para  $n = 1$  la densidad de  $Y^{-1}$  está dada por

$$p_{Y^{-1}}(y; \lambda, \psi) = \frac{\sqrt{\lambda}e^{\sqrt{\lambda\psi}}}{\sqrt{2\pi}}y^{-\frac{1}{2}}e^{-\frac{1}{2}(\lambda y + \psi y^{-1})}, \quad y, \psi, \lambda > 0. \quad (2.24)$$

Ahora sean  $Y_1, Y_2 \sim IG(\psi, \lambda)$ , de (2.24) y por métodos ordinarios de probabilidad obtenemos que la función de densidad de  $X = Y_1^{-1} + Y_2^{-1}$  es

$$p_X(x; \lambda, \psi) = \lambda e^{2\sqrt{\lambda\psi}} e^{-\frac{1}{2}\lambda x} \left(1 - \Phi\left(2\frac{\sqrt{\psi}}{\sqrt{x}}\right)\right). \quad (2.25)$$

**Caso 2. Parámetro de interés  $\theta = \sqrt{\psi}$  y  $\lambda$  conocido.**

Considere una muestra aleatoria  $y_1, \dots, y_n$  de (2.20) con  $\lambda$  conocido y  $\theta = \sqrt{\psi}$  el parámetro de interés, entonces

$$\ell(y, \theta; \lambda) = n\sqrt{\lambda}\theta - \frac{1}{2}\theta^2 \sum_{i=1}^n y_i,$$

de esta ecuación se sigue que  $\hat{\theta} = \sqrt{\lambda}\bar{y}^{-1}$ , de donde  $\bar{y} = \sqrt{\lambda}\hat{\theta}^{-1}$ , entonces la función de log-verosimilitud se puede expresar como

$$\ell(\theta; \hat{\theta}) = n\sqrt{\lambda}\theta - \frac{n}{2}\sqrt{\lambda}\theta^2\hat{\theta}^{-1}. \quad (2.26)$$

En este caso tenemos que

$$r = \lambda^{1/4} \frac{\sqrt{n}}{\sqrt{\hat{\theta}}} (\hat{\theta} - \theta) \quad \text{y} \quad u = \frac{\hat{\theta} + \theta}{2\hat{\theta}} r,$$

en consecuencia la verosimilitud dirigida modificada es

$$r^* = r + \frac{1}{r} \log\left(\frac{\hat{\theta} + \theta}{2\hat{\theta}}\right).$$

El factor de Bartlett es cero, por lo que la expansión de  $\frac{r}{u}$  alrededor de  $r = 0$  no contiene término cuadrático:

$$\frac{r}{u} = 1 + \frac{1}{2\lambda^{1/4}\sqrt{n\theta}}r - \frac{1}{16\lambda^{3/4}(n\theta)^{3/2}}r^3 + \dots$$

Es simple de verificar que  $Ee^{tr^2} = (1 - 2t)^{-1/2}$ , es decir  $r^2 \sim \chi_1^2$ . Shuster (1968) obtiene que  $r^2 \stackrel{D}{=} \chi_1^2$ , por un camino diferente. En este ejemplo se tiene la expresión de  $r^*$  en función de  $r$ , para  $\theta$  o  $\hat{\theta}$  fijo, como se ve en las siguientes expresiones

$$\frac{u(r, \hat{\theta})}{r} = 1 - \frac{1}{2\sqrt{n\hat{\theta}}}r \quad \text{y} \quad \frac{u(r, \theta)}{r} = \frac{1}{2} + \frac{2n\theta}{(r + \sqrt{r^2 + 4n\theta})^2}$$

En virtud de que  $r^2$  es  $\chi_1^2$  los intervalos de verosimilitud, con nivel de verosimilitud  $e^{-\frac{1}{2}z_{\alpha/2}^2 - \alpha/2}$ , son también intervalos de confianza, con nivel de confianza  $1 - \alpha$ , los cuales están dados por

$$\max \left\{ 0, \hat{\theta} - z_{\alpha/2} \sqrt{\frac{\hat{\theta}}{n}} \right\} \leq \theta \leq \hat{\theta} + z_{\alpha/2} \sqrt{\frac{\hat{\theta}}{n}}$$

En la Figura 2.2. Se muestran las gráficas de verosimilitud gaussiana inversa para  $\theta$ , con  $\hat{\theta} = 1/4$  y  $n = 3$ , los valores de los ejes izquierdo y derecho son exactos. La Figura 2.2-(b) es la gráfica de verosimilitud calibrada para  $\theta$ .

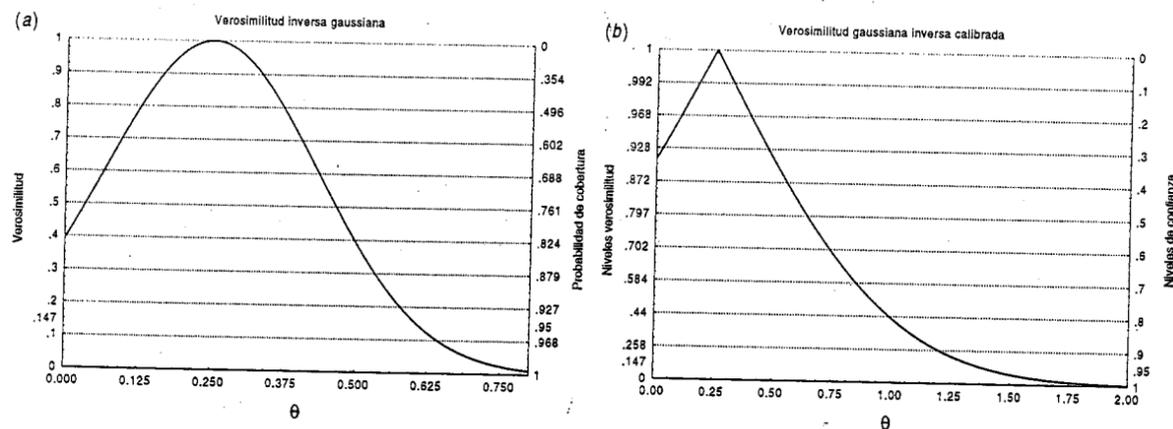


Figura 2.2. Funciones de verosimilitud relativa gaussiana inversa.

(a) Verosimilitud relativa. (b) Verosimilitud relativa calibrada

### 2.4.5 El modelo gaussiana inversa normal

La función de densidad de una variable aleatoria con función de distribución inversa gaussiana normal está dada por

$$p_Y(y; \alpha, \beta, \mu, \delta) = \frac{\delta\alpha}{\pi} \exp \left\{ \delta \sqrt{\alpha^2 - \beta^2} + \beta(y - \mu) \right\} \frac{K_1 \left( \alpha \sqrt{\delta^2 + (y - \mu)^2} \right)}{\sqrt{\delta^2 + (y - \mu)^2}},$$

donde  $y \in \mathbf{R}$ ,  $\mu \in \mathbf{R}$ ,  $\delta > 0$ ,  $0 \leq |\beta| \leq \alpha$  y  $K_1(z) = \frac{z}{4} \int_0^\infty t^{-2} e^{-t - \frac{z^2}{4t}} dt$ . Ver Apéndice C y Gradshteyn & Ryzhik (1994, ecuación 8.432-6).

Si tomamos  $\mu = 0$ ,  $\delta = 1$ ,  $\alpha$  conocido y  $\beta$  como el parámetro de interés entonces

$$p_Y(y; \beta) = \frac{\alpha K_1(\alpha \sqrt{1+y^2})}{\pi \sqrt{1+y^2}} e^{\sqrt{\alpha^2 - \beta^2} + \beta y}, \quad (2.27)$$

excepto por una constante aditiva la log verosimilitud de  $\beta$  para una muestra  $Y = (y_1, \dots, y_n)^T$  se expresa por

$$\ell(\beta; Y) = n \sqrt{\alpha^2 - \beta^2} + n\beta \bar{y},$$

$$\frac{\partial}{\partial \beta} \ell(\beta; Y) = -\frac{n\beta}{\sqrt{\alpha^2 - \beta^2}} + \sum_{i=1}^n y_i, \quad \text{de donde, } \hat{\beta} = \frac{\alpha \bar{y}}{\sqrt{1 + \bar{y}^2}}$$

Así, la log verosimilitud expresada en términos de  $(\beta, \hat{\beta})$  es

$$\ell(\beta; \hat{\beta}) = n \sqrt{\alpha^2 - \beta^2} + \frac{n\beta \hat{\beta}}{\sqrt{\alpha^2 - \hat{\beta}^2}}. \quad (2.28)$$

$$\frac{1}{2n} r^2(\beta; \hat{\beta}) = \frac{\alpha^2 - \beta \hat{\beta}}{\sqrt{\alpha^2 - \hat{\beta}^2}} - \sqrt{\alpha^2 - \beta^2}$$

$$\ell_{;\beta} = n \frac{\alpha^2 \beta}{(\alpha^2 - \hat{\beta}^2)^{3/2}} \Rightarrow u = \sqrt{n} \frac{\alpha (\hat{\beta} - \beta)}{(\alpha^2 - \hat{\beta}^2)^{3/4}}$$

Las cantidades útiles para calcular el factor de Bartlett son

$$\bar{\ell}_2 = -n \frac{\alpha^2}{(\alpha^2 - \beta^2)^{3/2}}, \quad \bar{\ell}_3 = -3n\beta \frac{\alpha^2}{(\alpha^2 - \beta^2)^{5/2}}, \quad \bar{\ell}_4 = -3n\alpha^2 \frac{4\beta^2 + \alpha^2}{(\alpha^2 - \beta^2)^{7/2}}$$

de donde

$$b(\beta, n) = -\frac{3}{8} \frac{\sqrt{\alpha^2 - \beta^2}}{\alpha^2 n}, \quad (2.29)$$

fácilmente se sigue que

$$b^+ = 0 \quad \text{y} \quad b^- = -\frac{3}{8\alpha n}.$$

El valor máximo es en  $\beta = 0$  y el mínimo es  $\beta = \alpha$ . En este caso  $b^- = 0$  entonces los intervalos de verosimilitud de la distribución gaussiana inversa normal son sobre-precisos si se usa la aproximación  $O(n^{-1})$ ,

$$\frac{1}{2n} r^2 + \sqrt{\alpha^2 - \beta^2} = \frac{\alpha^2 - \beta \hat{\beta}}{\sqrt{\alpha^2 - \hat{\beta}^2}}.$$

La log verosimilitud (2.28) nos permite calcular explícitamente los intervalos de verosimilitud para  $\beta$ ,

$$\beta = \hat{\beta} - \frac{\hat{\beta} r^2}{2\alpha^2 n} \sqrt{\alpha^2 - \hat{\beta}^2} \pm \frac{r}{2\alpha^2 n} (\alpha^2 - \hat{\beta}^2)^{3/4} \sqrt{4n\alpha^2 - r^2 \sqrt{\alpha^2 - \hat{\beta}^2}}$$

y también se puede ver explícitamente la forma de la región aleatoria, consistente de los extremos de  $\hat{\beta}$  tales que  $|r(\beta; \hat{\beta})| \leq z$ , así

$$\hat{\beta} = \alpha \frac{n^2 \beta \alpha \pm z \left( \frac{1}{2} z^2 + n \sqrt{\alpha^2 - \beta^2} \right) \sqrt{\frac{1}{4} z^2 + n \sqrt{\alpha^2 - \beta^2}}}{\frac{1}{4} z^4 + z^2 n \sqrt{\alpha^2 - \beta^2} + \alpha^2 n^2}$$

Si  $\beta = \alpha$

$$\hat{\beta} = \left( \alpha \frac{\alpha^2 n^2 - \frac{1}{4} z^4}{\alpha^2 n^2 + \frac{1}{4} z^4}, \alpha \right)$$

**Tabla 2.5.** Valores exactos y aproximados de la cobertura como función de  $\beta$  de los intervalos con nivel de verosimilitud .147. Los valores aproximados se calcularon con la ecuación (2.4). En este caso  $\alpha = 3, n = 1$ .

$r = 1.96$	$\beta = 0$	$\beta = \pm 2.95$	$\beta = \pm 2.9999$	$\beta = \pm 2.99999$
pca	.978636	.955206	.950234	.950074
pce	.967945	.9613	.959065	.958985
pca = Probabilidad de cobertura aproximada				
pce = Probabilidad de cobertura exacta				

## 2.5 Ejemplos con Parámetros de estorbo

### 2.5.1 El modelo normal

**Caso 1. Parámetro de interés  $\mu$ .**

Sea  $y = (y_1, \dots, y_n)^t$  una muestra de v.a.i.i.d. de una distribución  $N(\mu, \sigma^2)$ . Entonces la función de log-verosimilitud está dada por

$$\ell(\mu, \sigma; Y) = -n \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2,$$

los estimadores de máxima verosimilitud de  $\mu$  y  $\sigma$  son

$$\hat{\mu} = \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, \quad \text{y} \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2,$$

en consecuencia, la función de log-verosimilitud expresada en términos de  $(\hat{\mu}, \hat{\sigma})$  es

$$\ell(\mu, \sigma; \hat{\mu}, \hat{\sigma}) = -n \ln \sigma - \frac{n}{2\sigma^2} [\hat{\sigma}^2 + (\hat{\mu} - \mu)^2], \quad (2.30)$$

de donde obtenemos que

$$\hat{\sigma}_\mu^2 = \hat{\sigma}^2 + (\hat{\mu} - \mu)^2. \quad (2.31)$$

La verosimilitud dirigida y la estadística  $u$  son

$$r = \sqrt{n} \operatorname{sgn}(\hat{\mu} - \mu) \sqrt{\ln \frac{\hat{\sigma}_\mu^2}{\hat{\sigma}^2}} \quad \text{y} \quad u = \sqrt{n} \frac{\hat{\sigma}}{\hat{\sigma}_\mu} \frac{\hat{\mu} - \mu}{\hat{\sigma}_\mu}. \quad (2.32)$$

Por medio de (2.31) y (2.32) llegamos a que

$$u(r) = \sqrt{n} \operatorname{sgn}(r) e^{-\frac{1}{2n}r^2} \sqrt{1 - e^{-\frac{1}{n}r^2}},$$

y entonces

$$r^*(r) = r + \frac{1}{r} \ln \left( \sqrt{n} |r|^{-1} e^{-\frac{1}{2n}r^2} \sqrt{1 - e^{-\frac{1}{n}r^2}} \right).$$

La expansión de  $u(r)$  alrededor de  $r = 0$  es

$$u(r) = r - \frac{3}{4n}r^3 + O(n^{-2}),$$

de aquí se ve que  $t_3 = 0$  y que  $b = \frac{3}{2n}$ , entonces

$$r^*(r) = \left(1 - \frac{3}{4n}\right)r + O(n^{-2}).$$

De (2.32) obtenemos que el intervalo de verosimilitud, con nivel de verosimilitud  $c$  está dado

por

$$\mu = \hat{\mu} \pm \hat{\sigma} \sqrt{1 - e^{-\frac{1}{n}v^2}} = \hat{\mu} \pm \hat{\sigma} \sqrt{1 - c^{2/n}} \quad (2.33)$$

Tomando en cuenta que  $\hat{\sigma}^2 = \hat{\sigma}_\mu^2 - (\hat{\mu} - \mu)^2$ , se obtiene que los  $\hat{\mu}$  que satisfacen (2.33) son

$$\hat{\mu} = \mu \pm \hat{\sigma}_\mu \sqrt{1 - e^{-\frac{1}{n}r^2}} = \mu \pm \hat{\sigma}_\mu \sqrt{1 - c^{2/n}}.$$

Tabla 2.6. Niveles de verosimilitud calibrados para la media de la normal (ec. 2.10).

$1 - \alpha$	$n = 2$		$n = 5$		$n = 10$	
	$c(\alpha, \frac{3}{4})$	pce	$c(\alpha, \frac{3}{10})$	pce	$c(\alpha, \frac{3}{20})$	pce
.99	.002	.972	.012	.988	.022	.989
.95	.027	.895	.079	.943	.109	.948
.9	.078	.820	.167	.890	.209	.898
.8	.212	.695	.338	.786	.387	.797
.75	.286	.641	.417	.735	.466	.746
.5	.650	.403	.740	.486	.769	.496
.25	.908	.196	.935	.242	.943	.248
pce = probabilidad de cobertura exacta						

La Tabla 2.6. contiene los valores del nivel de verosimilitud calibrado para distintas probabilidades de cobertura. Observe que a partir de  $n = 5$  las probabilidades de cobertura verdaderas son muy cercanas a la nominal.

Las probabilidades exactas se calcularon con

$$p(\hat{\mu}, \mu | \hat{\sigma}_\mu^2) = \frac{\Gamma(\frac{1}{2}n)}{\sqrt{\pi}\Gamma(\frac{1}{2}n - \frac{1}{2})\hat{\sigma}_\mu} \left(1 - \left(\frac{\hat{\mu} - \mu}{\hat{\sigma}_\mu}\right)^2\right)^{\frac{n}{2} - \frac{3}{2}}$$

Así la probabilidad de cobertura del intervalo de verosimilitud con nivel de verosimilitud  $c$  es

$$\begin{aligned} \Pr(R_p(\psi) \geq c | \hat{\sigma}_\mu^2) &= \Pr(\mu - \hat{\sigma}_\mu \sqrt{1 - c^{2/n}} \leq \hat{\mu} \leq \mu + \hat{\sigma}_\mu \sqrt{1 - c^{2/n}} | \hat{\sigma}_\mu^2) \\ &= \frac{\Gamma(\frac{1}{2}n)}{\sqrt{\pi}\Gamma(\frac{1}{2}n - \frac{1}{2})\hat{\sigma}_\mu} \int_{\mu - \hat{\sigma}_\mu \sqrt{1 - c^{2/n}}}^{\mu + \hat{\sigma}_\mu \sqrt{1 - c^{2/n}}} \left(1 - \left(\frac{\hat{\mu} - \mu}{\hat{\sigma}_\mu}\right)^2\right)^{\frac{n}{2} - \frac{3}{2}} d\hat{\mu} \\ &= \frac{\Gamma(\frac{1}{2}n)}{\sqrt{\pi}\Gamma(\frac{1}{2}n - \frac{1}{2})} \int_{-\sqrt{1 - c^{2/n}}}^{\sqrt{1 - c^{2/n}}} (1 - t^2)^{\frac{1}{2}n - \frac{3}{2}} dt \end{aligned}$$

de donde obtenemos que  $\Pr(R_p(\psi) \geq c | \hat{\sigma}_\mu^2) = \Pr(R_p(\psi) \geq c)$ , no depende de  $\mu$  ni de  $\hat{\sigma}_\mu$ .

**Caso 2. Parámetro de interés  $\psi = \mu + \frac{1}{2}\sigma^2$ .**

La distribución log-normal tiene muchas aplicaciones en relación con variables aleatorias positivas y uno de los problemas más comunes es estimar la media de dicha distribución. Debido a que si  $Z$  tiene distribución log-normal se cumple que  $Y = \log Z$  tiene distribución normal  $\mathcal{N}(\mu, \sigma^2)$ , para estimar la media de la distribución log-normal es más simple trabajar con la distribución normal. Dado que  $EZ = Ee^Y$ , evaluando esta esperanza obtenemos que

$$\log Ee^Y = \mu + \frac{1}{2}\sigma^2.$$

En esta sección obtenemos intervalos de verosimilitud perfil para el parámetro  $\psi = \mu + \frac{1}{2}\sigma^2$  y calcular la probabilidad de cobertura de ellos. Inferencia para el parámetro la media de la distribución log-normal por método exacto se encuentra en Land (1971, 1972), por métodos de órdenes mayores en Jensen (1986a) y DiCiccio (1993). Un problema relacionado donde se hace inferencia sobre  $\psi = \mu + b\sigma + \frac{1}{2}c\sigma^2$  en el modelo de localización escala se encuentra en Barndorff-Nielsen & Cox (1994, págs. 204-205).

Consideremos la reparametrización  $\psi = \mu + \frac{1}{2}\sigma^2$ . Entonces por invariancia de la función de verosimilitud tenemos que  $\hat{\psi} = \hat{\mu} + \frac{1}{2}\hat{\sigma}^2$  y de (2.30) obtenemos que la función de log-verosimilitud en términos de  $\hat{\psi}$  y  $\hat{\sigma}$  es

$$l(\psi, \chi; \hat{\psi}, \hat{\chi}) = -n \ln \chi - \frac{n}{2\chi^2} \left[ \hat{\sigma}^2 + \left( \hat{\psi} - \psi - \frac{1}{2}\hat{\sigma}^2 + \frac{1}{2}\chi^2 \right)^2 \right]. \quad (2.34)$$

El estimador de máxima verosimilitud de  $\sigma^2$  restringido a  $\psi$  es

$$\hat{\sigma}_\psi^2 = -2 + 2\sqrt{1 + \hat{\sigma}^2 + \left( \hat{\psi} - \psi - \frac{1}{2}\hat{\sigma}^2 \right)^2}. \quad (2.35)$$

Si fijamos  $\hat{\sigma}_\psi$ ,  $\hat{\sigma}$  debe cumplir

$$\hat{\sigma}^2 = 2\hat{\psi} - 2\psi - 2 + 2\sqrt{2\psi - 2\hat{\psi} + \left(1 + \frac{1}{2}\hat{\sigma}_\psi^2\right)^2} \quad (2.36)$$

En virtud de que la log-verosimilitud (2.34) se puede reescribir como

$$l(\psi, \sigma; \hat{\psi}, \hat{\sigma}) = -n \ln \sigma - \frac{n}{2\sigma^2} \left[ \left(1 + \frac{1}{2}\hat{\sigma}_\psi^2\right)^2 - 1 + \sigma^2 \left( \hat{\psi} - \psi - \frac{1}{2}\hat{\sigma}^2 \right) + \frac{1}{4}\sigma^4 \right],$$

se tiene que la log-verosimilitud perfil se reduce a

$$l_p(\psi) = -n \ln \hat{\sigma}_\psi - \frac{n}{2} \left( \hat{\psi} - \psi + \frac{1}{2}\hat{\sigma}_\psi^2 - \frac{1}{2}\hat{\sigma}^2 + 1 \right). \quad (2.37)$$

Las derivadas necesarias para obtener la información de Fisher son

$$l_{\psi\psi} = -\frac{n}{\sigma^2}, \quad l_{\sigma\sigma} = -\frac{n}{\sigma^4} \left\{ \frac{1}{4}\sigma^4 - \sigma^2 + 3 \left(1 + \frac{1}{2}\hat{\sigma}_\psi^2\right)^2 - 3 \right\}, \quad l_{\psi\sigma} = -\frac{n}{\sigma^3} \left( 2\psi - 2\hat{\psi} + \hat{\sigma}^2 \right).$$

Cuando  $\sigma = \hat{\sigma}_\psi$  se cumple lo siguiente

$$l_{\sigma\sigma}(\psi, \hat{\sigma}_\psi; \hat{\psi}, \hat{\sigma}) = -\frac{2n}{\hat{\sigma}_\psi^2} \left( 1 + \frac{1}{2}\hat{\sigma}_\psi^2 \right), \quad (2.38)$$

y

$$\hat{l}_{\psi\psi} = -\frac{n}{\hat{\sigma}^2}, \quad \hat{l}_{\psi\sigma} = -\frac{n}{\hat{\sigma}}, \quad \hat{l}_{\sigma\sigma} = 2\frac{n}{\hat{\sigma}^2} - n \Rightarrow \hat{j} = \frac{2n^2}{\hat{\sigma}^4}$$

de (2.34) obtenemos que

$$l_{,\hat{\psi}} = -\frac{1}{2}\sigma l_{\sigma;\hat{\psi}}, \quad l_{,\hat{\sigma}} = -\frac{1}{2}\sigma l_{\sigma;\hat{\sigma}} - \frac{1}{2}n\hat{\sigma}, \quad l_{\sigma;\hat{\psi}} = \frac{2n}{\sigma^3} \left( \hat{\psi} - \psi - \frac{1}{2}\hat{\sigma}^2 \right)$$

$$l_{\sigma;\hat{\sigma}} = \frac{2n\hat{\sigma}}{\sigma^3} \left( \psi - \hat{\psi} + 1 + \frac{1}{2}\hat{\sigma}^2 \right).$$

Aplicando la fórmula  $\hat{\sigma}_{/\hat{\psi}} = -\tilde{l}_{\sigma;\hat{\psi}} \tilde{l}_{\sigma;\hat{\sigma}}^{-1}$  obtenemos que

$$\hat{\sigma}_{/\hat{\psi}} = -\frac{\hat{\psi} - \psi - \frac{1}{2}\hat{\sigma}^2}{\hat{\sigma} \left( \psi - \hat{\psi} + 1 + \frac{1}{2}\hat{\sigma}^2 \right)},$$

el mismo resultado se obtiene si se deriva (2.36) con respecto a  $\hat{\psi}$ .

De los cálculos anteriores encontramos que

$$\begin{aligned} |\hat{\ell}_{\hat{\sigma}} - \bar{\ell}_{\hat{\sigma}}| &= \frac{n^2 \hat{\sigma}}{\hat{\sigma}_\psi^3} + \frac{n \bar{\ell}_{\sigma; \hat{\psi}}}{\hat{\sigma}} \\ &= \frac{2n^2}{\hat{\sigma} \hat{\sigma}_\psi^3} (\hat{\psi} - \psi). \end{aligned}$$

Dado que  $\hat{j} = 2n^2/\hat{\sigma}^4$ , se tiene que  $\bar{j}_{\sigma\sigma}\hat{j} = \frac{4n^3}{\hat{\sigma}^4 \hat{\sigma}_\psi^2} (1 + \frac{1}{2}\hat{\sigma}^2)$  y así

$$u = \sqrt{n} \frac{\hat{\sigma} (\hat{\psi} - \psi)}{\hat{\sigma}_\psi^2 \sqrt{1 + \frac{1}{2}\hat{\sigma}^2}}. \quad (2.39)$$

En Barndorff-Nielsen & Cox (1994) se calcula  $u$  para  $\psi = \mu + b\sigma + \frac{1}{2}c\sigma^2$ .

### Aproximación del intervalo de verosimilitud

En lo siguiente presentamos las expansiones alrededor de  $\hat{\psi}$  con el fin de llegar a una expresión para los intervalos de verosimilitud perfil para la media de la log-normal  $\psi = \mu + \frac{1}{2}\sigma^2$ .

Usando el hecho de que  $\frac{\partial}{\partial \psi} \ell(\psi, \hat{\sigma}_\psi; \hat{\psi}, \hat{\sigma}) = \ell_\psi(\psi, \hat{\sigma}_\psi; \hat{\psi}, \hat{\sigma}) = n \hat{\chi}_\psi^{-2} [\hat{\psi} - \psi + \frac{1}{2}(\hat{\chi}_\psi^2 - \hat{\chi}^2)]$  obtenemos que

$$\check{\ell}_1 = 0, \check{\ell}_2 = -2n \frac{1}{\hat{\sigma}^2 (2 + \hat{\sigma}^2)}, \check{\ell}_3 = 8n \frac{2\hat{\sigma}^2 + 3}{\hat{\sigma}^2 (2 + \hat{\sigma}^2)^3}, \check{\ell}_4 = -48n \frac{3\hat{\sigma}^6 + 6\hat{\sigma}^4 - 4}{\hat{\sigma}^4 (2 + \hat{\sigma}^2)^5},$$

donde  $\check{\ell}_k = \frac{\partial^k}{\partial \psi^k} \bar{\ell} \Big|_{\psi=\hat{\psi}}$ . Así en vista de que  $r^2 = 2(\hat{\ell} - \bar{\ell})$  llegamos a

$$r^2 = 2n \frac{1}{\hat{\sigma}^2 (2 + \hat{\sigma}^2)} (\hat{\psi} - \psi)^2 + \frac{8}{3} n \frac{2\hat{\sigma}^2 + 3}{\hat{\sigma}^2 (2 + \hat{\sigma}^2)^3} (\hat{\psi} - \psi)^3 + 4n \frac{3\hat{\sigma}^6 + 6\hat{\sigma}^4 - 4}{\hat{\sigma}^4 (2 + \hat{\sigma}^2)^5} (\hat{\psi} - \psi)^4 + \dots$$

Aplicando raíz cuadrada tenemos

$$\frac{1}{\sqrt{2n}} r = \frac{1}{\hat{\sigma} \sqrt{2 + \hat{\sigma}^2}} (\hat{\psi} - \psi) + \frac{2}{3} \frac{2\hat{\sigma}^2 + 3}{\hat{\sigma} (2 + \hat{\sigma}^2)^{5/2}} (\hat{\psi} - \psi)^2 + \frac{1}{9} \frac{19\hat{\sigma}^6 + 30\hat{\sigma}^4 - 36 - 18\hat{\sigma}^2}{\hat{\sigma}^3 (2 + \hat{\sigma}^2)^{9/2}} (\hat{\psi} - \psi)^3 + \dots$$

Usando las expresiones para la inversa de la Sección A.1.4 se obtiene que los extremos del intervalo de verosimilitud son

$$\psi = \hat{\psi} \pm \frac{\hat{\sigma} \sqrt{2 + \hat{\sigma}^2}}{\sqrt{2n}} r - \frac{1}{3n} \frac{\hat{\sigma}^2 (2\hat{\sigma}^2 + 3)}{2 + \hat{\sigma}^2} r^2 \pm \frac{\sqrt{2}}{36n^{3/2}} \frac{\hat{\sigma} (13\hat{\sigma}^6 + 66\hat{\sigma}^4 + 36 + 90\hat{\sigma}^2)}{(2 + \hat{\sigma}^2)^{5/2}} r^3 + \dots$$

### Obtención del factor de Bartlett

Expansión de  $r$  alrededor de  $\psi$

$$\frac{1}{\sqrt{2n}} r = \frac{1}{\hat{\sigma}_\psi \sqrt{2 + \hat{\sigma}_\psi^2}} (\hat{\psi} - \psi) - \frac{2}{3} \frac{\hat{\sigma}_\psi}{(2 + \hat{\sigma}_\psi^2)^{5/2}} (\hat{\psi} - \psi)^2 + \frac{1}{9} \frac{7\hat{\sigma}_\psi^6 + 18\hat{\sigma}_\psi^4 + 72\hat{\sigma}_\psi^2 + 36}{\hat{\sigma}_\psi^3 (2 + \hat{\sigma}_\psi^2)^{9/2}} (\hat{\psi} - \psi)^3 + \dots$$

Por medio de técnicas de inversión de series (ver Sección A.1.4) la solución de  $r(\hat{\psi}; \hat{\psi}) = r$ , con respecto a  $\hat{\psi}$ , es

$$\hat{\psi} - \psi = \frac{1}{2} \sqrt{2} \hat{\sigma}_\psi \frac{\sqrt{2 + \hat{\sigma}_\psi^2}}{\sqrt{n}} r + \frac{1}{3} \frac{\hat{\sigma}_\psi^4}{(2 + \hat{\sigma}_\psi^2)n} r^2 + \frac{\sqrt{2} \hat{\sigma}_\psi (\hat{\sigma}_\psi^6 - 18\hat{\sigma}_\psi^4 - 72\hat{\sigma}_\psi^2 - 36)}{36(\sqrt{n})^3 (2 + \hat{\sigma}_\psi^2)^{5/2}} r^3 + \dots \quad (2.40)$$

En este ejemplo es más simple expandir la estadística  $u$  alrededor de  $\psi$  directamente de la expresión (2.39). Para esto necesitamos la expansión de  $\hat{\chi}$ , dado en (2.36), alrededor de  $\psi$ .

