



03061 2  
24.

**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**

**UNIDAD ACADÉMICA DE LOS CICLOS PROFESIONAL Y DE POSGRADO DEL  
COLEGIO DE CIENCIAS Y HUMANIDADES**

**“DESCRIPCIÓN DE ALGUNOS MÉTODOS RECIENTES PARA LA  
CONSTRUCCIÓN DE INTERVALOS DE CONFIANZA”**

**TESIS**

**QUE PARA OBTENER EL GRADO DE**

**MAESTRO EN ESTADÍSTICA E INVESTIGACIÓN DE OPERACIONES**

**PRESENTA**

**ACT. LUIS ANGEL BARRIENTOS HUESCA**

**DIRIGIDA POR**

**DR. FEDERICO O'REILLY TOGNO**

**TESIS CON  
FALLA DE ORIGEN**

**CD. UNIVERSITARIA, 1997.**



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

*“Nunca muestra un hombre tan claramente su carácter como cuando describe el carácter de otro hombre”*

**Jean Pablo Richter**

## **Dedicatoria**

A mis padres, Agustín y Leonor, por todo su cariño, confianza y ejemplo de superación.

A mi tía Susy, mi segunda madre, por su paciencia, cariño y cuidados.

A mis hermanos, Agustín Pépe y Arturo, por su entusiasmo y apoyo.

A mi Abuc por esta bonita familia.

A mis tías Ludy, Margarita, Lety y Edith, por su cariño.

A Vero por todo el cariño, apoyo y comprensión.

A Juanito simplemente por esta gran amistad.

A Chelita por su singular afecto.

A Aristóteles, Sandy, Ramón y Marisol por su entusiasmo y amistad.

## **Agradecimientos**

Al Dr. Federico O'Reilly Tognó por su apoyo, comentarios y humor para alcanzar esta meta.

A los profesores: Ignacio Méndez, Raúl Rueda, Silvia Ruiz-Velasco y Bladt Mogens, por sus comentarios y sugerencias.

A todos los profesores del proyecto por ayudar a superarme.

A la coordinación del proyecto: Juanita, Lulú, Alfredo, Violeta, Gaby y Ely, por su apoyo y amistad.

Al IIMAS por todas las facilidades prestadas y; particularmente, al personal de la biblioteca y la hemeroteca.

Al CONACyT por el apoyo económico, pues sin éste, hubiera sido muy difícil lograr este objetivo.

Especialmente a todos mis compañeros de generación: Saúl, Catín, Hortensia, Adriana, Gabriel, Miguel, Yolanda y Rosario, por hacer de ese tiempo momentos imborrables. También, a Salvador, Mara y Dolores.

## **Agradecimientos**

Al Dr. Federico O'Reilly Togno por su apoyo, comentarios y humor para alcanzar esta meta.

A los profesores: Ignacio Méndez, Raúl Rueda, Silvia Ruiz-Velasco y Bladt Mogens, por sus comentarios y sugerencias.

A todos los profesores del proyecto por ayudar a superarme.

A la coordinación del proyecto: Juanita, Lulú, Alfredo, Violeta, Gaby y Ely, por su apoyo y amistad.

Al IIMAS por todas las facilidades prestadas y; particularmente, al personal de la biblioteca y la hemeroteca.

Al CONACyT por el apoyo económico, pues sin éste, hubiera sido muy difícil lograr este objetivo.

Especialmente a todos mis compañeros de generación: Saúl, Catín, Hortensia, Adriana, Gabriel, Miguel, Yolanda y Rosario, por hacer de ese tiempo momentos imborrables. También, a Salvador, Mara y Dolores.

# Indice

<b>Introducción</b>	<b>i</b>
---------------------	----------

<b>Capítulo 1.</b> <b>Generalidades en la construcción de Intervalos de Confianza</b>	<b>1</b>
--	----------

1.1 Algunos conceptos de las Regiones de Confianza	2
1.2 Construcción de Intervalos de Confianza mediante cantidades pivotaes	5

<b>Capítulo 2.</b> <b>Algunas aproximaciones para construir Intervalos de Confianza</b>	<b>9</b>
--	----------

2.1 Transformaciones que aproximan las funciones de verosimilitud a la normalidad	10
2.2 Transformaciones del cuantil para construir Intervalos de Confianza	20
2.3 Una propuesta y ejemplos numéricos	28

### Capítulo 3.

## Algunas aproximaciones para construir Intervalos de Confianza en presencia de parámetros de ruido

37

- 3.1 Transformaciones que aproximan las funciones de verosimilitud a la normalidad en presencia de parámetros de ruido 37
- 3.2 Un enfoque bayesiano para la construcción de Intervalos de Confianza en presencia de parámetros de ruido 45

**Conclusiones** 55

**Anexo. Construcción de Intervalos de Confianza basada en el análisis de Sprott** 58

**Bibliografía** 62

## Introducción

La construcción de Intervalos de Confianza es un tema ampliamente discutido en la literatura estadística tanto por su interés teórico como por su trascendencia a través del uso de la estadística como herramienta de trabajo en otras disciplinas. Los Intervalos de Confianza al igual que las Pruebas de Hipótesis son herramientas que frecuentemente son utilizadas por los usuarios de la estadística; sin embargo, cuando es posible representar una prueba de hipótesis como un intervalo de confianza existe normalmente una ganancia adicional en la interpretación del resultado. Por ejemplo: suponga que se desea saber si no hay diferencia significativa entre dos proporciones  $p_0$  y  $p_1$ . Entonces, sea la hipótesis nula,  $H_0: p_0=p_1$ . Los resultados que se obtienen a un nivel de confianza del 95% podrían haber sido, por ejemplo:

- i.  $P [-0.6 \leq p_0-p_1 \leq 0.8]=0.95$
- ii.  $P [-0.01 \leq p_0-p_1 \leq 0.03]=0.95$
- iii.  $P [-0.001 \leq p_0-p_1 \leq 0.9]=0.95$
- iv.  $P [0.84 \leq p_0-p_1 \leq 0.86]=0.95$
- v.  $P [0.02 \leq p_0-p_1 \leq 0.8]=0.95$
- vi.  $P [0.02 \leq p_0-p_1 \leq 0.04]=0.95$

En los primeros tres ejemplos la hipótesis nula no se rechaza, en tanto que, en los tres últimos ejemplos  $H_0$  es rechazada. En principio, las anteriores conclusiones son las que ofrecería una prueba de hipótesis. En cambio vistos como intervalos de confianza, se puede observar, por ejemplo: que el intervalo de confianza construido es demasiado "grande" y quizás sea necesario obtener más información sobre este experimento (ejemplo i), o que aún cuando no se rechaza la hipótesis nula, esta decisión no se ve muy

respaldada (ejemplo iii) ya que a un nivel de confianza ligeramente menor, ésta podría ser rechazada, o que la diferencia entre ambas proporciones es lo suficientemente grande como para rechazar la hipótesis nula incluso para mayores niveles de confianza (ejemplo iv), etc..

En la literatura que trata sobre intervalos de confianza se encuentran avances importantes para la obtención de mejores aproximaciones a los resultados exactos, muchas veces apoyados por cómputo intensivo. Aún con todo este desarrollo, la mayoría sigue utilizando la regla de: la media más menos 2 desviaciones estándar, esto es, construir los intervalos de confianza a partir de la técnica estándar, sin más reparo en los supuestos que se deben cumplir para utilizarla correctamente. Esta poca observancia en el cumplimiento de supuestos<sup>1</sup>, en ocasiones puede llevar a resultados inconsistentes tales como construir intervalos de confianza conteniendo elementos que son imposibles para el parámetro de interés.

Este trabajo es una revisión bibliográfica de algunas técnicas para construir intervalos de confianza con y sin parámetros de ruido, sin embargo tiene algunos aportes como por ejemplo: un criterio para decidir sobre la presencia significativa del efecto de la curtosis en la función de verosimilitud en su aproximación a una función normal en el estudio de Sprott (1980) y, una modificación de una técnica propuesta por Dicciccio y Efron (1992) denominada  $ABC_q$ .

Las técnicas que se pretenden resumir aquí fueron desarrolladas por Sprott (1973, 1980, 1982), Dicciccio y Efron (1992) y Efron (1993). Estas técnicas, que tienen bases no triviales para aquella persona que no este muy relacionada con el tema, tienen por otro lado, una metodología directa y dan resultados cercanos a los exactos con solamente unas cuantas operaciones adicionales, comparándolas con las requeridas por la técnica estándar.

La distribución que se sigue en la presentación es la siguiente:

---

<sup>1</sup>Como ese del tamaño de muestra grande, ya que no hay un criterio absoluto que diga cuándo la muestra es grande. De ahí que la teoría de muestras grandes no indique con precisión cuando la función de verosimilitud tiende a la de una función normal, la cual es la base de la técnica estándar para construir intervalos de confianza.

En el primer capítulo se da un panorama general de las regiones de confianza; para entonces inducir el concepto de intervalos de confianza. También se muestran algunos criterios de optimalidad para la elección de los intervalos de confianza y se definen la cantidad pivotal exacta y la asintótica para la construcción de los intervalos.

En el segundo capítulo se muestra la relación que hace Sportt (1973, 1980, 1982) de las transformaciones en la verosimilitud a la normalidad y la aproximación resultante en la construcción de los intervalos de confianza obtenidos de la técnica estándar pero aplicada a los parámetros transformados. En este análisis, las transformaciones tienen como premisa eliminar la asimetría de la función de verosimilitud, si es que existe. Después, se expone y se propone un criterio para interpretar la función de verosimilitud en caso de que tenga problemas debido a su curtosis en su aproximación a la normal. Más adelante, se tratan de explicar otras metodologías recientes propuestas por Diccio y Efron en 1992, para construir intervalos de confianza para la familia exponencial y se da una sugerencia para mejorar el funcionamiento de la técnica denominada por ellos  $ABC_q$ ; además se muestran unos ejemplos numéricos. Se comenta que, de acuerdo a la teoría de Diccio y Efron (1992), estas aproximaciones son mejores que la de los intervalos de confianza estándar, pues sus cálculos tienen un orden de error respecto al cuantil exacto de  $O(n^{-1})$  mientras que el de los intervalos de confianza estándar es de  $O(n^{-1/2})$ .

El tercer capítulo trata sobre la construcción de intervalos de confianza en presencia de parámetros de ruido. Ahí se describen, algunos resultados del trabajo de Sprott (1980) así como una extensión a lo que él desarrolló para el caso uniparamétrico. Esta generalización, parte de calcular la función de verosimilitud pero maximizada en los parámetros de ruido para algún valor dado del parámetro de interés. Más adelante, se intentan mostrar algunos resultados del trabajo de Efron (1993) para construir intervalos de confianza (también conocidos como intervalos de máxima probabilidad a posteriori) en presencia de parámetros de ruido para la familia exponencial desde un punto de vista bayesiano.

En un anexo, se presenta un algoritmo para la construcción de intervalos de confianza siguiendo las recomendaciones de Sprott e incluyendo una propuesta de criterio sobre la curtosis de la función de verosimilitud.

# Capítulo 1

## Generalidades en la construcción de Intervalos de Confianza

Un problema importante en la inferencia estadística uniparamétrica es el que sigue: suponga que algunas características de los elementos de una población pueden ser representados por medio de una variable aleatoria  $X$ , cuya densidad es  $f(x;\theta)$  y la forma de la densidad se supone conocida excepto por algún parámetro desconocido<sup>1</sup>  $\theta$ ; además suponga que se puede obtener los valores  $x_1, x_2, \dots, x_n$  de una muestra aleatoria  $X_1, X_2, \dots, X_n$ . Entonces, de los valores de la muestra observada  $x_1, x_2, \dots, x_n$  se desea estimar el parámetro desconocido  $\theta$  ó alguna función de él, por ejemplo  $\tau(\theta)$ . A lo anterior se conoce como estimación y ésta puede ser de dos formas: “*Estimación Puntual*” y “*Estimación por Intervalos*”.

En la estimación puntual se busca obtener un estimador del parámetro de interés, en base a una estadística<sup>2</sup>, así pues, a cada muestra que se tome se obtendrán diferentes valores del parámetro estimado. Por lo tanto, aún cuando se obtiene una estimación del valor del parámetro hay que reconocer que éste puede diferir a cada muestra que se tome, esto es, existe un grado de incertidumbre en el verdadero valor del parámetro. Así que resultaría deseable que el estimador fuera acompañado por alguna medida de dispersión del posible error<sup>3</sup>, es decir, en vez de hacer estimaciones del verdadero valor del parámetro sería más conveniente hacer estimaciones de que el verdadero valor del parámetro esté contenido dentro de algún(os) intervalo(s), o lo que es lo mismo hacer estimaciones por intervalos de confianza. Existen dos vertientes en el problema de estimaciones por intervalos: 1.- Construir intervalos de confianza (IC) y; 2.- Determinar que los IC sean “óptimos” (en algún sentido).

---

<sup>1</sup>Si  $\theta$  fuera conocida, la función de densidad estaría completamente definida y no habría necesidad de hacer inferencias sobre ello.

<sup>2</sup>Una estadística es una función de las variables aleatorias observables, la cual es en sí misma una variable aleatoria observable que no contiene parámetros desconocidos (Mood et al 1974, p. 226).

<sup>3</sup>Por ejemplo, un estimador puntual puede ser acompañada por algún intervalo sobre éste, así junto con alguna medida de confianza, el verdadero valor del parámetro caerá dentro de tal intervalo.