Así, pues

$$\hat{\chi} = \hat{\chi}_\psi + \frac{\hat{\chi}_\psi}{\hat{\chi}_\psi^2 + 2} (\hat{\psi} - \psi) - \frac{1}{2} \frac{8 + \hat{\chi}_\psi^4 + 2\hat{\chi}_\psi^2}{\hat{\chi}_\psi (\hat{\chi}_\psi^2 + 2)^3} (\hat{\psi} - \psi)^2 + \frac{1}{2} \frac{-16 + 12\hat{\chi}_\psi^2 + 4\hat{\chi}_\psi^4 + \hat{\chi}_\psi^6}{\hat{\chi}_\psi (\hat{\chi}_\psi^2 + 2)^5} (\hat{\psi} - \psi)^3 + \dots \quad (2.41)$$

Después de algunas manipulaciones entre (2.40) y (2.41) llegamos a la siguiente expansión de la estadística  $u$

$$u = r + \frac{1}{6} \sqrt{2} \hat{\sigma}_\psi \frac{5\hat{\sigma}_\psi^2 + 6}{\sqrt{n} (2 + \hat{\sigma}_\psi^2)^{3/2}} r^2 + \frac{1}{36} \frac{17\hat{\sigma}_\psi^6 - 24\hat{\sigma}_\psi^4 - 252\hat{\sigma}_\psi^2 - 216}{n (2 + \hat{\sigma}_\psi^2)^3} r^3 + \dots$$

con ayuda de la fórmula para el recíproco de series de la Sección A.3 obtenemos

$$\frac{r}{u} = 1 - \frac{1}{6} \sqrt{2} \hat{\sigma}_\psi \frac{5\hat{\sigma}_\psi^2 + 6}{\sqrt{n} (2 + \hat{\sigma}_\psi^2)^{3/2}} r + \frac{1}{12} \frac{11\hat{\sigma}_\psi^6 + 48\hat{\sigma}_\psi^4 + 108\hat{\sigma}_\psi^2 + 72}{n (2 + \hat{\sigma}_\psi^2)^3} r^2 + \dots \quad (2.42)$$

finalmente se, llega a que

$$r^* = r + \frac{1}{6}\sqrt{2}\hat{\sigma}_\psi \frac{5\hat{\sigma}_\psi^2 + 6}{\sqrt{n}(\sqrt{(2+\hat{\sigma}_\psi^2)})^3} - \frac{1}{9} \frac{2\hat{\sigma}_\psi^6 + 21\hat{\sigma}_\psi^4 + 72\hat{\sigma}_\psi^2 + 54}{n(2+\hat{\sigma}_\psi^2)^3} r + \dots$$

De (2.42) se desprende que el factor de Bartlett en este caso es

$$b(\psi, \hat{\sigma}_\psi n) = b(\hat{\sigma}_\psi, n) = \frac{1}{6} \frac{11\hat{\sigma}_\psi^6 + 48\hat{\sigma}_\psi^4 + 108\hat{\sigma}_\psi^2 + 72}{n(2 + \hat{\sigma}_\psi^2)^3},$$

el cual tiene un mínimo en  $\hat{\sigma}_\psi^2 = 4/3$ , un máximo local  $\hat{\sigma}_\psi^2 = 0$  y un máximo global en  $\hat{\sigma}_\psi^2 = \infty$ .

Dado que  $b(\hat{\sigma}_\psi, n)$  depende de  $\psi$  sólo a través de  $\hat{\sigma}_\psi$  y para un valor fijo de  $\psi$ ,  $\hat{\sigma}_\psi$  puede variar desde 0 a  $\infty$ , podemos considerar a  $\hat{\sigma}_\psi$  independiente de  $\psi$  y minimizar el factor de Bartlett con respecto a  $\hat{\sigma}_\psi$ , así obtendremos que la probabilidad de cobertura del intervalo de verosimilitud es mínima y marginal. De este modo, con ayuda del factor de Bartlett podemos obtener dos tipos de probabilidades: si fijamos a  $\hat{\sigma}_\psi$  en por ejemplo  $\hat{\sigma}_\psi = \hat{\sigma}$ , tenemos probabilidad de cobertura condicional. Si minimizamos  $b(\hat{\sigma}_\psi, n)$  con respecto a  $\hat{\sigma}_\psi$ , obtenemos probabilidad de cobertura mínima marginal. La Tabla 2.7. muestra que el máximo global de  $b$ , con respecto a  $\psi$ , se alcanza suavemente en  $\infty$ . La Tabla 2.8. muestra los niveles de verosimilitud calibrados para  $\hat{\sigma}_\psi = 30$ . Entonces podemos considerar a las probabilidades exactas presentadas en la Tabla 2.8 como probabilidad de cobertura mínima marginal.

Las probabilidades de cobertura exacta se obtuvieron mediante la densidad condicional de  $\hat{\psi}$  dado  $\hat{\sigma}^2 + (\hat{\mu} - \psi)^2$ , la cual es proporcional a

$$p(\hat{\psi}; \psi | T_\psi) \propto \frac{(\hat{\psi} - \psi - 1 + \sqrt{1 + 2\psi - 2\hat{\psi} + T_\psi})^{\frac{3}{2} - \frac{3}{2}}}{\sqrt{1 + 2\psi - 2\hat{\psi} + T_\psi}} e^{\frac{\pi}{2} \sqrt{1 + 2\psi - 2\hat{\psi} + T_\psi}},$$

donde  $T_\psi = \hat{\sigma}^2 + (\hat{\mu} - \psi)^2$ ,  $-\sqrt{T_\psi} + \psi \leq \hat{\psi} \leq \frac{1}{2} + \frac{1}{2}T_\psi + \psi$ . Para un valor  $\hat{\sigma}_\psi$  se cumple que  $T_\psi = (1 + \frac{1}{2}\hat{\chi}_\psi^2)^2 - 1$ .

Tabla 2.7. Evaluaciones del factor de Bartlett para distintos valores de  $\hat{\sigma}_\psi$ .

$\hat{\sigma}_\psi$	0	1/2	$2/\sqrt{3} \cong 1.15$	5	10	30	100	$\infty$
$b(\hat{\sigma}_\psi, n)$	$\frac{3}{2n}$	$\frac{1.495}{n}$	$\frac{221}{150n} = \frac{1.473}{n}$	$\frac{1.7329}{n}$	$\frac{1.8047}{n}$	$\frac{1.83}{n}$	$\frac{1.833}{n}$	$\frac{11}{6n} = \frac{1.83333}{n}$

Tabla 2.8. Probabilidades de cobertura de los intervalo de verosimilitud perfil para la media de la log-normal, condicionadas a  $\hat{\sigma}_\psi = 30$ .

$\hat{\sigma}_\psi = 30$	$n = 2$		$n = 5$		$n = 10$		$n = 15$	
$1 - \alpha$	$c(\alpha, .916)$	pce	$c(\alpha, .367)$	pce	$c(\alpha, .183)$	pce	$c(\alpha, .122)$	pce
.99	.00086	**	.0096	.966	.019	.988	.024	.989
.95	.017	.503	.068	.934	.101	.948	.115	.949
.9	.056	.595	.15	.885	.199	.897	.218	.899
.8	.174	.627	.317	.784	.376	.796	.397	.798
.75	.245	.603	.396	.734	.455	.746	.475	.748
.5	.616	.399	.727	.485	.763	.496	.774	.498
.25	.898	.192	.931	.242	.941	.248	.944	.25
pce = probabilidad de cobertura exacta								

### 2.5.2 El modelo Gaussiana inversa

La densidad de la distribución Gaussiana inversa está dada por

$$p_Y(y; \lambda, \psi) = \frac{\sqrt{\lambda} e^{\sqrt{\lambda \psi}}}{\sqrt{2\pi}} y^{-\frac{3}{2}} e^{-\frac{1}{2}(\lambda y^{-1} + \psi y)}, \quad y, \psi, \lambda > 0.$$

Caso 1. Parámetro de interés  $\psi$  y  $\lambda$  desconocido.

Sean  $\bar{y} = \frac{1}{n} \sum y_i^{-1}$ ,  $\bar{y} = \frac{1}{n} \sum y_i$ , entonces la función de log-verosimilitud para una muestra de tamaño  $n$  es

$$\ell(\lambda, \psi, y) = \frac{n}{2} \log \lambda + n\sqrt{\lambda\psi} - \frac{n}{2}\lambda\bar{y} - \frac{n}{2}\psi\bar{y}.$$

Las ecuaciones de donde se obtienen  $\hat{\psi}$  y  $\hat{\lambda}$  son

$$\ell_{\psi} = \frac{1}{2} \frac{n}{\sqrt{\psi\lambda}} \lambda - \frac{1}{2} n\bar{y}, \quad \ell_{\lambda} = \frac{1}{2} \frac{n}{\lambda} + \frac{1}{2} \frac{n}{\sqrt{\psi\lambda}} \psi - \frac{1}{2} n\bar{y},$$

entonces

$$\hat{\psi}_{\lambda} = \frac{\lambda}{\bar{y}^2}, \quad y \quad \hat{\lambda}_{\psi} = 4 \left( -\sqrt{\psi} + \sqrt{\psi + 4\bar{y}} \right)^{-2},$$

de donde obtenemos que

$$\hat{\psi} = \bar{y}^{-1} (\bar{y}\bar{y} - 1)^{-1}, \quad \hat{\lambda} = \bar{y} (\bar{y}\bar{y} - 1)^{-1}, \quad \bar{y} = \sqrt{\frac{\hat{\lambda}}{\hat{\psi}}}, \quad \bar{y} = \sqrt{\frac{\hat{\psi}}{\hat{\lambda}}} + \hat{\lambda}^{-1}.$$

De lo anterior el estimador relativo de  $\lambda$  está dado por

$$\hat{\lambda}_{\psi} = 4 \left( -\sqrt{\psi} + \sqrt{\psi + 4 \left( \sqrt{\frac{\hat{\psi}}{\hat{\lambda}}} + \hat{\lambda}^{-1} \right)} \right)^{-2},$$

si  $\hat{\lambda}_{\psi}$  se fija, se debe tener que

$$\hat{\lambda} = \frac{1}{2\Lambda^2} \left( \hat{\psi} + 2\Lambda + \sqrt{\hat{\psi}^2 + 4\hat{\psi}\Lambda} \right), \quad (2.43)$$

donde  $\Lambda = \frac{1}{4} \left( 2\hat{\lambda}_{\psi}^{-1/2} + \sqrt{\psi} \right)^2 - \frac{1}{4}\psi = \sqrt{\frac{\hat{\psi}}{\hat{\lambda}}} + \frac{1}{\hat{\lambda}} = \bar{y}$ .

La función de log-verosimilitud en términos de  $(\hat{\psi}, \hat{\lambda})$  queda como

$$\ell(\psi, \lambda; \hat{\psi}, \hat{\lambda}) = \frac{n}{2} \log \lambda + n\sqrt{\lambda\psi} - n\frac{\lambda}{2} \left( \frac{1}{\hat{\lambda}} + \sqrt{\frac{\hat{\psi}}{\hat{\lambda}}} \right) - n\frac{\psi}{2} \sqrt{\frac{\hat{\lambda}}{\hat{\psi}}}, \quad (2.44)$$

y

$$\ell(\hat{\psi}, \hat{\lambda}, \hat{\psi}, \hat{\lambda}) = \frac{n}{2} (\ln \hat{\lambda} - 1).$$

Las cantidades necesarias para obtener la información de Fisher son

$$\bar{j}_{\psi\psi} = \frac{n}{4} \psi^{-3/2} \lambda^{1/2}, \quad \bar{j}_{\lambda\lambda} = -\frac{1}{4} \frac{n}{\lambda^2} (2 + \sqrt{\lambda\psi}), \quad \bar{j}_{\psi\lambda} = \bar{j}_{\lambda\psi} = -\frac{n}{4} (\psi\lambda)^{-1/2}, \quad (2.45)$$

de donde

$$\bar{j} = \frac{n^2}{8} (\psi\lambda)^{-3/2},$$

Tabla 2.9. Soluciones con respecto a  $\hat{\psi}$  de  $r(\psi; \hat{\psi}) = 1.96$ ,  $n = 7$ ,  $\hat{\lambda}_{\psi} = 1$ .

$r(\psi; \hat{\psi}) = 1.96$					
$\psi$	.1	1	2	5	10
Raíces	.0072, 1.496	.178, 6.486	.427, 11.115	1.291, 23.895	2.885, 44.035
$b^*$	.191	.125	.114	.1088	.1089

y por (1.17) la estadística  $u$  es

$$u_{\psi} = \frac{\sqrt{n}}{\sqrt{2}} \frac{\hat{\lambda}_{\psi} (\hat{\psi}\hat{\lambda})^{-3/4}}{\sqrt{2 + (\psi\hat{\lambda}_{\psi})^{1/2}}} (\hat{\psi} - \psi).$$

En este ejemplo es más simple resolver numéricamente la relación  $r(\psi; \hat{\psi}) = r_0$  y evaluar directamente (2.2) que calcular exactamente el factor de Bartlett. La Tabla 2.9. contiene las

raíces de  $r(\psi; \hat{\psi}) = 1.96$ , con respecto a  $\hat{\psi}$ , se eligió arbitrariamente  $r_0 = 1.96$ , dado que  $b^*$  de (2.13) depende levemente del valor  $r_0$  que se use para resolver la ecuación  $r(\psi; \hat{\psi}) = r_0$ . La Tabla 2.10 Presenta valores de los niveles de verosimilitud calibrados con  $b^*$  para distintos valores de la probabilidad de cobertura. Observe que la aproximación numérica del factor de Bartlett dada en (2.13) trabaja bien. Las probabilidades de cobertura exacta se obtuvieron de (2.46).

Tabla 2.10. Niveles de verosimilitud calibrados con la aproximación del factor de Bartlett

dada en (2.13).

$n = 7$		
$1 - \alpha$	$c(\alpha, .125)$	pce
.99	.015	.9897
.95	.088	.948
.9	.18	.896
.8	.354	.794
.75	.433	.744
.5	.650	.493
.25	.908	.246
pce = Prob. de cob. exc.		

**Densidad de  $\hat{\psi}|\hat{\lambda}_\psi$ .**

Dado que la fórmula  $p^*$  para la densidad de  $(\hat{\psi}, \hat{\lambda})$  es exacta en este caso tenemos que

$$p(\hat{\psi}, \hat{\lambda}; \psi, \lambda) = p^*(\hat{\psi}, \hat{\lambda}; \psi, \lambda) \\ \propto \frac{n}{\hat{\psi}^{3/4} \lambda^{1/2} n + 3/4} \exp \left( \frac{n}{2} \log \lambda + n \sqrt{\lambda \psi} - \frac{n}{2} \lambda \left( \sqrt{\frac{\hat{\psi}}{\lambda}} + \hat{\lambda}^{-1} \right) - \frac{n}{2} \psi \left( \sqrt{\frac{\hat{\lambda}}{\hat{\psi}}} \right) \right)$$

De  $\ell_{\lambda\lambda} = -\frac{1}{4} \frac{n}{\lambda^2} (2 + \sqrt{\psi \lambda})$ ,  $\ell_{\lambda; \hat{\lambda}} = \frac{1}{4} \frac{n}{\lambda^2} (2 + \sqrt{\hat{\psi} \hat{\lambda}})$  y de  $\frac{\partial \hat{\lambda}}{\partial \lambda_\psi} = -\tilde{\ell}_{\lambda\lambda} \tilde{\ell}_{\lambda; \hat{\lambda}}^{-1}$  se tiene que

$$\frac{\partial \hat{\lambda}}{\partial \lambda_\psi} = \frac{\hat{\lambda}^2 (2 + \sqrt{\psi \hat{\lambda}_\psi})}{\hat{\lambda}_\psi^2 (2 + \sqrt{\hat{\psi} \hat{\lambda}})}$$

De donde se tiene que

$$p(\hat{\psi}, \psi | \hat{\lambda}_\psi) \propto \frac{2 + \sqrt{\psi \hat{\lambda}_\psi}}{2 + \sqrt{\hat{\psi} \hat{\lambda}_\psi}} \frac{n}{\hat{\lambda}_\psi^2 \hat{\psi}^{3/4} \lambda^{1/2} n - 5/4} \exp \left( -\frac{n}{2} \psi \sqrt{\frac{\hat{\lambda}}{\hat{\psi}}} \right) \quad (2.46)$$

Donde  $\hat{\lambda} = \hat{\lambda}(\hat{\psi}, \psi; \hat{\lambda}_\psi)$  está dado en (2.43).

**Expansión alrededor de  $\hat{\psi}$ .**

Para calcular los extremos de un intervalo de verosimilitud a un nivel especificado necesitamos obtener la expansión de  $r$  alrededor de  $\hat{\psi}$ .

Las derivadas con respecto a  $\psi$  y  $\lambda$  de la verosimilitud  $\ell(\psi, \lambda)$  dada en (2.44), son

$$\begin{aligned} \hat{\ell}_\psi &= 0, \quad \hat{\ell}_{\psi\psi} = -\frac{1}{4\sqrt{\hat{\psi}\hat{\lambda}\hat{\psi}}}n\hat{\lambda}, \quad \hat{\ell}_{\psi\psi\psi} = \frac{3}{8\hat{\psi}^{5/2}}\sqrt{\hat{\lambda}}n, \quad \hat{\ell}_{\psi\psi\psi\psi} = -\frac{15}{16\hat{\psi}^{7/2}}\sqrt{\hat{\lambda}}n, \\ \hat{\ell}_{\psi\lambda} &= \frac{1}{4\sqrt{\hat{\psi}\hat{\lambda}}}n, \quad \hat{\ell}_{\psi\lambda\lambda} = -\frac{1}{8\sqrt{\hat{\psi}\hat{\lambda}}}\frac{n}{\hat{\lambda}}, \quad \hat{\ell}_{\psi\lambda\lambda\lambda} = \frac{3}{16\sqrt{\hat{\psi}\hat{\lambda}}}\frac{n}{\hat{\lambda}^2}, \quad \hat{\ell}_{\psi\psi\lambda} = -\frac{1}{8\sqrt{\hat{\psi}\hat{\lambda}\hat{\psi}}}n, \\ \hat{\ell}_{\psi\psi\lambda\lambda} &= \frac{1}{16\sqrt{\hat{\psi}\hat{\lambda}\hat{\psi}}}\frac{n}{\hat{\lambda}}, \quad \hat{\ell}_{\psi\psi\psi\lambda} = \frac{3}{16\sqrt{\hat{\psi}\hat{\lambda}\hat{\psi}^2}}n, \quad \frac{\partial}{\partial\psi}\hat{\lambda}_\psi\Big|_{\psi=\hat{\psi}} = \frac{\hat{\lambda}^{3/2}}{(\sqrt{\hat{\psi}\hat{\lambda}}+2)\sqrt{\hat{\psi}}}, \\ \frac{\partial^2}{\partial\psi^2}\hat{\lambda}_\psi\Big|_{\psi=\hat{\psi}} &= -\frac{2}{(\sqrt{\hat{\psi}\hat{\lambda}}+2)^3}\left(\frac{\hat{\lambda}}{\hat{\psi}}\right)^{3/2}, \quad \frac{\partial^3}{\partial\psi^3}\hat{\lambda}_\psi\Big|_{\psi=\hat{\psi}} = 6\frac{\hat{\psi}\hat{\lambda}+2\sqrt{\hat{\psi}\hat{\lambda}}+2}{(\sqrt{\hat{\psi}})^5(\sqrt{\hat{\psi}\hat{\lambda}}+2)^5}\hat{\lambda}^{3/2}. \end{aligned}$$

Así, de las expresiones de la Sección A.3 tenemos que las derivadas con respecto a  $\psi$  de la log-verosimilitud perfil  $\ell_p(\psi) = \ell(\psi, \hat{\lambda}_\psi)$  son

$$\hat{\ell}_{p/\psi\psi} = -\frac{n\hat{\lambda}}{2\hat{\psi}}\frac{1}{\hat{\lambda}\hat{\psi}+2\sqrt{\hat{\lambda}\hat{\psi}}}, \quad \text{y} \quad \hat{\ell}_{p/\psi\psi\psi} = n\sqrt{\hat{\lambda}}\frac{\hat{\lambda}\hat{\psi}+3\sqrt{\hat{\psi}\hat{\lambda}}+3}{\hat{\psi}^{5/2}(\sqrt{\hat{\psi}\hat{\lambda}}+2)^3},$$

y si  $w_1 = \sqrt{\hat{\psi}\hat{\lambda}}$  y  $w_2 = \sqrt{\hat{\lambda}/\hat{\psi}} \Rightarrow \hat{\lambda}/w_1 = w_2$ ,  $w_3 = w_1 + 2$  se obtiene que

$$n^{-1}\hat{\ell}_{p/\psi\psi\psi} = \frac{3}{16}\frac{w_2^4}{w_1w_3^4}(12+w_1w_3+w_3^2) + \frac{3}{16}\frac{w_2^3}{\hat{\psi}w_1w_3^5}(16+4w_3^2+3w_3^4) - \frac{15}{16\hat{\psi}^3}w_2.$$

Con las expresiones anteriores se obtiene la expansión de la verosimilitud dirigida alrededor de  $\hat{\psi}$ , y se puede obtener una expresión aproximada de orden  $O(n^{-2})$  para los extremos del intervalo de verosimilitud al nivel de verosimilitud que se desee. Los cálculos pueden hacerse siguiendo las sugerencias del apéndice.

Para encontrar los valores de  $\hat{\psi}$  tales que  $r(\psi; \hat{\psi}) = r_0$  es necesario expandir  $\hat{\ell}_{p/\psi\psi}$  y  $\hat{\ell}_{p/\psi\psi\psi}$  alrededor de  $\psi$ . Debido a la inmanejabilidad de las expresiones no se calcularán aquí.

**Caso 2. Parámetro de interés  $\lambda$  y  $\psi$  desconocido.**

Para este caso el cálculo de  $u$  es más simple si se utiliza como parámetro de estorbo a  $\gamma = EY_i$ . Entonces, la función de log-verosimilitud de  $(\lambda, \gamma)$  queda de la siguiente manera

$$\ell(\lambda, \gamma) = \frac{n}{2}\ln\lambda + n\frac{\lambda}{\gamma} - \frac{\lambda n}{2}(\bar{y} + \bar{y}/\gamma^2). \quad (2.47)$$

Dado que  $\ell_\lambda = -\frac{n}{2\lambda\gamma^2}(-\gamma^2 - 2\lambda\gamma + \lambda\bar{y}\gamma^2 + \lambda\bar{y})$  se tiene que  $\hat{\lambda}_\gamma = \gamma^2(\bar{y}\gamma^2 - 2\gamma + \bar{y})^{-1}$  y por lo tanto  $\hat{\lambda} = \bar{y}(\bar{y}\bar{y} - 1)^{-1}$  así  $\bar{y} = \hat{\gamma}^{-1} + \hat{\lambda}^{-1}$ . Ahora de  $\ell_\gamma = \frac{n\lambda}{\gamma^3}(-\gamma + \bar{y})$  se sigue que  $\hat{\gamma}_\lambda = \hat{\gamma} = \bar{y}$ . En consecuencia la función de log-verosimilitud expresada en términos de  $(\hat{\lambda}, \hat{\gamma})$  es

$$\ell(\lambda, \gamma, \hat{\lambda}, \hat{\gamma}) = \frac{n}{2}\ln\lambda + n\frac{\lambda}{\gamma} - \frac{\lambda n}{2}(\hat{\gamma}^{-1} + \hat{\lambda}^{-1} + \hat{\gamma}/\gamma^2).$$

Las cantidades para evaluar  $u$  son:  $\ell_{\hat{\lambda}} = \frac{1}{2}\lambda\frac{n}{\hat{\lambda}^2}$ ,  $\ell_{\gamma\gamma} = -n\lambda\frac{-2\hat{\lambda}+3\hat{\gamma}}{\gamma^4}$ ,  $\ell_{\lambda\lambda} = -\frac{1}{2}\frac{n}{\hat{\lambda}^2}$  por lo tanto

$$u = -\sqrt{\frac{n}{2}}\left(\frac{\lambda}{\hat{\lambda}}\right)^{3/2}(1 - \lambda/\hat{\lambda}).$$

Se puede probar que  $\Sigma(y_i^{-1} - \bar{y}^{-1}) \sim \Gamma(\frac{n-1}{2}, \lambda/2)$ , entonces dado que  $\hat{\lambda}^{-1} = \frac{1}{n}\Sigma(y_i^{-1} - \bar{y}^{-1})$  se desprende que  $\lambda n\hat{\lambda}^{-1} \sim \chi_{n-1}^2$ .

La verosimilitud dirigida y la estadística  $u_\lambda$  se pueden escribir en función de  $t = \lambda/\hat{\lambda}$  de la siguiente manera:

$$r = \sqrt{n}\sqrt{-\ln t + t - 1} \quad \text{y} \quad u = \sqrt{\frac{n}{2}}t^{3/2}(1 - t).$$

**Observación.** La verosimilitud de  $t$  es proporcional al recíproco de una  $\chi_{n+2}^2$ , puesto que  $R(t) = e^{-\frac{1}{2}r^2} = t^{-\frac{n}{2}}e^{\frac{n}{2}t}e^{-\frac{1}{2}n}$ .

La expansión de Taylor de  $r$  alrededor de  $t = 1$  es

$$r = \frac{\sqrt{2n}}{2} (t - 1) - \frac{\sqrt{2n}}{6} (t - 1)^2 + \frac{7\sqrt{2n}}{72} (t - 1)^3 + \dots,$$

de donde, mediante las manipulaciones algebraicas con series asintóticas que se encuentran en el Apéndice A llegamos a

$$t - 1 = \frac{\sqrt{2}}{\sqrt{n}} r + \frac{2}{3n} r^2 + \frac{\sqrt{2}}{18n^{3/2}} r^3 + \dots, \tag{2.48}$$

así

$$u = r + \frac{11}{6} \frac{\sqrt{2}}{\sqrt{n}} r^2 + \frac{101}{36n} r^3 + \dots,$$

y por último

$$\frac{r}{u} = 1 - \frac{11}{6} \frac{\sqrt{2}}{\sqrt{n}} r + \frac{47}{12n} r^2 + \dots,$$

de la última expresión obtenemos que  $b = \frac{47}{12n}$ .

Tabla 2.11. Niveles de verosimilitud aproximados y probabilidades de cobertura de los intervalos de verosimilitud obtenidos con (2.47).

	$n = 10$		$n = 20$	
	$c(\alpha, \frac{47}{120})$	pce	$c(\alpha, \frac{47}{240})$	pce
.95	.024	.987	.064	.959
.5	.939	.606	.722	.558

pce = probabilidad de cobertura exacta.

La Tabla 2.11. muestra que la aproximación con el factor de Bartlett no calibra bien. Las probabilidades exceden al valor que se les fija. También se ve que la aproximación  $O(n^{-1})$  es sub-precisa, lo que se muestra de la probabilidad exacta del intervalo de verosimilitud con nivel de verosimilitud .147. En este ejemplo conviene evaluar la probabilidad de cobertura del hecho que  $t = \lambda/\hat{\lambda}$  es exactamente  $\chi_{n-1}^2$  y usar una aproximación más precisa de  $t$  en (2.48), como lo es la siguiente expansión

$$t - 1 = \frac{\sqrt{2}}{\sqrt{n}} r + \frac{2}{3n} r^2 + \frac{\sqrt{2}}{18n^{3/2}} r^3 - \frac{2}{135n^2} r^4 + \frac{\sqrt{2}}{1080n^{5/2}} r^5 + O(n^{-3}).$$

### 2.5.3 El modelo gama

Consideremos una muestra aleatoria de tamaño  $n$  de la distribución gama, cuya densidad está dada por

$$p_Y(y; \alpha, \lambda) = \frac{\alpha^\lambda}{\Gamma(\lambda)} y^{\lambda-1} e^{-\alpha y},$$

donde  $\alpha, \lambda, y > 0$  y  $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$ .

Supongamos que el parámetro de interés es  $\mu = \lambda/\alpha$

$$p_Y(y; \mu, \lambda) = \frac{\lambda^\lambda}{\mu^\lambda \Gamma(\lambda)} y^{\lambda-1} e^{-\lambda y/\mu}. \tag{2.49}$$

Inferencia sobre el parámetro  $\mu$  ha sido discutida en Fraser, Reid & Wong (1997), Jensen & Kristensen(1991) y Jensen (1986b). Denotemos por  $\check{y}$  a la media geométrica, i.e.,  $\check{y} = (\prod_{i=1}^n y_i)^{1/n}$  y  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ ; entonces la función de log-verosimilitud es

$$\ell(\mu, \lambda) = n\lambda \ln \lambda - n\lambda \ln \mu - n \ln \Gamma(\lambda) + n\lambda \ln \check{y} - n\lambda \frac{\bar{y}}{\mu}.$$

Los estimadores de máxima verosimilitud se obtienen de

$$\frac{\partial \ell(\mu, \lambda)}{\partial \mu} = -n\lambda \frac{1}{\mu} + n\lambda \frac{\bar{y}}{\mu^2} \Rightarrow \hat{\mu} = \hat{\mu}_\lambda = \bar{y}, \tag{2.50}$$

$$\frac{\partial \ell(\mu, \lambda)}{\partial \lambda} = n\vartheta(\lambda) + n \ln \frac{\check{y}}{\mu} - n \frac{\bar{y}}{\mu} + n \Rightarrow \vartheta(\hat{\lambda}) = \ln \frac{\check{y}}{\bar{y}}, \quad (2.51)$$

donde  $\vartheta(z) = \ln z - \Psi(z)$  y  $\Psi(z)$  es la función digama, i.e.,  $\Psi(z) = \frac{d}{dz} \ln \Gamma(z) = \Gamma'(z) / \Gamma(z)$ .