En seguida se presenta, en la primera sección, un panorama general de las regiones de confianza y entonces se induce el concepto de intervalos de confianza. También, se muestran algunos criterios para la elección de IC óptimos. Después, en la segunda sección se definen las cantidades pivotaes exactas y asintóticas para la construcción de intervalos de confianza. En esta sección se da una muy breve semblanza acerca de las cantidades pivotaes. Si el interés es mayúsculo en este tópico se recomienda revisar como una primera lectura la tesis de Góngora (1988) capítulos 1 y 2, en donde hace una presentación de las cantidades pivotaes, sus variantes y alcances así como también de sus limitaciones.

### 1.1 Algunos conceptos de las Regiones de Confianza

En seguida se definen algunos conceptos de las regiones de confianza (RC) de acuerdo a la estadística clásica o también llamada estadística frecuentista.

Se define a la RC con nivel de confianza del  $100 \times (1-\alpha)\%$  o simplemente nivel de confianza del  $(1-\alpha)$ , denotada por  $C_\alpha(X_1, X_2, \dots, X_n)$ , como la región aleatoria del espacio paramétrico  $\Theta$ , tal que

$$P[\theta \in C_\alpha(X_1, X_2, \dots, X_n); \theta] = 1-\alpha \quad \text{para cada } \theta \in \Theta \text{ y } \alpha \in (0, 1), \quad (1)$$

donde  $(X_1, X_2, \dots, X_n)$ <sup>4</sup> es un vector aleatorio. La ecuación (1) significa que  $C_\alpha(X_1, X_2, \dots, X_n)$  contendrá al verdadero valor de  $\theta$  en un  $100 \times (1-\alpha)\%$  de las veces en un esquema de muestreo aleatorio<sup>5</sup>; a pesar de que en la práctica la idea de muestreo repetido puede ser hipotética.

Para determinar de mejor manera al parámetro  $\theta$  es ventajoso considerar a la familia de RC

$$\{C_\alpha(\underline{X})\}, \quad (2)$$

<sup>4</sup> $(X_1, X_2, \dots, X_n)$  también se denotará por  $\underline{X}$ , cuya función de densidad conjunta será denotada como  $f(\underline{x}; \theta) = \prod f(x_i; \theta)$ ; además,  $(x_1, x_2, \dots, x_n) = \underline{x}$  será el vector observado de  $\underline{X}$ .

<sup>5</sup>En el resto de este trabajo cuando se mencione que se hace un muestreo, éste se supondrá que será aleatorio.

que se construyen variando el nivel de confianza  $1-\alpha$  en el intervalo  $(0, 1)$ . La ecuación (2) debe gozar de una propiedad conocida como *anidamiento*, la cual se define así:

Si  $\alpha_1, \alpha_2 \in (0,1)$  y  $\alpha_1 > \alpha_2$ , entonces las correspondientes RC deben cumplir

$$C_{\alpha_1}(\underline{X}) \subset C_{\alpha_2}(\underline{X}), \quad (3)$$

lo que significa en palabras que, corresponderá a un mayor nivel de confianza ( $1-\alpha_1 < 1-\alpha_2$ ) una región más grande y; además, deberá contener a la construida con los mismos datos pero de menor nivel de confianza. Ahora bien, si se generan todos los elementos de (2), es decir, se calcula para cada  $\alpha \in (0, 1)$  las respectivas RC, entonces lo que se obtiene es la llamada *distribución de confianza* o también conocida como *distribución confidencial* del parámetro  $\theta$ , la cual consiste en una sucesión infinita de RC en la que por la propiedad de anidamiento, cada una está incluida en la anterior conforme el valor de  $\alpha$  decrece de 1 a 0.

Una acotación sobre las RC es que en caso de que los datos  $\underline{x}$  sean consistentes con valores parametrales adyacentes (funciones unimodales), se puede sintetizar dicha información en un sólo IC; pero, si sucede que los datos  $\underline{x}$  son consistentes con valores parametrales no contiguos (funciones poliimodales), el compendio de la información muestral se hace a través de una RC consistente con la unión de intervalos de confianza ajenos.

Comúnmente se utilizan las llamadas RC basadas en verosimilitud. Esto consiste en que todos los valores parametrales en  $C_{\alpha}(\underline{X})$  tengan mayor verosimilitud que aquellos fuera de  $C_{\alpha}(\underline{X})$ .

Existen diversas formas de construir las RC por lo que es necesario elegir aquellas con "mejores" propiedades. Se presentan dos criterios para la elección de las RC óptimas.

1.- Se dice que una RC de nivel de confianza del  $100 \times (1-\alpha)\%$  para  $\theta$  es *insesgada* si

$$P[\theta' \in C_{\alpha}(X_1, X_2, \dots, X_n); \theta] \leq 1-\alpha \quad \forall \theta' \neq \theta \in \Theta \quad \text{y} \quad \alpha \in (0, 1),$$

esto es, será menor o igual el nivel de confianza de la región  $C_{\alpha}(\underline{X})$  de incluir a un valor parametral equivocado  $\theta'$  que aquella confianza de contener al verdadero valor del parámetro  $\theta$ .

2.- Sean  $C_\alpha^1(\underline{X})$  y  $C_\alpha^2(\underline{X})$  dos RC distintas sobre el parámetro de interés  $\theta$  y sea  $\alpha$  algún valor entre 0 y 1, entonces se dice que  $C_\alpha^1(\underline{X})$  es preferible a  $C_\alpha^2(\underline{X})$  si

$$P[\theta' \in C_\alpha^1(\underline{X}) ; \theta] \leq P[\theta' \in C_\alpha^2(\underline{X}) ; \theta] \quad \forall \theta' \neq \theta \in \Theta \quad \text{y} \quad \alpha \in (0, 1), \quad (4)$$

esto es, el nivel de confianza de incluir un valor equivocado  $\theta'$  en  $C_\alpha^1(\underline{X})$  es probabilísticamente menor o igual a que se encuentre en  $C_\alpha^2(\underline{X})$ .

Ahora bien, en referencia a las funciones unimodales comentadas antes, esto es, las situaciones donde es posible resumir la información muestral sobre el parámetro escalar  $\theta$  en un solo IC, es factible determinar a las RC de la siguiente manera: se define al IC de dos colas para el parámetro  $\theta$  de nivel de confianza  $1-\alpha$ , denotado por  $[h_1(\underline{x}), h_2(\underline{x})]$ , como el intervalo aleatorio del espacio parametral  $\Theta$  tal que

$$P[h_1(\underline{x}) \leq \theta \leq h_2(\underline{x}); \theta] = 1-\alpha \quad \text{para cada } \theta \in \Theta \quad \text{y} \quad \alpha \in (0, 1), \quad (5)$$

es decir, el intervalo  $[h_1(\underline{x}), h_2(\underline{x})]$  contiene al verdadero valor del parámetro  $\theta$  en un  $100 \times (1-\alpha)\%$  de las veces que se tomen muestras de tamaño  $n$  de la misma población. A las estadísticas  $h_1$  y  $h_2$  se les conoce como límite de confianza inferior (LCI) y límite de confianza superior (LCS), respectivamente.

También existen otras variantes de los IC los cuales solamente están acotados por uno de sus lados o colas<sup>6</sup>.

Se define al IC de cola a la izquierda para el parámetro  $\theta$  de nivel de confianza  $1-\alpha_2$  por  $(-\infty, h_2(\underline{x}))$ , como el intervalo de  $\Theta$  tal que

$$P[\theta \leq h_2(\underline{x}); \theta] = 1-\alpha_2 \quad \text{para cada } \theta \in \Theta \quad \text{y} \quad \alpha_2 \in (0, 1). \quad (6)$$

La ecuación (6) tiene una interpretación semejante a la (5) pero ahora para el intervalo  $(-\infty, h_2(\underline{x}))$ . A  $h_2(\underline{x})$  se le llama límite de confianza superior, como antes.

Igualmente, se define el IC de cola a la derecha para el parámetro  $\theta$  de nivel de confianza  $1-\alpha_1$  por  $[h_1(\underline{x}), \infty)$ , como el intervalo de  $\Theta$  tal que

<sup>6</sup>Estos pueden verse como casos particulares del IC acotado por ambas colas. De igual manera, este último puede obtenerse como un caso particular de los IC acotados por una sola cola.

$$P [h_1(\underline{x}) \leq \theta; \theta] = 1 - \alpha_1 \quad \text{para cada } \theta \in \Theta \quad \text{y} \quad \alpha_1 \in (0, 1), \quad (7)$$

lo cual expresa que  $[h_1(\underline{x}), \infty)$  contiene al verdadero valor de  $\theta$  en un  $100 \times (1 - \alpha_1)\%$  de las veces en un esquema de muestreo repetido, donde  $h_1(\underline{x})$  es el LCI.

Ahora bien, es posible clasificar a los IC de dos colas en centrales y no centrales. Para aclarar este concepto suponga que la distribución muestral en que los IC están basados es simétrica y, tomando iguales desviaciones de la media los niveles de confianza obtenidos en (6) y (7) son iguales. En general, lo anterior no se puede aceptar con iguales desviaciones (por la posible asimetría de la distribución), *i.e.*, si la distribución no es simétrica y se toman a ambos lados de la media iguales desviaciones no se cumple que los respectivos niveles de confianza en (6) y (7) sean iguales. Luego, si  $\alpha_1 = \alpha_2$  se dice que los IC son los IC centrales, en cuyo caso se tiene

$$P[h_2(\underline{x}) < \theta] = P[\theta < h_1(\underline{x})] = \alpha/2,$$

en caso contrario se dice que los IC son IC no centrales. Debe observarse que centralidad de los intervalos de confianza de dos colas no significa que los límites de confianza sean equidistantes de la media estimada, a menos que la distribución muestral sea simétrica.

### 1.2 Construcción de Intervalos de Confianza mediante cantidades pivotaes

Existen diversos métodos que construyen regiones de confianza en el caso uniparamétrico, los cuales están basados en las *cantidades pivotaes*. En las siguientes secciones se revisan algunas técnicas para la construcción de los IC basados precisamente en las cantidades pivotaes.

Las cantidades pivotaes se pueden dividir en: *cantidades pivotaes exactas* y *cantidades pivotaes asintóticas*. Las cantidades pivotaes exactas son cantidades aleatorias que dependen de la muestra  $X_1, \dots, X_n$  y del parámetro  $\theta$  y que cuentan con la particularidad de tener una distribución única, *i.e.*, independiente del parámetro  $\theta$ . Las cantidades pivotaes asintóticas expresan que al hacer crecer el tamaño de la muestra

su función de distribución, con parámetro  $\theta$ , tiende en el límite a una función de distribución independiente de dicho parámetro. En seguida se definen formalmente estas cantidades.

Se define la cantidad pivotal exacta  $U=U(\underline{X};\theta)$  como una función de la muestra  $\underline{X}$  y del parámetro  $\theta$  con distribución  $F_0(u)$  independiente del parámetro  $\theta$

$$P[U(\underline{X};\theta)\leq u]=F_0(u) \quad \forall \theta \in \Theta,$$

es decir, la función de distribución se mantiene constante para toda  $\theta \in \Theta$ .

Dado que la función de distribución de  $U$ ,  $F_0(u)$ , es independiente de  $\theta$  entonces, ninguno de sus momentos dependen de este parámetro. En el siguiente ejemplo se muestran: la forma de la función de distribución y los dos primeros momentos (la media y la varianza) de  $F_0(u)$  los cuales son cantidades fijas e independientes del parámetro  $\theta$ .

Ejemplo 1.1 Sean  $X_1, \dots, X_n$  variables aleatorias independientes e idénticamente distribuidas (i.i.d.) normales con media  $\mu=\theta$  y varianza unitaria,  $N(\mu, 1)$ . El estimador de la media  $\mu$  es  $\bar{X} = \frac{1}{n} \sum X_i$  el cual se distribuye como una  $N(\mu, 1/n)$ . Así que, para este caso una cantidad pivotal exacta es

$$U = \sqrt{n} (\bar{X} - \mu),$$

la cual tan solo depende de la muestra  $\underline{X}$  y del parámetro  $\mu$ . Su función de distribución es entonces

$$P[U(\underline{X};\mu) \leq u] = P[\sqrt{n} (\bar{X} - \mu) \leq u] = F_0(u),$$

donde  $F_0(u)$  es una  $N(0,1)$  para cualquier valor de  $\mu \in \mathfrak{R}$ .

Para el caso de las cantidades pivotaes asintóticas. Se define la cantidad pivotal asintótica  $U_n=U(\underline{X}_n;\theta)$  como una función del parámetro y de la muestra, cuando el tamaño de ésta crece al infinito y la distribución  $F_0(u)$  en el límite es independiente del parámetro  $\theta$ , esto es

$$P[U(\underline{X}_n; \theta) \leq u] \rightarrow F_0(u) \quad \text{cuando } n \rightarrow \infty.$$

En algunos casos es posible encontrar funciones de distribución de las cantidades pivotaes asintóticas con sus dos primeros momentos exactos, *i.e.*, independientes del parámetro  $\theta$ ; en otros casos solamente son exactos asintóticamente. En el siguiente ejemplo se construye una cantidad pivotal asintótica.

Ejemplo 1.2 Sean  $X_1, \dots, X_n$  variables aleatorias i.i.d. con media y varianza comunes  $\mu = \theta$  y 1, respectivamente. Del teorema central del límite se tiene que la cantidad

$$U_n = \sqrt{n}(\bar{X}_n - \mu)$$

tiene como función de distribución límite a

$$P\left[U\left(\underline{X}_n; \theta\right) \leq u\right] \rightarrow \Phi(u),$$

donde  $\Phi(u)$  denota la distribución normal estándar,  $N(0, 1)$ . Así que la cantidad pivotal asintótica,  $U_n$ , tiene como distribución límite a la distribución normal estándar.

La técnica estándar para la construcción de intervalos de confianza toma como base la cantidad pivotal de este ejemplo. La construcción de los IC de dos colas con nivel de confianza  $1-\alpha$  con la técnica estándar se obtiene así, de (5) y la cantidad pivotal anterior:

$$P\left[h_1^* \leq U\left(\underline{X}_n; \theta\right) \leq h_2^*\right] = \Phi\left(h_2^*\right) - \Phi\left(h_1^*\right) = 1 - \alpha,$$

luego se pivotea, esto es, se despeja el parámetro  $\theta$  de la desigualdad probabilística construida sobre la cantidad pivotal  $U_n$  con su propia distribución

$$P\left[h_1 \leq \theta \leq h_2\right] = \Phi\left(h_2\right) - \Phi\left(h_1\right) = 1 - \alpha,$$

donde  $h_1 = U^{-1}\left(\underline{X}_n; h_1^*\right)$  y  $h_2 = U^{-1}\left(\underline{X}_n; h_2^*\right)$ ,  $h_1$  y  $h_2$  son los LCI y LCS, respectivamente para  $\theta$ .

Dado que el supuesto primordial de esta técnica es: "conforme el tamaño de la muestra aumenté la función de distribución tenderá a la de una normal estándar" entonces, el problema de esta técnica radica en que en la práctica no existe una cota mínima para decidir cuando una muestra es grande como para suponer que su función se aproxime a la de una normal, esto es, en ocasiones una muestra de 40 datos será suficiente para suponer que la distribución de  $U_n$  será aproximadamente una normal y; en ocasiones una

muestra de 1000 no precisa que la distribución de la cantidad pivotal mencionada en el ejemplo 1.2 tienda a la de una normal.

En el siguiente capítulo se pretende profundizar un poco más en este problema exhibiendo otras técnicas para construir los IC y, comparándolos con la técnica estándar por medio de ejemplos.

## Capítulo 2

### Algunas aproximaciones para construir Intervalos de Confianza

El concepto de funciones de verosimilitud normales esta muy relacionado con las distribuciones asintóticas normales de la Teoría de Muestras Grandes (TMG). En este sentido, Sprott y Kalbfleish (1969) exponen que el utilizar estimadores puntuales y varianzas de acuerdo a aproximaciones normales de muestras grandes puede llevar a resultados inconsistentes, cuando la función de verosimilitud (FV) es más bien no-normal. Así que, un primer paso y antes de aplicar la TMG, Sprott (1973) recomienda: analizar la forma de la FV para ver su cercanía con la de una función normal; si la semejanza es "buena" entonces utilizar la TMG sin mayor reflexión; en caso contrario, la propuesta es hacer una reparametrización, tal que la función de verosimilitud resultante sea aproximadamente normal. Siguiendo estas ideas, años después Sprott (1980) revisa el efecto que tiene la curtosis de la FV en su aproximación a la normalidad.

En lo que sigue se presenta la relación entre: las transformaciones en la verosimilitud a la normalidad y la aproximación resultante de la construcción de los IC obtenidos de la TMG aplicado a los parámetros transformados (Sprott, 1973). En este análisis, las transformaciones tienen como premisa eliminar la asimetría de la función de verosimilitud. Después de esta presentación, se muestra un estudio y se da un criterio para la FV en caso de que tenga problemas de curtosis en su aproximación a la normalidad.

Más adelante, se presenta otra metodología propuesta por Diccio y Efron en 1992 para construir los IC. Ellos comentan que ésta es una mejor aproximación que la de los IC estándar, pues, sus cálculos tienen un orden de error respecto al cuantil exacto de  $O(n^{-1})$  mientras que el del IC estándar es de  $O(n^{-1/2})$ .

Las ideas de Diccio y Efron toman como base la expansión de Cornish-Fisher y el remuestreo (bootstrap) y a diferencia de Sprott, para ellos es importante el tamaño de la muestra.

### 2.1 Transformaciones que aproximan las funciones de verosimilitud a la normalidad

Suponga que la función de densidad de las observaciones i.i.d. es  $f(\underline{x}; \theta) = f(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta)$  y que  $f$  satisface las condiciones usuales de regularidad, las cuales entre otras cosas permiten intercambiar el orden de integración y diferenciación. Se define la función de verosimilitud relativa de  $\theta$ , (FVR), como

$$R(\theta; \underline{x}) = f(\underline{x}; \theta) / \sup_{\theta} f(\underline{x}; \theta) \\ = f(\underline{x}; \theta) / f(\underline{x}; \hat{\theta}),$$

donde  $\hat{\theta}$  es el estimador máximo verosímil de  $\theta$  (EMV). Si se observan muestras de tamaño  $n$ , lo que se tiene es una familia de las posibles FVR; de la cual la FVR observada es la calculada con el valor observado  $\underline{x}$ ; y ésta es tan sólo un elemento de dicha familia. De aquí que la atención se fije en la familia completa de posibles FVR, en vez de tan sólo el miembro que fue observado,  $\underline{x}$ . Esto es porque se estará relacionando la función de verosimilitud a los IC, los cuales involucran la familia de verosimilitudes y no únicamente a la observada. Nótese, además, que la FVR es simplemente una estandarización de la FV, por lo que indistintamente se hará referencia a las características de la FV o de la FVR entendiéndose que expresan lo mismo para ambas.

Ahora bien, las bases para eliminar la asimetría de la FV y corregir el efecto de la curtosis, vienen al desarrollar el  $\log R(\theta; \underline{x})$  en serie de Taylor alrededor de  $\hat{\theta}$

$$\log R(\theta; \underline{x}) = -\frac{1}{2}(\theta - \hat{\theta})^2 I_0 + \frac{1}{6}(\theta - \hat{\theta})^3 \frac{\partial^3 \log R(\theta; \underline{x})}{\partial \hat{\theta}^3} + \frac{1}{24}(\theta - \hat{\theta})^4 \frac{\partial^4 \log R(\theta; \underline{x})}{\partial \hat{\theta}^4} + \dots, \tag{1}$$

donde  $\log$  es el logaritmo natural,  $\partial f / \partial \hat{\theta} = \partial f / \partial \theta|_{\theta = \hat{\theta}}$  e  $I_0$  es la información observada de Fisher

$$I_0 = - \partial^2 \log R(\theta; \underline{x}) / \partial \hat{\theta}^2; \tag{2}$$

además, se supone que el denominador esta acotado y  $R$  converge rápidamente.

Sí la función de verosimilitud que se tiene es una función normal o casi normal, los momentos de tercer orden en adelante estarán cercanos a cero, esto es, la función de verosimilitud se encontrará en la familia de verosimilitudes normales sí, al expandir el  $\log R(\theta; \underline{x})$  en serie de Taylor la contribución del término cúbico (la asimetría), el término de cuarto orden (la curtosis) y los mayores son despreciables en tal expansión<sup>1</sup>

$$R_N(\theta; \underline{x}) \approx \exp\left[-\frac{1}{2}(\theta - \hat{\theta})^2 I_0\right]. \tag{3}$$

En ocasiones<sup>2</sup>, es más conveniente utilizar  $E\{\log R(\theta; \underline{X})\}$  que  $\log R(\theta; \underline{x})$ . Bajo condiciones de rápida convergencia se tiene la expansión de  $E\{\log R(\theta; \underline{X})\}$  en serie de Taylor alrededor de  $\theta = \theta_0$ .

$$E\{\log R(\theta; \underline{X})\} = -\frac{1}{2}(\theta - \theta_0)^2 I_E(\theta_0) + \frac{1}{6}(\theta - \theta_0)^3 E\left\{\frac{\partial^3 \log R(\theta; \underline{X})}{\partial \theta_0^3}\right\} + \frac{1}{24}(\theta - \theta_0)^4 E\left\{\frac{\partial^4 \log R(\theta; \underline{X})}{\partial \theta_0^4}\right\} + \dots, \tag{4}$$

donde  $I_E(\theta_0)$  denota la información esperada de Fisher y se calcula como

$$I_E(\theta_0) = -E\left\{\partial^2 \log R(\theta; \underline{X}) / \partial \theta_0^2\right\}. \tag{5}$$

En caso de que no haya confusión se denotará  $I_E(\theta_0) = I_E$ .

Como una primera aproximación a las verosimilitudes normales, se intenta corregir la asimetría de la FV y la idea de Sprott es reparametrizar esta función, para esto, se tomará en cuenta hasta el término de orden tres de la expansión de  $\log R$ , suponiendo los términos de mayor orden nulos; entonces de (1) y (2)

<sup>1</sup>Una demostración sobre la aseveración de que los términos de tercer y cuarto orden son medidas de la asimetría y la curtosis, respectivamente, se presenta al final del capítulo en el apéndice 2.A.

<sup>2</sup>Aunque el uso de valores observados puede tener algún beneficio en un caso numérico dado, puesto que utilizando la verosimilitud se reflejan los datos obtenidos; el uso de valores esperados, frecuentemente resulta más conveniente ya que las fórmulas resultantes pueden ser calculadas una vez y utilizadas para cualquier situación; además, los intervalos de confianza resultantes están más próximos a los exactos. (Sprott, 1980).

$$\log R(\theta; \underline{x}) \approx -\frac{1}{2}(\theta - \hat{\theta})^2 I_{\theta} \left[ 1 - \frac{(\theta - \hat{\theta}) \partial^3 \log R(\theta; \underline{x}) / \partial \hat{\theta}^3}{3 I_{\theta}} \right].$$

El segundo término dentro de los corchetes es una medida de la desviación de la aproximación normal, (3), en cuanto a asimetría; si ésta es cero entonces se obtiene (3). Denote a la correspondiente medida estandarizada de asimetría por

$$|F_0| = \left| \frac{\partial^3 \log R(\theta; \underline{x}) / \partial \hat{\theta}^3}{I_{\theta}^{3/2}} \right|. \tag{6}$$

De igual manera, se sigue de (4) y (5), para las expresiones con valores esperados,

$$|F_E(\theta)| = \left| \frac{E \left\{ \partial^3 \log R(\theta; \underline{X}) / \partial \theta_0^3 \right\}}{I_E^{3/2}(\theta)} \right|. \tag{7}$$

Se utilizará  $F_3(\theta) = F_3$  para representar  $F_0$  ó  $F_E(\theta)$  en caso de que no exista confusión. Luego si se hace una reparametrización  $\phi = g(\theta)$ , entonces  $F_3(\phi)$  se puede calcular así<sup>3</sup>

$$|F_3(\phi)| = \left| F_3(\theta) + \frac{3 \, d^2 \phi / d\hat{\theta}^2}{(d\phi / d\hat{\theta}) I^{1/2}} \right|, \tag{8}$$

donde  $I$  será  $I_0$  ó  $I_E$  dependiendo si es  $F_0$  ó  $F_E(\theta)$ . La idea de (8) es examinar transformaciones que puedan eliminar -o al menos reducir- el efecto de la asimetría de la FV, *i.e.*, hacer  $|F_3(\phi)| = 0$ ; entonces, ahora el problema tan sólo radica en encontrar una  $\phi = g(\theta)$  que lo haga. Afortunadamente, este problema fue resuelto por Anscombe (1964) quien obtuvo una función  $\phi$ , tal que  $F_3(\phi) = 0$

$$d\phi/d\theta = c \exp \left\{ \frac{1}{3} \int \frac{\partial^3 \log R(\theta) / \partial \theta^3}{\partial^2 \log R(\theta) / \partial \theta^2} d\theta \right\}.$$

Un problema de este resultado es que en la práctica, a veces, existen dificultades para calcular  $\phi$ . Esto se ilustra adelante con el parámetro  $\theta = \sigma$  de la distribución lognormal.

<sup>3</sup>Una demostración de la ecuación (8) se presenta en el apéndice 2.B.

A continuación se presenta un ejemplo relacionado con el ejemplo 2.3 de Sprott y Kalbfleish (1969).