De las ecuaciones de verosimilitud (2.50) y (2.51) se sigue que

$$\bar{y} = \hat{\mu}, \quad \check{y} = \hat{\mu} e^{-\vartheta(\hat{\lambda})}, \quad (2.52)$$

además, el estimador de máxima verosimilitud,  $\hat{\lambda}_\mu$ , de  $\lambda$  relativo a un valor de  $\mu$  satisface

$$\vartheta(\hat{\lambda}_\mu) = \vartheta(\hat{\lambda}) + \frac{\hat{\mu}}{\mu} - \ln \frac{\hat{\mu}}{\mu} - 1, \quad (2.53)$$

en consecuencia para  $\hat{\lambda}_\mu$  fijo se cumple que

$$\vartheta(\hat{\lambda}) = \vartheta(\hat{\lambda}_\mu) - \frac{\hat{\mu}}{\mu} + \ln \frac{\hat{\mu}}{\mu} + 1 \quad (2.54)$$

Observemos que no se requiere de ninguna estadística auxiliar, debido a que  $(\hat{\mu}, \hat{\lambda})$  es minimal suficiente. Por lo tanto, la función de log-verosimilitud expresada en términos de  $\mu, \lambda, \hat{\mu}, \hat{\lambda}$  queda como

$$\begin{aligned} \ell(\mu, \lambda; \hat{\mu}, \hat{\lambda}) &= n\lambda \ln \lambda - n\lambda \ln \mu - n \ln \Gamma(\lambda) + n\lambda \ln(\hat{\mu} e^{-\vartheta(\hat{\lambda})}) - n\lambda \frac{\hat{\mu}}{\mu} \\ &= n\lambda \ln \lambda - n \ln \Gamma(\lambda) - n\lambda \left( \vartheta(\hat{\lambda}) + \frac{\hat{\mu}}{\mu} - \ln \frac{\hat{\mu}}{\mu} - 1 \right) - n\lambda \\ &= n\lambda \ln \lambda - n \ln \Gamma(\lambda) - n\lambda \vartheta(\hat{\lambda}_\mu) - n\lambda. \end{aligned}$$

Las derivadas necesarias para evaluar  $r^*$  son

$$\begin{aligned} \ell_\mu &= -n \frac{\lambda}{\mu} \left( 1 - \frac{\hat{\mu}}{\mu} \right), & \ell_{\mu\mu} &= n \frac{\lambda}{\mu^2} \left( 1 - 2 \frac{\hat{\mu}}{\mu} \right), & \ell_{\mu\lambda} &= -\frac{n}{\mu} \left( 1 - \frac{\hat{\mu}}{\mu} \right), \\ \ell_\lambda &= n - n \ln \mu + n\vartheta(\lambda) + n \left[ \ln \hat{\mu} - \vartheta(\hat{\lambda}) \right] - n \frac{\hat{\mu}}{\mu}, & \ell_{\lambda\lambda} &= n\vartheta'(\lambda), \\ \ell_{;\hat{\mu}} &= n\lambda \left( \frac{1}{\hat{\mu}} - \frac{1}{\mu} \right) - \frac{n}{\hat{\mu}}, & \ell_{;\hat{\lambda}} &= -n(\lambda - 1)\vartheta'(\hat{\lambda}), & \ell_{\mu;\hat{\mu}} &= n \frac{\lambda}{\mu^2}, \\ \ell_{\mu;\hat{\lambda}} &= 0, & \ell_{\lambda;\hat{\mu}} &= n \left( \frac{1}{\hat{\mu}} - \frac{1}{\mu} \right), & \ell_{\lambda;\hat{\lambda}} &= -n\vartheta'(\hat{\lambda}), \end{aligned}$$

dato que  $\hat{\ell}_{\mu\mu} = -n \frac{\lambda}{\mu^2}$ ,  $\hat{\ell}_{\mu\lambda} = 0$  y  $\hat{\ell}_{\lambda\lambda} = n\vartheta'(\hat{\lambda})$  se obtiene que  $\hat{j} = -n^2 \frac{\lambda}{\mu^2} \vartheta'(\hat{\lambda})$ . Se necesitan las siguientes evaluaciones

$$\begin{aligned} \hat{\ell}_{;\hat{\mu}} &= -\frac{n}{\hat{\mu}}, & \hat{\ell}_{;\hat{\lambda}} &= -n(\hat{\lambda} - 1)\vartheta'(\hat{\lambda}), & \bar{\ell}_{;\hat{\mu}} &= n\hat{\lambda}_\mu \left( \frac{1}{\hat{\mu}} - \frac{1}{\mu} \right) - \frac{n}{\hat{\mu}}, \\ \bar{\ell}_{;\hat{\lambda}} &= -n(\hat{\lambda}_\mu - 1)\vartheta'(\hat{\lambda}), & \bar{\ell}_{\lambda;\hat{\mu}} &= \frac{n}{\hat{\mu}} - \frac{n}{\mu}, & \bar{\ell}_{\lambda;\hat{\lambda}} &= -n\vartheta'(\hat{\lambda}), \end{aligned}$$

de donde

$$\hat{\lambda}_{/\hat{\mu}} = \frac{\mu - \hat{\mu}}{\hat{\mu}\mu\vartheta'(\hat{\lambda})} = [\vartheta'(\hat{\lambda})]^{-1} \left( \frac{1}{\hat{\mu}} - \frac{1}{\mu} \right) \Rightarrow \hat{\lambda}_{\mu/\hat{\mu}} \Big|_{\hat{\mu}=\mu} = 0. \quad (2.55)$$

El determinante necesario para obtener  $u$  es

$$\left| \begin{array}{cc} \hat{\ell}_{;\hat{\theta}} - \bar{\ell}_{;\hat{\theta}} & \bar{\ell}_{\lambda;\hat{\theta}} \\ \bar{\ell}_{\lambda;\hat{\theta}} & \bar{\ell}_{\lambda;\hat{\lambda}} \end{array} \right| = \left| \begin{array}{cc} -n\hat{\lambda}_\mu \left( \frac{1}{\hat{\mu}} - \frac{1}{\mu} \right) & n \left( \frac{1}{\hat{\mu}} - \frac{1}{\mu} \right) \\ n\vartheta'(\hat{\lambda}) (\hat{\lambda}_\mu - \hat{\lambda}) & -n\vartheta'(\hat{\lambda}) \end{array} \right| = n^2 \hat{\lambda} \vartheta'(\hat{\lambda}) \left( \frac{1}{\hat{\mu}} - \frac{1}{\mu} \right),$$

y puesto que

$$\bar{j}_{\lambda\lambda} \hat{j} = n^3 \frac{\hat{\lambda}}{\hat{\mu}^2} \vartheta'(\hat{\lambda}_\mu) \vartheta'(\hat{\lambda}),$$

llegamos a la siguiente expresión de  $u$ :

$$u = \sqrt{\hat{\lambda}n} \sqrt{\frac{\vartheta'(\hat{\lambda})}{\vartheta'(\hat{\lambda}_\mu)}} \left( 1 - \frac{\hat{\mu}}{\mu} \right). \quad (2.56)$$

De la fórmula (B.8) del Apéndice B, con las siguientes notaciones

$$\Sigma = n \begin{bmatrix} \frac{\lambda}{\alpha^2} & -\frac{1}{\alpha} \\ -\frac{1}{\alpha} & \Psi'(\lambda) \end{bmatrix},$$

$$\begin{aligned} \alpha(\mu, \lambda) &= \lambda/\mu, & \hat{\Sigma}_{11} &= n\hat{\lambda}/\hat{\alpha}^2 = n\hat{\mu}^2/\hat{\lambda}, & \tilde{\Sigma}_{11}^{-1} \tilde{\Sigma}_{12} &= -\alpha/\lambda = -1/\mu, \\ \Delta_{22} &= n^{-1} [\Psi'(\lambda) - \frac{1}{\lambda}]^{-1} = -1/[n\vartheta'(\lambda)], \end{aligned}$$

se llega a que

$$u = \sqrt{n\hat{\lambda}} \sqrt{\frac{\vartheta'(\hat{\lambda})}{\vartheta'(\hat{\lambda}_\psi)}} \left[ 1 - \frac{\hat{\mu}}{\mu} \right]. \quad (2.57)$$

Para que el signo de  $u$  coincida con el de  $r$  debemos multiplicar (2.57) por  $-1$ , entonces

$$u = \sqrt{n\hat{\lambda}} \sqrt{\frac{\vartheta'(\hat{\lambda})}{\vartheta'(\hat{\lambda}_\psi)}} \left[ \frac{\hat{\mu}}{\mu} - 1 \right]. \quad (2.58)$$

Aplicando la fórmula para el recíproco de series (ver Sección A.1.3) a la ecuación (2.61)

llegamos a

$$\frac{r}{u} = 1 - \frac{1}{3\sqrt{n\hat{\lambda}_\mu}} r + \frac{1}{4n\hat{\lambda}_\mu} \left( \frac{1}{3} + \frac{\vartheta''(\hat{\lambda}_\mu)}{\vartheta'(\hat{\lambda}_\mu)^2} + \frac{1}{2\hat{\lambda}_\mu \vartheta'(\hat{\lambda}_\mu)} \right) r^2 + \dots,$$

de donde obtenemos que

$$b(\psi, \hat{\lambda}_\mu, n) = b(\hat{\lambda}_\mu, n) = \frac{1}{2n\hat{\lambda}_\mu} \left( \frac{1}{3} + \frac{\vartheta''(\hat{\lambda}_\mu)}{\vartheta'(\hat{\lambda}_\mu)^2} + \frac{1}{2\hat{\lambda}_\mu \vartheta'(\hat{\lambda}_\mu)} \right).$$

En particular se cumple que

$$\lim_{\hat{\lambda}_\mu \rightarrow 0} b = \infty \quad \text{y} \quad \lim_{\hat{\lambda}_\mu \rightarrow \infty} b = \frac{3}{2n}.$$

Observe que el factor de Bartlett converge suavemente a  $\frac{3}{2n}$  cuando  $\hat{\lambda}_\mu \rightarrow \infty$ . Además para  $\hat{\lambda}_\mu \geq 1$  el factor de Bartlett varía poco. De donde las probabilidad de cobertura condicionadas a un valor fijo de  $\hat{\lambda}_\mu$  son muy precisas y se puede obtener el mínimo de la probabilidad de cobertura para  $\hat{\lambda}_\mu \geq 1$ .

Tabla 2.12. Evaluaciones del factor de Bartlett, modelo gama.

$\hat{\lambda}_\mu$	.0001	.001	.01	.1	1	5	10	20
$b$	$\frac{1667.4}{n}$	$\frac{167.41792}{n}$	$\frac{17.429}{n}$	$\frac{2.307}{n}$	$\frac{1.4669}{n}$	$\frac{1.4976}{n}$	$\frac{1.4994}{n}$	$\frac{1.4998}{n}$

**Obtención del factor de Bartlett.**

Las operaciones presentadas a continuación son útiles para evaluar  $r^*$  en función de  $r$ , una expresión del factor de Bartlett o para encontrar los  $\hat{\mu}$  tales que  $r(\mu; \hat{\mu}) = r_0$ . Con este fin es necesario expandir alrededor de  $\mu$  las cantidades  $r$  y  $u$ .

Gracias a que  $\hat{\ell}_{\mu\lambda} = 0$  se tiene que  $\hat{\lambda}_{/\hat{\mu}}|_{\hat{\mu}=\mu} = 0$  y de la Sección A.1.4. vemos que no es necesario calcular  $\hat{\lambda}_{\mu 3}$ .

De (2.54) se obtiene que

$$\vartheta''(\hat{\lambda}) \hat{\lambda}_{/\hat{\mu}}^2 + \vartheta'(\hat{\lambda}) \hat{\lambda}_{/\hat{\mu}\hat{\mu}} = -\frac{1}{\hat{\mu}^2} \Rightarrow \hat{\lambda}_{/\hat{\mu}\hat{\mu}}|_{\hat{\mu}=\mu} = -\frac{1}{\mu^2} [\vartheta'(\hat{\lambda}_\mu)]^{-1}. \quad (2.59)$$

Expandiendo alrededor de  $\mu$  los coeficientes que acompañan las potencias de  $\hat{\mu} - \mu$  en la ecuación (2.62) y después de muchas manipulaciones llegamos a que

$$r^2 = n \frac{\hat{\lambda}_\mu}{\mu^2} (\hat{\mu} - \mu)^2 - \frac{2}{3} n \frac{\hat{\lambda}_\mu}{\mu^3} (\hat{\mu} - \mu)^3 + \frac{1}{4} n \frac{2\hat{\lambda}_\mu - \vartheta'(\hat{\lambda}_\mu)^{-1}}{\mu^4} (\hat{\mu} - \mu)^4 + \dots$$

Aplicando raíz (Sección A.1.2) a la expresión anterior sigue que

$$r = \frac{1}{\mu} \sqrt{n\hat{\lambda}_\mu} (\hat{\mu} - \mu) - \frac{1}{3\mu^2} \sqrt{n\hat{\lambda}_\mu} (\hat{\mu} - \mu)^2 + \frac{1}{\mu} \sqrt{n\hat{\lambda}_\mu} \left( \frac{1}{8} \frac{2\hat{\lambda}_\mu - \vartheta'(\hat{\lambda}_\mu)^{-1}}{\mu^2 \hat{\lambda}_\mu} - \frac{1}{18\mu^2} \right) (\hat{\mu} - \mu)^3 + \dots$$

Mediante la inversa de la Sección A.1.4 se obtiene que

$$\hat{\mu} - \mu = \frac{\mu}{\sqrt{n\hat{\lambda}_\mu}} r + \frac{1}{3} \frac{\mu}{n\hat{\lambda}_\mu} r^2 + \frac{1}{72n} \left( 2\hat{\lambda}_\mu + 9\vartheta'(\hat{\lambda}_\mu)^{-1} \right) \frac{\mu}{\hat{\lambda}_\mu^2 \sqrt{n\hat{\lambda}_\mu}} r^3 + \dots \quad (2.60)$$

De (2.60) se obtienen los extremos del intervalo de valores para  $\hat{\mu}$  que cumplen  $r(\mu; \hat{\mu}) = r$ , de aquí ya se puede evaluar  $r^*(\mu, r) = r^*(\mu, \mu(r))$ .

Para encontrar el factor de Bartlett se requiere de expandir  $u$  alrededor de  $\mu$ .

Expandiendo alrededor de  $\mu$  la estadística  $u$  dada en (2.58) y utilizando (2.60) llegamos a que

$$u = r + \frac{1}{3\sqrt{n\hat{\lambda}_\mu}}r^2 - \frac{1}{4n\hat{\lambda}_\mu} \left( -\frac{1}{9} + B^2C + \frac{1}{2\hat{\lambda}_\mu}B \right) r^3 + \dots \quad (2.61)$$

donde  $B = \vartheta'(\hat{\lambda}_\mu)^{-1}$  y  $C = \vartheta''(\hat{\lambda}_\mu)$ .

### Obtención del intervalo de verosimilitud

Con el fin de dar los extremos del intervalo de verosimilitud es necesario expandir la verosimilitud dirigida alrededor de  $\hat{\mu}$ .

En virtud de que  $\hat{\ell}_{\mu\lambda} = 0$  los cálculos para expandir alrededor de  $\hat{\mu}$  la función de log-verosimilitud perfil se reducen considerablemente (ver Sección A.3). En particular no se necesita  $\lambda_{\psi 3}$  y las primeras tres derivadas de la log-verosimilitud perfil evaluadas en  $\mu = \hat{\mu}$  se reducen a

$$\hat{\ell}_{p/\mu} = \hat{\ell}_\mu = 0, \quad \hat{\ell}_{p/\mu\mu} = \hat{\ell}_{\mu\mu}, \quad \hat{\ell}_{p/\mu\mu\mu} = \hat{\ell}_{\mu\mu\mu}.$$

De la relación (2.53), al derivar con respecto a  $\mu$  se tiene

$$\vartheta'(\hat{\lambda}_\mu) \hat{\lambda}_{\mu/\mu} = -\frac{\hat{\mu}}{\mu^2} + \frac{1}{\mu} \Rightarrow \hat{\lambda}_{\mu/\mu} \Big|_{\mu=\hat{\mu}} = 0.$$

$$\vartheta''(\hat{\lambda}_\mu) \hat{\lambda}_{\mu/\mu}^2 + \vartheta'(\hat{\lambda}_\mu) \hat{\lambda}_{\mu/\mu\mu} = 2\frac{m}{\mu^3} - \frac{1}{\mu^2} \Rightarrow \hat{\lambda}_{\mu/\mu\mu} \Big|_{\mu=\hat{\mu}} = \frac{1}{\hat{\mu}^2} \vartheta'(\hat{\lambda})^{-1}.$$

Haciendo cálculos se tiene

$$\hat{\ell}_{\mu\mu\mu\mu} = -\frac{18}{\hat{\mu}^4}n\hat{\lambda}, \quad \hat{\ell}_{\mu\mu\mu} = \frac{4}{\hat{\mu}^3}n\hat{\lambda}, \quad \hat{\ell}_{\mu\mu} = -\frac{1}{\hat{\mu}^2}n\hat{\lambda}, \quad \hat{\ell}_{\mu\lambda} = 0, \quad \hat{\ell}_{\mu\mu\lambda} = -\frac{1}{\hat{\mu}^2}n.$$

De la Sección A.3 y del hecho que  $\hat{\lambda}_{\mu/\mu} \Big|_{\mu=\hat{\mu}} = 0$  y  $\hat{\ell}_{\mu\lambda} = 0$  se desprende que

$$\hat{\ell}_{p/\mu\mu\mu\mu} = \hat{\ell}_{\mu\mu\mu\mu} + 3\hat{\ell}_{\mu\mu\lambda} \lambda_{\mu/\mu\mu} = -\frac{18}{\hat{\mu}^4}n\hat{\lambda} - 3n\frac{1}{\hat{\mu}^4} \vartheta'(\hat{\lambda})^{-1}.$$

Así, pues la verosimilitud perfil expandida alrededor de  $\hat{\mu}$  está dada por

$$\bar{\ell} = \hat{\ell} - \frac{1}{2\hat{\mu}^2}n\hat{\lambda}(\hat{\mu} - \mu)^2 - \frac{1}{6} \frac{4}{\hat{\mu}^3}n\hat{\lambda}(\hat{\mu} - \mu)^3 - \frac{1}{8} \left( \frac{6}{\hat{\mu}^4}n\hat{\lambda} + n\frac{1}{\hat{\mu}^4} \vartheta'(\hat{\lambda})^{-1} \right) (\hat{\mu} - \mu)^4,$$

de donde la expansión de  $r^2$  es igual a

$$r^2 = 2(\hat{\ell} - \bar{\ell}) = n\frac{\hat{\lambda}}{\hat{\mu}^2}(\hat{\mu} - \mu)^2 + n\frac{4}{3}\frac{\hat{\lambda}}{\hat{\mu}^3}(\hat{\mu} - \mu)^3 + \frac{n}{4} \left( 6\frac{\hat{\lambda}}{\hat{\mu}^4} + \frac{1}{\hat{\mu}^4}B \right) (\hat{\mu} - \mu)^4 + \dots, \quad (2.62)$$

de la fórmula para la raíz de series de la Sección A.1.2. llegamos a la siguiente expansión de la verosimilitud dirigida

$$r = \frac{\sqrt{n\hat{\lambda}}}{\hat{\mu}}(\hat{\mu} - \mu) + \frac{2}{3} \frac{\sqrt{n\hat{\lambda}}}{\hat{\mu}^2}(\hat{\mu} - \mu)^2 + \frac{\sqrt{n}}{72\hat{\mu}^3\sqrt{\hat{\lambda}}} \left( 38\hat{\lambda} + 9\vartheta'(\hat{\lambda})^{-1} \right) (\hat{\mu} - \mu)^3 + \dots,$$

invirtiendo la serie (ver Sección A.1.4) tenemos

$$\hat{\mu} - \mu = \frac{\hat{\mu}}{\sqrt{n\hat{\lambda}}}r - \frac{2\hat{\mu}}{3n\hat{\lambda}}r^2 + \frac{1}{72} \frac{\hat{\mu}}{n^{3/2}\hat{\lambda}^{5/2}} \left( 26\hat{\lambda} - 9\vartheta'(\hat{\lambda})^{-1} \right) r^3 + \dots$$

### 2.5.4 Regresión exponencial

Sea  $y = (y_1, \dots, y_n)^T$  una muestra de *v.a.i.i.d.*, tal que  $y_i = \lambda e^{-\psi z_i} w_i$  y  $p(w_i) = e^{-w_i}$ . Los valores de  $z_i$  son valores conocidos que satisfacen  $\sum_{i=1}^n z_i = 0$ . La función de log-verosimilitud

es

$$\ell(\psi, \lambda; y) = -n \ln \lambda - \lambda^{-1} \sum_{i=1}^n y_i e^{\psi z_i}$$

En este caso, la estadística auxiliar es

$$a_i = \log y_i - \log \hat{\lambda} + \hat{\psi} z_i, \quad i = 1, \dots, n.$$

Sea

$$A_k(t, a, z) = \sum_{i=1}^n z_i^k e^{a_i - t z_i}, \quad k \geq 0, \quad (2.63)$$

en particular, los primeros tres valores de (2.63) cumplen que  $A_0(0, a, z) = n$ ,  $A_1(0, a, z) = 0$  y  $A_2(0, a, z) > 0$ .

La función de log-verosimilitud, expresada en términos de  $\hat{\psi}, \hat{\lambda}, a$ , está dada por

$$\ell(\psi, \lambda; \hat{\psi}, \hat{\lambda}, a) = -n \log \lambda - \lambda^{-1} \hat{\lambda} A_0(\hat{\psi} - \psi, a, z).$$

Las derivadas de la log-verosimilitud con respecto a  $\psi$  y  $\lambda$  son

$$\ell_\psi = -\frac{\hat{\lambda}}{\lambda} A_1(\hat{\psi} - \psi, a, z), \quad \ell_\lambda = -\frac{n}{\lambda} + \lambda^{-2} \hat{\lambda} A_0(\hat{\psi} - \psi, a, z),$$

así, pues

$$\hat{\lambda}_\psi = \frac{\hat{\lambda}}{n} A_0(\hat{\psi} - \psi, a, z). \quad (2.64)$$

La función de log-verosimilitud se puede reescribir como

$$\ell(\psi, \lambda; \hat{\psi}, \hat{\lambda}, a) = -n \log \lambda - n \lambda^{-1} \hat{\lambda}_\psi, \quad (2.65)$$

en consecuencia

$$\ell_p(\psi) = -n \log \lambda_\psi - n$$

de donde se desprende que

$$r = \sqrt{2n} \operatorname{sign}(\hat{\psi} - \psi) \sqrt{\log(\hat{\lambda}_\psi / \hat{\lambda})}. \quad (2.66)$$

En virtud de que  $\ell_\psi(\hat{\theta}) = \ell_\lambda(\hat{\theta}) = 0$  se siguen las siguientes relaciones

$$A_0(0, a, z) = \sum_{i=1}^n e^{a_i} = n, \quad \text{y} \quad A_1(0, a, z) = \sum_{i=1}^n z_i e^{a_i} = 0.$$

Los cálculos necesarios para obtener  $r^*$  son

$$\begin{aligned} \ell_\psi &= -\frac{\hat{\lambda}}{\lambda} A_1(\hat{\psi} - \psi, a, z), \quad \ell_\lambda = n \lambda^{-2} (\hat{\lambda}_\psi - \lambda), \\ \ell_{\psi\psi} &= -\frac{\hat{\lambda}}{\lambda} \sum_{i=1}^n z_i^2 e^{a_i + (\psi - \hat{\psi}) z_i} \Rightarrow \hat{j}_{\psi\psi} = A_2(0, a, z), \\ \ell_{\lambda\lambda} &= -n \lambda^{-3} (2\hat{\lambda}_\psi - \lambda) \Rightarrow \hat{j}_{\lambda\lambda} = n \hat{\lambda}^{-2}, \\ \ell_{\psi\lambda} &= \ell_{\lambda\psi} = -\lambda^{-1} \ell_\psi \Rightarrow \hat{j}_{\psi\lambda} = \hat{j}_{\lambda\psi} = 0 \Rightarrow \hat{j}_p = \hat{j}_{\psi\psi} \\ \ell_{\psi\hat{\psi}} &= -\ell_\psi, \Rightarrow \ell_{\psi\hat{\psi}} = -\ell_{\psi\psi}, \quad \ell_{\lambda\hat{\psi}} = -\ell_{\psi\lambda} = \lambda^{-1} \ell_\psi, \\ \ell_{\lambda\hat{\lambda}} &= -n \lambda^{-1} \hat{\lambda}_\psi \hat{\lambda}^{-1} \Rightarrow \ell_{\lambda\hat{\lambda}} = n \lambda^{-2} \hat{\lambda}^{-1} \hat{\lambda}_\psi, \end{aligned}$$

de donde se llega a que

$$\hat{j} = n \hat{\lambda}^{-2} \sum_{i=1}^n z_i^2 e^{a_i}, \quad \tilde{\ell}_{\lambda\lambda} = -n \hat{\lambda}_\psi^{-2}, \quad \hat{\ell}_{\lambda\lambda} = -n \hat{\lambda}^{-2}, \quad \tilde{\ell}_{\lambda\hat{\lambda}} = n \hat{\lambda}_\psi^{-1} \hat{\lambda}^{-1}.$$

En este modelo se cumple que  $C = 1$ , puesto que  $\tilde{\ell}_{\lambda\lambda} \hat{\ell}_{\lambda\lambda} = \tilde{\ell}_{\lambda\hat{\lambda}}^2$ . Entonces

$$\begin{aligned} u &= \hat{u} = \hat{j}_p^{-1/2} \tilde{\ell}_{p/\hat{\psi}}(\psi) \\ &= n A_2^{-1/2}(0, a, z) \frac{A_1(\hat{\psi} - \psi, a, z)}{A_0(\hat{\psi} - \psi, a, z)} \end{aligned} \quad (2.67)$$

$$= \left( \sum_{i=1}^n z_i^2 e^{a_i} \right)^{-1/2} \frac{\hat{\lambda}}{\hat{\lambda}_\psi} \sum_{i=1}^n z_i e^{a_i - (\hat{\psi} - \psi) z_i}. \quad (2.68)$$

La verosimilitud dirigida (2.66) y la estadística  $u$  (2.68) dependen de  $\hat{\lambda}$  a través del cociente  $\hat{\lambda}_\psi/\hat{\lambda}$ , el cual no depende de  $\hat{\lambda}$ . De donde tenemos que la distribución de  $r$  y  $r^*$  es independiente de  $\hat{\lambda}$ . Además  $u$  y  $r$  dependen  $\psi$  y  $\hat{\psi}$  a través de  $t = \hat{\psi} - \psi$ , por lo que en este ejemplo con parámetros de estorbo  $r^*$  es función uno-a-uno de  $r$ , para los valores fijos de  $a$  y  $z$ .

La Tabla 2.13 muestra los niveles de verosimilitud calibrados con (2.10). Para este ejemplo se supuso  $\psi = \lambda = 5$  y las probabilidades de cobertura exactas se calcularon de

$$p(\hat{\psi} - \psi | a) = \bar{c}(a, z) \left[ \sum_{i=1}^n e^{a_i + (\hat{\psi} - \psi) z_i} \right]^{-n},$$

donde  $\bar{c}(a, z) = \int_{-\infty}^{\infty} \left[ \sum_{i=1}^n e^{a_i + (\hat{\psi} - \psi) z_i} \right]^{-n} d\hat{\psi}$ .

Tabla 2.13. Niveles de verosimilitud calibrados.

$n$	$c(.95, a)$	pce
3	.1094708	0.9513
4	.1305851	0.9505
5	.1253001	0.9504
6	.1343893	0.9502
pce = Prob. de cob. exc.		

$n = 3: z = (-0.6325715, -0.2153551, 0.8479266)$

$a = (0.17224929, -0.30344866, 0.07116489)$

$n = 4: z = (-0.71089948, -0.08897221, -0.10864187, 0.90851356)$

$a = (-0.3627202, -1.5351118, 0.8125926, -0.1802709)$

$n = 5: z = (0.86492108, -0.68463955, -0.08212786, 0.21093973, -0.30909340)$

$a = (0.38109442, 0.64534847, -1.95173833, -0.01821946, -0.68222625)$

$n = 6$

$z = (0.12623859, -0.04544833, -0.64130889, -0.09866623, 0.25421908, 0.40496578)$

$a = (-2.8546947, 0.4774899, -0.4437598, 0.2925090, 0.8180419, -2.4896246)$

Expansiones de  $r$  y  $u$

Mediante las expansiones de  $A_0(\hat{\psi} - \psi, a, z)$  y  $A_1(\hat{\psi} - \psi, a, z)$  alrededor de  $\hat{\psi} = \psi$  obtenemos las expansiones de la verosimilitud dirigida y la estadística  $u$ .

En virtud de que

$$\frac{\partial^k}{\partial \hat{\psi}^k} A_0(\hat{\psi} - \psi, a, z) = (-1)^k \sum_{i=1}^n z_i^k e^{a_i - t z_i + \psi z_i} = (-1)^k A_k(\hat{\psi} - \psi, a, z)$$

se llega a que

$$A_0(\hat{\psi} - \psi, a, z) = n + \frac{1}{2} A_2(\hat{\psi} - \psi)^2 - \frac{1}{6} A_3(\hat{\psi} - \psi)^3 + \frac{1}{24} A_4(\hat{\psi} - \psi)^4 + \dots,$$

dado que  $A_0(0, a, z) = n$ .