**Ejemplo 2.1** Suponga que se observa una muestra con  $n=2$  cuya familia de densidades es una exponencial, i.e.,  $f(x, \lambda) = \lambda \exp(-\lambda x)$  y sean los datos de dicha muestra  $\underline{x}=(0.6342, 1.2419)$ . Entonces, utilizando la técnica presentada arriba se tiene que

$$L(\lambda; \underline{x}) = \exp\left\{-\lambda \sum_{i=1}^n x_i + n \log \lambda\right\} \Rightarrow \log L(\lambda; \underline{x}) = l(\lambda; \underline{x}) = -\lambda \sum_{i=1}^n x_i + n \log \lambda,$$

donde  $L(\lambda; \underline{x})$  denota la función de verosimilitud de  $\lambda$  para una muestra dada,  $\underline{x}$ . Al calcular la primera derivada del logaritmo de la función de verosimilitud con respecto a  $\lambda$

$$\partial l(\lambda; \underline{x}) / \partial \lambda = -\sum_{i=1}^n x_i + n/\lambda,$$

e igualando a cero el resultado anterior, se obtiene el estimador máximo verosímil de  $\lambda$ .

$$\hat{\lambda} = 1/\bar{x} \quad \text{donde} \quad \bar{x} = \sum_{i=1}^n x_i / n;$$

calculando la segunda y tercera derivada de  $\log L$  con respecto a  $\lambda$

$$\partial^2 l(\lambda; \underline{x}) / \partial \lambda^2 = -n/\lambda^2 \quad \text{y}, \quad \partial^3 l(\lambda; \underline{x}) / \partial \lambda^3 = 2n/\lambda^3,$$

ya que  $\partial^m \log R(\theta; \underline{x}) / \partial \theta^m = \partial^m l(\theta; \underline{x}) / \partial \theta^m$ , para  $m=1, 2, 3, \dots$ . El cálculo de  $|F_3(\lambda)| = |2/n^{1/2}|$  indica que, cuando sea  $n=2$  la asimetría será importante en la FV; así que será necesario reparametrizar (según las recomendaciones de Sprott). Para encontrar la transformación que elimine la asimetría de la FV se sigue del resultado de Anscombe (1964)

$$\frac{\partial^3 \log R(\lambda; \underline{x}) / \partial \lambda^3}{\partial^2 \log R(\lambda; \underline{x}) / \partial \lambda^2} = \frac{2n/\lambda^3}{-n/\lambda^2} = -\frac{2}{\lambda}$$

y por tanto

$$\phi' = g'(\lambda) = c \exp\left\{\frac{1}{3} \int \frac{\partial^3 \log R(\lambda; \underline{x}) / \partial \lambda^3}{\partial^2 \log R(\lambda; \underline{x}) / \partial \lambda^2} d\lambda\right\} = c \exp\left\{-\frac{2}{3} \log \lambda\right\} = c\lambda^{-2/3},$$

donde  $\phi' = g'(*)$  es la primera derivada de  $g(*)$ . Por tanto, resolviendo esta diferencial se llega a la transformación  $g(\lambda) = \frac{1}{3}c\lambda^{-1/3}$  y, tomando por conveniencia  $c=3$

$$\phi = g(\lambda) = \lambda^{1/3}.$$

Ahora bien, para verificar la corrección que hizo esta transformación a la asimetría de la FV, de (8)

$$|F_3(\phi)| = \left| \frac{2}{n^{1/2}} - \frac{2\lambda^{-5/3}}{n^{1/2}\lambda^{-5/3}} \right| = 0,$$

es decir, si se utiliza esta transformación en la serie de Taylor de la FVR el término de orden cúbico será nulo.

Para aclarar estas ideas observe, en las siguientes tablas, el contraste en el comportamiento de la FVR, alrededor de la estimada MV y su transformada.

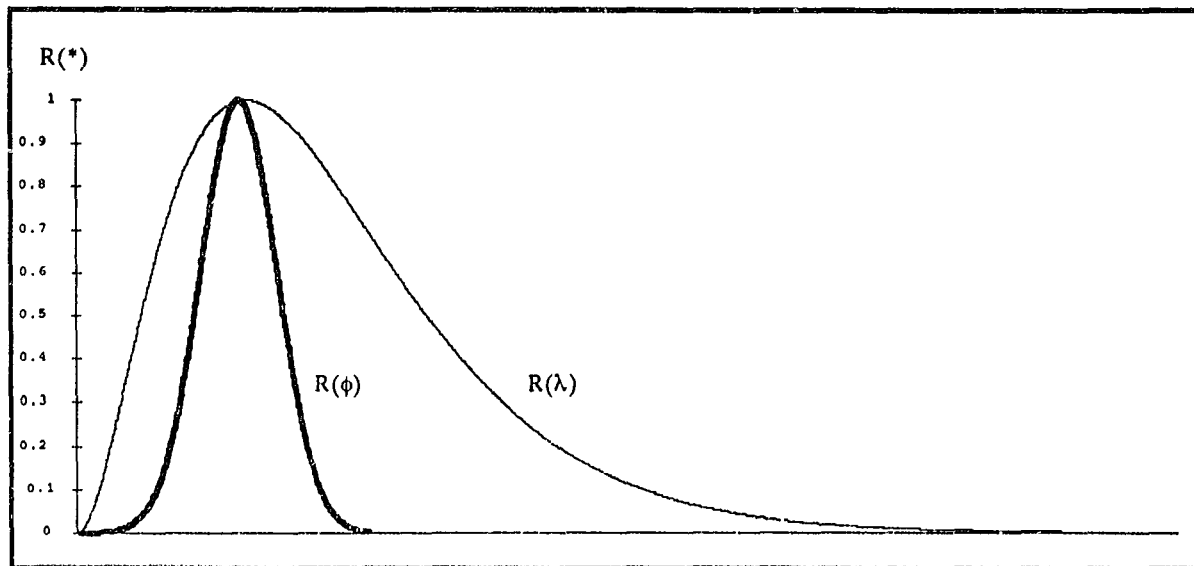
R( $\lambda; \underline{x}$ )	0.0243	0.8211	0.9600	1	0.9694	0.8934	0.5759
$\lambda$	0.066	0.666	0.866	<b>1.066</b>	1.266	1.466	2.066

Tabla 1. Forma del FVR alrededor del MV para el caso exponencial.

R( $\phi; \underline{x}$ )	0.1163	0.4595	0.9173	1	0.9177	0.4753	0.1158
$\phi$	0.5215	0.7215	0.9215	<b>1.0215</b>	1.125	1.3215	1.5125

Tabla 2. Forma del FVR alrededor de la transformada del MV para el caso exponencial.

Nótese, en la tabla 1, que al alejarse la misma distancia del máximo verosímil se obtienen muy diferentes verosimilitudes (estandarizadas), esto indica que sería muy difícil suponer normalidad en la forma de FV, pero quizá, uno podría suponer que el tamaño de la muestra es la causa del problema; en este caso, obsérvese, en la tabla 2, que -con el mismo tamaño de muestra- se presenta un comportamiento más simétrico alrededor de su máximo verosímil estandarizado, *i.e.*, bajo esta transformación la función de verosimilitud tiende más a una función normal. Lo anterior se ilustra en la siguiente gráfica.



Gráfica 1. Formas de la FVR para  $\lambda$  y  $\phi$  (caso exponencial) con  $n=2$ .

Obsérvese como  $R(\phi)$  se distingue por un comportamiento semejante a la de una función normal, en tanto que la  $R(\lambda)$  se caracteriza por ser más bien una función asimétrica.

A continuación otro ejemplo, esta vez de una gamma (con  $r=0.08$ , conocida).

Ejemplo 2.2 Suponga que se tiene una muestra i.i.d. de 40 datos ( $n=40$ ) de una gamma,  $f(x; \lambda, r) = \lambda^r x^{r-1} \exp(-\lambda x) / \Gamma(r)$ , que arrojó una suma de 250. Para construir un IC para  $\lambda$ , se sigue de la técnica de Sprott:

$$\partial l(\lambda; \underline{x}, r) / \partial \lambda = - \sum_{i=1}^n x_i + nr / \lambda,$$

igualando a cero la derivada anterior, se obtiene el estimador máximo verosímil de  $\lambda$

$$\hat{\lambda} = r / \bar{x} \quad \text{donde} \quad \bar{x} = \sum_{i=1}^n x_i / n;$$

calculando la segunda y tercera derivada de  $\log L$  con respecto a  $\lambda$

$$\partial^2 l(\lambda; \underline{x}, r) / \partial \lambda^2 = -nr / \lambda^2 \quad \text{y} \quad \partial^3 l(\lambda; \underline{x}, r) / \partial \lambda^3 = 2nr / \lambda^3.$$

Al calcular  $|F_3(\lambda)| = |2 / (nr)^{1/2}|$ , se observa que la asimetría de la función no es pequeña para  $r=0.08$  y  $n=40$ . Por lo tanto, una reparametrización como la propuesta por Anscombe (1964) ayudaría a eliminar el efecto de la asimetría en la FV. La transformación de Anscombe (1964) para este caso sería

$$\frac{\partial^3 \log R(\lambda; \underline{x}, r) / \partial \lambda^3}{\partial^2 \log R(\lambda; \underline{x}, r) / \partial \lambda^2} = \frac{2nr/\lambda^3}{-nr/\lambda^2} = -\frac{2}{\lambda}$$

y, por tanto

$$\phi' = g'(\lambda) = c \exp\left\{\frac{1}{3} \int \frac{\partial^3 \log R(\lambda; \underline{x}, r) / \partial \lambda^3}{\partial^2 \log R(\lambda; \underline{x}, r) / \partial \lambda^2} d\lambda\right\} = c \exp\left\{-\frac{2}{3} \log \lambda\right\} = c\lambda^{-2/3}.$$

Resolviendo la diferencial se llega a la transformación  $g(\lambda) = \frac{1}{3}c\lambda^{1/3}$  y tomando  $c=3$

$$\phi = g(\lambda) = \lambda^{1/3}.$$

Para verificar la corrección que hizo esta transformación a la asimetría de la FV, de (8)

$$|F_3(\phi)| = \left| \frac{2}{(nr)^{1/2}} - \frac{\frac{2}{3}\lambda^{-5/3}}{\frac{1}{3}(nr)^{1/2}\lambda^{-5/3}} \right| = 0.$$

Así que, si se utiliza esta transformación en la serie de Taylor de la FVR el término de orden cúbico desaparece.

Nótese, cómo el supuesto de conocer o no a  $r$ , no influyó para obtener la transformación del parámetro de interés,  $\lambda$ ; esto en general no se cumple. También, obsérvese que al comparar esta transformación y la del ejemplo anterior, ¡ambas resultaron iguales!; hecho que no necesariamente se cumple para toda función de verosimilitud. Por ejemplo, el parámetro  $\theta = \sigma$  de la distribución lognormal

tiene como transformación (de Anscombe 1964) a:  $\phi = C \int \frac{[n\sigma^2 - 3 \sum (\log x_i)^2]^{1/3}}{\sigma^{4/3}} d\sigma$ .

Haciendo un análisis semejante al del ejercicio anterior, observe el contraste en la forma de la FVR, alrededor del estimador MV y para su transformada cuando el tamaño de muestra es “relativamente grande”.

$R(\lambda; \underline{x}, r)$	0.4294	0.8539	0.9652	1.0000	0.9716	0.8987	0.6893
$\lambda$	0.0056	0.0092	0.0110	0.0128	0.0146	0.0164	0.0200

Tabla 3. Forma del FVR alrededor del MV para el caso gamma.

$R(\phi; \underline{x}, r)$	0.2268	0.5167	0.8482	1.0000	0.8482	0.5178	0.2279
$\phi$	0.1589	0.1839	0.2089	0.2339	0.2539	0.2839	0.3089

Tabla 4. Forma del FVR alrededor de la transformada del MV para el caso gamma.

Nótese que la aproximación a una función simétrica en la tabla 4 es mejor que aquella donde no se hizo ninguna reparametrización (tabla 3); esto a pesar del tamaño de la muestra que ya no es pequeña. Se presenta abajo la confianza real de encontrar al parámetro  $\lambda$  dentro de los LC construidos con las técnicas estándar y de Sprott.

Exacto	0.68	0.75	0.90	0.95	0.99
Estándar	0.710	0.790	0.930	0.954*	0.980*
Sprott	0.669	0.740	0.895	0.947	0.990

Tabla 5. Comparación de los niveles de confianza real de los IC estándar y de Sprott de dos colas, con  $\alpha = 0.16, 0.125, 0.05, 0.025$  y  $0.005$

Los espacios de la tabla 5 marcados con "\*" indican que el IC construido para  $\lambda (>0)$  toma valores fuera de su rango, por lo que se tomó al cero como su LCI. En este ejemplo, los límites de confianza inferior calculados con el técnica estándar fueron negativos a niveles de confianza del 0.95 y 0.99 (-0.0012 y -0.0056, respectivamente), *i.e.*, en un muestreo aleatorio, de 100 muestras de tamaño  $n=40$  en 95 (99) de ellas  $\lambda$  se encontrará dentro del intervalo  $[-0.0012, 0.0268]$  ( $[-0.0056, 0.0312]$ ), este resultado da un IC no óptimo, en el sentido que puede existir otro IC con el mismo nivel de confianza pero de menor longitud:  $(0, 0.268]$  ( $(0, 0.0312]$ ), esto es porque la probabilidad de encontrar una muestra con  $\lambda \leq 0$  es nula. Ahora bien, obsérvese que las construcciones de los IC calculados con la técnica de Sprott (1973) resultan en una mejor aproximación a los IC exactos. Por ejemplo, las aproximaciones que se obtienen con ambas técnicas a un nivel de confianza de 0.68, difieren de las exactas en: 0.031 para el caso estándar y 0.011 para la

técnica de Sprott; a un nivel de confianza de 0.75 la diferencia del IC exacto con respecto a la estándar es aún mayor: 0.04, en tanto que para la técnica de Sprott es de 0.01. Para las demás aproximaciones, las construcciones de los IC con la técnica de Sprott resultaron bastante cercanas a las exactas.

Sprott comenta que la transformación de Anscombe  $\phi=g(\theta)$  produce normalidad o casi normalidad para la familia de verosimilitudes si el estimador del parámetro no toma los valores extremos en su rango, por ejemplo: para el caso binomial, cuya transformación es  $\phi = \int^{\theta} \frac{dt}{[t(1-t)]^{2.3}}$ , existen problemas de aproximación a una verosimilitud normal cuando  $\hat{\theta}=0$  ó 1. Aunque lo anterior no se puede generalizar para toda transformación (obtenida siguiendo el resultado de Anscombe) es recomendable al momento de calcularla, analizar si no se tienen problemas cuando  $\hat{\theta}$  tome valores extremos. En estos casos, será mejor proponer otras transformaciones -quizás, por medio de ensayo y error- que, disminuyan el efecto de asimetría y no tenga problemas para tales valores de  $\hat{\theta}$  o proponer alguna modificación de la transformación en los puntos de discontinuidad.

Hasta el momento se ha expuesto una metodología para reducir la asimetría de la FVR mediante una reparametrización, para esto se ha supuesto que los términos de orden mayor al cúbico son pequeños y por tanto de poco efecto en la descripción de la verosimilitud. Habrá veces que suponer lo anterior llevará a resultados inconsistentes<sup>4</sup> debido a que la curtosis de esta función es muy importante. Si se esta en dicha situación Sprott en diversos artículos (1978, 1980 y 1982, por ejemplo) propone: el término cúbico de (1) ó (4) se analiza como antes, *i.e.*, si éste es importante entonces reparametrizar (o sea, eliminar la asimetría de la FVR), en caso contrario no hay necesidad de transformar el parámetro; luego, se calcula el término de cuarto orden para examinar si hay necesidad de aproximar la FVR a una distribución t de Student. Para esto, denótese por

$$F_4(\theta) = \frac{\partial^4 \log R(\theta; \underline{x}) / \partial \hat{\theta}^4}{I_{\theta}^2}$$

que, proporciona la principal contribución de la curtosis a la FV. Además, el cuarto momento de la expansión de la FVR con respecto a  $\phi$  se puede calcular así<sup>5</sup>

<sup>4</sup>Por ejemplo: valores fuera del rango del parámetro.

<sup>5</sup>Una demostración de la igualdad de  $F_4(\phi)$  se presenta en el apéndice 2.C.

$$F_4(\phi) = F_4(\theta) - 6F_3(\theta)I_0^{-1/2} \ddot{\phi}/\dot{\phi} - I_0^{-1} \left\{ 3(\ddot{\phi}/\dot{\phi})^2 - 4\ddot{\phi}/\dot{\phi} \right\},$$

donde  $\dot{\phi} = J\phi / d\hat{\theta}$ ,  $\ddot{\phi} = d^2\phi / d\hat{\theta}^2$ , etc.. Sea  $\varpi$  igual a  $\phi$  ó  $\theta$  dependiendo si hubo reparametrización o no. Entonces, haciendo  $|F_3(\varpi)|=0$ , se define  $m$  como

$$|F_4(\varpi)| = 6/(m + 1), \tag{9a}$$

o la correspondiente  $F_{E(4)}$  satisfaciendo  $|F_{E(3)}(\varpi)|=0$

$$|F_{E(4)}(\varpi)| = 6(m + 2)(m + 3)^2 / [m(m + 1)(m + 5)(m + 7)] \tag{9b}$$

y

$$\nu\sqrt{m/(m + 1)} = t_{(m)}, \tag{10}$$

donde  $\nu = \nu(\varpi; \underline{x}) = (\hat{\varpi} - \varpi)I_c^{1/2}$ . Despejando<sup>6</sup> a  $m$  de (9) se concluye que  $\nu$  sigue una distribución<sup>7</sup>  $t$  de Student con  $m$  grados de libertad (*g.l.*). Así que, si  $F_3(\varpi)$  esta cerca a cero, Sprott sugiere utilizar (10) como una  $t$  con  $m$  *g.l.*, donde  $m$  se calcula de (9); además, si  $F_4(\varpi)$  es pequeño,  $m$  será grande y  $\nu$  será aproximadamente una  $N(0,1)$ . También recomienda utilizar a  $m$  como el número entero más próximo.

Dado que una  $t$  con  $m$  *g.l.* es prácticamente una normal si  $m$  es 25 ó más, al tener

$$|F_4(\varpi)| \in [0.231, 3] \quad |F_{E(4)}(\varpi)| \in [5.7, 12]$$

se deduce que sería viable la sugerencia de Sprott de utilizar la  $t$  de Student, pero si  $F_4(\varpi) \in (-0.231, 0.231)$  ó  $F_{E(4)}(\varpi) \in (-5.7, 5.7)$  se recomienda usar la aproximación normal.

Ejemplo 2.3 Suponga que en los ejemplos 2.1 y 2.2 se quiere analizar el efecto que provoca la curtosis en la FVR. Para esto, primero se calcula dicho efecto con los parámetros y después con sus respectivas transformaciones.

- Para el caso exponencial: sin reparametrización  $F_4(\lambda) = -6/n$  y para su respectiva transformación  $F_4(\phi) = -2/(9n)$

<sup>6</sup>Para calcular  $m$  en (9b) se recomienda utilizar métodos numéricos.

<sup>7</sup>Sprott (1982).

- Para la gamma, con r conocida: sin reparametrización  $F_4(\lambda) = -6/(nr)$  y para su transformación  $F_4(\phi) = -2/(9nr)$ .

Donde  $\phi = \lambda^{1/3}$ ; obsérvese que  $F_4(\lambda)$  es importante cuando n es 2 para el ejemplo exponencial y 40 para el caso gamma con r=0.08; por lo que resultaría errado asumir que su presencia es insignificante para la aproximación. Ahora, revisando el efecto en  $F_4(\phi)$  de las transformaciones, resulta interesante observar que en ambos casos además de suprimir problemas de asimetría, elimina el efecto de curtosis, de acuerdo al criterio comentado arriba.

## 2.2 Transformaciones del cuantil para construir Intervalos de Confianza

En seguida se presentan otras aproximaciones a la construcción de los IC, las cuales son propuestas por Diccio y Efron (1992) y cuya teoría esta basada en la expansión de Cornish-Fisher y el remuestreo. Aunque esta teoría no es restrictiva, por cuestiones de cálculo se realiza esta presentación sólo en la familia exponencial multiparamétrica, la cual es por sí sola una familia muy basta. Así pues, se empieza con un breviarío de la familia exponencial, la expansión de Cornish-Fisher y la notación utilizada en esta sección, para entonces dar paso a las transformaciones propuestas.

Sea  $\underline{X}' = (X_1, X_2, \dots, X_n)$  una muestra aleatoria,  $\underline{x}' = (x_1, x_2, \dots, x_n)$  la muestra observada y  $\underline{y}' = (y_1, \dots, y_p)$  el vector estadístico suficiente obtenido de  $\underline{x}$  y cuya familia de posibles funciones de densidad,  $g_{\underline{\mu}}(\underline{y})$ , esta inducida por un vector de parámetros  $\underline{\mu}' = (\mu_1, \dots, \mu_p)$ . Sea  $\theta = t(\underline{\mu})$  el parámetro escalar de interés, estimado por su EMV  $\hat{\theta} = t(\hat{\underline{\mu}})$ , donde  $\hat{\underline{\mu}}$  es el EMV de  $\underline{\mu}$ . La función  $t(\underline{\mu})$  debe tener segundas derivadas continuas. Se supondrá que la familia paramétrica de densidades de  $\underline{y}$  es una familia exponencial con  $\underline{\eta}$ , el vector de parámetros naturales de dimensión p; así que la densidad se puede expresar como

$$g_{\underline{\eta}}(\underline{y}) \propto \exp[\underline{\eta}'\underline{y} - r'\underline{\eta}]. \tag{11}$$

El vector de parámetros de esperanza  $\underline{\mu} = E_{\underline{\eta}}(\underline{y})$  indexa a la familia exponencial exactamente igual que  $\underline{\eta}$  dado que  $\underline{\mu}$  y  $\underline{\eta}$  están relacionadas biunívocamente. Se utilizará la notación  $g_{\underline{\mu}}(\underline{y})$  y  $E_{\underline{\mu}}(\underline{y})$  cuando sea conveniente, significando  $g_{\underline{\eta}}(\underline{y})$  y  $E_{\underline{\eta}}(\underline{y})$  para el valor de  $\underline{\eta}$  correspondiente a  $\underline{\mu}$ . La matriz de covarianza de  $\underline{y}$ , cuya dimensión es de  $p \times p$ , se denotará por  $V_{\underline{\mu}} = V(\underline{\mu})$ . El EMV de  $\underline{\mu}$  es  $\underline{y}$  ( $\hat{\underline{\mu}} = \underline{y}$ ), de aquí que el EMV de  $\theta = t(\underline{\mu})$  es  $\hat{\theta} = t(\hat{\underline{\mu}}) = t(\underline{y})$ . El error estándar de  $\hat{\theta}$  es aproximadamente

$$V_0^{1/2} = \sigma = \left[ \dot{t}(\underline{\mu})' V_{\underline{\mu}} \dot{t}(\underline{\mu}) \right]^{1/2}, \tag{12}$$

donde  $\dot{t}(\underline{\mu})$  es el vector gradiente  $(\partial t(\underline{\mu}) / \partial \mu_i)'$ ; a (12) también se le conoce como la cota inferior de Crámer-Rao para la desviación estándar de un estimador insesgado de  $\theta$ . Ahora bien el estimador usual para (12) es

$$V_0^{1/2} = \hat{\sigma} = \left[ \dot{t}(\hat{\underline{\mu}})' V_{\hat{\underline{\mu}}} \dot{t}(\hat{\underline{\mu}}) \right]^{1/2}.$$

La función de verosimilitud en  $\underline{y}$  será  $L(\underline{\eta}; \underline{y}) \propto \exp\{\underline{\eta}' \underline{y} - \psi(\underline{\eta})\}$  y su logaritmo

$$\log L(\underline{\eta}; \underline{y}) = l(\underline{\eta}; \underline{y}) \propto \underline{\eta}' \underline{y} - \psi(\underline{\eta}).$$

La función generatriz de cumulantes para  $\underline{y}$  (que es el logaritmo de la función generatriz de momentos) es<sup>8</sup>

$$K(\underline{s}) = \log E_{\underline{\eta}}\{\exp(\underline{s}' \underline{y})\} = \psi(\underline{\eta} + \underline{s}) - \psi(\underline{\eta}),$$

donde  $\underline{s} = (s_1, \dots, s_p)'$ . Los cumulantes<sup>9</sup> conjuntos de  $y_1, \dots, y_p$  se obtienen de la siguiente manera:

$$\begin{aligned} \left. \frac{\partial K(\underline{s})}{\partial s_i} \right|_{\underline{s}=\underline{0}} &= \left. \frac{\partial}{\partial s_i} \left\{ \log E_{\underline{\eta}} \left( e^{\underline{s}' \underline{y}} \right) \right\} \right|_{\underline{s}=\underline{0}} = E_{\underline{\eta}}(y_i) = \mu_i, \\ \left. \frac{\partial^2 K(\underline{s})}{\partial s_i \partial s_j} \right|_{\underline{s}=\underline{0}} &= \left. \frac{\partial^2}{\partial s_i \partial s_j} \left\{ \log E_{\underline{\eta}} \left( e^{\underline{s}' \underline{y}} \right) \right\} \right|_{\underline{s}=\underline{0}} = E_{\underline{\eta}}(y_i y_j) - E_{\underline{\eta}}(y_i) E_{\underline{\eta}}(y_j) = \text{cov}(y_i, y_j), \end{aligned}$$

<sup>8</sup>Se presenta una demostración de este resultado en el apéndice 2.D.

<sup>9</sup>El orden de error del k-ésimo cumulante será de  $O(n^{1-k})$ , Kendall y Stuart (1979a, p. 176).

$$\begin{aligned} \left. \frac{\partial^3 K(\underline{s})}{\partial s_i \partial s_j \partial s_k} \right|_{\underline{s}=\underline{0}} &= \frac{\partial^3}{\partial s_i \partial s_j \partial s_k} \left\{ \log E_{\underline{\eta}}(e^{\underline{s}^T \underline{y}}) \right\} \Big|_{\underline{s}=\underline{0}} = E_{\underline{\eta}}(y_i y_j y_k) - E_{\underline{\eta}}(y_i) E_{\underline{\eta}}(y_j y_k) - \\ &- E_{\underline{\eta}}(y_j) E_{\underline{\eta}}(y_i y_k) - E_{\underline{\eta}}(y_k) E_{\underline{\eta}}(y_i y_j) + 2 E_{\underline{\eta}}(y_i) E_{\underline{\eta}}(y_j) E_{\underline{\eta}}(y_k) - \\ &= \text{cum}(y_i, y_j, y_k) \end{aligned}$$

$$i, j, k \in \{1, \dots, p\}.$$

Para el caso de la familia exponencial, también es posible calcular estos cumulantes a partir de

$$\psi(\underline{\eta}) = \log \int_{\mathcal{Y}} h(\underline{y}) e^{\underline{\eta}^T \underline{y}} d y_1 \dots d y_p$$

$$\psi_i(\underline{\eta}) = \frac{\partial \psi(\underline{\eta})}{\partial \eta_i} = E_{\underline{\eta}}(y_i) = \mu_i,$$

$$\psi_{ij}(\underline{\eta}) = \frac{\partial^2 \psi(\underline{\eta})}{\partial \eta_i \partial \eta_j} = E_{\underline{\eta}}(y_i y_j) - E_{\underline{\eta}}(y_i) E_{\underline{\eta}}(y_j) = \text{cov}(y_i, y_j),$$

$$\begin{aligned} \psi_{ijk}(\underline{\eta}) &= \frac{\partial^3 \psi(\underline{\eta})}{\partial \eta_i \partial \eta_j \partial \eta_k} = E_{\underline{\eta}}(y_i y_j y_k) - E_{\underline{\eta}}(y_i) E_{\underline{\eta}}(y_j y_k) - E_{\underline{\eta}}(y_j) E_{\underline{\eta}}(y_i y_k) - \\ &- E_{\underline{\eta}}(y_k) E_{\underline{\eta}}(y_i y_j) + 2 E_{\underline{\eta}}(y_i) E_{\underline{\eta}}(y_j) E_{\underline{\eta}}(y_k) = \\ &= \text{cum}(y_i, y_j, y_k) \end{aligned}$$

$$i, j, k \in \{1, \dots, p\},$$

donde  $h(\underline{y})$  es una constante de normalización.