La verosimilitud dirigida se expresa como

$$r = \text{sgn}(\hat{\psi} - \psi) \sqrt{2 \log \left( \frac{1}{n} A_0(\hat{\psi} - \psi, a, z) \right)}.$$

Utilizando el resultado de que  $\log(1+x) = x - \frac{1}{2}x^2 + \dots$  se sigue que

$$\begin{aligned} \log \left( \frac{1}{n} A_0(\hat{\psi} - \psi, a, z) \right) &= \ln \left( 1 + \frac{1}{2n} A_2(\hat{\psi} - \psi)^2 - \frac{1}{6n} A_3(\hat{\psi} - \psi)^3 + \frac{1}{24n} A_4(\hat{\psi} - \psi)^4 + \dots \right) \\ &= \frac{1}{2n} A_2(\hat{\psi} - \psi)^2 - \frac{1}{6n} A_3(\hat{\psi} - \psi)^3 + \left( \frac{1}{24n} A_4 - \frac{1}{8n^2} A_2^2 \right) (\hat{\psi} - \psi)^4 + \dots, \end{aligned}$$

usando las expresiones para la raíz de series de la Sección A.1.2 llegamos a que

$$r = A_2^{1/2}(\hat{\psi} - \psi) - \frac{1}{6} \frac{A_3}{\sqrt{A_2}} (\hat{\psi} - \psi)^2 - \frac{1}{72n A_2^{3/2}} (A_3^2 n - 3A_2 A_4 n + 9A_2^3) (\hat{\psi} - \psi)^3 + \dots,$$

aplicando la inversa de series obtenemos que

$$\hat{\psi} - \psi = A_2^{-1/2} r + \frac{1}{6} \frac{A_3}{A_2^2} r^2 + \frac{1}{72 A_2^{7/2}} (5A_3^2 - 3A_2 A_4 + 9n^{-1} A_2^3) r^3 + \dots \quad (2.69)$$

Para expandir  $u$  alrededor de  $\hat{\psi} = \psi$  se requiere expandir  $A_1(\hat{\psi} - \psi, a, z)$ . Observando que

$$\frac{\partial^k}{\partial \hat{\psi}^k} A_1(\hat{\psi} - \psi, a, z) = (-1)^k \sum_{i=1}^n z_i^{k+1} e^{a_i - tz_i + \psi z_i} = (-1)^k A_{k+1}(\hat{\psi} - \psi, a, z),$$

obtenemos que

$$A_1(\hat{\psi} - \psi, a, z) = -A_2(\hat{\psi} - \psi) + \frac{1}{2} A_3(\hat{\psi} - \psi)^2 - \frac{1}{6} A_4(\hat{\psi} - \psi)^3 + \frac{1}{24} A_5(\hat{\psi} - \psi)^4 + \dots$$

Mediante la expresión para el cociente de series (Sección A.1.1) tenemos que

$$\frac{A_1(\hat{\psi} - \psi, a, z)}{A_0(\hat{\psi} - \psi, a, z)} = -\frac{A_2}{n}(\hat{\psi} - \psi) + \frac{1}{2} \frac{A_3}{n}(\hat{\psi} - \psi)^2 - \frac{1}{6n^2} (A_4 n - 3A_2^2) (\hat{\psi} - \psi)^3 + \dots$$

Así, de la ecuación (2.67) se deduce que

$$u = -A_2^{1/2}(\hat{\psi} - \psi) + \frac{A_3}{2A_2^{1/2}}(\hat{\psi} - \psi)^2 - \frac{1}{6nA_2^{1/2}}(A_4 n - 3A_2^2)(\hat{\psi} - \psi)^3 + \dots,$$

substituyendo  $u$  por  $-u$  tenemos

$$u = A_2^{1/2}(\hat{\psi} - \psi) - \frac{A_3}{2A_2^{1/2}}(\hat{\psi} - \psi)^2 + \frac{1}{6nA_2^{1/2}}(A_4 n - 3A_2^2)(\hat{\psi} - \psi)^3 + \dots \quad (2.70)$$

Luego substituyendo (2.69) en (2.70) tenemos que

$$u = r - \frac{A_3}{3A_2^{3/2}} r^2 + \frac{1}{72A_2^3 n} (-7A_3^2 n + 9A_2 A_4 n - 27A_2^3) r^3 + \dots$$

y por lo tanto

$$\frac{r}{u} = 1 + \frac{1}{3} \frac{A_3}{A_2^{3/2}} r - \frac{1}{24A_2^3 n} (-5A_3^2 n + 3A_2 A_4 n - 9A_2^3) r^2 + \dots$$

### 2.5.5 Distribución sobre una esfera

La densidad de la distribución de Fisher está dada por

$$p(x, y; \theta, \phi, \kappa) = \frac{\kappa}{2 \sinh \kappa} \exp[\kappa \{\cos \theta \cos y + \sin \theta \sin y \cos(x - \phi)\}] \sin y, \quad (2.71)$$

donde  $0 \leq x < 2\pi$ ,  $0 \leq y < \pi$ ,  $0 \leq \phi < 2\pi$ ,  $0 \leq \theta < \pi$ ,  $\kappa \geq 0$ , ver Mardia (1972, Sec. 8.5.4) y Fisher (1953). En la Sección 2.5.5.1 se muestra en general que (2.71) integra 1.

Para una muestra  $(x_1, y_1), \dots, (x_n, y_n)$  de  $n$  observaciones sobre la esfera la función de log-verosimilitud es

$$l(\theta, \phi, \kappa) = n \ln \kappa - n \ln \sinh \kappa + \kappa \cos \theta \sum_{i=1}^n \cos y_i + \kappa \sin \theta \sum_{i=1}^n \sin y_i \cos(x_i - \phi),$$

los estimadores de máxima verosimilitud son

$$\cos \hat{\theta} = \frac{1}{R} \sum_{i=1}^n \cos y_i, \quad \hat{\phi} = \arctan \left( \frac{\sum \sin y_i \sin x_i}{\sum \sin y_i \cos x_i} \right) \quad \text{y} \quad \coth \hat{\kappa} - \frac{1}{\hat{\kappa}} = \frac{R}{n}$$

donde  $R^2 = (\sum_{i=1}^n \cos y_i)^2 + (\sum_{i=1}^n \sin y_i \cos(x_i - \hat{\phi}))^2$ .

Después de algunas manipulaciones obtenemos la expresión de la función de verosimilitud en términos de  $\hat{\theta}$ ,  $\hat{\phi}$  y  $\hat{\kappa}$  es

$$l(\theta, \phi, \kappa; \hat{\theta}, \hat{\phi}, \hat{\kappa}) = n \ln \kappa - n \ln \sinh \kappa + n \kappa \left( \coth \hat{\kappa} - \frac{1}{\hat{\kappa}} \right) \left( \cos \theta \cos \hat{\theta} + \sin \theta \sin \hat{\theta} \cos(\hat{\phi} - \phi) \right). \quad (2.72)$$

Los estimadores de máxima verosimilitud relativos son

$$\tan \hat{\theta}_{\phi, \kappa} = \tan \hat{\theta}_{\phi} = \tan(\hat{\theta}) \cos(\hat{\phi} - \phi), \quad \hat{\phi}_{\theta, \kappa} = \hat{\phi},$$

y

$$\coth \hat{\kappa}_{\theta, \phi} - \frac{1}{\hat{\kappa}_{\theta, \phi}} = \frac{R}{n} \left( \cos \theta \cos \hat{\theta} + \sin \theta \sin \hat{\theta} \cos(\hat{\phi} - \phi) \right).$$

Consideraremos a  $\theta$  como el parámetro de interés,  $\phi$  el de estorbo y  $\kappa$  conocida. Así la función de verosimilitud (2.72) se reduce a

$$\ell(\theta, \phi; \hat{\theta}, \hat{\phi}) = \kappa R \left( \cos \theta \cos \hat{\theta} + \sin \theta \sin \hat{\theta} \cos(\hat{\phi} - \phi) \right),$$

en virtud de que  $\hat{\phi}_\theta = \hat{\phi}$  la función de log-verosimilitud perfil de  $\theta$  se expresa por

$$\ell_p(\theta; \hat{\theta}, \hat{\phi}) = \ell(\theta, \hat{\phi}_\theta; \hat{\theta}, \hat{\phi}) = \ell(\theta, \hat{\phi}; \hat{\theta}, \hat{\phi}) = \kappa R \cos(\hat{\theta} - \theta).$$

La verosimilitud dirigida se expresa como

$$r = \sqrt{2\kappa R} \operatorname{sgn}(\hat{\theta} - \theta) \sqrt{1 - \cos(\hat{\theta} - \theta)}. \quad (2.73)$$

Para este modelo se tiene una expresión simple de los intervalos de verosimilitud, a saber

$$\hat{\theta} - \theta = \arccos\left(1 - \frac{r^2}{2\kappa R}\right). \quad (2.74)$$

Puesto que  $0 \leq \theta \leq \pi$  los extremos del intervalo de verosimilitud son

$$\max\left(0, \hat{\theta} - \arccos\left(1 - \frac{r^2}{2\kappa R}\right)\right), \quad \min\left(\pi, \hat{\theta} + \arccos\left(1 - \frac{r^2}{2\kappa R}\right)\right).$$

Las cantidades necesarias para obtener la estadística  $u$  de (1.18) son

$$\bar{\ell}_{p/\hat{\theta}} = -k \sin(\hat{\theta} - \theta), \quad \hat{j}_p = \kappa R, \quad \Rightarrow \dot{u} = -\sqrt{\kappa R} \sin(\hat{\theta} - \theta), \quad \text{además } C = \sqrt{\frac{\sin \theta}{\sin \hat{\theta}}},$$

la función  $C(\theta)$  siempre es positiva, ya que  $0 < \theta, \hat{\theta} < \pi$ . Así, pues

$$u = \sqrt{\frac{\sin \theta}{\sin \hat{\theta}}} \sqrt{\kappa R} \sin(\hat{\theta} - \theta). \quad (2.75)$$

Para este modelo podemos expresar  $u$  en término de  $r$  de manera exacta. Simplemente se debe substituir  $\hat{\theta}$  de (2.74) en (2.75).

De manera sencilla se llega a

$$\frac{u}{r} = \frac{1}{\sqrt{2}} \sqrt{\frac{\sin \theta}{\sin \hat{\theta}}} \sqrt{1 + \cos(\hat{\theta} - \theta)},$$

donde  $\operatorname{sgn}(\hat{\theta} - \theta)$  desaparece porque se usó la igualdad  $\sin(x) = \operatorname{sgn}(x) \sqrt{1 - \cos^2 x}$ , para  $-\pi \leq x \leq \pi$ .

La expansión de  $\frac{r}{u}$  alrededor de  $\theta$ , se obtiene después de muchos cálculos, y es

$$\frac{r}{u} = 1 + \frac{1 \cot \theta}{2 \sqrt{\kappa R}} r - \frac{1}{8 \sin^2 \theta \kappa R} r^2 + \frac{1}{16 (\kappa R)^{3/2}} \frac{2 - \cos^2 \theta}{\sin^2 \theta} r^3 + \dots,$$

de donde obtenemos que el factor de Bartlett es

$$b(\theta, \kappa R) = -\frac{1}{4} \frac{1}{\sin^2 \theta \kappa R},$$

el cual tiene un máximo en  $\theta = \frac{\pi}{2}$  y explota a  $-\infty$  en  $\theta = 0$  y  $\theta = \pi$ .

En este ejemplo la aproximación del factor de Bartlett es no acotada, sin embargo podemos evaluar analíticamente la probabilidad de cobertura mínima, con ayuda de las fórmulas de la sección 2.5.5.2. La aproximación de la probabilidad de cobertura para  $\theta = \frac{\pi}{2}$  es precisa con cualquiera de las ecuaciones que se utilicen para aproximarla (ecuaciones 2.3, 3.40, 2.4 y 3.36).

Tenemos tres mínimos locales de la probabilidad de cobertura de los intervalos de verosimilitud. Los tres mínimos se alcanzan en  $\theta = 0, \frac{\pi}{2}, \pi$ . El valor numérico de la probabilidad de cobertura en  $\theta = \frac{\pi}{2}$  se calcula numéricamente mediante la ecuación (2.77). Para  $\theta = 0$  ó  $\theta = \pi$  se calcula analíticamente mediante (2.77) y en este caso obtenemos que

$$\Pr(R_p \geq c) = \begin{cases} 1, & \text{si } 0 \leq c \leq e^{-2\kappa R}, \\ \frac{1-c}{1-e^{-2\kappa R}}, & \text{si } e^{-2\kappa R} < c \leq 1. \end{cases} \quad (2.76)$$

Dado que  $1 - e^{-2R\kappa}$  es muy cercano a 1 para  $\kappa R \geq 2$  el resultado expresado en la ecuación (2.76) es similar al que se obtiene en el caso de variables aleatorias uniformes en  $(0, \theta)$  donde  $\Pr(R(\theta) \geq c) = 1 - c$ . Las relaciones (2.76) y (2.73) dependen de  $\kappa$  y  $R$  sólo a través del producto  $\kappa R$ , por lo que en los ejemplos numéricos que daremos sólo especificaremos el producto  $\kappa R$ .

La Figura 2.3 muestra la precisión obtenida con  $r^*$  al medir la probabilidad de cobertura del intervalo con nivel de verosimilitud .147 y  $\kappa R = 7$ , lo mismo se ve en la Figura 2.4 donde  $\kappa = 35.092$ ,  $R = 8.77203$  y  $\kappa R = 307.83$ , lo cual se obtiene de los datos de los nueve especímenes de Fisher (1953).

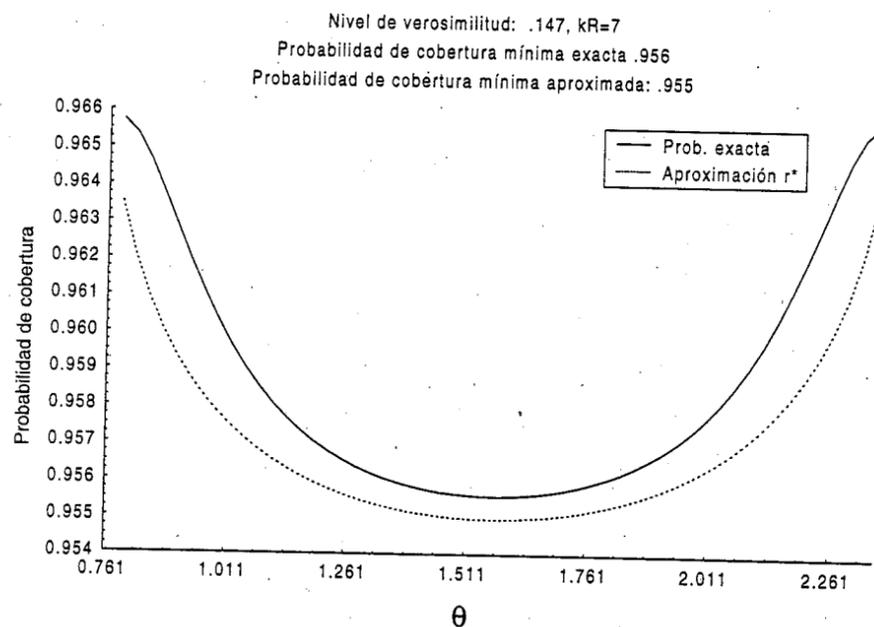


Figura 2.3. Probabilidades de cobertura exacta y aproximada del intervalo de verosimilitud con nivel de verosimilitud .147 y  $\kappa R = 7$ . Línea continua: probabilidad exacta. Línea punteada: probabilidad de cobertura aproximada con  $r^*$ .

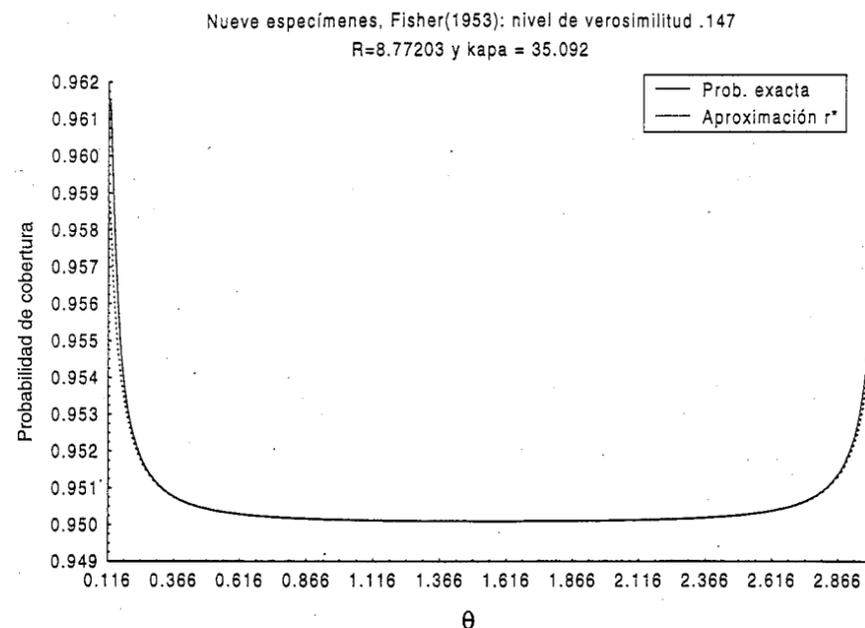


Figura 2.4. Probabilidades de cobertura exacta y aproximada del intervalo de verosimilitud con nivel de verosimilitud .147 y  $\kappa = 35$ ,  $R = 307.8$ . Línea continua: probabilidad exacta. Línea punteada: probabilidad de cobertura aproximada con  $r^*$ .

La Figura 2.5 muestra los valores de la probabilidad de cobertura exacta de intervalos con nivel de verosimilitud .147. Se utiliza la fórmula (2.77) con los valores de  $\kappa$  más plausibles de acuerdo con los datos de los nueve especímenes en Fisher (1953).

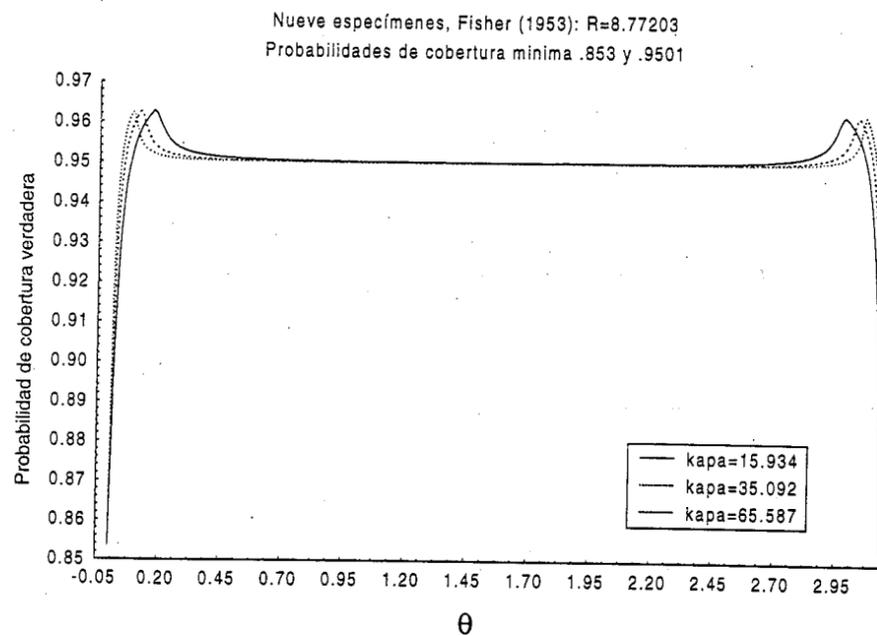


Figura 2.5. Probabilidades de cobertura exacta del intervalo de verosimilitud con nivel de verosimilitud .147 y  $\kappa$  evaluada en los extremos del intervalo de verosimilitud .147 para  $\kappa$ .

2.5.5.1 La constante de integración de la distribución de Fisher

$$\begin{aligned} \int_0^\pi \int_0^{2\pi} p(x, y; \theta, \phi, \kappa) dx dy &= \\ &= \frac{\kappa}{2 \sinh \kappa} \int_0^\pi \int_0^{2\pi} \exp[\kappa \{\cos \theta \cos y + \sin \theta \sin y \cos(x - \phi)\}] \sin y dx dy \\ &= \frac{\kappa}{2 \sinh \kappa} \int_0^\pi \left( \int_0^{2\pi} \exp[\kappa \sin \theta \sin y \cos(x - \phi)] dx \right) \exp[\kappa \cos \theta \cos y] \sin y dy \\ &= \frac{\kappa}{2 \sinh \kappa} \int_0^\pi I_0(\kappa \sin \theta \sin y) \exp[\kappa \cos \theta \cos y] \sin y dy \\ &= \frac{1}{2} \frac{\kappa}{\sinh \kappa} \int_{-1}^1 I_0[\kappa (\sin \theta) \sqrt{1-t^2}] e^{\kappa (\cos \theta) t} dt \\ &= \frac{\kappa}{2 \sinh \kappa} \frac{2 \sinh(\sqrt{\kappa^2 \cos^2 \theta + \kappa^2 \sin^2 \theta})}{\sqrt{\kappa^2 \cos^2 \theta + \kappa^2 \sin^2 \theta}} \\ &= 1. \end{aligned}$$

Los cálculos anteriores se obtienen con ayuda de las fórmulas

$$I_0(x) = \frac{1}{2\pi} \int_0^{2\pi} e^{x \cos t} dt = \frac{1}{2\pi} \int_{-\phi}^{2\pi-\phi} e^{x \cos t} dt, \quad \forall \phi$$

y

$$\int_{-1}^1 e^{-ax} I_0(b\sqrt{1-x^2}) dx = 2(a^2 + b^2)^{-1/2} \sinh(\sqrt{a^2 + b^2}) \quad a > 0, \quad b > 0,$$

(ver Gradshteyn & Ryzhik 1994, ecuaciones 8.431-3 y 6.616-5).

2.5.5.2 Densidad de  $\hat{\theta}|R$

En virtud de que la distribución de Fisher pertenece a la familia de transformación, mediante la fórmula- $p^*$  (ver Sección 3.2) tenemos que la densidad exacta de  $(\hat{\theta}, \hat{\phi})$  está dada por

$$\begin{aligned} p(\hat{\theta}, \hat{\phi}; \theta, \phi | R) &= p^*(\hat{\theta}, \hat{\phi}; \theta, \phi | R) \\ &= \frac{\kappa R}{4\pi \sinh \kappa R} \sin \hat{\theta} \exp \left\{ \kappa R \left[ \cos \theta \cos \hat{\theta} + \sin \theta \sin \hat{\theta} \cos(\hat{\phi} - \phi) \right] \right\}, \end{aligned}$$

si integramos esta densidad con respecto a  $\hat{\phi}$ , obtenemos la densidad marginal de  $\hat{\theta}$ . Entonces

$$\begin{aligned} p(\hat{\theta}; \theta | R) &= \frac{\kappa R}{4\pi \sinh \kappa R} \sin \hat{\theta} \int_0^{2\pi} e^{\kappa R (\cos \theta \cos \hat{\theta} + \sin \theta \sin \hat{\theta} \cos(u-\phi))} du \\ &= \frac{\kappa R}{4\pi \sinh \kappa R} \sin \hat{\theta} e^{\kappa R \cos \theta \cos \hat{\theta}} \int_{-\phi}^{2\pi-\phi} e^{\kappa R \sin \theta \sin \hat{\theta} \cos v} dv \\ &= \frac{\kappa R}{2 \sinh \kappa R} \sin \hat{\theta} I_0(\kappa R \sin \theta \sin \hat{\theta}) e^{\kappa R \cos \theta \cos \hat{\theta}}, \end{aligned} \quad (2.77)$$

donde se usó fórmula  $I_0(x) = \frac{1}{2\pi} \int_0^{2\pi} e^{x \cos t} dt = \frac{1}{2\pi} \int_{-\phi}^{2\pi-\phi} e^{x \cos t} dt, \quad \forall \phi$  (ver Gradshteyn & Ryzhik 1994, ecuación 8.431-3).

Los valores de esta densidad en los extremos de  $\Omega_\theta = (0, \pi)$  son: Para  $\theta = 0$ ,

$$p(\hat{\theta}; 0 | R) = \frac{\kappa R}{2 \sinh \kappa R} \sin \hat{\theta} e^{\kappa R \cos \hat{\theta}},$$

y para  $\theta = \pi$

$$p(\hat{\theta}; \pi | R) = \frac{\kappa R}{2 \sinh \kappa R} \sin \hat{\theta} e^{-\kappa R \cos \hat{\theta}}.$$

## 2.6 Discusión

Hemos visto que para obtener una aproximación analítica con error de orden  $O(n^{-2})$  de la probabilidad de cobertura de un intervalo de verosimilitud podemos proceder por lo menos de cinco maneras distintas. Sin embargo, para calibrar analíticamente un intervalo de verosimilitud con error de orden  $O(n^{-2})$ , sólo tenemos un procedimiento: el de los factores de Bartlett. Esto nos da una justificación o interpretación natural para los factores de Bartlett, dado que originalmente fueron introducidos de manera *ad hoc* para "corregir" la distribución de la estadística log razón de verosimilitudes. En esta tesis ilustramos que los factores de Bartlett emergen de manera natural al resolver de manera analítica el problema de calibración de intervalos de verosimilitud.

Tenemos por lo menos cinco procedimientos para evaluar la probabilidad de cobertura de un intervalo de verosimilitud. Cuatro de éstas son aproximaciones analíticas y están ligadas entre sí. Si se logra evaluar cualquiera de ellas, inmediatamente se puede evaluar cualquiera de las otras tres. Además, si resolvemos el problema de calibración analíticamente se pueden obtener expresiones analíticas de los extremos de los intervalos de verosimilitud observado y muestral. Sólo es cuestión de almacenar todas las derivadas conforme se vayan evaluando. Esto no es complicado, porque para calcular la derivada  $n$ -ésima es necesario conocer la derivada  $(n-1)$ -ésima.

La importancia de la verosimilitud dirigida,  $r$ , radica en que para obtener analíticamente

los valores de  $R(\theta) \geq c$  o  $W \leq -2 \ln c$  es más simple resolver todo para la verosimilitud dirigida. Además en esta tesis siempre que se utiliza el factor de Bartlett en presencia de parámetros de estorbo nos basamos en la expansión analítica de  $u/r$ . La verosimilitud dirigida modificada,  $r^*$ , juega un papel importante porque con base en ella obtenemos expresiones aproximadas del factor de Bartlett. En general el procedimiento que utilizamos en todos los ejemplos para obtener el factor de Bartlett es por medio de  $r^*$ . Si el procedimiento es numérico utilizamos  $r^*$  evaluada en las raíces de  $r$ . Si es analítico, estudiamos la simplicidad de  $u$  y  $r$  y decidimos si es factible expandir estas cantidades usando aritmética de series de potencias. En general no es necesario calcular el factor de Bartlett para evaluar (2.3) como lo ilustran los ejemplos 2.4.4, 2.4.5 y 2.5.5.

Se descartaron los procedimientos  $O(n^{-1})$  por su carencia de información, no por que puedan ser poco precisos. Utilizamos procedimientos de altos órdenes con el fin de poder evaluar la probabilidad de cobertura de un intervalo de verosimilitud con mayor precisión para así conocer si ésta depende de  $\theta$ . Cabe mencionar que si un procedimiento es preciso, con error de orden  $O(n^{-2})$ , no es garantía de que funcione bien en todos los casos. Observe que en el modelo gaussiana inversa con parámetro de interés  $\lambda$  (Ejemplo 2.5.2), se requiere de muestras de tamaño moderadamente grande ( $n > 20$ ) para que la aproximación (2.3) (o 2.4) trabaje bien.

Cuando la probabilidad de cobertura de un intervalo de verosimilitud depende de  $\theta$  y suponemos completo desconocimiento sobre  $\theta$ , el valor más objetivo de la incertidumbre del intervalo es la probabilidad de cobertura mínima del intervalo sobre todo el espacio paramétrico. Se puede obtener también la probabilidad de cobertura máxima, pero no es nuestro interés primordial calcularla, sobre todo si se requiere simulación. También puede

sucedir que tengamos preferencias por alguna región de valores para  $\theta$ . En esta situación es conveniente calcular la probabilidad de cobertura mínima dentro de dicha región.

El factor de Bartlett puede diverger a  $+\infty$  o  $-\infty$  (ver Ejemplos 2.5.2, 2.5.3 y 2.5.5). El ejemplo de la esfera es un ejemplo donde el factor de Bartlett diverge a  $-\infty$  y sin embargo se puede calcular analíticamente la cota mínima de la probabilidad de cobertura. En otros ejemplos, como los casos especiales de los modelos gama y gaussiana inversa, se debe hacer por simulación porque el procedimiento analítico o numérico es complicado incluso para  $n = 1$ . Si el factor de Bartlett diverge la aproximación falla, por lo que debemos usar otro procedimiento para calibrar regiones de verosimilitud, por ejemplo simulación. Para lo cual nos podemos ayudar con el factor de Bartlett aunque diverja. Es decir si se obtiene el factor de Bartlett y éste es no acotado todavía nos sirve porque nos ayuda en el proceso de minimizar la probabilidad de cobertura.

Si  $\dim(\psi) = d > 1$  hay dos maneras de proceder para medir o calibrar la probabilidad de cobertura de una región de verosimilitud. La primera es trabajar con  $\psi = (\psi_1, \dots, \psi_d)$  de manera conjunta, para lo cual se requiere obtener el factor de Bartlett para  $\dim \psi > 1$ . La segunda consiste en trabajar separadamente con cada  $\psi_i$ ,  $i = 1, \dots, d$  y aplicar los resultados para cada  $\psi_i$  considerando el resto de los parámetros como de estorbo.

Una de las condiciones de regularidad, tanto para procedimientos con error de orden  $O(n^{-1})$ , como para los de error de orden  $O(n^{-2})$  es que  $\theta \in \text{int } \Omega$ . En general si  $\theta$  esta cerca de la frontera de  $\Omega$  se necesita que  $n$  sea muy grande para que la aproximación dé buenos resultados. Cabe señalar que si suponemos completo desconocimiento sobre  $\theta$  no hay porque suponer que  $\theta$  está lejos de la frontera de  $\Omega$  por lo que puede ser pésima la aproximación y si el procedimiento es  $O(n^{-1})$  (ecuaciones 1.6 y 2.1) no lo podremos saber. En cambio si el

procedimiento es  $O(n^{-2})$  (ecuaciones 2.3, 2.4) también es obligatoria que  $\theta \in \text{int } \Omega$  y también se necesita  $n$  grande. Sin embargo en este caso por lo menos tenemos noción de que puede fallar y dónde.