Como se ha mencionado, algunas veces será más conveniente trabajar con  $\underline{\mu}$ , cuyas componentes son  $\mu_i = E(y_i) = \psi_i(\underline{\eta})$   $i=1, \dots, p$ . Las ecuaciones resultantes  $y_i = \psi_i(\underline{\eta})$  de la primera derivada de la función generatriz de cumulantes lleva a  $\hat{\underline{\mu}} = \underline{y}$ .

Las ideas de Diccio y Efron (1992) se fundamentan, entre otras cosas, en la expansión de Cornish-Fisher<sup>10</sup>, que es la inversa de la expansión de Edgeworth. Para aclarar estas ideas, se presenta someramente la expansión de Edgeworth y su relación con la de Cornish-Fisher.

Edgeworth (1904) desarrolló una manera de aproximar una v.a. normal estándar a una v.a.  $X$ , para esto utilizó la llamada "forma de Edgeworth de la serie tipo A"

$$\zeta(x) = \exp\left(-\frac{\kappa_1 - m}{1!}D + \frac{\kappa_2 - \sigma^2}{2!}D^2 - \frac{\kappa_3}{3!}D^3 + \frac{\kappa_4}{4!}D^4 - \dots\right)\beta(x),$$

donde  $\kappa_r$  es el  $r$ -ésimo cumulante de  $X$ ,  $D$  es el operador derivada,  $\beta(x)$  es la función de densidad de la normal y  $m$  y  $\sigma$  serán, en ocasiones, ligeramente diferentes de la media y la desviación estándar de  $X$  ya que esto puede mejorar su aproximación<sup>11</sup>. Si  $\beta(x)$  se estandariza,  $\beta(x) = \phi\{(x-m)/\sigma\}/\sigma = \phi(u)$  donde  $u = (x-m)/\sigma$  es una pivotal y  $\phi(u)$  es la función de densidad estándar de la normal, entonces se puede probar que esta estandarización lleva a

$$f\left(\frac{x-m}{\sigma}\right) = \exp\left\{-\frac{\kappa_3}{3!}D^3 + \frac{\kappa_4}{4!}D^4 - \dots\right\}\phi(u). \quad (13)$$

Nótese que si  $X$  fuera una v.a. normal el término en llaves sería 0 y por tanto  $f\left(\frac{x-m}{\sigma}\right) = f(u) = \phi(u)$ . En caso de que existan desviaciones de la normalidad de la forma de  $f(u)$ , éstas serán corregidas por el factor en llaves<sup>12</sup>.

Ahora bien, Edgeworth expande la función exponencial de (13) en serie de Taylor, para luego obtener la función de distribución de  $U$ ,  $F(u)$ . Su siguiente paso fue igualar  $F(u)$  con la función de distribución de una v.a. normal estándar  $Z$ . El resultado de esto fue la llamada expansión de Edgeworth que, expresa la v.a. normal estándar  $Z$  en términos de  $U$ . Más tarde, Cornish y Fisher (1937) desarrollan la idea de exhibir la v.a.  $U$  en función de la v.a. normal estándar  $Z$ ; esta es la razón por la que se denomina a

<sup>10</sup>Para más detalles véase por ejemplo: Hall (1988), Withers (1984), Barndorff-Nielsen y Cox (1979), Fisher y Cornish (1960), Kendall y Stuart (1979a); éste último recomendado como una primera lectura.

<sup>11</sup>Kendall (1991).

<sup>12</sup>Kendall y Stuart (1979a) extraen del artículo de Crámer (1925) las condiciones que debe cumplir  $f(u)$  para que sea una función de densidad.

la expansión de Cornish-Fisher la inversa de la expansión de Edgeworth. Una representación de la expansión de Cornish-Fisher del orden  $O(n^{-1})$  es

$$u = z - E(U) - \frac{1}{6} \text{ASIMET}(U) (z^2 - 1) + O(n^{-1}),$$

donde  $\text{ASIMET}(U) = \kappa_3/\sigma^3$  denota una medida de asimetría de la v.a.  $U$  (en (15) se da una caracterización cercana de esta medida). Nótese que la aproximación de  $u$  está tomando en cuenta parte del efecto de la asimetría de  $U$ , *i.e.*, esta aproximación corrige algunas posibles desviaciones de la normal provocadas por la asimetría de la función de  $U$ .

En seguida se presenta una justificación de las técnicas que desarrollaron Diccio y Efron (1992) para construir IC.

Se pueden definir los límites del IC exacto en términos de la cantidad pivotal  $v = (\hat{\theta} - \theta) / \hat{\sigma}$  - al menos teóricamente<sup>13</sup>-, donde  $\theta = t(\underline{\mu})$  es una función con segundas derivadas continuas. Ahora bien, defínase  $G_{\underline{\eta}}(v) = P(V \leq v; \underline{\eta})$  y

$$\hat{\theta}_{Ex}^{(1-\alpha)} = \hat{\theta} - \hat{\sigma} G_{\underline{\eta}}^{-1}(\alpha), \tag{14}$$

la cota superior "exacta" para  $\theta$  con confianza  $1-\alpha$ , *i.e.*,  $P(\theta \leq \hat{\theta}_{Ex}^{(1-\alpha)}; \underline{\eta}) = 1 - \alpha$ . En ocasiones no podrá calcularse (14) explícitamente (o más bien, desde un punto de vista práctico será muy complicado hacerlo), pues, está en función de  $\underline{\eta}$  a través de  $G_{\underline{\eta}}^{-1}(\alpha)$ . Empero, si se utiliza la noción de "correcto a segundo

orden"<sup>14</sup> en términos de  $\hat{\theta}_{Ex}^{(1-\alpha)}$ , por medio de la expansión de Cornish-Fisher, será posible conseguir una aproximación a los límites del IC; además, eludirá algunos problemas de cálculo. Para esto, en Diccio y Efron (1992) aparecen las siguientes cantidades que allí se afirma, son aproximaciones de orden de  $O(n^{-1})$ ,

$$E(V) = -m_1/2 + m_2/2 - m_3, \quad \text{VAR}(V) = 1, \quad \text{ASIMET}(V) = -2m_2 - 3m_3, \tag{15}$$

que en su notación (ver pp. 242-245)

<sup>13</sup>Góngora (1988) exhibe algunas dificultades que se presentan al momento de construir IC exactos.

<sup>14</sup>Una aproximación al límite del IC a un nivel  $1-\alpha$ ,  $\hat{\theta}^{(1-\alpha)}$ , se le dirá "second-order correct" o correcto a segundo orden, (SOC), si difiere de  $\hat{\theta}_{Ex}^{(1-\alpha)}$  por  $O_p(n^{-3/2})$ . Por otro lado, un límite cerca de  $\hat{\theta}^{(1-\alpha)}$  se le dirá "second-order accurate" o aproximado a segundo orden, (SOA), si se tiene una cobertura de error del orden  $O(n^{-1})$ , *i.e.*, si  $P(\theta \leq \hat{\theta}^{(1-\alpha)}; \underline{\eta}) = 1 - \alpha + O(n^{-1})$ . Además, generalmente SOC implica SOA. Hall (1988)

$$m_1 = \psi_{ijk} t_i t_j t_k / \sigma^3, \quad m_2 = \psi_{ij} t_{ij} / \sigma, \quad m_3 = \psi_{ik} \psi_{jl} t_k t_l / \sigma^3, \quad i, j, k, l = 1, \dots, p,$$

por simplicidad notacional ellos eliminaron el signo de suma, el cual corre sobre los subíndices que aparecen dos veces en cada formula. Los ordenes de  $m_1$ ,  $m_2$  y  $m_3$  son  $O(n^{-1/2})$ ; todos los otros cumulantes de  $V$  mayores a tres son  $O(n^{-1})$ . El motivo de mencionar los ordenes de los errores es porque esta construcción del IC se aproxima con un orden del error de  $O(n^{-3/2})$ . Utilizando la expansión de Cornish-Fisher para  $-G_{\eta}^{-1}(\alpha)$  con  $O_p(n^{-1})$

$$-G_{\eta}^{-1}(\alpha) = z_{\alpha} - E(V) - \frac{1}{6} \text{ASIMET}(V)(z_{\alpha}^2 - 1) + O_p(n^{-1});$$

sustituyendo (15) en  $G_{\eta}^{-1}(\alpha)$

$$-G_{\eta}^{-1}(\alpha) = z_{\alpha} + \left(\frac{1}{6}m_1 - \frac{1}{2}m_2 + \frac{1}{2}m_3\right) + \left(\frac{1}{3}m_1 + \frac{1}{2}m_3\right)z_{\alpha}^2 + O_p(n^{-1}), \quad (16)$$

en donde  $z_{\alpha}$  es el cuantil de una distribución normal estándar a nivel  $1-\alpha$ . Obsérvese que se pide que  $G_{\eta}^{-1}(\alpha)$  sea de orden  $O_p(n^{-1})$ ; esto es así, porque al hacer el producto de  $G_{\eta}^{-1}(\alpha)$  y  $\hat{\sigma}$  (cuyo orden de error es  $O(n^{-1/2})$ ) en (14) se obtiene una aproximación con error de  $O(n^{-3/2})$  al límite de confianza exacto.

Se denotará por  $\hat{m}_i = m_i(\hat{\eta})$   $i=1, 2, 3$  y

$$a = \frac{1}{6}\hat{m}_1, \quad z_0 = \frac{1}{6}\hat{m}_1 - \frac{1}{2}\hat{m}_2 + \frac{1}{2}\hat{m}_3, \quad c_q = \frac{1}{2}\hat{m}_3,$$

así que  $\hat{m}_i = m_i(\hat{\eta}) + O_p(n^{-1})$   $i=1, 2, 3$ ; además,  $a$ ,  $z_0$  y  $c_q$  son del orden de  $O_p(n^{-1/2})$ . Por tanto, de (16)

$$\hat{\theta}_{\text{Ex}}^{(1-\alpha)} = \hat{\theta} + \hat{\sigma} \left\{ z_{\alpha} + z_0 + (2a + c_q)z_{\alpha}^2 \right\} + O_p(n^{-3/2}). \quad (17)$$

Defínase  $\hat{\gamma} = \frac{1}{2}(\hat{m}_2 - \hat{m}_3)$ , entonces  $z_0 = a - \hat{\gamma}$ . Si  $p=1$ , entonces  $m_2$  y  $m_3$  son iguales,  $\hat{\gamma} = 0$  y  $z_0 = a$ .

También, si  $\theta$  es una combinación lineal de  $\mu_1, \dots, \mu_p$ ,  $m_2$  y  $m_3$  desaparecen y  $z_0 = a$ .

Una desventaja de (14) para la construcción del IC exacto es que si se considera algún otro parámetro escalar  $\tau = h(\theta)$ , donde  $h$  es una función monótona creciente, entonces  $\hat{\tau}_{\text{Ex}}^{(1-\alpha)}$  y  $h\left(\hat{\theta}_{\text{Ex}}^{(1-\alpha)}\right)$  no son necesariamente los mismos.

Hasta ahora, se han presentado las bases para construir IC de acuerdo a las propuestas de Diccio y Efron (1992).

Para que esta técnica sea computacionalmente sencilla, Diccio y Efron (1992) desarrollaron un procedimiento, cuyas cantidades  $a$ ,  $\hat{b} = \frac{1}{2} \hat{\sigma} \hat{m}_2$  y  $c_q$  se expresan en términos de derivadas que se calculan numéricamente. Sea

$$\hat{t}_i = dt(\hat{\mu} + h\epsilon^i)/dh|_{h=0} \quad i = 1, \dots, p,$$

donde  $\epsilon^i$  es el vector unitario  $p$ -dimensional con un 1 en la  $i$ -ésima posición. De la relación  $\mu_i = \psi_i(\underline{\eta})$  se sigue

$$\hat{\psi}_{ij} = d\mu_i(\hat{\eta} + \pi \epsilon^j) / d\pi|_{\pi=0} \quad i, j = 1, \dots, p,$$

denotando  $\hat{\underline{t}} = (\hat{t}_1, \dots, \hat{t}_p)$  y  $\hat{\sigma} = (\hat{\psi}_{ij} \hat{t}_i \hat{t}_j)^{1/2}$

$$a = \frac{1}{6} \frac{\hat{\psi}_{ijk} \hat{t}_i \hat{t}_j \hat{t}_k}{\hat{\sigma}^3} = \frac{1}{6} \frac{d^2 \hat{t}_i \mu_i(\hat{\eta} + h\hat{\rho})/dh^2|_{h=0}}{\hat{\sigma}^3},$$

donde  $\hat{\psi}_{ij} = \psi_{ij}(\underline{\eta})|_{\underline{\eta}=\hat{\eta}}$  y  $\hat{\psi}_{ijk} = \psi_{ijk}(\underline{\eta})|_{\underline{\eta}=\hat{\eta}}$   $i, j, k \in \{1, \dots, p\}$ .