Cuando las condiciones de regularidad fallan lo más conveniente es estudiar cada modelo de manera individual. Por ejemplo si tenemos observaciones de la distribución uniforme en  $(0, \theta)$  se puede probar que  $\Pr(R(\theta) \geq c) = 1 - c$ . Por lo tanto el problema de calibración es muy simple, tómesese  $c = \alpha$  para un intervalos de verosimilitud con  $100(1 - \alpha)\%$  de nivel confianza (ver también Ejemplo 2.5.5).

Los procedimientos que aquí utilizamos vienen de la teoría de altos órdenes. Reid & Fraser (2000) hablan de los costos y beneficios de estos procedimientos y dan algunas expresiones para la cantidad  $u$  que utiliza  $r^*$  y la aproximación de Lugannani-Rice.

Para calibrar intervalos de verosimilitud en modelos con censura se puede proceder de manera similar a esta tesis mediante el uso del factor de Bartlett para modelos con censura (Jensen 1987).

En general recomendamos simulación. En algunos modelos no regulares se puede medir y calibrar numéricamente o analíticamente sin muchas complicaciones (ver Kalbfleisch 1985, sec. 11.2).

## Capítulo 3

### Justificación de resultados

En este capítulo se presentan las demostraciones de los resultados presentados en los Capítulos 1 y 2, las demostraciones se basan principalmente en la fórmula- $p^*$  y expansiones asintóticas, alrededor del parámetro de interés  $\psi$ , de cantidades relacionadas con la función de log-verosimilitud. Una serie asintótica, en general es una serie divergente, que al parar la suma en un número dado de términos, el residuo es de orden menor que el último término que se haya incluido. En el Apéndice A se muestran más detalles sobre series asintóticas y se presentan expansiones alrededor de  $\hat{\theta}$  con el fin de obtener expresiones analíticas aproximadas para intervalos de verosimilitud e intervalos de confianza obtenidos con  $r^*$ .

Se requiere que  $l_{j,k}$  exista para  $j+k \leq 6$ ,  $j, k \geq 0$ , aunque en relación con las aplicaciones sólo se necesita mostrar explícitamente derivadas de la forma  $l_{j,k}$ , para  $j+k \leq 4$ ,  $j, k \geq 0$ .

**Advertencia:** Las condiciones de regularidad que se presentan en este capítulo son las necesarias para que se cumplan los resultados relacionados con  $r^*$ . Sin embargo, se puede medir exacta o aproximadamente la cobertura de intervalos de verosimilitud en algunos ejemplos

donde fallen varias de estas condiciones como en: Cauchy  $(\mu, \sigma)$ ,  $e^{x-\theta}$ ,  $\theta > x$ , Uniforme  $(0, \theta)$ , etc. Por ejemplo para que se cumpla la fórmula- $p^*$  en los modelos de transformación es suficiente con que exista  $j$  y que el máximo verosímil sea único.

### 3.1 Condiciones de regularidad

El modelo estadístico paramétrico regular  $\mathcal{M} = \{p_Y(y; \theta), \theta \in \Omega \subset \mathbf{R}^d\}$ ,  $d \geq 1$  que consideramos en esta tesis satisface lo siguiente:

El soporte de  $Y$  no depende de  $\theta$ ,  $\Omega$  es un conjunto abierto,  $\hat{\theta} \in \Omega$  es único, consistente y cumple las siguientes condiciones

$$\nabla \ell(\theta; y)|_{\theta=\hat{\theta}} = 0 \quad \text{y} \quad |\nabla_k^T \nabla_k \ell(\theta; y)|_{\theta=\hat{\theta}} < 0, \quad k = 1, \dots, n. \quad (3.1)$$

donde  $\nabla = (\partial_1, \dots, \partial_d)^T$  y  $\nabla_k = (\partial_1, \dots, \partial_k)$ . (ver Marsden & Hoffman 1998, p. 365). Si  $d = 1$

$$\frac{\partial}{\partial \psi} \ell(\theta; y) \Big|_{\psi=\hat{\psi}} = 0 \quad \text{y} \quad \frac{\partial^2}{\partial \theta^2} \ell(\theta; y) \Big|_{\psi=\hat{\psi}} < 0. \quad (3.2)$$

Sprott (2000, Sec. 2.6) presenta varios ejemplos del cálculo numérico de  $\hat{\theta}$  y Kolassa (1997, Sec. 4.9) contiene sugerencias numéricas para obtener el  $\hat{\theta}$  en modelos donde no funciona bien Newton-Rapson.

Si el modelo es exponencial todo se resume a que el modelo es un modelo exponencial regular, el cual es un modelo exponencial completo  $(k, k)$  perfecto.

Cuando se tiene suficiencia observe que la función de log-verosimilitud es función únicamente de una estadística suficiente  $s$  y el parámetro  $\theta$ , excepto por algún término que no dependa de  $\theta$ . Si la dimensión de  $s$  es la misma que la de  $\theta$  usualmente el estimador de

máxima verosimilitud es función uno-a-uno de  $s$  y entonces  $\hat{\theta}$  es minimal suficiente si, y sólo si,  $s$  es minimal suficiente. En consecuencia podemos tomar la función de log-verosimilitud de la forma  $\ell(\theta; \hat{\theta})$ , esta será la misma como si los datos observados consistieran solamente de  $s$  o  $\hat{\theta}$ . Si  $s = (t, a)$  donde la dimensión de  $t$  es la de  $\theta$  y  $a$  es auxiliar, entonces generalmente se puede escribir la log-verosimilitud como  $\ell(\theta; \hat{\theta}, a)$ .

El estimador de máxima verosimilitud de  $(\psi, \chi)$  se obtiene por invariancia  $(\hat{\psi}, \hat{\chi}) = (\psi(\hat{\theta}), \chi(\hat{\theta}))$ . Considere la función de log-verosimilitud expresada como  $\ell(\psi, \chi; \hat{\psi}, \hat{\chi}, a)$ . El estimador de máxima verosimilitud relativo  $\hat{\chi}_\psi$  se obtiene como solución de

$$\ell_\chi(\psi, \hat{\chi}_\psi; \hat{\psi}, \hat{\chi}, a) = 0. \quad (3.3)$$

La función de log-verosimilitud es de orden  $O(n)$ , para valores fijos de  $(\hat{\psi}, \hat{\chi})$  la función de log-verosimilitud perfil permite expansión de Taylor, en una vecindad de  $\psi_0$ , hasta el quinto orden y el valor absoluto del término residual uniformemente acotado. Además para valores fijos de  $(\psi, \hat{\chi}_\psi, a)$  la log-verosimilitud perfil también permite expansión de Taylor, en una vecindad de  $\psi$ , hasta el quinto orden y el valor absoluto del término residual uniformemente acotado. Vea Sección A.3 para detalles de la expansión alrededor de  $\psi$ .

### 3.2 Parámetro $\theta$ escalar

Utilizaremos la aproximación a la distribución condicional del estimador de máxima verosimilitud propuesta por Barndorff-Nielsen (1983)

$$p^*(\hat{\theta}; \theta|a) = c_0 |j|^{1/2} e^{\ell-\hat{\ell}}, \quad (3.4)$$

donde  $c_0 = c_0(\theta, a)$  es la constante de integración. La fórmula- $p^*$  es una aproximación de orden  $O(n^{-3/2})$  en el siguiente sentido

$$p(\hat{\theta}; \theta|a) = p^*(\hat{\theta}; \theta|a) [1 + O(n^{-3/2})]. \quad (3.5)$$

Debido a que requerimos llegar a expresiones de orden  $O(n^{-2})$  expresaremos el término de error de la forma

$$p(\hat{\theta}; \theta|a) = p^*(\hat{\theta}; \theta|a) \left[1 + \frac{1}{n} c_1(\theta) (\hat{\theta} - \theta) + O(n^{-2})\right], \quad (3.6)$$

donde  $c_1(\theta)$  es de orden  $O(1)$ .

### 3.2.1 Densidad de la verosimilitud dirigida

Para valores fijos de  $\theta$  y  $a$ , tenemos que

$$\frac{\partial}{\partial \theta} r = \frac{\hat{\ell}_{;1} - \ell_{;1}}{r},$$

entonces, suponiendo que  $r$  es función uno-a-uno de  $\hat{\theta}$ , se sigue de (3.4) que una aproximación a la densidad de la verosimilitud dirigida está dada por

$$p^*(r; \theta|a) = c_0(\theta, a) \frac{r}{u} e^{-\frac{1}{2}r^2}, \quad (3.7)$$

donde  $c_0(\theta, a)$  es la constante de integración de la fórmula- $p^*$  y  $u = \hat{j}^{-1/2} (\hat{\ell}_{;1} - \ell_{;1})$ .

En virtud de que

$$\frac{r}{u} = 1 - \bar{t}_3 + \frac{1}{2}br^2 + \dots, \quad (3.8)$$

donde  $\bar{t}_3 = O(n^{-1/2})$  y  $b = O(n^{-1})$  se definen en (A.10), se obtienen las siguientes relaciones

$$p(r; \theta) = \phi(r) [1 - \bar{t}_3 r + O(n^{-1})],$$

luego,

$$\Pr(r \leq x) = \Phi(x) [1 + O(n^{-1})] + \phi(x) \bar{t}_3, \quad (3.9)$$

y para  $x > 0$

$$\begin{aligned} \Pr(-x \leq r \leq x) &= [\Phi(x) - \Phi(-x)] [1 + O(n^{-1})] \\ &= [2\Phi(x) - 1] [1 + O(n^{-1})], \end{aligned} \quad (3.10)$$

La relación (3.10) nos dice que  $r^2 \sim \chi_1^2$  con error relativo de orden  $O(n^{-1})$ .

### 3.2.2 Probabilidad de cobertura de un intervalo de verosimilitud

El objetivo de esta sección es demostrar el Teorema 3.1, el cual es uno de los resultados principales de esta tesis. Este Teorema se obtiene con ayuda de los Lemas 3.1 y 3.2 presentados a continuación.

**Lema 3.1.** La función de densidad de  $r^*$  es de la forma

$$p(r^*; \theta) = \phi(r^*) [1 + Hr^* + O(n^{-2})], \quad (3.11)$$

donde  $H$  es  $O(n^{-3/2})$ .

*Demostración.* Mediante el cambio de variable  $r \mapsto r^*$ , la densidad de  $r^*$  debe ser

$$p(r^*; \theta) = c_0(\theta, a) \frac{r(r^*)}{u(r^*)} e^{-\frac{1}{2}r(r^*)^2} \frac{\partial r}{\partial r^*}. \quad (3.12)$$

Necesitamos la expresión explícita de  $r^*$  en función de  $r$ , para valores de  $a$  y  $\theta$  fijos, la cual está dada por

$$r^* = r + t_3 - \frac{1}{2}(b - \bar{t}_3^2)r + \bar{d}_3 + \dots \quad (3.13)$$

donde,  $\bar{d}_k = O(n^{-k/2})$  y  $\bar{d}_2 = -\frac{1}{2}(b - \bar{t}_3^2) = O(n^{-1})$ , (ver ecuación 3.13). Por medio de la técnica de inversión de series de la Sección A.1.4, tenemos que

$$r = (1 - d_2)(r^* - \bar{t}_3) - d_3(r^* - \bar{t}_3)^2 + O(n^{-2}). \quad (3.14)$$

Para evaluar (3.12) es conveniente observar que

$$\frac{r}{u} e^{-\frac{1}{2}r^2} = e^{-\frac{1}{2}r^2 - \log \frac{r}{u}} = e^{\frac{1}{2}r^2 - r^*r} = e^{-\frac{1}{2}(r^*)^2 + \frac{1}{2}(r^* - r)^2} = \phi(r^*) e^{\frac{1}{2}(r^* - r)^2}, \quad (3.15)$$

además de (3.14), junto con la expansión  $e^x = 1 + x + \frac{1}{2}x^2 + \dots$ , se obtiene que

$$e^{\frac{1}{2}(r^* - r)^2} = 1 + \frac{1}{2}\bar{t}_3^2 + \bar{t}_3 d_2 (r^* - \bar{t}_3) + O(n^{-2}).$$

derivando (3.14) con respecto a  $r^*$

$$\frac{\partial r}{\partial r^*} = 1 - d_2 - 2d_3(r^* - \bar{t}_3) + O(n^{-2}),$$

entonces

$$e^{\frac{1}{2}(r^* - r)^2} \frac{\partial r}{\partial r^*} = 1 + \frac{1}{2}\bar{t}_3^2 - d_2 - H(r^* - \bar{t}_3) + O(n^{-2}), \quad (3.16)$$

sustituyendo  $d_2 = -\frac{1}{2}(b - \bar{t}_3^2)$ , se sigue que

$$e^{\frac{1}{2}(r^* - r)^2} \frac{\partial r}{\partial r^*} = 1 + \frac{1}{2}b - H(r^* - \bar{t}_3) + O(n^{-2}),$$

Luego, de (3.15) y (3.16), obtenemos que (3.12) se reduce a

$$\begin{aligned} \Pr(r^*; \theta) &= c_0 \phi(r^*) e^{\frac{1}{2}(r^* - r)^2} \frac{\partial r}{\partial r^*} \\ &= c_0 \phi(r^*) [1 + \frac{1}{2}b - H(r^* - \bar{t}_3) + O(n^{-2})] \\ &= c_0 \phi(r^*) [1 + \frac{1}{2}b - Hr^* + O(n^{-2})]. \end{aligned} \quad (3.17)$$

dado que  $\int \Pr(r^*; \theta) dr^* = 1$  se sigue que

$$c_0 = 1 - \frac{1}{2}b + O(n^{-2}), \quad (3.18)$$

del producto de (3.17) con (3.18) se obtiene (3.11). ■

**Lema 3.2:** Para valores fijos de  $a$  y  $\theta$ , y considerando  $r^* = r^*(\theta, r)$ , se sigue para  $r_0 > 0$  que

$$\int_{r^*(-r_0)}^{r^*(r_0)} s \phi(s) ds = O(n^{-1/2}).$$

*Demostración:* De (3.13) se obtiene que  $r^*(r_0) = r_0 + \bar{t}_3 + O(n^{-1})$  y  $r^*(-r_0) = -r_0 + \bar{t}_3 + O(n^{-1})$ . Entonces

$$\int_{-r_0 + \bar{t}_3 + O(n^{-1})}^{r_0 + \bar{t}_3 + O(n^{-1})} x \phi(x) dx = -\phi(x) \Big|_{-r_0 + \bar{t}_3 + O(n^{-1})}^{r_0 + \bar{t}_3 + O(n^{-1})} = O(n^{-1/2}),$$

el resultado se sigue, de la relación  $\phi(x+a) - \phi(-x+a) = -2ax\phi(x) + O(a^3)$ .

De los Lemas 3.1 y 3.2 se desprende uno de los resultados mas importantes de esta tesis. El Teorema 3.1 presenta la manera de evaluar la probabilidad de cobertura de un intervalo de verosimilitud, con error de orden  $O(n^{-2})$  en muestras repetidas ordinarias.

**Resultado 3.1.** Para  $r_0 > 0$  se cumple que

$$\begin{aligned} \Pr(R \geq c) &= \Pr(r^*(-v_c) \leq r^* \leq r^*(v_c)) \\ &= \Phi(r^*(v_c)) - \Phi(r^*(-v_c)) + O(n^{-2}). \end{aligned} \quad (3.19)$$

*Demostración.* Sea  $v_c = \sqrt{-2 \ln c}$ ,  $0 < c < 1$ . Entonces

$$\begin{aligned} \Pr(R \geq c) &= \Pr(-v_c \leq r \leq v_c) \\ &= \Pr(r^*(-v_c) \leq r^* \leq r^*(v_c)) \\ &= \int_{r^*(-v_c)}^{r^*(v_c)} \phi(s) [1 + Hs + O(n^{-2})] ds \quad (\text{por Lema 3.1}) \\ &= \Phi(r^*(v_c)) - \Phi(r^*(-v_c)) + H \int_{r^*(-v_c)}^{r^*(v_c)} s \phi(s) ds + O(n^{-2}) \\ &= \Phi(r^*(v_c)) - \Phi(r^*(-v_c)) + HO(n^{-1/2}) + O(n^{-2}) \quad (\text{por Lema 3.2}) \\ &= \Phi(r^*(v_c)) - \Phi(r^*(-v_c)) + O(n^{-2}). \end{aligned}$$

En la Sección 3.3.3. se obtiene una aproximación de (3.19), la cual es

$$\Phi(r^*(v_c)) - \Phi(r^*(-v_c)) = \Phi(v_c) - \Phi(-v_c) - 2bv_c \phi(v_c) + O(n^{-2}). \quad (3.20)$$

### Con parámetros de estorbo

En este caso se tiene que  $\Pr(r^* \leq x) = \Phi(x) [1 + M]$ , donde  $M = O(n^{-3/2})$ . Pero no se conoce la expresión analítica de  $M$ . Si el término de error fuera de la forma  $Hr^* + O(n^{-2})$ , donde  $H$  es  $O(n^{-3/2})$ , se siguen los mismos resultados que en el caso escalar. Sin embargo, puesto que no conocemos la forma de  $M$  se sigue que

$$\begin{aligned} \Pr(R_p(\psi) \geq c) &= \Pr(r^*(-r) \leq r^* \leq r^*(r)) \\ &= [\Phi(r^*(r)) - \Phi(r^*(-r))] [1 + O(n^{-3/2})]. \end{aligned}$$

Las demostración es similar que la de la sección anterior, utilizando las sugerencias para expandir  $r^*$  alrededor de  $r$ .

### 3.2.3 Resultados auxiliares

**Lema 3.3:** Sean  $z > 0$ ,  $k > 0$  y  $a = O(n^{-k})$  números reales. Entonces

$$\Phi(z+a) = \Phi(z) + a\phi(z) - \frac{1}{2}a^2 z \phi(z) + O(n^{-3k})$$

*Demostración:*

La serie de Taylor para  $\Phi(x+a)$  alrededor de  $a=0$  es

$$\Phi(x+a) = \Phi(x) + a\phi(x) - \frac{1}{2}x\phi(x)a^2 + \frac{1}{6}((x+t)^2 - 1)\phi(x+t)t^3, \quad t \in (-|a|, |a|). \quad \blacksquare$$

Antes de pasar al Lema 3.4, es necesario obtener una expresión asintótica simple de  $z_{1-\alpha/2}$ . Con este fin usaremos la expresión de  $\Phi(x)$  dada en Abramowitz & Stegun (1965),

$$\Phi(x) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{n! 2^n (2n+1)} \quad (3.21)$$

$$= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \left( x - \frac{1}{6}x^3 + \frac{1}{40}x^5 - \frac{1}{336}x^7 + \dots \right) \quad (3.22)$$

El cuantil  $1 - \alpha/2$  satisface que

$$2\Phi(z_{1-\alpha/2}) - 1 = 1 - \alpha.$$

Entonces, la expresión para el cuantil  $1 - \alpha/2$  se obtiene al resolver

$$1 - \alpha = \frac{\sqrt{2}}{\sqrt{\pi}} \left( z_{1-\alpha/2} - \frac{1}{6}z_{1-\alpha/2}^3 + \frac{1}{40}z_{1-\alpha/2}^5 - \frac{1}{336}z_{1-\alpha/2}^7 + \dots \right) \quad (3.23)$$

mediante las técnicas de la Sección A.1.4. Así, con  $C = \sqrt{\frac{\pi}{2}}(1 - \alpha)$ , se obtiene la expresión buscada

$$z_{1-\alpha/2} = C + \frac{1}{6}C^3 + \frac{7}{120}C^5 + \dots \quad (3.24)$$

**Lema 3.4.:** Sean  $x, k > 0$  y  $d = O(n^{-k})$  dos números reales. Entonces la solución de

$$\Phi(x) - \Phi(-x) - 2dx\phi(x) = 1 - \alpha \quad (3.25)$$

está dada por

$$\begin{aligned} x &= z_{1-\alpha/2} (1-d)^{-1} + O(n^{-2k}) \\ &= z_{1-\alpha/2} (1+d) + O(n^{-2k}) \end{aligned}$$

*Demostración:* Mediante derivación, término a término, de (3.21) se obtiene una expansión de  $\phi(x) = \frac{d}{dx} \Phi(x)$ . Entonces

$$\begin{aligned} \Phi(x) - \Phi(-x) - 2xd\phi(x) &= \frac{2}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{n!2^n(2n+1)} - \frac{2d}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{n!2^n} \\ &= \frac{2}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{n!2^n} \left( \frac{1}{2n+1} - d \right) \\ &= \frac{\sqrt{2}}{\pi} \left( x(1-d) - \frac{1}{2} \left( \frac{1}{3} - d \right) x^3 + \frac{1}{8} \left( \frac{1}{5} - d \right) x^5 + \dots \right). \end{aligned}$$

De la última igualdad, se desprende que resolver (3.25) es equivalente a resolver, con respecto a  $x$ , la siguiente igualdad

$$1 - \alpha = \frac{\sqrt{2}}{\pi} \left( x(1-d) - \frac{1}{2} \left( \frac{1}{3} - d \right) x^3 + \frac{1}{8} \left( \frac{1}{5} - d \right) x^5 + \dots \right). \quad (3.26)$$

Por medio de las expresiones de la Sección A.1.4, con  $C = \frac{\sqrt{\pi}}{\sqrt{2}} (1-\alpha)$ , se desprende que

$$\begin{aligned} x &= \frac{C}{1-d} + \frac{1}{6} \frac{1-3C}{(1-d)^4} C^3 + \frac{1}{120} \frac{7-42C+75C^2}{(1-d)^7} C^5 + \dots \\ &= \frac{1}{1-d} \left( C + \frac{1}{6} \frac{1-3A}{(1-d)^3} C^3 + \frac{1}{120} \frac{7-42A+75A^2}{(1-d)^6} C^5 + \dots \right) \\ &= \frac{1}{1-d} \left( C + \frac{1}{6} C^3 + \frac{7}{120} C^5 + \dots \right) + O(n^{-2k}). \end{aligned}$$

Por lo tanto de (3.24) se llega a

$$\begin{aligned} x &= \frac{1}{1-d} z_{1-\alpha/2} + O(n^{-2k}) \\ &= z_{1-\alpha/2} (1+d) + O(n^{-2k}). \end{aligned}$$

### 3.3 Verificación de cinco equivalencias

En esta sección se prueba que los cinco procedimientos para medir la probabilidad de cobertura de intervalos de verosimilitud son equivalentes con error de orden  $O(n^{-2})$ . El procedimiento es como sigue:

Damos la siguiente expresión de la probabilidad de cobertura de intervalos de verosimilitud

$$\Pr(R(\theta) \geq c) = 2\Phi(v_c) - 1 - bv_c\phi(v_c) + O(n^{-2}), \quad (3.27)$$

donde  $v_c = \sqrt{-2 \ln c}$ . Esta fórmula se obtiene directamente en el modelo de localización de Efron & Hinkley (1978) y DiCiccio (1986) y en Jensen (1988) en un contexto general.

Después probamos que (3.27) es equivalente, con error de orden  $O(n^{-2})$ , a los cuatro procedimientos dados para medir la probabilidad de cobertura de intervalos de verosimilitud. Los cuatro procedimientos son la *fórmula-p\**, los *factores de Bartlett*, la *fórmula-r\** y la *aproximación de Lugannani-Rice*.

#### 3.3.1 Fórmula $p^*$

La *fórmula-p\** es una aproximación a la distribución condicional del estimador de máxima verosimilitud propuesta por Barndorff-Nielsen (1983) y está dada por

$$p^*(\hat{\theta}; \theta|a) = k |\hat{j}|^{1/2} e^{\ell - \hat{\ell}},$$

donde  $k = k(\theta, a)$  es la constante de integración, la fórmula  $p^*$  es una aproximación de orden  $O(n^{-3/2})$  en el siguiente sentido

$$p(\hat{\theta}; \theta|a) = p^*(\hat{\theta}; \theta|a) [1 + O(n^{-3/2})]. \quad (3.28)$$

Debido a que requerimos llegar a resultados de orden  $O(n^{-2})$ , necesitamos expresar el término  $O(n^{-3/2})$  con más detalle. Haciendo esto obtenemos que

$$p(\hat{\theta}; \theta|a) = p^*(\hat{\theta}; \theta|a) \left[ 1 + \frac{1}{n} c_1(\theta) (\hat{\theta} - \theta) + O(n^{-2}) \right]. \quad (3.29)$$

La constante de integración es de la forma

$$k = 1 - \frac{1}{2} b(\theta, a) + O(n^{-2}), \quad (3.30)$$

donde  $b(\theta, a)$  es un factor de Bartlett (Sección 3.3.2).

En el resto de las demostraciones se suprimirá en la notación la estadística auxiliar  $a$ , la cual se debe tener siempre presente.

Con ayuda de la aproximación (3.29) se puede calcular la función de densidad de cualquier función de  $\hat{\theta}$ . En particular si se supone que la verosimilitud dirigida es función uno-a-uno de  $\hat{\theta}$ , para  $a$  fijo, se puede calcular la densidad de  $r$  mediante técnicas ordinarias de cambio de variable.

La derivada de  $r(\theta; \hat{\theta})$  con respecto a  $\hat{\theta}$  está dada por

$$\frac{\partial}{\partial \hat{\theta}} r(\theta; \hat{\theta}) = \frac{\ell_{;\hat{\theta}}(\hat{\theta}; \hat{\theta}) - \ell_{;\hat{\theta}}(\theta; \hat{\theta})}{r(\theta; \hat{\theta})},$$

donde  $\ell_{;\hat{\theta}}(\theta; \hat{\theta}) = \frac{\partial}{\partial \hat{\theta}} \ell(\theta; \hat{\theta})$ . Sea

$$u(\theta; \hat{\theta}) = \hat{j}^{1/2} \left[ \ell_{;\hat{\theta}}(\hat{\theta}; \hat{\theta}) - \ell_{;\hat{\theta}}(\theta; \hat{\theta}) \right], \quad (3.31)$$

donde  $\hat{j} = -\hat{\ell}_{\theta\theta}$  es la información de Fisher observada.

Así, por medio de la ecuación (3.29) la función de densidad de la verosimilitud dirigida

se puede calcular aproximadamente mediante

$$\begin{aligned} P(r; \theta) &= \frac{k}{\sqrt{2\pi} u} r e^{-\frac{1}{2} r^2} \left[ 1 + \frac{1}{n} \bar{c}_1 \left[ \hat{\theta}(r; \theta) - \theta \right] \right] + O(n^{-2}) \\ &= \frac{k}{\sqrt{2\pi} u} r e^{-\frac{1}{2} r^2} \left[ 1 + \frac{1}{n} \bar{c}_1 \bar{j}^{-1/2} r \right] + O(n^{-2}) \end{aligned} \quad (3.32)$$

donde  $\hat{\theta}(r, \theta)$  es el estimador de máxima verosimilitud expresado como función de  $r$ , para  $\theta$  fijo. En particular  $\hat{\theta}(r; \theta) = \theta + \bar{j}^{-1/2} r + O(n^{-1})$ , ver ecuación (A.7). Mediante (3.32) obtenemos que

$$\Pr(R(\theta; \hat{\theta}) \geq c) = \Pr(-v_c \leq r \leq v_c) = \frac{k}{\sqrt{2\pi}} \int_{-v_c}^{v_c} \left( \frac{r}{u} e^{-\frac{1}{2} r^2} \left( 1 + \frac{1}{n} \bar{c}_1 \bar{j}^{-1/2} r \right) + O(n^{-2}) \right) dr. \quad (3.33)$$

En virtud de que  $\frac{r}{u} = 1 - \bar{t}_3 r + \frac{1}{2} b r^2 + c_2 r^3 + O(n^{-2})$ , (ver ec. A.10) se sigue que

$$\begin{aligned} \Pr(R(\theta; \hat{\theta}) \geq c) &= \frac{k}{\sqrt{2\pi}} \int_{-v_c}^{v_c} e^{-\frac{1}{2} r^2} \left( 1 - \bar{t}_3 r + \frac{1}{2} b r^2 + c_2 r^3 + O(n^{-2}) \right) \left( 1 + \frac{1}{n} \bar{c}_1 \bar{j}^{-1/2} r \right) dr + O(n^{-2}) \\ &= \frac{k}{\sqrt{2\pi}} \int_{-v_c}^{v_c} e^{-\frac{1}{2} r^2} \left[ 1 + \left( \frac{1}{n} \bar{c}_1 \bar{j}^{-1/2} - \bar{t}_3 \right) r + \frac{1}{2} b r^2 + c_2 r^3 \right] dr + O(n^{-2}) \\ &= \left( 1 - \frac{1}{2} b \right) \left\{ 2\Phi(v_c) - 1 + \frac{1}{2} b [2\Phi(v_c) - 1 - 2v_c \phi(v_c)] \right\} + O(n^{-2}) \\ &= 2\Phi(v_c) - 1 - 2bv_c \phi(v_c) + O(n^{-2}). \end{aligned} \quad (3.34)$$

El penúltimo paso de las igualdades anteriores es válido porque para  $x > 0$  se cumple lo siguiente:  $\frac{1}{\sqrt{2\pi}} \int_{-x}^x t^2 e^{-\frac{1}{2} t^2} dt = 2\Phi(x) - 1 - 2x\phi(x)$ .

Veamos ahora como se puede medir la cobertura de un intervalo de verosimilitud por medio de ajustes de Bartlett.

### 3.3.2 Factores de Bartlett

Como se muestra en (3.10)  $r^2$  sigue una distribución aproximada  $\chi_1^2$  con error de orden  $O(n^{-1})$  en muestras repetidas ordinarias. Una forma natural para acelerar la aproximación

$\chi_1^2$  de  $r^2$  es considerar la estadística

$$W = \frac{r^2}{Er^2}. \quad (3.35)$$

En general no es posible calcular explícitamente  $Er^2$  y en lugar de usar (3.35) se utiliza el siguiente hecho

$$Er^2 = 1 + B + O(n^{-2}), \quad B = B(\theta) = O(n^{-1})$$

y la estadística "corregida" es la estadística ajustada de Bartlett

$$W^* = \frac{r^2}{1+B},$$

la cual cumple que

$$\Pr(W^* \leq x) = 2\Phi(x) - 1 + O(n^{-2}),$$

Esta idea fue explotada primeramente por Bartlett (1937). Ver también Lawley (1956), Barndorff-Nielsen & Cox (1984), Barndorff-Nielsen & Hall (1988) y Jensen (1991).

Sea  $B = E(r^2) - 1$ . Por medio de (3.32) se puede calcular directamente la esperanza de  $r^2$ . Así, pues

$$\begin{aligned} Er^2 &= \int_{-\infty}^{\infty} r^2 p(r; \theta) dr \\ &= \frac{k}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{r^3}{u} e^{-\frac{1}{2}r^2} \left[1 + \frac{1}{n} \bar{c}_1 J^{-1/2} r\right] dr \\ &= \frac{k}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}r^2} \left[r^2 + \left(\frac{1}{n} \bar{c}_1 J^{-1/2} - \bar{t}_3\right) r^3 + \frac{1}{2} b r^4 + c_2 r^5\right] dr \\ &= \left[1 - \frac{1}{2}b + O(n^{-2})\right] \left[1 + \frac{3}{2}b + O(n^{-2})\right] \\ &= 1 + b + O(n^{-2}) \end{aligned}$$

Se desprende que  $B = b$ .

Puesto que  $(1+b)^{-1} = 1 - b + O(n^{-2})$  se sigue también que  $(1-b)r^2 \sim \chi_1^2 + O(n^{-2})$ .

Para  $0 < c < 1$  sea  $v_c = \sqrt{-2 \ln c}$ , la cobertura del intervalo de verosimilitud es

$$\begin{aligned} \Pr(R(\theta; \hat{\theta}) \geq c) &= \Pr(r^2 < v_c^2) \\ &= \Pr((1-b)r^2 < (1-b)v_c^2) \\ &= 2\Phi\left(\sqrt{1-b}v_c\right) - 1 + O(n^{-2}). \end{aligned} \quad (3.36)$$

$$\begin{aligned} &= 2\Phi\left(\left(1 - \frac{1}{2}b\right)v_c\right) - 1 + O(n^{-2}) \\ &= 2\Phi(v_c) - 1 - b v_c \phi(v_c) + O(n^{-2}) \end{aligned} \quad (3.37)$$

Las últimas dos igualdades se siguen del Lema 3.3 y por el hecho de que  $\sqrt{1-b} = 1 - \frac{1}{2}b + O(n^{-2})$ . ■

### 3.3.3 Fórmula- $r^*$

La expresión de  $r^*$  como función de  $r$  ayuda a calcular la cobertura de intervalos de verosimilitud con error de orden  $O(n^{-2})$ . Sea  $v_c = \sqrt{-2 \ln c}$ , entonces se cumple lo siguiente

$$\begin{aligned} \Pr(R(\theta; \hat{\theta}) \geq c) &= \Pr(-v_c \leq r \leq v_c) \\ &= \Pr(r^*(-v_c) \leq r^* \leq r^*(v_c)) \\ &= \Phi(r^*(v_c)) - \Phi(r^*(-v_c)) + O(n^{-2}). \end{aligned}$$

De (A.11) se tiene que  $r^* = r + \bar{t}_3 + c_1 r + c_2 r^2 + O(n^{-2})$ , donde  $c_1 = \frac{1}{2}(\bar{t}_3^2 - b)$ . Así, con ayuda del Lema 3.1 se llega a

$$\Pr(R(\theta; \hat{\theta}) \geq c) = \Phi(v_c + \bar{t}_3 + c_1 v_c) - \Phi(-v_c + \bar{t}_3 + c_1 v_c) + O(n^{-2}).$$

De nuevo, por medio del Lema 3.3 se obtiene que

$$\begin{aligned} \Phi(v_c + t_3 + c_1 v_c) &= \Phi(v_c) + (\bar{t}_3 + c_1 v_c) \phi(v_c) - \frac{1}{2} (\bar{t}_3^2 + 2t_3 c_1 v_c) v_c \phi(v_c) \\ &\quad + \frac{1}{6} \bar{t}_3^3 (v_c^2 - 1) \phi(v_c) + O(n^{-2}), \end{aligned}$$

y que

$$\begin{aligned} \Phi(-v_c + t_3 - c_1 v_c) &= \Phi(-v_c) + (\bar{t}_3 - c_1 v_c) \phi(v_c) + \frac{1}{2} (\bar{t}_3^2 - 2\bar{t}_3 c_1 v_c) v_c \phi(v_c) \\ &\quad + \frac{1}{6} \bar{t}_3^3 (v_c^2 - 1) \phi(v_c) + O(n^{-2}). \end{aligned}$$

Así

$$\begin{aligned} \Pr(R(\theta; \hat{\theta}) \geq c) &= \Phi(v_c + \bar{t}_3 + c_1 v_c) - \Phi(-v_c + \bar{t}_3 + c_1 v_c) + O(n^{-2}) \\ &= 2\Phi(v_c) - 1 + 2(c_1 - \frac{1}{2}\bar{t}_3^2) v_c \phi(v_c) + O(n^{-2}) \\ &= 2\Phi(v_c) - 1 - b v_c \phi(v_c) + O(n^{-2}), \end{aligned} \quad (3.38)$$

la última igualdad es válida porque  $b = \bar{t}_3^2 - 2c_1$  como se ilustra en (A.11).

### 3.3.4 Aproximación de Lugannani-Rice

Hemos visto tres procedimientos para calcular la cobertura de un intervalo de verosimilitud, con las tres anteriores se ha visto la manera de calcular directamente la probabilidad  $\Pr(R \geq c)$ . Ahora con la aproximación de Lugannani-Rice veremos una variante.

La fórmula de Lugannani-Rice se utiliza en el cálculo de probabilidades de eventos cola, ésta fórmula es

$$\begin{aligned} \Pr(\hat{\theta} \geq x; \theta) &= 1 - \Phi(r) + \phi(r) \left( \frac{1}{u} - \frac{1}{r} \right) + O(n^{-3/2}) \\ &= 1 - \Phi(r) + \frac{1}{r} \phi(r) \left( \frac{r}{u} - 1 \right) + O(n^{-3/2}) \end{aligned} \quad (3.39)$$

Para calcular con la aproximación de Lugannani-Rice la cobertura de intervalos de verosimilitud procederemos de manera diferente a los tres procedimientos anteriores.

Los valores de  $\hat{\theta}$  que satisfacen  $R(\theta; \hat{\theta}) \geq c$ , también cumplen la relación  $-v_c \leq r(\theta; \hat{\theta}) \leq v_c$ ,  $v_c = \sqrt{-2 \ln c}$ . Bajo el supuesto de que  $r$  es función uno-a-uno de  $\hat{\theta}$ , se pueden encontrar los  $h_1(\theta, c)$  y  $h_2(\theta, c)$  tales que  $h_1(\theta, c) \leq \hat{\theta} \leq h_2(\theta, c)$  y  $r(\theta; h_1(\theta, c)) = -r(\theta; h_2(\theta, c))$ . Así

$$\begin{aligned} \Pr(R(\theta; \hat{\theta}) \geq c) &= \Pr(h_1(\theta, c) \leq \hat{\theta} \leq h_2(\theta, c)) \\ &= \Pr(\hat{\theta} \geq h_1(\theta, c)) - \Pr(\hat{\theta} \geq h_2(\theta, c)) \\ &= 1 - \Phi(-v_c) + \frac{1}{v_c} \phi(v_c) \left( \frac{v_c}{u(-v_c)} + 1 \right) \\ &\quad - 1 + \Phi(v_c) - \frac{1}{v_c} \phi(v_c) \left( \frac{v_c}{u(v_c)} - 1 \right) + O(n^{-3/2}) \\ &= 2\Phi(v_c) - 1 - \frac{1}{v_c} \phi(v_c) \left[ -\frac{v_c}{u(-v_c)} + \frac{v_c}{u(v_c)} - 2 \right] \\ &= 2\Phi(v_c) - 1 - \frac{1}{v_c} \phi(v_c) \left[ 1 + \bar{t}_3 v_c + \frac{1}{2} b v_c^2 + c_2 v_c^3 + 1 \right. \\ &\quad \left. - \bar{t}_3 v_c + \frac{1}{2} v_c^2 - c_2 v_c^3 + O(n^{-2}) \right] + O(n^{-3/2}) \\ &= 2\Phi(v_c) - 1 - b v_c \phi(v_c) + O(n^{-3/2}) \end{aligned} \quad (3.40)$$

$$= 2\Phi(v_c) - 1 - b v_c \phi(v_c) + O(n^{-3/2}) \quad (3.41)$$

La penúltima igualdad es porque  $\frac{r}{u} = 1 - \bar{t}_3 r + b r^2 + c_2 r^3 + \dots$  (ver A.10).

De aquí tenemos que sin necesidad de estudiar con detalle el término de error de la aproximación Lugannani-Rice, la cual es  $O(n^{-3/2})$ , se llega a que debe ser  $O(n^{-2})$  para intervalos de verosimilitud, puesto que (3.34), (3.37) y (3.38) son verdaderas. Es decir, por

Lugannani-Rice se tiene

$$\Pr \left( R(\theta; \hat{\theta}) \geq c \right) = 2\Phi(v_c) - 1 - 2bv_c\phi(v_c) + O(n^{-2}). \quad (3.42)$$

## Capítulo 4

### Precisión de $(r^*)^2$

En este capítulo demostramos que con la verosimilitud dirigida modificada  $r^*$  se pueden construir intervalos de confianza para  $\psi$  escalar, con error de orden  $O(n^{-2})$  en muestras repetidas ordinarias. En el caso  $\theta$  escalar la demostración es analítica y cuando  $\theta = (\psi, \chi)$ ,  $\psi$  escalar, se presenta evidencia numérica que avala la afirmación anterior. Este orden de precisión no se había observado desde la propuesta de la verosimilitud dirigida modificada (Barndorff-Nielsen, 1986). En dicho trabajo Barndorff-Nielsen comenta que con la verosimilitud dirigida modificada se pueden construir intervalos de confianza aproximados con error del mismo orden que  $r^*$ , es decir de orden  $O(n^{-3/2})$ . La anterior afirmación es válida para cualquier intervalo construido mediante los valores de  $\psi$  tales que  $t_1 \leq r^*(\psi) \leq t_2$ ,  $t_1 < t_2$ . En este Capítulo se demuestra que si tomamos  $t_1 = -t_2$ ,  $t_2 > 0$  el nivel de confianza aproximado es más preciso, ahora el error es de orden  $O(n^{-2})$ .

### 4.1 $\theta$ escalar

Utilizaremos la aproximación a la distribución condicional del estimador de máxima verosimilitud propuesta por Barndorff-Nielsen (1983)

$$p^*(\hat{\theta}; \theta|a) = c_0 |\hat{j}|^{1/2} e^{\ell - \hat{\ell}}, \quad (4.1)$$

donde  $c_0 = c_0(\theta, a)$  es la constante de integración. La fórmula- $p^*$  es una aproximación de orden  $O(n^{-3/2})$  en el siguiente sentido

$$p(\hat{\theta}; \theta|a) = p^*(\hat{\theta}; \theta|a) [1 + O(n^{-3/2})]. \quad (4.2)$$

Debido a que requerimos llegar a resultados de orden  $O(n^{-2})$ , es necesario expandir el término de error  $O(n^{-3/2})$ , de la fórmula (4.2). Así

$$p(\hat{\theta}; \theta|a) = p^*(\hat{\theta}; \theta|a) \left[ 1 + \frac{1}{n} c_1(\theta) (\hat{\theta} - \theta) + O(n^{-2}) \right]. \quad (4.3)$$

En la Sección 3.2 se demostró que la densidad de la verosimilitud dirigida modificada tiene el siguiente desarrollo

$$p(r^*; \theta) = \phi(r^*) [1 + H_3 r^* + O(n^{-2})]. \quad (4.4)$$

Directamente de (4.4) se demuestra el Teorema 4.1, el cual afirma que la distribución de  $(r^*)^2$  es  $\chi^2$ -cuadrada con error, en muestras repetidas ordinarias, de orden  $O(n^{-2})$ .

**Teorema 4.1.** Para  $x > 0$  se cumple que

$$\Pr((r^*)^2 \leq x^2) = [2\Phi(x) - 1] [1 + O(n^{-2})]. \quad (4.5)$$

*Demostración.* Mediante la densidad de  $r^*$ , dada en (4.4), obtenemos, para  $x > 0$ , que

$$\begin{aligned} \Pr((r^*)^2 \leq x^2) &= \Pr(-x \leq r^* \leq x) = \int_{-x}^x p(r^*; \psi, \chi) dr^* \\ &= \int_{-x}^x \phi(r^*) [1 + Hr^* + O(n^{-2})] dr^* \\ &= \int_{-x}^x \phi(r^*) [1 + O(n^{-2})] dr^* + H \int_{-x}^x r^* \phi(r^*) dr^* \\ &= [2\Phi(x) - 1] [1 + O(n^{-2})]. \end{aligned}$$

### 4.2 Parámetros de estorbo: $\psi$ escalar

En general, cuando el parámetro de interés es escalar, la verosimilitud dirigida modificada satisface que

$$\Pr(r^* \leq x) = \Phi(x) [1 + M],$$

donde  $M = O(n^{-3/2})$ . El problema, para obtener error de orden  $O(n^{-2})$  para la distribución de  $(r^*)^2$ , estriba en que no se tiene de manera explícita la cantidad  $M$ . Si  $M$  es de la forma  $Hr^* + O(n^{-2})$ , donde  $H$  es  $O(n^{-3/2})$ , se siguen los mismos resultados que en el caso escalar, principalmente que

$$\Pr((r^*)^2 \leq x) = [\Phi(x) - \Phi(-x)] [1 + O(n^{-2})]$$

y que la probabilidad de cobertura de un intervalo de verosimilitud es

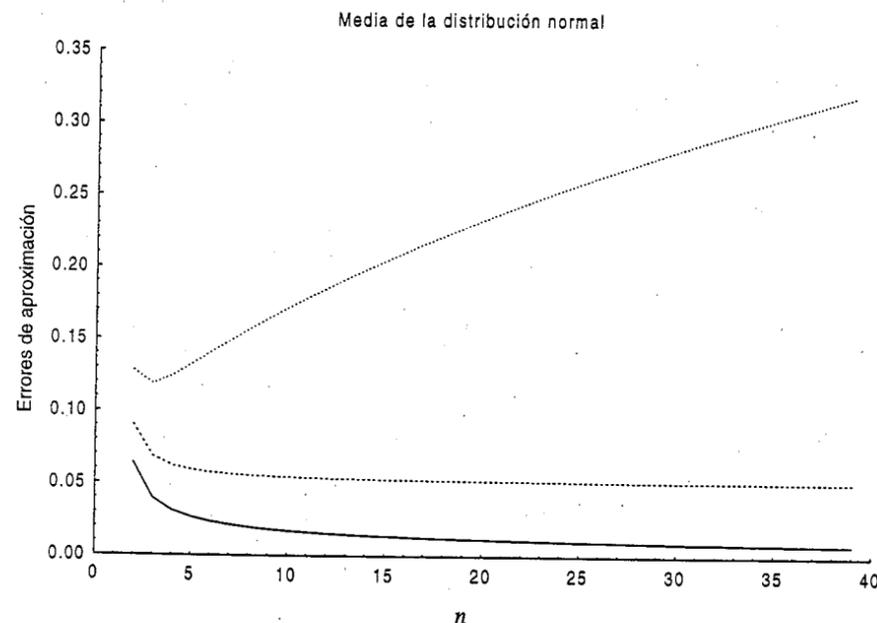
$$\Pr(R_p(\psi) \geq c) = \Pr(r^*(-v_c) \leq r^* \leq r^*(v_c)) = [\Phi(r^*(v_c)) - \Phi(r^*(-v_c))] [1 + O(n^{-2})],$$

donde  $v_c = \sqrt{-2 \ln c}$ .

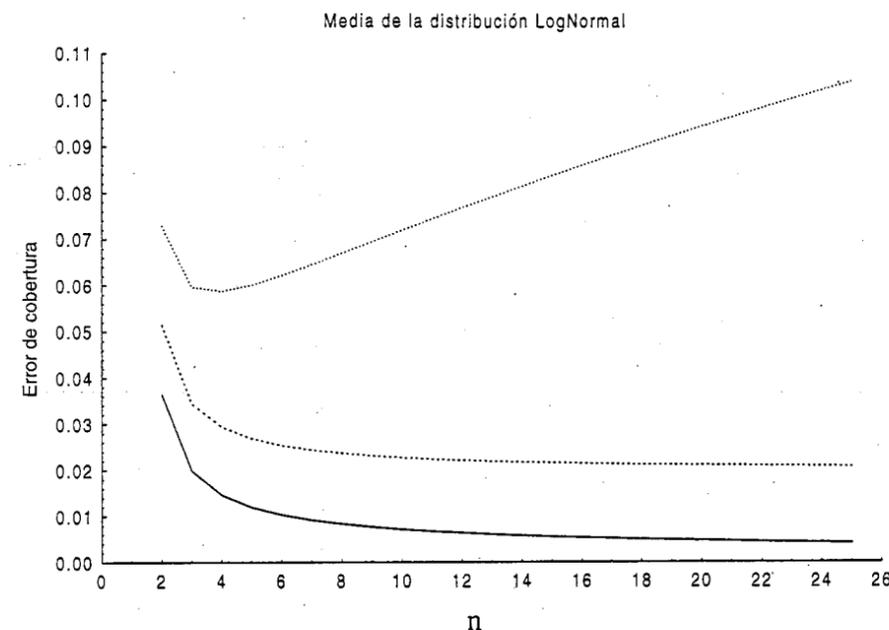
La demostraciones son similares a la de la sección anterior y a las del Capítulo 3.

Consideremos la distribución normal para ilustrar numéricamente que  $(r^*)^2 \sim \chi_1^2 + O(n^{-2})$ .

Sea  $y = (y_1, \dots, y_n)^t$  una muestra de v.a.i.i.d. de una distribución  $N(\mu, \sigma^2)$ . Mediante la fórmula (2.32) obtenemos  $r^*(\mu; \hat{\mu})$ . La Figuras 4.1 y 4.2 muestran las gráficas de los errores de aproximación de la verdadera probabilidad de  $-1.96 \leq r^* \leq 1.96$  con respecto a  $n$ , para los modelos normal y la media de la log-normal. La línea constante indica el orden de aproximación.



**Figura 4.1.** Errores de la aproximación de los intervalos de confianza obtenidos con  $-1.96 \leq r^* \leq 1.96$  para la media de la distribución normal con  $\mu = 0$ . De abajo hacia arriba : gráfica de  $n$  vs  $n^{k/2} |P - .95|$ ,  $k = 3, 4, 5$ . Donde  $P = \Pr[-1.96 \leq r^*(\mu; \hat{\mu}; n) \leq 1.96]$ .



**Figura 4.2.** Errores de la aproximación de los intervalos de confianza obtenidos con  $-1.96 \leq r^* \leq 1.96$  para la media de la distribución log-normal con  $\psi = 0, \hat{\chi}_\psi = 3$ . De abajo hacia arriba: primer línea es la gráfica de  $n$  vs  $n^{k/2} |P - .95|$ ,  $k = 3, 4, 5$ . Donde  $P = \Pr[-1.96 \leq r^*(\psi; \hat{\psi}; n) \leq 1.96]$ .

### 4.3 Discusión

La verosimilitud dirigida modificada,  $r^*$ , surge directamente de la fórmula  $p^*$  como se muestra en la Sección 3.2.2 (ver también Barndorff-Nielsen & Cox 1994, Sec. 6.6.1 y Reid, 1993). Efron (1998) le llama "fórmula mágica" a la fórmula  $p^*$ . A este respecto es conveniente traer a colación lo que se discute en ese artículo:

En Efron (1998) p. 105 se puede leer: "In particular the magic formula<sup>1</sup> can be used to generate approximate confidence intervals that are more accurate than the standard intervals<sup>2</sup>. These intervals agree to second order with the bootstrap intervals. If this were not true, then one or both of them would not be second order correct"

En los comentarios D.A.S. Fraser dice:

"Brad mentions that the magic formula can be used to generate approximate confidence intervals [with] at least second order accuracy. In fact for continuous models third order confidence intervals are available in wide generality but require additional theory for approximate ancillaries..."

Es decir en 1998 todavía no se observaba que con  $r^*$  se pueden construir intervalos de confianza con error de orden  $O(n^{-2})$ .

<sup>1</sup>La fórmula mágica mencionada por Efron es  $f_{\theta}(\hat{\theta}|A) = c \frac{L(\theta)}{L(\hat{\theta})} \left[ \frac{d^2}{d\theta^2} \log L(\theta) \Big|_{\theta=\hat{\theta}} \right]^{1/2}$

<sup>2</sup>Los intervalos estándares son  $\theta = \hat{\theta} \pm 1.645 se$

## Apéndice A

### Expansiones asintóticas

Este apéndice contiene los pormenores para el cálculo de las expansiones asintóticas que se utilizaron en los Capítulos 3 y 4.

#### A.1 Álgebra de series asintóticas

En este Apéndice supondremos que  $a_j = O(n^{-k_1j})$  y  $b_j = O(n^{-k_2j})$ ,  $d = \min(k_1, k_2)$ .

Todas las fórmulas son válidas aunque las  $a_j$ 's y las  $b_j$ 's sean constantes que no dependan de  $n$ . La diferencia radica en que el error de la aproximación no se cumple.

##### A.1.1 Cociente de series

$$\frac{\sum_{n=0}^{\infty} a_n x^n}{\sum_{n=0}^{\infty} b_n x^n} = \sum_{n=0}^{m-1} c_n x^n + O(n^{-dm}) \tag{A.1}$$

Para  $m = 5$ , los primeros cuatro valores de  $c_i$  que satisfacen (A.1) son

$i$	$c_i$
0	$a_0 b_0^{-1}$
1	$(-b_1 a_0 + a_1 b_0) b_0^{-2}$
2	$(b_1^2 a_0 - b_1 a_1 b_0 - b_2 a_0 b_0 + a_2 b_0^2) b_0^{-3}$
3	$(-b_1^3 a_0 + b_1^2 a_1 b_0 + 2b_2 b_1 a_0 b_0 - b_1 a_2 b_0^2 - b_2 a_1 b_0^2 - b_3 a_0 b_0^2 + a_3 b_0^3) b_0^{-4}$

Ver Gradshteyn & Ryzhik (1994, pág. 18) para una expresión general.

**Ejemplo. Recíproco de series**

Tomando  $a_0 = 1$ , y  $a_n = 0, n \geq 1$ , de la fórmula para cocientes (A.1) se obtiene que

$i$	$c_i$
0	$b_0^{-1}$
1	$-b_1 b_0^{-2}$
2	$(-b_0 b_2 + b_1^2) b_0^{-3}$
3	$-(b_0^2 b_3 - 2b_0 b_2 b_1 + b_1^3) b_0^{-4}$
4	$(-3b_0 b_2 b_1^2 - b_0^3 b_4 + 2b_0^2 b_3 b_1 + b_0^2 b_2^2 + b_1^4) b_0^{-5}$
5	$-(3b_0^2 b_3 b_1^2 + b_0^4 b_5 - 2b_0^3 b_4 b_1 - 2b_3 b_0^3 b_2 + 3b_0^2 b_2^2 b_1 - 4b_0 b_2 b_1^3 + b_1^5) b_0^{-6}$

**A.1.2 Raíz de series**

Considere  $a_0 > 0$ .

$$\sqrt{\sum_{n=0}^{\infty} a_n x^n} = \sum_{n=0}^{m-1} b_n x^n + O(n^{-dm}). \tag{A.2}$$

Para  $m = 6$  se tiene que

$i$	$b_i$
0	$\sqrt{a_0}$
1	$\frac{1}{2} a_1 a_0^{-1/2}$
2	$\frac{1}{8} (-a_1^2 + 4a_0 a_2) a_0^{-3/2}$
3	$\frac{1}{16} (a_1^3 - 4a_1 a_0 a_2 + 8a_0^2 a_3) a_0^{-5/2}$
4	$\frac{1}{128} (-5a_1^4 + 24a_1^2 a_0 a_2 - 32a_1 a_0^2 a_3 - 16a_0^2 a_2^2 + 64a_0^3 a_4) a_0^{-7/2}$
5	$\frac{1}{256} (7a_1^5 - 40a_1^3 a_0 a_2 + 48a_1^2 a_0^2 a_3 + 48a_0^2 a_2^2 a_1 - 64a_1 a_0^3 a_4 - 64a_0^3 a_2 a_3 + 128a_0^4 a_5) a_0^{-9/2}$

**Ejemplo. Recíproco de raíz de series**

Sea  $b_0 > 0$

$$\left( \sum_{n=0}^{\infty} b_n x^n \right)^{-1/2} = \sum_{n=0}^{m-1} c_n x^n + O(n^{-dm}),$$

entonces de las expresiones para el recíproco (Sección A.1.1) y de (A.2) obtenemos que

$i$	$c_i$
0	$b_0^{-1/2}$
1	$-\frac{1}{2} b_1 b_0^{-3/2}$
2	$-\frac{1}{8} (-3b_1^2 + 4b_0 b_2) b_0^{-5/2}$
3	$-\frac{1}{16} (5b_1^3 - 12b_0 b_2 b_1 + 8b_0^2 b_3) b_0^{-7/2}$
4	$-\frac{1}{128} (-35b_1^4 + 120b_1^2 b_0 b_2 - 96b_1 b_0^2 b_3 - 48b_0^2 b_2^2 + 64b_0^3 b_4) b_0^{-9/2}$
5	$-\frac{1}{256} (63b_1^5 - 280b_1^3 b_0 b_2 + 240b_1^2 b_0^2 b_3 + 240b_0^2 b_2^2 b_1 - 192b_1 b_0^3 b_4 - 192b_0^3 b_2 b_3 + 128b_0^4 b_5) b_0^{-11/2}$

### A.1.3 Inversa de Series.

Sea

$$y = \sum_{n=1}^{\infty} a_n x^n,$$

se desea encontrar

$$x = \sum_{n=1}^{m-1} b_n y^n + O(n^{-dm}).$$

Por ejemplo, los primeros cinco valores de  $b_i$  son

$i$	$c_i$
1	$a_1^{-1}$
2	$-a_2 a_1^{-3}$
3	$-(a_3 a_1 - 2a_2^2) a_1^{-5}$
4	$-(-5a_3 a_2 a_1 + 5a_2^3 + a_4 a_1^2) a_1^{-7}$
5	$-(21a_3 a_2^2 a_1 - 14a_2^4 - 6a_4 a_2 a_1^2 - 3a_3^2 a_1^2 + a_5 a_1^3) a_1^{-9}$

## A.2 Expansiones asintóticas de $r$ y $r^*$

En esta sección se aplican los resultados de la sección anterior para obtener expansiones asintóticas de cantidades relacionadas con la función de verosimilitud.

### A.2.1 Parámetro $\theta$ escalar

Consideremos la siguiente notación

$$l_{j;k}(\theta; \hat{\theta}, a) = \frac{\partial^{j+k}}{\partial \theta^j \partial \hat{\theta}^k} \ell(\theta; \hat{\theta}, a), \quad j, k \geq 0$$

con las siguientes convenciones

$$l_i \equiv l_{i;0}, \quad l_{i;j} = l_{0;j}, \quad \bar{l}_{i;j} \equiv l_{i;j}(\theta; \theta) = l_{i;j}(\theta; \hat{\theta}) \Big|_{\hat{\theta}=\theta}, \quad i, j \geq 1.$$

#### A.2.1.1 Expansiones asintóticas alrededor de $\hat{\theta}$

$$\ell = \hat{\ell} + \frac{1}{2} \hat{\ell}_2 (\hat{\theta} - \theta)^2 - \frac{1}{6} \hat{\ell}_3 (\hat{\theta} - \theta)^3 + \frac{1}{24} \hat{\ell}_4 (\hat{\theta} - \theta)^4 + \dots$$

Dado que  $j = -\ell_2$  tenemos que

$$\hat{\ell} - \ell = \frac{1}{2} \hat{j} (\hat{\theta} - \theta)^2 + \frac{1}{6} \hat{\ell}_3 (\hat{\theta} - \theta)^3 - \frac{1}{24} \hat{\ell}_4 (\hat{\theta} - \theta)^4 + \dots \quad (\text{A.3})$$

#### Expansión asintótica de $r$

Dado que  $r^2 = 2(\hat{\ell} - \ell)$  se sigue que

$$r^2 = \hat{j} (\hat{\theta} - \theta)^2 + \frac{1}{3} \hat{\ell}_3 (\hat{\theta} - \theta)^3 - \frac{1}{12} \hat{\ell}_4 (\hat{\theta} - \theta)^4 + \dots$$

Por medio de las expresiones para la raíz de series de la Sección A.1.2 se obtiene que

$$r = \hat{j}^{1/2} (\hat{\theta} - \theta) + \frac{1}{6} \hat{\ell}_3 \hat{j}^{-1/2} (\hat{\theta} - \theta)^2 - \frac{1}{72} (\hat{\ell}_3^2 + 3\hat{j}\hat{\ell}_4) \hat{j}^{-3/2} (\hat{\theta} - \theta)^3 + \dots$$

la cual es válida porque el primer término de la serie para  $r^2$  es  $\hat{j} > 0$ .

Mediante inversión de series (ver Sección A.1.3) se obtiene que los extremos del intervalo de verosimilitud con nivel de verosimilitud  $c$  son

$$\hat{\theta} - \theta = \hat{j}^{-1/2} v_c + \frac{1}{6} \hat{\ell}_3 \hat{j}^{-2} v_c^2 \pm \frac{1}{72} (5\hat{\ell}_3^2 + 3\hat{j}\hat{\ell}_4) \hat{j}^{-7/2} v_c^3 + \dots \quad (\text{A.4})$$

*Expansión asintótica de u y r\**

Derivando (A.3) con respecto a  $\hat{\theta}$  se sigue que

$$\begin{aligned} \hat{\ell}_{;1} - \ell_{;1} = & -\frac{1}{2} (\hat{\ell}_3 + \hat{\ell}_{2;1}) (\hat{\theta} - \theta)^2 - \hat{\ell}_2 (\hat{\theta} - \theta) + \frac{1}{6} (\hat{\ell}_4 + \hat{\ell}_{3;1}) (\hat{\theta} - \theta)^3 + \frac{1}{2} \hat{\ell}_3 (\hat{\theta} - \theta)^2 \\ & - \frac{1}{24} (\hat{\ell}_5 + \hat{\ell}_{4;1}) (\hat{\theta} - \theta)^4 - \frac{1}{6} \hat{\ell}_4 (\hat{\theta} - \theta)^3 + \dots, \end{aligned}$$

de donde, acomodando términos, se tiene que

$$\hat{\ell}_{;1} - \ell_{;1} = \hat{j} (\hat{\theta} - \theta) - \frac{1}{2} \hat{\ell}_{2;1} (\hat{\theta} - \theta)^2 + \frac{1}{6} \hat{\ell}_{3;1} (\hat{\theta} - \theta)^3 - \frac{1}{24} \hat{\ell}_{4;1} (\hat{\theta} - \theta)^4 + \dots$$

La expansión anterior también se obtiene si se expande  $\ell_{;1}$  alrededor de  $\hat{\theta}$ .