Dado que  $V_{\hat{\underline{\mu}}} = (\hat{\psi}_{ij})$  es una matriz de  $p \times p$ , sea  $D$  la matriz diagonal de valores propios de  $V_{\hat{\underline{\mu}}}$  y  $\Gamma$  la matriz ortogonal cuyas columnas consisten de los correspondientes vectores propios de  $V_{\hat{\underline{\mu}}}$ . En particular, sea  $d_i$  el  $i$ -ésimo elemento de la diagonal de  $D$  y  $\zeta^i = (\zeta_1^i, \dots, \zeta_p^i)'$  la  $i$ -ésima columna de  $\Gamma$ , entonces de  $V_{\hat{\underline{\mu}}} = \Gamma D \Gamma'$  se sigue

$$\hat{\psi}_{ij} = \sum_{k=1}^p d_k \zeta_i^k \zeta_j^k \quad i, j = 1, \dots, p,$$

luego

$$\hat{b} = \frac{1}{2} \hat{\psi}_{ij} \hat{t}_i \hat{t}_j = \frac{1}{2} \sum_{k=1}^p d_k^2 t(\hat{\underline{\mu}} + \pi d_k^{1/2} \zeta^k) / d\pi^2|_{\pi=0},$$

con  $\hat{t}_{ij} = \partial^2 t(\underline{\mu}) / \partial \mu_i \partial \mu_j |_{\underline{\mu}=\hat{\underline{\mu}}}$ . Finalmente sea  $\hat{\delta}_i = \hat{\psi}_{ij} \hat{t}_j$   $i = 1, \dots, p$  y  $\hat{\underline{\delta}} = (\hat{\delta}_1, \dots, \hat{\delta}_p)$  entonces

$$c_q = \frac{1}{2} \frac{\hat{\psi}_{ik} \hat{\psi}_{jl} \hat{t}_{ij} \hat{t}_{kl} \hat{t}_l}{\hat{\sigma}^3} = \frac{1}{2} \frac{d^2 t(\hat{\mu} + \pi \hat{\delta} / \hat{\sigma}) / d\pi^2}{\hat{\sigma}} \Big|_{\pi=0}$$

Una vez que se calcula  $a$ ,  $\hat{b}$  y  $c_q$  se obtiene  $\hat{\gamma} = \hat{b} / \hat{\sigma} - c_q$  y  $z_0 = a - \hat{\gamma}$ . Una manera más directa de aproximar a  $z_0$ , es la que sigue:  $z_0 = \Phi^{-1} \{ 2 \Phi(a) \Phi(-\hat{\gamma}) \}$ ; la cual lleva a resultados próximos a los exactos cuando el tamaño de muestra es pequeño. Este resultado tiene como bases teoría de remuestreo<sup>15</sup>. Nótese que si  $\hat{\gamma} = 0$  entonces  $z_0 = a$ , como antes.

Con las acotaciones anteriores es posible calcular (17) con menos cálculos. Las aportaciones principales de Dicciccio y Efron (1992) son definir dos formas para calcular los límites del IC (cada una con sus propias ventajas y desventajas), las cuales tienen su fundamento en que el miembro derecho de (17) es una aproximación de orden  $O(n^{-3/2})$  de los IC.

Definición.- El límite del IC "quadratic ABC"<sup>16</sup> para  $\theta$ , denotado por  $\hat{\theta}_{ABC_q}^{(1-\alpha)}$  se construye calculando las cantidades  $(\hat{\theta}, \hat{\sigma}, a, z_0, c_q)$ ,  $z_{1-\alpha} = \Phi^{-1}(1-\alpha)$  y empleando el siguiente algoritmo<sup>17</sup>

$$w = z_0 + z_{1-\alpha} \quad \lambda = w / (1 - aw)^2 \quad \xi = \lambda + c_q \lambda^2 \quad \hat{\theta}_{ABC_q}^{(1-\alpha)} = \hat{\theta} + \hat{\sigma} \xi \tag{18}$$

Definición.- El límite del IC "ABC" para  $\theta$ , denotado por  $\hat{\theta}_{ABC}^{(1-\alpha)}$ , se construye calculando las cantidades  $(\hat{\theta}, \hat{\sigma}, a, z_0)$ ,  $z_{1-\alpha} = \Phi^{-1}(1-\alpha)$  y aplicando el algoritmo siguiente

$$w = z_0 + z_{1-\alpha} \quad \lambda = w / (1 - aw)^2 \quad \underline{\hat{\delta}} = V_{\hat{\mu}} i(\hat{\mu}) \quad \hat{\theta}_{ABC}^{(1-\alpha)} = t(\hat{\mu} + \lambda \underline{\hat{\delta}} / \hat{\sigma}) \tag{19}$$

<sup>15</sup>Efron (1985) llega a tal conclusión después de correr simulaciones del error medio cuadrático de  $z_0 = \Phi^{-1} \{ 2 \Phi(a) \Phi(-\hat{\gamma}) \}$  con  $z_0 = a - \hat{\gamma}$ . Los cambios de notación son:  $z_{01} = a$  y  $z_{02} = -\hat{\gamma}$ .

<sup>16</sup>ABC significa Approximate Bootstrap Confidence. Las iniciales ABC de estas técnicas se deben a que sus precursoras eran precisamente construcciones de IC por medio del bootstrap.

<sup>17</sup>La cantidad  $\lambda$  que se utiliza en las definiciones cumple, bajo SOC:  $\lambda = z_{1-\alpha} + z_0 + 2 a z_{1-\alpha}^2 + O_p(n^{-1})$  y  $\lambda^2 = z_{1-\alpha}^2 + 4 a z_{1-\alpha}^3 + 2 z_0 z_{1-\alpha} + O_p(n^{-1})$ .

La solución dada en (19) tiene la propiedad de invarianza, esto es, suponga que el parámetro de interés  $\theta=t(\underline{\mu})$  se transforma por alguna función monótona creciente de  $\theta$ ,  $\tau=h(\theta)$ , entonces el límite  $\hat{\theta}_{ABC}$  se transforma<sup>18</sup> en  $\hat{\tau}_{ABC}^{(1-\alpha)} = h(\hat{\theta}_{ABC}^{(1-\alpha)})$  con el mismo orden de error que  $\hat{\theta}_{ABC}$ ; en tanto que esta propiedad no necesariamente se cumple para  $\hat{\theta}_{ABC,q}$ . Además, obsérvese que para obtener  $\hat{\theta}_{ABC}$  no es preciso calcular  $c_q$ . En cambio las ventajas de  $\hat{\theta}_{ABC,q}$  son más bien computacionales, ya que los cálculos de  $t(\cdot)$  son locales, en el sentido que solamente requieren de la evaluación de  $t(\underline{\mu})$  para  $\underline{\mu}$  cerca de  $\hat{\underline{\mu}}$ ; lo anterior no se cumple para el límite  $\hat{\theta}_{ABC}$  ya que de acuerdo a como se define, precisa de la valuación de  $t(\underline{\mu})$  en un punto  $\underline{\mu}$  alejado de  $\hat{\underline{\mu}}$  y esto puede acarrear problemas computacionales, sobre todo cuando  $\hat{\underline{\mu}}$  está en la frontera del espacio paramétrico correspondiente.

### 2.3 Una propuesta y ejemplos numéricos

Fuera de este trabajo se han revisado numerables ejemplos utilizando (18) y (19) para la construcción de IC y se ha observado que generalmente ABC funciona más eficientemente que “quadratic” ABC. Para explorar una posible mejora en la eficiencia de “quadratic” ABC se propone en este trabajo que en lugar de  $\xi$  en (18) se utilice

$$\xi' = \lambda + \lambda^2 c_q - 2z_\alpha c_q (2az_\alpha^2 + z_0). \tag{20}$$

El tercer término del miembro de la derecha de (20) es de orden  $O(n^{-1})$  y por tanto formaría parte del error en (17) pero, cuando la muestra es “pequeña”, la presencia de este tercer término mejora la aproximación al límite exacto, por lo que se sugiere tomarlo en cuenta en tales casos. Por otro lado, cuando la muestra crece la importancia de este término se reduce y su aportación se vuelve insignificante, por lo que llegará a ser indistinto tomar  $\xi$  ó  $\xi'$  en (18) para la construcción de los IC. A esta alternativa se le denominará “segunda variante cuadrática ABC” y será denotada por  $\hat{\theta}_{ABC_{2,q}}$ . También resultará

---

<sup>18</sup>Diccio y Efron (1992).

indistinto utilizar  $\xi$  ó  $\xi'$  cuando  $c_q=0$ , pues, ambas serían iguales a  $\lambda$  en (18). Además, se obtendrá casi el mismo resultado con cualquiera de las anteriores tres técnicas ABC sí  $c_q=0$ .

Nótese que

$$\hat{\theta}_{ABC_{2q}}^{(1-\alpha)} = \hat{\theta}_{ABC_q}^{(1-\alpha)} - 2\hat{\sigma}z_\alpha c_q (2az_\alpha^2 + z_0).$$

A continuación se presentan dos ejemplos empleando las propuestas anteriores.

Ejemplo 2.4 Suponga que una muestra de ocho datos ( $n=8$ ) de una gamma expresada de la siguiente forma  $f(x;\beta,\alpha) = x^{\alpha-1} \exp(-x/\beta)/\Gamma(\alpha)\beta^\alpha$ , arrojó una suma de 9.3644. Para construir los IC para  $\theta=\beta$  con las propuestas anteriores (ABC, ABC<sub>q</sub> y ABC<sub>2q</sub>) con  $\alpha=0.25$  conocida,

$$L(\beta; \underline{x}, \alpha) \propto \exp\left\{-\frac{1}{\beta} \sum_{i=1}^n x_i - n\alpha \log \beta\right\} \Rightarrow L(\eta; \underline{x}, \alpha) \propto \exp\left\{\eta y - n\alpha \log\left(-\frac{1}{\eta}\right)\right\},$$

donde  $\eta=-\beta^{-1}$ ,  $y=\sum_{i=1}^n x_i$  y  $n\alpha \log(-\eta^{-1})=\psi(\eta)$ . Como  $\mu=\psi'(\eta) = -n\alpha/\eta$  y el interés es construir un IC para  $\beta$ , de la relación anterior se infiere que  $t(\mu)=n\alpha/\mu=\beta$ ; en lo que sigue del ejemplo se tomará  $\alpha=r$  para distinguirlo del  $\alpha$  ocupado en el nivel de confianza  $1-\alpha$ . Haciendo las operaciones necesarias se obtienen las cantidades  $(\hat{\theta}, \hat{\sigma}, a, z_0, c_q)$ :

$$\hat{\theta} = \frac{\bar{x}}{r}, \quad \hat{\sigma} = \frac{\bar{x}}{(nr^3)^{1/2}}, \quad a = z_0 = \frac{1}{3(nr)^{1/2}} \quad \text{y} \quad c_q = 0.$$

En este caso  $a=z_0$  ya que se supone  $r$  conocido, *i.e.*, sólo existe un parámetro desconocido. Aplicando (18), (19) y (20) se calculan los intervalos ABC, ABC<sub>q</sub> y ABC<sub>2q</sub>

$$\hat{\theta}_{ABC}^{(1-\alpha)} = \frac{n\bar{x} + \frac{\left[\frac{1}{3(nr)^{1/2}} + z_\alpha\right] \left[\frac{\bar{x}}{r}\right]^2}{\left[1 - \frac{1}{3(nr)^{1/2}} \left(\frac{1}{3(nr)^{1/2}} + z_\alpha\right)\right]^2} \left(\frac{r^3}{n}\right)^{1/2} \frac{1}{r}}{nr},$$

$$\hat{\theta}_{ABC_q}^{(1-\alpha)} = \frac{\bar{x}}{r} + \frac{\left[ \frac{1}{3(nr)^{1/2}} + Z_\alpha \right] \left[ \frac{\bar{x}}{(nr)^{1/2}} \right]}{\left\{ 1 - \frac{1}{3(nr)^{1/2}} \left[ \frac{1}{3(nr)^{1/2}} + Z_\alpha \right] \right\}^2}$$

y como  $c_q=0$  entonces

$$\hat{\theta}_{ABC_{2q}}^{(1-\alpha)} = \hat{\theta}_{ABC_q}^{(1-\alpha)}$$

Para la construcción del IC con la técnica ABC también se calcula la cantidad  $\hat{\delta} = (\bar{x}/r)^2$ .

En la siguiente tabla se compara el comportamiento de los IC: Estándar, ABC, ABC<sub>q</sub> y ABC<sub>2q</sub>.