En virtud de que  $u = \hat{j}^{-1/2} (\hat{\ell}_{;1} - \ell_{;1})$  la expansión de  $u$  es

$$u = \hat{j}^{1/2} (\hat{\theta} - \theta) - \frac{1}{2} \hat{\ell}_{2;1} \hat{j}^{-1/2} (\hat{\theta} - \theta)^2 + \frac{1}{6} \hat{\ell}_{3;1} \hat{j}^{-1/2} (\hat{\theta} - \theta)^3 - \frac{1}{24} \hat{\ell}_{4;1} \hat{j}^{-1/2} (\hat{\theta} - \theta)^4 + \dots$$

y

$$\begin{aligned} \frac{u}{r} = & 1 - \frac{1}{6} (3\ell_{2;1} + \ell_3) \hat{j}^{-1} (\hat{\theta} - \theta) + \frac{1}{24} (\ell_3^2 + 2\ell_{2;1}\ell_3 + 4\ell_{3;1}\hat{j} + \hat{j}\ell_4) \hat{j}^{-2} (\hat{\theta} - \theta)^2 + \dots \\ = & 1 + \hat{t}_3 \hat{j}^{1/2} (\hat{\theta} - \theta) + \hat{c} (\hat{\theta} - \theta)^2 + \dots \end{aligned} \quad (A.5)$$

Usando la expansión  $\ln(1+x) = x - \frac{1}{2}x^2 + \dots$  se desprende que

$$\frac{1}{r} \ln \frac{u}{r} = \hat{t}_3 - \frac{1}{6} (-6\hat{c}\hat{j}^{1/2} + 3\hat{t}_3^2\hat{j}^{3/2} + \hat{t}_3\hat{\ell}_3) \hat{j}^{-1} A + \dots$$

Así, pues

$$r^* = \hat{j}^{1/2} A + \hat{t}_3 - \frac{1}{6} (-6\hat{c}\hat{j}^{1/2} + 3\hat{t}_3^2\hat{j}^{3/2} + \hat{t}_3\hat{\ell}_3) \hat{j}^{-1} A + \dots$$

Entonces la región  $-z_{1-\alpha/2} \leq r^* \leq z_{1-\alpha/2}$  se calcula mediante

$$\theta = \hat{\theta} - \frac{\pm z_{1-\alpha/2} - \hat{t}_3}{\hat{j}^{1/2} - \frac{1}{6} (-6\hat{c}\hat{j}^{1/2} + 3\hat{t}_3^2\hat{j}^{3/2} + \hat{t}_3\hat{\ell}_3) \hat{j}^{-1}}$$

Al substituir (A.4) en (A.5) se obtiene

$$\frac{u}{r} = 1 - \frac{1}{6} (\hat{\ell}_3 + 3\hat{\ell}_{2;1}) \hat{j}^{-3/2} r + \frac{1}{72} ((5\hat{\ell}_3 + 12\hat{\ell}_{2;1}) \hat{\ell}_3 + 3\hat{j} (4\hat{\ell}_{3;1} + \hat{\ell}_4)) \hat{j}^{-3} r^2 + \dots,$$

mediante la fórmula para el recíproco de series de la Sección A.1.1 llegamos a

$$\frac{r}{u} = 1 + \frac{1}{6} (\ell_3 + 3\ell_{2;1}) \hat{j}^{-3/2} r - \frac{1}{24} (\ell_3^2 + \hat{j}\ell_4 + 4\ell_{3;1}\hat{j} - 6\ell_{2;1}^2) \hat{j}^{-3} r^2 + \dots$$

**A.2.1.2 Expansiones asintóticas alrededor de  $\theta$**

*Expansión asintótica de r*

Al expandir cada  $\hat{\ell}_j$  alrededor de  $\theta$  y substituir en (A.3), se obtiene

$$\hat{\ell} - \ell = -\frac{1}{2} \bar{\ell}_2 (\hat{\theta} - \theta)^2 - \frac{1}{6} (2\bar{\ell}_3 + 3\bar{\ell}_{2;1}) (\hat{\theta} - \theta)^3 - \frac{1}{24} (3\bar{\ell}_4 + 8\bar{\ell}_{3;1} + 6\bar{\ell}_{2;2}) (\hat{\theta} - \theta)^4 + \dots \quad (A.6)$$

Entonces

$$r^2 = \bar{j} (\hat{\theta} - \theta)^2 - \frac{1}{3} (2\bar{\ell}_3 + 3\bar{\ell}_{2;1}) (\hat{\theta} - \theta)^3 - \frac{1}{12} (3\bar{\ell}_4 + 8\bar{\ell}_{3;1} + 6\bar{\ell}_{2;2}) (\hat{\theta} - \theta)^4 + \dots,$$

por medio de las expresiones para la raíz de serie (ver Sección A.1.2) se llega a

$$\begin{aligned} r = & \bar{j}^{1/2} (\hat{\theta} - \theta) - \frac{1}{6} (2\bar{\ell}_3 + 3\bar{\ell}_{2;1}) \bar{j}^{-1/2} (\hat{\theta} - \theta)^2 \\ & - \frac{1}{72} ((2\bar{\ell}_3 + 3\bar{\ell}_{2;1})^2 + 3(3\bar{\ell}_4 + 8\bar{\ell}_{3;1} + 6\bar{\ell}_{2;2}) \bar{j}) \bar{j}^{-3/2} (\hat{\theta} - \theta)^3 + \dots \end{aligned}$$

y por la Sección A.1.3, la inversa es

$$\begin{aligned} \hat{\theta} - \theta &= \bar{j}^{-1/2} r + \frac{1}{6} (2\bar{\ell}_3 + 3\bar{\ell}_{2,1}) \bar{j}^{-2} r^2 \\ &+ \frac{1}{72} \left( 5 (2\bar{\ell}_3 + 3\bar{\ell}_{2,1})^2 + 3 (3\bar{\ell}_4 + 8\bar{\ell}_{3,1} + 6\bar{\ell}_{2,2}) \bar{j} \right) \bar{j}^{-7/2} r^3 + \dots, \end{aligned} \quad (\text{A.7})$$

mediante (A.7) se obtiene una expresión útil para obtener los valores de  $\hat{\theta}$  tales que  $|r(\theta; \hat{\theta})| \leq z_{1-\alpha/2}$ , y para evaluar rápidamente  $r^*(\pm r)$ .

### Expansión asintótica de $r^*$

Derivando (A.6) con respecto a  $\hat{\theta}$  se tiene que

$$\hat{\ell}_{,1} - \ell_{,1} = \bar{j} (\hat{\theta} - \theta) - \frac{1}{2} (2\bar{\ell}_3 + 3\bar{\ell}_{2,1}) (\hat{\theta} - \theta)^2 - \frac{1}{6} (3\bar{\ell}_4 + 8\bar{\ell}_{3,1} + 6\bar{\ell}_{2,2}) (\hat{\theta} - \theta)^3 + \dots \quad (\text{A.8})$$

Después de expandir  $\hat{\ell}_2$  alrededor de  $\theta$  se obtiene que

$$\begin{aligned} \hat{j} &= \bar{j} - (\bar{\ell}_3 + \bar{\ell}_{2,1}) (\hat{\theta} - \theta) - \frac{1}{2} (\bar{\ell}_4 + 2\bar{\ell}_{3,1} + \bar{\ell}_{2,2}) (\hat{\theta} - \theta)^2 \\ &- \frac{1}{6} (\bar{\ell}_5 + 3\bar{\ell}_{4,1} + 3\bar{\ell}_{3,2} + \bar{\ell}_{2,3}) (\hat{\theta} - \theta)^3 + \dots, \end{aligned} \quad (\text{A.9})$$

mediante las expresiones para el recíproco de la raíz de series de la Sección A.1.2, obtenemos que

$$\begin{aligned} \hat{j}^{-1/2} &= \bar{j}^{-1/2} + \frac{1}{2} (\bar{\ell}_3 + \bar{\ell}_{2,1}) \bar{j}^{-3/2} (\hat{\theta} - \theta) \\ &+ \frac{1}{8} \left( 3 (\bar{\ell}_3 + \bar{\ell}_{2,1})^2 + 2 (\bar{\ell}_4 + 2\bar{\ell}_{3,1} + \bar{\ell}_{2,2}) \bar{j} \right) \bar{j}^{-5/2} (\hat{\theta} - \theta)^2 + \dots \end{aligned}$$

Multiplicando (A.8) y (A.9) se obtiene la expresión para  $u$ , i.e.,

$$\begin{aligned} u &= \sqrt{\bar{j}} (\hat{\theta} - \theta) - \frac{1}{2} (\bar{\ell}_3 + 2\bar{\ell}_{2,1}) \bar{j}^{-1/2} (\hat{\theta} - \theta)^2 \\ &- \frac{1}{24} \left( 3 (\bar{\ell}_3^2 + \bar{\ell}_{2,1} (3\bar{\ell}_{2,1} + 4\bar{\ell}_3)) + 2 (3\bar{\ell}_4 + 10\bar{\ell}_{3,1} + 9\bar{\ell}_{2,2}) \bar{j} \right) \bar{j}^{-3/2} (\hat{\theta} - \theta)^3 + \dots \end{aligned}$$

Sustituyendo  $A$  por (A.7) se sigue que

$$\begin{aligned} u &= r - \frac{1}{6} (\bar{\ell}_3 + 3\bar{\ell}_{2,1}) \bar{j}^{-3/2} r^2 \\ &- \frac{1}{72} (13\bar{\ell}_3^2 + 6\bar{\ell}_{2,1} (9\bar{\ell}_{2,1} + 10\bar{\ell}_3) + 9 (\bar{\ell}_4 + 4\bar{\ell}_{3,1} + 4\bar{\ell}_{2,2}) \bar{j}) \bar{j}^{-3} r^3 + \dots \end{aligned}$$

entonces, por Sección A.1.3 se tiene que,

$$\begin{aligned} \frac{r}{u} &= 1 + \frac{1}{6} (\bar{\ell}_3 + 3\bar{\ell}_{2,1}) \bar{j}^{-3/2} r \\ &+ \frac{1}{24} (5\bar{\ell}_3^2 + 24\bar{\ell}_{2,1} (\bar{\ell}_{2,1} + \bar{\ell}_3) + (3\bar{\ell}_4 + 12 (\bar{\ell}_{3,1} + \bar{\ell}_{2,2})) \bar{j}) \bar{j}^{-3} r^2 + \dots \\ &= 1 - \bar{t}_3 + \frac{1}{2} b r^2 + \dots \end{aligned} \quad (\text{A.10})$$

Por lo tanto

$$\begin{aligned} r^* &= r - \frac{1}{r} \ln (1 - \bar{t}_3 r + \frac{1}{2} b r^2 + \dots) \\ &= r + \bar{t}_3 - \frac{1}{2} (b - \bar{t}_3^2) r + \dots \end{aligned} \quad (\text{A.11})$$

### A.2.2 Parámetro $\theta = (\psi, \chi)$ , $\psi$ escalar

En esta sección consideramos la expansión asintótica de la verosimilitud dirigida alrededor de  $\psi$  y alrededor de  $\hat{\psi}$ , así como las indicaciones necesarias para obtener el factor de Bartlett cuando el modelo contiene parámetros de estorbo. Los cálculos son similares a los del caso  $\theta$  escalar (Sección A.2.1). Sin embargo hay que tomar en cuenta las indicaciones al inicio de las siguientes dos secciones. Las herramientas del análisis que más se utilizan aquí son la Regla de la cadena y el Teorema de la función implícita (Marsden & Hoffman 1998, secciones 6.5 y 7.2).

### Expansión de la verosimilitud dirigida alrededor de $\hat{\psi}$

Para tener aproximaciones los extremos del intervalo de verosimilitud de tal manera que la probabilidad de cobertura de dicha aproximación sea de orden  $O(n^{-2})$ , es necesario calcular las primeras cuatro derivadas de la función de log-verosimilitud perfil. A continuación daremos indicaciones para evaluar estas derivadas.

Mediante la Regla de la cadena obtenemos que

$$\frac{\partial}{\partial \psi} l_p(\psi) = \frac{\partial}{\partial \psi} l(\psi, \hat{\chi}_\psi) = \frac{\partial}{\partial \psi} l(\psi, \chi) \Big|_{(\psi, \chi) = (\psi, \hat{\chi}_\psi)} + \nabla_x^T l(\psi, \chi) \Big|_{(\psi, \chi) = (\psi, \hat{\chi}_\psi)} \frac{\partial}{\partial \psi} \hat{\chi}_\psi \quad (\text{A.12})$$

donde  $\frac{\partial}{\partial \psi} \hat{\chi}_\psi = \left( \frac{\partial}{\partial \psi} \hat{\chi}_{1,\psi}, \dots, \frac{\partial}{\partial \psi} \hat{\chi}_{d-1,\psi} \right)^T$ . Para calcular  $\frac{\partial}{\partial \psi} \hat{\chi}_\psi$  se requiere del el Teorema de la función implícita aplicado a la ecuación que define a  $\hat{\chi}_\psi$ , la cual es

$$l_x(\psi, \hat{\chi}_\psi) = 0, \quad (\text{A.13})$$

derivando (A.13) con respecto a  $\psi$  obtenemos que

$$\frac{\partial}{\partial \psi} l_x(\psi, \hat{\chi}_\psi) = \frac{\partial}{\partial \psi} l_x(\psi, \chi) \Big|_{(\psi, \chi) = (\psi, \hat{\chi}_\psi)} + \left[ \nabla^T l_x(\psi, \chi) \Big|_{(\psi, \chi) = (\psi, \hat{\chi}_\psi)} \right] \left[ \frac{\partial}{\partial \psi} \hat{\chi}_\psi \right] = 0,$$

en consecuencia

$$\frac{\partial}{\partial \psi} \hat{\chi}_\psi = - \left[ \frac{\partial}{\partial \psi} l_x(\psi, \chi) \Big|_{(\psi, \chi) = (\psi, \hat{\chi}_\psi)} \right] \left[ \nabla^T l_x(\psi, \chi) \Big|_{(\psi, \chi) = (\psi, \hat{\chi}_\psi)} \right]^{-1}, \quad (\text{A.14})$$

utilizando la notación de la Sección 1.3.1 la ecuación (A.14) se reduce a

$$\begin{aligned} \frac{\partial}{\partial \psi} \hat{\chi}_\psi &= -l_{xx}(\psi, \hat{\chi}_\psi)^{-1} l_{\psi x}(\psi, \hat{\chi}_\psi) \\ &= -\tilde{l}_{xx}^{-1} \tilde{l}_{\psi x}. \end{aligned}$$

Por medio de la notación simplificada y con ayuda de (A.13) la ecuación (A.12) se reduce a

$$\frac{\partial}{\partial \psi} l_p(\psi) = l_\psi(\psi, \hat{\chi}_\psi) = \tilde{l}_\psi,$$

Mediante el mismo procedimiento anterior, tenemos que la segunda derivada de la función de log-verosimilitud es

$$\begin{aligned} \frac{\partial^2}{\partial \psi^2} l_p(\psi) &= \frac{\partial}{\partial \psi} \left[ \frac{\partial}{\partial \psi} l_p(\psi) \right] \\ &= \tilde{l}_{\psi\psi} + \tilde{l}_{\psi x}^T \frac{\partial}{\partial \psi} \hat{\chi}_\psi \\ &= \tilde{l}_{\psi\psi} - \tilde{l}_{\psi x}^T \tilde{l}_{xx}^{-1} \tilde{l}_{\psi x} \end{aligned}$$

De manera similar obtenemos la siguientes derivadas.

Sean

$$\hat{j}_p = -\hat{l}_2 = -\frac{\partial^2}{\partial \psi^2} l_p(\psi) \Big|_{\psi = \hat{\psi}}, \quad \hat{l}_3 = -\frac{\partial^3}{\partial \psi^3} l_p(\psi) \Big|_{\psi = \hat{\psi}}, \quad \hat{l}_4 = \frac{\partial^4}{\partial \psi^4} l_p(\psi) \Big|_{\psi = \hat{\psi}},$$

entonces

$$l_p = \hat{l}_p - \frac{1}{2} \hat{j}_p (\psi - \hat{\psi})^2 + \frac{1}{6} \hat{l}_3 (\psi - \hat{\psi})^3 + \frac{1}{24} \hat{l}_4 (\psi - \hat{\psi})^4 + \dots,$$

de donde obtenemos que

$$r^2 = 2(\hat{l}_p - l_p) = \hat{j}_p (\hat{\psi} - \psi)^2 + \frac{1}{3} \hat{l}_3 (\hat{\psi} - \psi)^3 - \frac{1}{12} \hat{l}_4 (\hat{\psi} - \psi)^4 + \dots$$

Mediante la raíz de series (Sección A.1.2) obtenemos que

$$r = j^{1/2} (\hat{\psi} - \psi) + \frac{1}{6} \hat{l}_3 \hat{j}_p^{-1/2} (\hat{\psi} - \psi)^2 - \frac{1}{72} (3\hat{l}_4 \hat{j}_p + \hat{l}_3^2) \hat{j}_p^{-3/2} (\hat{\psi} - \psi)^3 + \dots$$

Invirtiendo la serie anterior tenemos que los extremos del intervalo de verosimilitud son

$$\hat{\psi} - \psi = \pm \hat{j}_p^{-1/2} r - \frac{1}{6} \hat{l}_3 \hat{j}_p^{-2} r^2 \pm \frac{1}{72} (3\hat{l}_4 \hat{j}_p + 5\hat{l}_3^2) \hat{j}_p^{-7/2} r^3 + \dots$$

Expansión de  $r$  y  $u$  alrededor de  $\psi$

En este caso los valores de  $\psi$ ,  $\hat{\chi}_\psi$  y  $a$  se mantienen constantes lo que obliga a  $\hat{\chi}$  a ser función de  $\hat{\psi}$ . La relación funcional de  $\hat{\chi}$  con  $\hat{\psi}$  se define igual que (A.13), pero a hora se debe resolver con respecto a  $\hat{\chi}$ .

Necesitamos derivar la función de log-verosimilitud perfil con respecto a  $\hat{\psi}$ . Mediante la regla de la cadena obtenemos que

$$\frac{\partial}{\partial \hat{\psi}} \ell(\psi, \hat{\chi}_\psi; \hat{\psi}, \hat{\chi}) = \ell_{;\hat{\psi}}(\psi, \hat{\chi}_\psi; \hat{\psi}, \hat{\chi}) + \ell_{;\hat{\chi}}^T(\psi, \hat{\chi}_\psi; \hat{\psi}, \hat{\chi}) \frac{\partial}{\partial \hat{\psi}} \hat{\chi}$$

donde  $\frac{\partial}{\partial \hat{\psi}} \hat{\chi}$  se obtiene del Teorema de la función implícita mediante derivación de

$$\ell_{;\hat{\chi}}(\psi, \hat{\chi}_\psi; \hat{\psi}, \hat{\chi}) = 0$$

con respecto a  $\hat{\psi}$ , manteniendo  $\psi$ ,  $\hat{\chi}_\psi$  y  $a$  fijos. Entonces

$$\bar{\ell}_{\hat{\chi};\hat{\psi}} + \bar{\ell}_{\hat{\chi};\hat{\chi}} \frac{\partial}{\partial \hat{\psi}} \hat{\chi} = 0$$

lo que implica que

$$\frac{\partial}{\partial \hat{\psi}} \hat{\chi} = -\bar{\ell}_{\hat{\chi};\hat{\chi}}^{-1} \bar{\ell}_{\hat{\chi};\hat{\psi}}$$

De manera similar que en la Sección A.2.1.2 iniciamos con la expresión

$$r^2 = 2(\hat{\ell}_p - \ell_p) = \hat{j}_p (\hat{\psi} - \psi)^2 + \frac{1}{3} \hat{l}_3 (\hat{\psi} - \psi)^3 - \frac{1}{12} \hat{l}_4 (\hat{\psi} - \psi)^4 + \dots,$$

y procedemos a expandir alrededor de  $\psi$  los coeficientes de las potencias de  $\hat{\psi} - \psi$ . Para hacer esto es necesario primero definir las siguientes cantidades:

$$\begin{aligned} \bar{j}_p &= \hat{j}_p|_{\hat{\psi}=\psi}, \quad \bar{l}_3 = \hat{l}_3|_{\hat{\psi}=\psi}, \quad \bar{l}_4 = \hat{l}_4|_{\hat{\psi}=\psi}, \\ \bar{l}_{2;1} &= \frac{\partial}{\partial \hat{\psi}} \hat{\ell}_2|_{\hat{\psi}=\psi}, \quad \bar{l}_{2;2} = \frac{\partial^2}{\partial \hat{\psi}^2} \hat{\ell}_2|_{\hat{\psi}=\psi}, \quad \bar{l}_{3;1} = \frac{\partial}{\partial \hat{\psi}} \hat{\ell}_3|_{\hat{\psi}=\psi} \end{aligned}$$

Entonces

$$\begin{aligned} r^2 &= \left( \bar{j}_p - \bar{l}_{2,1} (\hat{\psi} - \psi) - \frac{1}{2} \bar{l}_{2,2} (\hat{\psi} - \psi)^2 \right) (\hat{\psi} - \psi)^2 \\ &\quad + \frac{1}{3} (\bar{l}_3 + \bar{l}_{3,1} (\hat{\psi} - \psi)) (\hat{\psi} - \psi)^3 - \frac{1}{12} \bar{l}_4 (\hat{\psi} - \psi)^4 + \dots \\ &= \bar{j}_p (\hat{\psi} - \psi)^2 + \frac{1}{3} (\bar{l}_3 - 3\bar{l}_{2,1}) (\hat{\psi} - \psi)^3 - \frac{1}{12} (6\bar{l}_{2,2} - 4\bar{l}_{3,1} + \bar{l}_4) (\hat{\psi} - \psi)^4 + \dots \end{aligned}$$

Mediante raíz de series (Sección A.1.2) obtenemos que

$$\begin{aligned} r &= j_p^{1/2} (\hat{\psi} - \psi) + \frac{1}{6} (\bar{l}_3 - 3\bar{l}_{2,1}) \bar{j}_p^{-1/2} (\hat{\psi} - \psi)^2 \\ &\quad - \frac{1}{72} \left[ 3(6\bar{l}_{2,2} - 4\bar{l}_{3,1} + \bar{l}_4) \bar{j}_p + (\bar{l}_3 - 3\bar{l}_{2,1})^2 \right] \bar{j}_p^{-3/2} (\hat{\psi} - \psi)^3 + \dots \end{aligned} \quad (A.15)$$

Sea

$$d_1 = \frac{1}{6} (\bar{l}_3 - 3\bar{l}_{2,1}) \bar{j}_p^{-1/2} \quad \text{y} \quad d_2 = -\frac{1}{72} \left[ 3(6\bar{l}_{2,2} - 4\bar{l}_{3,1} + \bar{l}_4) \bar{j}_p + (\bar{l}_3 - 3\bar{l}_{2,1})^2 \right] \bar{j}_p^{-3/2}.$$

Al invertir (A.15) (ver Sección A.1.3) se llega a

$$\hat{\psi} - \psi = \bar{j}_p^{-1/2} r - d_1 \bar{j}_p^{-3/2} r^2 - (d_2 \bar{j}_p^{1/2} - 2d_1^2) \bar{j}_p^{-5/2} r^3 + \dots \quad (A.16)$$

La expansión de  $u$  alrededor de  $\psi$  es de la forma

$$u = \bar{j}_p^{1/2} (\hat{\psi} - \psi) + u_2 (\hat{\psi} - \psi)^2 + u_3 (\hat{\psi} - \psi)^3 + \dots, \quad u_k = O(n^{1/2}), \quad k \geq 1, \quad (A.17)$$

de donde, al sustituir (A.16) en (A.17) se desprende que

$$u = r + (u_2 - d_1) \bar{j}_p^{-1} r^2 + (2d_1(d_1 - u_2) + (u_3 - d_2) \bar{j}_p^{1/2}) \bar{j}_p^{-2} r^3 + \dots$$

Mediante la fórmula de recíproco de la Sección A.1 la expansión de  $r/u$ , alrededor de  $r = 0$ , se expresa como

$$\frac{r}{u} = 1 + (d_1 - u_2) \bar{j}_p^{-1} r + [u_2^2 - d_1^2 + (d_2 - u_3) \bar{j}_p^{1/2}] \bar{j}_p^{-2} r^2 + \dots$$

Entonces, el factor de Bartlett en este caso es

$$b = 2 [u_2^2 - d_1^2 + (d_2 - u_3) \bar{j}_p^{1/2}] \bar{j}_p^{-2}.$$

## Apéndice B

### Obtención de $u$ en el modelo exponencial- $(m, m)$

Considere el modelo exponencial- $(m, m)$  con función de log-verosimilitud  $\ell = s^T \phi - k(\phi)$ ,  $\phi$  es el parámetro canónico,  $\eta = E(S; \phi)$  y la función cumulante es  $k(\phi)$ .

Las ecuaciones de verosimilitud satisfacen que

$$\ell_\psi = (s - \eta)^T \phi_{/\psi} = 0,$$

$$\ell_x = (s - \eta)^T \phi_{/x} = 0,$$

bajo condiciones de regularidad estándar se tiene que  $s = \nabla k(\hat{\phi})$ . Entonces la función de log-verosimilitud se expresa como

$$\ell = \nabla k(\hat{\phi})^T \phi - k(\phi) \quad (\text{B.1})$$

de (B.1) se obtienen las siguientes expresiones

$$\left. \begin{aligned} l_{\hat{\psi}} &= \hat{\phi}_{/\hat{\psi}}^T \hat{\Sigma} \phi, & l_{\chi; \hat{\psi}} &= \hat{\phi}_{/\hat{\psi}}^T \hat{\Sigma} \phi_{/\chi}, & l_{\hat{\chi}} &= \hat{\phi}_{/\hat{\chi}}^T \hat{\Sigma} \phi, & l_{\chi; \hat{\chi}} &= \hat{\phi}_{/\hat{\chi}}^T \hat{\Sigma} \phi_{/\chi}, \\ l_{\psi\psi} &= (s - \eta)^T \phi_{/\psi\psi} - \phi_{/\psi}^T \Sigma \phi_{/\psi}, & l_{\chi\chi} &= (s - \eta)^T \phi_{/\chi\chi} - \phi_{/\chi}^T \Sigma \phi_{/\chi}, \\ l_{\psi\chi} &= l_{\chi\psi} = (s - \eta)^T \phi_{/\psi\chi} - \phi_{/\psi}^T \Sigma \phi_{/\chi}. \end{aligned} \right\} \quad (\text{B.2})$$

A partir de ahora usaremos las siguientes notaciones

$$\nabla \nabla^T k(\phi) = \text{var}(S) = \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \quad \text{and} \quad \Delta = \Sigma^{-1} = \begin{bmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22} \end{bmatrix}. \quad (\text{B.3})$$

En las siguientes secciones calcularemos  $u$  de (1.17) para diferentes funciones del parámetro canónico.

I.-Si el parámetro de interés es  $\psi = h(\phi)$ , y tomamos el parámetro de estorbo igual a  $\chi = \phi_2$  las expresiones en (B.2) son

$$\begin{aligned} l_{\hat{\psi}} &= \hat{\phi}_{1/\hat{\psi}} \hat{\Sigma}_{11} \phi_1 + \hat{\phi}_{1/\hat{\psi}} \hat{\Sigma}_{12} \phi_2, & l_{\hat{\chi}} &= \hat{\phi}_{1/\hat{\chi}} \hat{\Sigma}_{11} \phi_1 + \hat{\Sigma}_{21} \phi_1 + \hat{\phi}_{1/\hat{\chi}} \hat{\Sigma}_{12} \phi_2 + \hat{\Sigma}_{22} \phi_2, \\ l_{\chi; \hat{\psi}} &= \hat{\phi}_{1/\hat{\psi}} \hat{\Sigma}_{11} \phi_{1/\chi} + \hat{\phi}_{1/\hat{\psi}} \hat{\Sigma}_{12}, & l_{\chi; \hat{\chi}} &= \hat{\phi}_{1/\hat{\chi}} \hat{\Sigma}_{11} \phi_{1/\chi} + \hat{\Sigma}_{21} \phi_{1/\chi} + \hat{\phi}_{1/\hat{\chi}} \hat{\Sigma}_{12} + \hat{\Sigma}_{22}, \\ l_{\psi\psi} &= (s_1 - \eta_1) \phi_{1/\psi\psi} - \phi_{1/\psi}^2 \Sigma_{11}, & l_{\psi\chi} &= l_{\chi\psi} = (s_1 - \eta_1) \phi_{1/\psi\chi} - \phi_{1/\psi} (\Sigma_{11} \phi_{1/\chi} + \Sigma_{12}), \\ l_{\chi\chi} &= (s_1 - \eta_1) \phi_{1/\chi\chi} - \phi_{1/\chi} (\phi_{1/\chi} \Sigma_{11} + \Sigma_{21} + \phi_{1/\chi} \Sigma_{12}) + \Sigma_{22}. \end{aligned} \quad (\text{B.4})$$

Por ejemplo si  $\psi = h(\phi) = \phi_1$  y  $\chi = \phi_2$ , las relaciones en (B.4) se reducen a

$$\begin{aligned} l_{\psi} &= s_1 - k_{/\psi}, & l_{\chi} &= s_2 - k_{/\chi}, & l_{\psi\psi} &= -\Sigma_{11}, \\ l_{\chi\chi} &= -\Sigma_{22}, & l_{\psi\chi} &= -\Sigma_{12} = l_{\chi\psi} = -\Sigma_{21}, & l_{\chi; \hat{\chi}} &= \hat{\Sigma}_{22}, \\ l_{\chi; \hat{\psi}} &= \hat{\Sigma}_{21}, & l_{\hat{\psi}} &= \hat{\Sigma}_{11} \psi + \hat{\Sigma}_{12} \chi, & l_{\hat{\chi}} &= \hat{\Sigma}_{12} \psi + \hat{\Sigma}_{22} \chi, \end{aligned}$$

se sigue que

$$\hat{j} = \hat{\Sigma}_{11} \hat{\Sigma}_{22} - \hat{\Sigma}_{12} \hat{\Sigma}_{21} \quad \text{and} \quad \hat{j}_{\chi\chi} = \Sigma_{22}. \quad (\text{B.5})$$

Dado que

$$\begin{aligned} \begin{vmatrix} \hat{l}_{\hat{\psi}} - \bar{l}_{\hat{\psi}} & \bar{l}_{\chi; \hat{\psi}} \\ \hat{l}_{\hat{\chi}} - \bar{l}_{\hat{\chi}} & \bar{l}_{\chi; \hat{\chi}} \end{vmatrix} &= (\hat{\Sigma}_{22} \hat{\Sigma}_{11} - \hat{\Sigma}_{21} \hat{\Sigma}_{12}) (\hat{\psi} - \psi) \\ &= \hat{j} (\hat{\psi} - \psi), \end{aligned}$$

se desprende que

$$u = \left| \frac{\hat{j}}{\hat{j}_{\chi\chi}} \right|^{1/2} (\hat{\psi} - \psi) \quad (\text{B.6})$$

II.-Si  $\psi = \phi_1$ ,  $\chi = h(\phi)$ , donde  $h$  es una función uno-a-uno de  $\phi$ , se sigue que las relaciones de (B.2) son ahora

$$\begin{aligned} l_{\psi\psi} &= (s_2 - \eta_2) \phi_{2/\psi\psi} - \Sigma_{11} - (\Sigma_{12} + \Sigma_{21}) \phi_{2/\psi} - \Sigma_{22} \phi_{2/\psi}^2, \\ l_{\chi\chi} &= (s_2 - \eta_2) \phi_{2/\chi\chi} - \Sigma_{22} \phi_{2/\chi}^2, & l_{\psi\chi} &= l_{\chi\psi} = (s_2 - \eta_2) \phi_{2/\psi\chi} - \Sigma_{12} \phi_{2/\chi} - \Sigma_{22} \phi_{2/\psi} \phi_{2/\chi}, \\ l_{\chi; \hat{\chi}} &= \hat{\phi}_{/\hat{\chi}}^T \hat{\Sigma} \phi_{/\chi}, & l_{\chi; \hat{\psi}} &= \hat{\phi}_{/\hat{\chi}}^T \phi_{/\chi}, & l_{\hat{\psi}} &= \hat{\phi}_{/\hat{\psi}}^T \hat{\Sigma} \phi, & l_{\hat{\chi}} &= \hat{\phi}_{/\hat{\chi}}^T \hat{\Sigma} \phi, \end{aligned}$$

y

$$\begin{aligned} s_{/\hat{\psi}} &= \hat{\Sigma} \hat{\phi}_{/\hat{\psi}}, & s_{/\hat{\chi}} &= \hat{\Sigma} \hat{\phi}_{/\hat{\chi}}, & \phi_{/\psi} &= (1, \phi_{2/\psi})^T, & \phi_{/\chi} &= (0, \phi_{2/\chi})^T, \\ \hat{\phi}_{/\hat{\psi}}^T \hat{\Sigma} &= \hat{\Sigma} \hat{\phi}_{/\hat{\psi}} = (\hat{\Sigma}_{11} + \hat{\Sigma}_{12} \hat{\phi}_{2/\hat{\psi}}, \hat{\Sigma}_{21} + \hat{\Sigma}_{22} \hat{\phi}_{2/\hat{\psi}})^T, & \hat{\phi}_{/\hat{\chi}}^T \hat{\Sigma} &= \hat{\Sigma} \hat{\phi}_{/\hat{\chi}} = (\hat{\Sigma}_{12} \hat{\phi}_{2/\hat{\chi}}, \hat{\Sigma}_{22} \hat{\phi}_{2/\hat{\chi}})^T. \end{aligned}$$

III.-Por ejemplo si  $\psi = \phi_1$  y  $\chi = h(\phi) = \eta_2(\phi)$  es una parametrización mixta (Barndorff-Nielsen & Cox 1994, pág. 62). El estimador de máxima verosimilitud  $\hat{\eta}_{2\psi}$  restringido a un valor fijo de  $\psi$  cumple que  $\hat{\eta}_{2\psi} = \hat{\eta}_2$ . Así, pues

$$\begin{aligned} s_{1/\hat{\psi}} &= \hat{\Delta}_{11}^{-1}, & s_{1/\hat{\chi}} &= \hat{\Sigma}_{12} \hat{\Sigma}_{22}^{-1}, & s_{2/\hat{\psi}} &= \hat{\Sigma}_{21}, & s_{2/\hat{\chi}} &= I, & \phi_{2/\chi} &= \Sigma_{22}^{-1}, \\ \phi_{2/\psi} &= -\Sigma_{22}^{-1} \Sigma_{21}, & l_{\hat{\psi}} &= \hat{\Delta}_{11}^{-1} \psi, & l_{\hat{\chi}} &= \hat{\Sigma}_{12} \hat{\Sigma}_{22}^{-1} \psi + \chi, & l_{\chi; \hat{\chi}} &= \Sigma_{22}^{-1}, & l_{\chi; \hat{\psi}} &= 0, \\ \bar{l}_{\psi\chi} &= 0, & \bar{l}_{\chi\chi} &= -\bar{\Sigma}_{22}^{-1}, & \bar{l}_{\psi\psi} &= -\bar{\Delta}_{11}^{-1}, \end{aligned}$$

además  $|j| = |\hat{\Delta}_{11}^{-1} \hat{\Sigma}_{22}^{-1}|$ . entonces

$$u = \frac{|\hat{\Sigma}_{22}|^{1/2}}{|\hat{\Delta}_{11}|^{1/2}} |\hat{\Delta}_{11}|^{1/2} (\hat{\psi} - \psi). \quad (B.7)$$

IV.-Si  $\psi = h(\phi) = \eta_1(\phi)$  y  $\chi = \phi_2$  y notando que  $\hat{j} = |\hat{\Sigma}_{11}^{-1} \hat{\Delta}_{22}^{-1}|$  y

$$\begin{aligned} s_{1/\hat{\psi}} &= I, & s_{1/\hat{\chi}} &= \hat{\Sigma}_{12}, & s_{2/\hat{\psi}} &= -\hat{\Sigma}_{21} \Sigma_{11}^{-1}, & s_{2/\hat{\chi}} &= \hat{\Sigma}_{22}, & \phi_{1/\psi} &= \Sigma_{11}^{-1}, \\ \phi_{1/\chi} &= -\Sigma_{11}^{-1} \Sigma_{12}, & \ell_{\hat{\psi}} &= \phi_1(\psi, \chi) + \hat{\Sigma}_{21} \hat{\Sigma}_{11}^{-1} \chi, & \ell_{\hat{\chi}} &= \hat{\Delta}_{22}^{-1} \chi, & \tilde{\ell}_{\chi/\hat{\psi}} &= \hat{\Sigma}_{11}^{-1} \hat{\Sigma}_{12} - \tilde{\Sigma}_{11}^{-1} \tilde{\Sigma}_{12}, \\ \ell_{\chi/\hat{\chi}} &= \hat{\Delta}_{22}^{-1}, & \tilde{j}_{\chi\chi} &= \tilde{\Delta}_{22}^{-1}, & \ell_{\chi\psi} &= 0, & \tilde{j}_{\psi\psi} &= \Sigma_{11}^{-1}, \end{aligned}$$

se llega a que

$$u = \frac{|\tilde{\Delta}_{22}|^{1/2} |\hat{\Sigma}_{11}|}{|\hat{\Delta}_{22}|^{1/2}} \left[ \phi_1(\hat{\psi}, \hat{\chi}) - \phi_1(\psi, \hat{\chi}_\psi) + \tilde{\Sigma}_{11}^{-1} \tilde{\Sigma}_{12} (\hat{\chi} - \hat{\chi}_\psi) \right]. \quad (B.8)$$

## Apéndice C

### Distribución gaussiana inversa normal

En esta sección se ilustra como surge la función de densidad de la distribución gaussiana inversa normal.

Una de las representaciones de la función de Bessel del tercer tipo es

$$K_\nu(z) = \frac{1}{2} \left(\frac{z}{2}\right)^\nu \int_0^\infty t^{-\nu-1} e^{-t-\frac{z^2}{4t}} dt. \quad (C.1)$$

(Gradshteyn & Ryzhik 1994, ecuación 8.432-6). La densidad de la distribución gaussiana inversa está dada por

$$p_z(z; \chi, \psi) = \frac{\sqrt{\chi} e^{\sqrt{\chi\psi}}}{\sqrt{2\pi}} z^{-3/2} e^{-\frac{1}{2}(\frac{z}{z} + \psi z)}.$$

Sea  $Y$  una variable aleatoria tal que la densidad de  $Y|Z$  es

$$p_{Y|Z}(y; \mu, \beta, z) = \frac{1}{\sqrt{2\pi z}} e^{-\frac{1}{2z}(y-\mu-\beta z)^2}.$$

Nos interesa encontrar la función de densidad marginal de  $Y$ . Con este fin, consideremos la reparametrización  $\chi = \delta^2$  y  $\psi = \alpha^2 - \beta^2$ . Entonces

$$p_z(z; \delta^2, \alpha^2 - \beta^2) = \frac{\delta e^{\delta\sqrt{\alpha^2 - \beta^2}}}{\sqrt{2\pi}} z^{-3/2} e^{-\frac{1}{2}[\frac{\delta^2}{z} + (\alpha^2 - \beta^2)z]}.$$

Así, la densidad marginal de  $Y$  se calcula al efectuar la siguiente integral

$$\int_0^{\infty} p_z(z; \delta^2, \alpha^2 - \beta^2) p_{Y|Z}(y; \mu, \beta, z) dz,$$

luego,

$$\begin{aligned} p_Y(y; \mu, \delta, \alpha, \beta) &= \frac{\delta}{2\pi} e^{\delta\sqrt{\alpha^2-\beta^2}} \int_0^{\infty} z^{-2} e^{-\frac{1}{2}(\delta^2 z^{-1} + (\alpha^2 - \beta^2)z) - \frac{1}{2z}(y - \mu - \beta z)^2} dz \\ &= \frac{\delta}{2\pi} e^{\delta\sqrt{\alpha^2-\beta^2}} e^{\beta(y-\mu)} \int_0^{\infty} z^{-2} e^{-\frac{1}{2z}(\delta^2 + (y-\mu)^2) - \frac{1}{2}z\alpha^2} dz \\ &= \frac{\alpha^2 \delta}{4\pi} e^{\delta\sqrt{\alpha^2-\beta^2}} e^{\beta(y-\mu)} \int_0^{\infty} t^{-2} e^{-t - \frac{1}{4t}\alpha^2(\delta^2 + (y-\mu)^2)} dt, \end{aligned}$$

en el último paso se realizó el cambio de variable  $t = \frac{1}{2}z\alpha^2$ . Ahora, utilizando (C.1) llegamos a que

$$p_Y(y; \mu, \delta, \alpha, \beta) = \frac{\delta\alpha}{\pi} e^{\delta\sqrt{\alpha^2-\beta^2} + \beta(y-\mu)} \frac{K_1\left(\alpha\sqrt{\delta^2 + (y-\mu)^2}\right)}{\sqrt{\delta^2 + (y-\mu)^2}}. \quad (C.2)$$

La densidad (C.2) se conoce como la *densidad de la distribución gaussiana inversa normal*.

## Bibliografía

- Abramowitz, M. y Stegun, I. A. (1965). *Handbook of Mathematical Functions*. New York, Dover.
- Barnard, G. A. (1965). The use of the likelihood function in statistical practice. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* 1, 27-40.
- Barnard, G. A. (1996). Scientific practice and statistical inference. *Symposium on the foundation of statistical inference, with special emphasis on applications. In honour of Professor David A. Sprott, October 3-4, 1996. University of Waterloo, Waterloo, Ontario, Canada*.
- Barnard, G. A. y Sprott, D. A. (1983). Likelihood. *Encyclopedia of Statistical Sciences*. Vol. 4. pp. 639-644. Editors-in-chief S. Kotz & N.L. Johnson. Associate editor: C.B. Read.
- Barndorff-Nielsen, O. E. (1977). Exponentially decreasing distributions for the logarithm of particle size. *Proc. Roy. Soc. London. A* Vol. 353, pp. 401-419.
- Barndorff-Nielsen, O. E. (1982). Hyperbolic likelihood. *Statistic and Probability. Essays in Honour of C.R. Rao*. G. Kallianpur, P.R. Krishnaiah, and J.K. Gosh, editors, North-Holland Publishing Company, pp. 67-76.

- Barndorff-Nielsen, O. E. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika*, Vol. 70, pp. 343-365.
- Barndorff-Nielsen, O. E. (1985). Properties of modified profile likelihood. In *Contributions to Probability and Statistics in Honour of Gunnar Blom* (eds. J. Lanke and G. Lindgren). Dept. Math. Statist., Lund University, pp. 25-38.
- Barndorff-Nielsen, O. E. (1986). Inference on full and partial parameters based on the standardized signed log likelihood ratio. *Biometrika*, Vol. 73, pp. 307-322.
- Barndorff-Nielsen, O. E. (1990). Likelihood theory. In *Statistical Theory and Modelling. Essays in Honour of Sir David Cox, FRS*. Edited by D.V. Hinkley, N. Reid and E.J. Snell. Chapman and Hall.
- Barndorff-Nielsen, O. E. (1991). Modified signed log likelihood ratio. *Biometrika*, Vol. 78, pp. 557-563.
- Barndorff-Nielsen, O. E. y Cox, D. R. (1984). Bartlett adjustments to the likelihood ratio statistic and the distribution of the maximum likelihood estimator. *J. Roy. Statist. Soc. Ser. B* Vol. 46, pp. 483-495.
- Barndorff-Nielsen, O. E. y Cox, D. R. (1994). *Inference and Asymptotics*. Monographs on Statistics and Applied Probability 52. Chapman & Hall.
- Barndorff-Nielsen, O. E. y Hall, P. (1988). On the level-error after Bartlett adjustment of the likelihood ratio statistic. *Biometrika*, Vol. 75, pp. 374-378.
- Bartlett (1937) Properties of sufficiency and statistical tests. *Proc. Roy. Soc. (London) A*, Vol. 160, pp. 268-282.

- Chamberlin, S.R. y Sprott, D. A. (1991). Inferential estimation, likelihood, and maximum likelihood linear estimation functions. *Estimating Functions*, pp. 225-266, Editor, V. P. Godambe, Oxford Press.
- Díaz-Francis, E. (1998). *Scientific Application of Maximum Likelihood in Multiparametric Problems*. Doctor of Science dissertation, Centro de Investigación en Matemáticas, Guanajuato, Gto. México.
- Díaz-Francis, E. y Sprott, D. A. (2000). The Use of the Likelihood Function in the Analysis of Environmental Data. *Environmetrics*, Vol. 11, pp. 75-97.
- DiCiccio, T. J. (1986). Approximate conditional inference for location families. *Canadian J. of Statist.* Vol. 14 pp. 5-18.
- DiCiccio, T. J. (1993). Simple modifications for signed roots of likelihood ratio statistics. *J. R. Statist. Soc. B*. Vol. 55, No 1, pp. 305-316.
- Efron B. (1998). R. A. Fisher in the 21st century. *Statistical Science*, Vol. 13 No 2.
- Efron, B. y Hinkley, D. V. (1978) Assessing the accuracy of the maximum likelihood estimator: observed versus expected Fisher Information. *Biometrika* Vol. 65, pp. 457-487.
- Fisher, R. A. (1953). Dispersion on a sphere. *Proy. Roy. Soc. Lond. A* Vol. 217, pp. 295-305.
- Fisher, R. A. (1993, posthumous edition). *Statistical Methods and Scientific Inference*, 3rd ed. This book is in *Statistical Methods, Experimental Design, and Scientific Inference*. Oxford Science Publications.

- Land, C. E. (1971). Confidence intervals for linear functions of normal mean and variance. *Ann. math. Statist.* Vol. 42, pp. 1187-1205.
- Land, C. E. (1972). An evaluation of approximate confidence interval estimation methods for lognormal means. *Technometrics* Vol. 14, pp. 145-158.
- Lawley, D. N. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika*. Vol. 43, pp. 295-303.
- Lugannani, R. y Rice, S. (1980). Saddlepoint approximation for the distribution of the sum of indeendent random variables. *Adv. Appl. Prob.*, Vol. 12, pp. 475-490.
- Mardia, K. V. (1972). *Statistics of Directional Data*. Academic Press inc.
- Marsden, J. E. y Hoffman, M. J. (1998). *Análisis Clásico Elemental*. Addison-Wesley Iberoamericana.
- Meeker, W. Q. y Escobar, L. A. (1998). *Statistical Methods for Reliability Data*. Wiley series in Probability and Statistics.
- Reid, N. (1993). Making higher order asymptotics useful. Proceedings of the Congrese on Statistical Inference and Biostatistics. CIMAT, Guanajuato, March pp. 22-26, 1993.
- Reid, N. (1995). The roles of conditioning in inference. *Statistical Science*. Vol 10, No. 2, pp. 138-199.
- Reid, N. y Fraser, D. A. S. (2000). Higher order asymptotics: costs and benefits. *Statistics for the 21st Century*. Edited by C.R. Rao and G.J. Székely, pp. 351-365.

- Land, C. E. (1971). Confidence intervals for linear functions of normal mean and variance. *Ann. math. Statist.* Vol. 42, pp. 1187-1205.
- Land, C. E. (1972). An evaluation of approximate confidence interval estimation methods for lognormal means. *Technometrics* Vol. 14, pp. 145-158.
- Lawley, D. N. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika*. Vol. 43, pp. 295-303.
- Lugannani, R. y Rice, S. (1980). Saddlepoint approximation for the distribution of the sum of indeendent random variables. *Adv. Appl. Prob.*, Vol. 12, pp. 475-490.
- Mardia, K. V. (1972). *Statistics of Directional Data*. Academic Press inc.
- Marsden, J. E. y Hoffman, M. J. (1998). *Análisis Clásico Elemental*. Addison-Wesley Iberoamericana.
- Meeker, W. Q. y Escobar, L. A. (1998). *Statistical Methods for Reliability Data*. Wiley series in Probability and Statistics.
- Reid, N. (1993). Making higher order asymptotics useful. Proceedings of the Congrese on Statistical Inference and Biostatistics. CIMAT, Guanajuato, March pp. 22-26, 1993.
- Reid, N. (1995). The roles of conditioning in inference. *Statistical Science*. Vol 10, No. 2, pp. 138-199.
- Reid, N. y Fraser, D. A. S. (2000). Higher order asymptotics: costs and benefits. *Statistics for the 21st Century*. Edited by C.R. Rao and G.J. Székely, pp. 351-365.

Shuster J. J. (1968). On the inverse Gaussian distribution function. *J. Amer. Statist. Assoc.* Vol. **63**, pp. 1514-1516.

Sprott, D. A. (1973). Normal likelihood and their relation to large sample theory of estimation. *Biometrika*, Vol. **60**, pp. 457-465.

Sprott, D. A. (1975). Application of maximum likelihood methods to finite samples. *Sankhyā B*, Vol. **37**, pp. 259-270.

Sprott, D. A. (1980). Maximum likelihood in small samples: Estimation in the presence of nuisance parameters. *Biometrika*, Vol. **67**, pp. 515-523.

Sprott, D. A. (1984). Likelihood and maximum likelihood estimation. *Mathematical Reports of the Academy of Science*. Volume VI, No. 5.

Sprott, D. A. (1996). The likelihood function in scientific inference. *Symposium on the foundation of statistical inference, with special emphasis on applications. In honour of Professor David A. Sprott, October 3-4, 1996. University of Waterloo, Waterloo, Ontario, Canada.*

Sprott, D. A. (2000). *Statistical Inference in Science*. Springer Series in Statistics.

Viveros, R. (1985). *Estimation in Small Samples*. Ph. D. thesis, University of Waterloo.

Viveros, R. y Sprott, D. A. (1987). Allowance for skewness in maximum-likelihood estimation with application to the location-scale model. *The Canadian Journal of Statistics*, Vol. **15**, No. 4, pp. 349-361.

## Glosario

### Parametro $\theta$ escalar

En general se supone que existe una estadística  $a$  tal que  $y \longleftrightarrow (\hat{\theta}, a)$ . Si la distribución de  $a$  no depende de  $\theta$  se dice que  $a$  es auxiliar.

La función de verosimilitud se puede escribir como  $L(\theta; y) = L(\theta; \hat{\theta}, a)$ .

Función de log-verosimilitud:  $\ell(\theta) = \log L(\theta)$ .

Función de verosimilitud relativa:  $R(\theta) = L(\theta) / L(\hat{\theta})$ .

Función de verosimilitud dirigida:  $r(\theta; \hat{\theta}) = \text{sgn}(\hat{\theta} - \theta) \sqrt{-2 \ln R_p(\theta; \hat{\theta})}$ .

$\ell_{j;k}(\theta; \hat{\theta}, a) = \frac{\partial^{j+k}}{\partial \theta^j \partial \hat{\theta}^k} \ell(\theta; \hat{\theta}, a)$ ,  $j, k \geq 0$ , con las siguientes convenciones  $\ell_i \equiv \ell_{i;0}$ ,  $\ell_{;j} = \ell_{0;j}$ .

Además se utiliza un circunflejo “^” sobre  $\ell_{j;k}$  o  $j$  cuando se evalúa en  $\theta = \hat{\theta}$ ; i.e.,

$$\hat{\ell}_{j;k} = \ell_{j;k}(\hat{\theta}; \hat{\theta}, a), \quad \hat{j} = -\hat{\ell}_2,$$

y usamos una barra “-” para indicar que evaluamos en  $\hat{\theta} = \theta$ ; i.e.,

$$\bar{\ell}_{j;k} = \ell_{j;k}(\theta; \theta), \quad \bar{j} = -\bar{\ell}_2.$$

$$r^* = r + \frac{1}{r} \log(u/r).$$

$$u = j^{-1/2} (\bar{\ell}_{;1} - \ell_{;1}).$$

$$\frac{r}{u} = 1 - \bar{t}_3 r + \frac{1}{2} b r^2 + \dots$$

$$\bar{t}_3 = \frac{1}{6} (\bar{\ell}_3 + 3\bar{\ell}_{2;1}) \bar{j}^{-3/2}.$$

$$b(\theta, a, n) = \frac{1}{24n} \{5\bar{\ell}_3^2 + 24\bar{\ell}_{2;1}(\bar{\ell}_3 + \bar{\ell}_{2;1}) + [3\bar{\ell}_4 + 12(\bar{\ell}_{3;1} + \bar{\ell}_{2;2})] \bar{j}\} \bar{j}^{-3}.$$

$$b^-(a, \hat{\chi}_\psi, n) = \inf_{\omega \in \Omega_\psi} b(\omega, a, \hat{\chi}_\psi, n).$$

$$b^+(a, \hat{\chi}_\psi, n) = \sup_{\omega \in \Omega_\psi} b(\omega, a, \hat{\chi}_\psi, n).$$

**Parametro**  $\theta = (\psi, \chi)$ ;  $\psi$  escalar

Si  $\theta = (\psi, \chi)$ ,  $\psi$  es el parámetro de interés y  $\chi$  es el parámetro de estorbo. En este caso consideramos que el espacio paramétrico se descompone de la forma  $\Omega = \Omega_\psi \times \Omega_\chi$ .

En este caso  $L(\theta) = L(\psi, \chi) = L(\psi, \chi; \hat{\psi}, \hat{\chi}, a)$ .

Función de verosimilitud perfil de  $\psi$ :  $L_p(\psi) = L(\psi, \hat{\chi}_\psi; \hat{\psi}, \hat{\chi}, a)$ , donde  $\hat{\chi}_\psi$  es tal que

$$L(\psi; \hat{\chi}_\psi) \geq L(\psi, \chi) \text{ para todo } \chi \in \Omega_\chi.$$

Función de log-verosimilitud perfil de  $\psi$ :  $\ell_p(\psi) = \log L_p(\psi)$ .

Función de verosimilitud relativa perfil de  $\psi$ :  $R_p(\psi) = L_p(\psi) / L_p(\hat{\psi})$ .

Función de verosimilitud dirigida:  $r(\psi; \hat{\psi}) = \text{sgn}(\hat{\psi} - \psi) \sqrt{-2 \ln R_p(\psi; \hat{\psi})}$ .

$\theta^T$  = vector transpuesto de  $\theta$ .

$$\ell_\psi = \frac{\partial}{\partial \psi} \ell(\psi, \chi), \quad \ell_{\psi\psi} = \frac{\partial^2}{\partial \psi^2} \ell(\psi, \chi), \text{ etc.}$$

$$\ell_\chi = \nabla_\chi \ell(\psi, \chi) = (\ell_{\chi_1}, \dots, \ell_{\chi_{d-1}}) = \left( \frac{\partial}{\partial \chi_1} \ell(\psi, \chi), \dots, \frac{\partial}{\partial \chi_{d-1}} \ell(\psi, \chi) \right).$$

$$\ell_{\chi\chi} = \nabla_\chi^T \nabla_\chi \ell = (\ell_{\chi_i \chi_j})_{i,j=1}^{d-1} = \left( \frac{\partial^2}{\partial \chi_i \partial \chi_j} \ell_{\chi_i \chi_j}(\psi, \chi) \right)_{i,j=1}^{d-1}.$$

$$\ell_{;\hat{\psi}} = \frac{\partial}{\partial \hat{\psi}} \ell(\psi, \chi; \hat{\psi}, \hat{\chi}); \quad \ell_{;\hat{\chi}} = \nabla_{\hat{\chi}} \ell(\psi, \chi; \hat{\psi}, \hat{\chi}); \quad \ell_{\chi;\hat{\psi}} = \nabla_\chi \ell_{;\hat{\psi}}(\psi, \chi; \hat{\psi}, \hat{\chi}).$$

$$\ell_{\chi;\hat{\chi}} = \nabla_\chi^T \nabla_{\hat{\chi}} \ell(\psi, \chi; \hat{\psi}, \hat{\chi}), \quad \ell_{\psi\chi;\hat{\chi}} = \nabla_\chi^T \nabla_{\hat{\chi}} \ell_\psi(\psi, \chi; \hat{\psi}, \hat{\chi}), \text{ etc.}$$

$$\ell_{\theta\theta} = \nabla_\theta^T \nabla_\theta \ell = (\ell_{\theta_i \theta_j})_{i,j=1}^d = \left( \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ell_{\theta_i \theta_j}(\psi, \chi) \right)_{i,j=1}^d.$$

Se usa, sobre una función de  $(\psi, \chi)$ , una tilde “~” o un circunflejo “^” de acuerdo a si se evalúa en  $\hat{\chi}_\psi$  o en el  $\hat{\theta}$ . Así  $\bar{\ell} = \ell(\psi, \hat{\chi}_\psi; \hat{\theta}, a)$  y  $\hat{\ell} = \ell(\hat{\theta}; \hat{\theta}, a) = \ell(\hat{\psi}, \hat{\chi}; \hat{\psi}, \hat{\chi}, a)$ .

Cuando consideramos la función de verosimilitud  $\hat{\psi}$ , utilizamos una barra “-” sobre las derivadas de  $\ell$  para indicar evaluación en  $\hat{\psi} = \psi$ . Por ejemplo,  $\bar{\ell} = \ell(\psi, \hat{\chi}_\psi; \psi, \hat{\chi}_\psi)$ . La última expresión es porque cuando consideramos la función de verosimilitud en función de  $\hat{\psi}$ , tenemos que  $\hat{\chi}$  es función de  $\hat{\psi}$  y cuando  $\hat{\psi} = \psi$  se cumple que  $\hat{\chi}(\psi) = \hat{\chi}_\psi$ .

$$\hat{\chi}_{/\hat{\psi}} = -\bar{\ell}_{\chi;\hat{\psi}} \left( \bar{\ell}_{\chi;\hat{\chi}} \right)^{-1}.$$

$$j = -\ell_{\chi\chi} = \nabla \nabla^T \ell(\psi, \chi).$$

$$j_p = \frac{\partial^2}{\partial \psi^2} \ell_p(\psi).$$

$$r^* = r + \frac{1}{r} \log(u/r).$$

$$u = \frac{|\bar{\ell}_{;\hat{\theta}} - \bar{\ell}_{;\hat{\psi}} \quad \bar{\ell}_{\chi;\hat{\theta}}|}{|j_{\chi\chi}|^{1/2} |\bar{j}|^{1/2}}.$$

$$b^*(\psi, \hat{\chi}_\psi, a, n; c) = \frac{\Phi(r^*(\psi; v_c)) - \Phi(r^*(\psi; -v_c)) - \Phi(v_c) + \Phi(-v_c)}{2v_c \phi(v_c)}.$$

**Otras cantidades**

v.a.i.i.d.: variables aleatorias independientes e idénticamente distribuidas.

Función de densidad normal estándar:  $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ .

Función de distribución normal estándar:  $\Phi(x) = \int_{-\infty}^x \phi(t) dt$ .

Función modificada de Bessel del primer tipo:  $I_\nu(z) = \frac{\left(\frac{z}{2}\right)^\nu}{\Gamma(\nu + \frac{1}{2}) \Gamma(\frac{1}{2})} \int_0^\pi e^{\pm z \cos \theta} \sin^{2\nu} \theta d\theta$ ,  
 $\text{Re}(\nu + \frac{1}{2}) > 0$ .

de donde  $I_0(x) = \frac{1}{\Gamma(\frac{1}{2}) \Gamma(\frac{1}{2})} \int_0^\pi e^{\pm x \cos \theta} d\theta = \frac{1}{\pi} \int_0^\pi e^{\pm x \cos \theta} d\theta = \frac{1}{2\pi} \int_0^{2\pi} e^{\pm x \cos t} dt, \quad \forall \phi$ .

Función modificada de Bessel del tercer tipo:  $K_\nu(z) = \frac{1}{2} \left(\frac{z}{2}\right)^\nu \int_0^\infty \frac{e^{-t - \frac{z^2}{4t}}}{t^{\nu+1}} dt, \quad |\arg z| < \frac{\pi}{2}$ ,  
 $\text{Re } z^2 > 0$ .

de donde  $K_1(z) = \frac{z}{4} \int_0^\infty t^{-2} e^{-t - \frac{z^2}{4t}} dt$ .