Exacto	0.68	0.75	0.90	0.95	0.99
Estándar	0.663	0.698	0.764*	0.795*	0.841*
ABC's	0.669	0.737	0.887	0.938	0.982

Tabla 6. NIVELES REALES DE CONFIANZA. Comparación de los IC estándar y ABC's<sup>19</sup> de dos colas, con  $\alpha=0.16$ , 0.125, 0.05, 0.025 y 0.005

Los espacios de la tabla 6 marcados con "\*" indican que el IC construido para  $\beta (>0)$  toma valores fuera de su rango. En este ejemplo, los límites de confianza inferiores calculados con la técnica estándar arrojaron valores negativos a niveles de confianza del 0.90, 0.95 y 0.99 (-0.781, -1.807 y -3.843, respectivamente), *i.e.*, para el caso de un nivel de confianza del 90% y en un muestreo aleatorio, de 100 muestras de tamaño  $n=8$  en 90 de ellas  $\beta$  se encontrará dentro del intervalo [-0.781, 10.145]. De igual manera que en la interpretación de la tabla 5 este resultado da un IC no óptimo, en el sentido que puede existir otro IC con el mismo nivel de confianza pero de menor longitud: (0, 10.145]. Esta es la interpretación que se podría hacer cuando hay este tipo de inconsistencias. Ahora bien, obsérvese que las construcciones de los IC calculados con las técnicas ABC resultan en una mejor aproximación a los IC exactos. Por ejemplo, las aproximaciones que se obtienen con ambas técnicas, a un nivel de confianza de 0.68, difieren de las exactas en: 0.017 para el caso estándar y 0.011 para las ABC's. A un nivel de

<sup>19</sup>Dado que las tres técnicas ABC obtuvieron los mismos IC sólo se presenta en la tabla una fila para las tres, denotada por ABC's.

confianza de 0.75 la diferencia del IC exacto con respecto a la estándar es aún mayor: 0.052, en tanto que para las técnicas ABC's es de 0.013; lo anterior se puede interpretar así: el IC construido con la técnica estándar contendrá al verdadero valor del parámetro en 698 de 1000 muestras de tamaño  $n=8$ , *i.e.*, el IC construido con la técnica estándar es más pequeño del que debiera ser para contener el verdadero valor del parámetro en 750 de las 1000 muestras; en tanto que con las ABC's, éstas contienen al verdadero valor del parámetro en 737 de las 1000 muestras hipotéticamente tomadas. Las demás construcciones con las ABC's también resulta, on cercanas a las exactas.

Ejemplo 2.5 Analizar el ejemplo 2.1 con la técnica ABC (ecuación 19).

$$L(\lambda; \underline{x}) = \exp\left(-\lambda \sum_{i=1}^n x_i + n \log \lambda\right) \Rightarrow L(\eta; y) = \exp[\eta y + n \log(-\eta)],$$

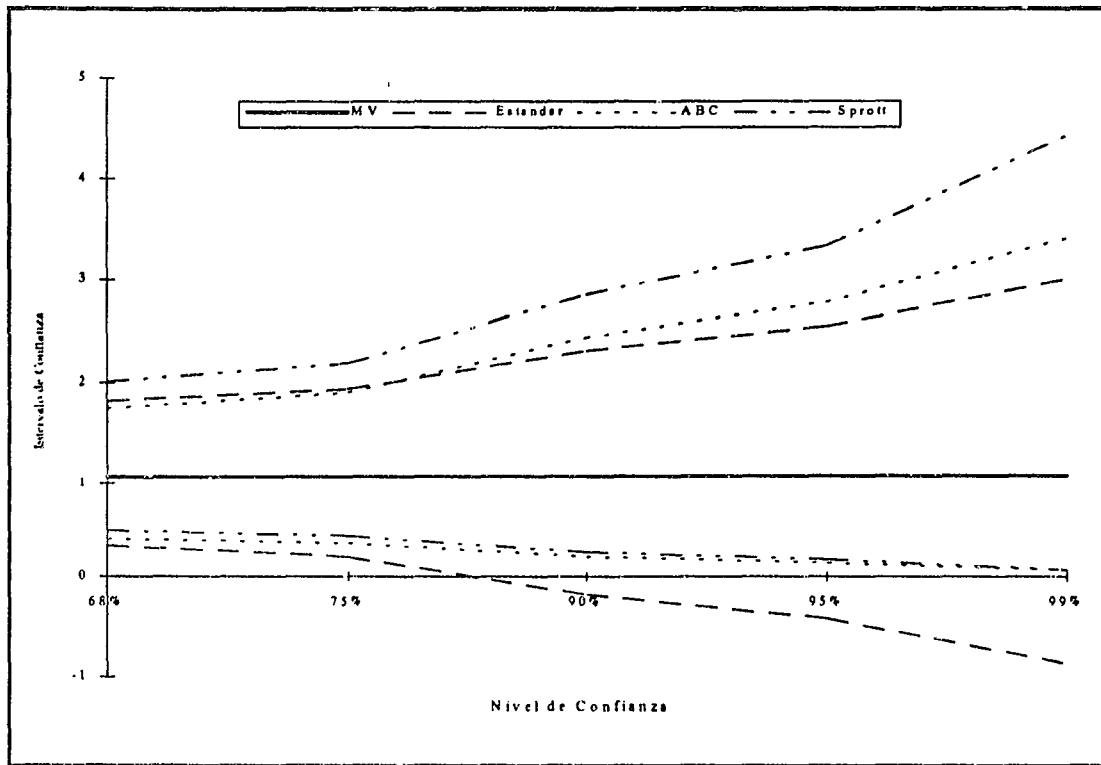
donde  $\eta = -\lambda$ ,  $y = \sum_{i=1}^n x_i$  y  $\phi(\eta) = -n \log(-\eta)$ . Ahora bien, debido a que  $\mu = \psi'(\eta) = -n/\eta$  y lo que se intenta es construir un IC para  $\theta = \lambda$ , entonces, se deduce que la transformación apropiada de  $t^*$  es  $t(\mu) = n/\mu = \lambda$ . Haciendo las operaciones necesarias se derivan las cantidades  $(\hat{\theta}, \hat{\sigma}, a, z_0)$

$$\hat{\theta} = \frac{1}{\bar{x}}, \quad \hat{\sigma} = \frac{1}{n^{1/2} \bar{x}}, \quad a = z_0 = -\frac{1}{3n^{1/2}}.$$

En este caso  $a = z_0$  pues existe sólo un parámetro en la función,  $p=1$ . Ahora de (19) se calculan los límites de confianza de la técnica ABC a un nivel de confianza  $(1-\alpha)$ :

$$\hat{\theta}_{ABC}^{(1-\alpha)} = \frac{n}{n\bar{x} + \frac{\frac{1}{3n^{1/2}} - z_\alpha}{\left[1 - \frac{1}{3n^{1/2}} \left(\frac{1}{3n^{1/2}} - z_\alpha\right)\right]} n^{1/2} \bar{x}}.$$

En las siguientes gráfica y tabla se compara el comportamiento de los IC: Estándar, Sprott y ABC a diferentes niveles de confianza.



Gráfica 2. Comparación de los IC de las técnicas: Estándar, ABC y Sprott.

En la gráfica 2 se presentan los LC construidos con las técnicas estándar, ABC y de Sprott y la estimación máximo verosímil (MV). Nótese que conforme aumenta el nivel de confianza, en la técnica estándar, el IC construido tiende a contener elementos negativos del parámetro, lo cual es inconsistente con la definición del parámetro.

Por lo que respecta a los límites de confianza de las técnicas ABC y de Sprott, éstos son muy parecidos; los límites inferiores están muy próximos, en tanto que los respectivos límites superiores están más bien alejados uno del otro, pero eso se explica porque la confianza real de la ABC es inferior a la de Sprott. Estas diferencias se marcan en la tabla 7 al momento de calcular los verdaderos IC obtenidos a partir de los LC estimados con las técnicas anteriores.

Exacto	0.68	0.75	0.90	0.95	0.99
Estándar	0.734	0.822	0.930*	0.951*	0.976*
ABC	0.669	0.737	0.887	0.938	0.982
Sprott	0.663	0.734	0.893	0.946	0.991

Tabla 7. Comparación de los IC de dos colas construidos con las técnicas Estándar, ABC y Sprott.  $\alpha=0.16, 0.125, 0.05, 0.025$  y  $0.005$

La tabla 7 establece que los verdaderos IC construidos (a los niveles de confianza ahí definidos) con los LC calculados utilizando la técnica estándar son mayores a los que se buscaban o, sus IC contienen elementos con probabilidad cero de ocurrir. Por otro lado, los verdaderos IC obtenidos a partir de los LC calculados con las técnicas ABC y de Sprott tienen una importante aproximación con respecto a los exactos.

### APÉNDICE 2.A

Para probar que los términos de tercer y cuarto orden son medidas de la asimetría y la curtosis, respectivamente, se hace lo siguiente: sean  $f(\underline{x};\theta)=L(\theta)$  la función de densidad,  $L(\theta)$  la función de verosimilitud,  $l=l(\theta)=\log L(\theta)$  y por tanto  $L(\theta)=\exp\{l(\theta)\}$ . Dado que la expansión de  $l(\theta+\tau)$  alrededor de  $\theta$  es

$$l(\theta + \tau) = l(\theta) + \tau \frac{\partial l}{\partial \theta} + \frac{\tau^2}{2!} \frac{\partial^2 l}{\partial \theta^2} + \frac{\tau^3}{3!} \frac{\partial^3 l}{\partial \theta^3} + \frac{\tau^4}{4!} \frac{\partial^4 l}{\partial \theta^4} + \dots,$$

entonces

$$\begin{aligned} L(\theta + \tau) &= \exp\{l(\theta + \tau)\} = \exp\left\{l(\theta) + \tau \frac{\partial l}{\partial \theta} + \frac{\tau^2}{2!} \frac{\partial^2 l}{\partial \theta^2} + \frac{\tau^3}{3!} \frac{\partial^3 l}{\partial \theta^3} + \frac{\tau^4}{4!} \frac{\partial^4 l}{\partial \theta^4} + \dots\right\} \\ &= L(\theta) \exp\left\{\tau \frac{\partial l}{\partial \theta} + \frac{\tau^2}{2!} \frac{\partial^2 l}{\partial \theta^2} + \frac{\tau^3}{3!} \frac{\partial^3 l}{\partial \theta^3} + \frac{\tau^4}{4!} \frac{\partial^4 l}{\partial \theta^4} + \dots\right\}. \end{aligned}$$

Además

$$1 = \int L(\theta + \tau) = \int L(\theta) \exp\left\{\tau \frac{\partial}{\partial \theta} + \frac{\tau^2}{2!} \frac{\partial^2}{\partial \theta^2} + \frac{\tau^3}{3!} \frac{\partial^3}{\partial \theta^3} + \frac{\tau^4}{4!} \frac{\partial^4}{\partial \theta^4} + \dots\right\},$$

donde  $\int L(\theta + \tau)$  denota la integral de Stieltjes sobre todas las posibles muestras. Entonces de

$$\int L(\theta + \tau) = \int L(\theta) \exp\left\{\tau \frac{\partial}{\partial \theta}\right\} \exp\left\{\frac{\tau^2}{2!} \frac{\partial^2}{\partial \theta^2}\right\} \exp\left\{\frac{\tau^3}{3!} \frac{\partial^3}{\partial \theta^3}\right\} \exp\left\{\frac{\tau^4}{4!} \frac{\partial^4}{\partial \theta^4}\right\} \dots,$$

expandiendo las exponenciales de la segunda integral e igualando coeficientes, se puede obtener una infinidad de relaciones, de las cuales las cuatro primeras son (Bartlett, 1953)

$$\begin{aligned} l_1 &= 0, \\ l_2 + l_1^{(2)} &= 0, \\ l_3 + 3(l_1 l_2) + l_1^{(3)} &= 0, \\ l_4 + 3l_2^{(2)} + 4(l_1 l_3) + 6(l_1^{(2)} l_2) + l_1^{(4)} &= 0, \end{aligned}$$

donde

$$l_i^{(j)} = E\left\{\left(\frac{\partial}{\partial \theta}\right)^j\right\} \text{ y } (l_i^{(j)} l_k^{(m)}) = E\left\{\left(\frac{\partial}{\partial \theta}\right)^j \left(\frac{\partial}{\partial \theta}\right)^k\right\}^m.$$

Bajo condiciones de normalidad es posible probar (Kendall et al, 1987) que  $(l_i^{(j)} l_j) = 0$ ,  $i > 0$  y  $j > 1$ . Además, de la relación

$${}_1(l_1 l_2) = l_2^{(2)} + (l_1 l_3) + (l_1^{(2)} l_2) \Rightarrow l_2^{(2)} = {}_1(l_1 l_2) - (l_1 l_3) - (l_1^{(2)} l_2),$$

donde  ${}_1(l_1 l_2) = \frac{\partial}{\partial \theta} (l_1 l_2)$ . Por tanto se sigue de las relaciones anteriores que

$$l_1^{(3)} = -l_3, \text{ i.e., } E\left\{\left(\frac{\partial}{\partial \theta}\right)^3\right\} = -E\left\{\frac{\partial^3 l}{\partial \theta^3}\right\} \text{ y}$$

$$l_1^{(4)} = -l_4, \text{ i.e., } E\left\{\left(\frac{\partial}{\partial \theta}\right)^4\right\} = -E\left\{\frac{\partial^4 l}{\partial \theta^4}\right\}.$$

## APÉNDICE 2.B

Para obtener la igualdad (8) sólo es necesario aplicar la regla de la cadena al  $\log R(\theta, \underline{x})$  y valorarla en  $\hat{\theta}$  (y por tanto en  $\hat{\phi} = g(\hat{\theta})$ ):

$$\frac{\partial \log R(\phi)}{\partial \hat{\phi}} = \left[ \frac{\partial \log R(\theta)}{\partial \hat{\theta}} \right] \frac{d\theta}{d\hat{\phi}} = 0,$$

$$\frac{\partial^2 \log R(\phi)}{\partial \hat{\phi}^2} = \left[ \frac{\partial^2 \log R(\theta)}{\partial \hat{\theta}^2} \right] \frac{d\theta}{d\hat{\phi}} = -I_{\phi},$$

$$\partial^3 \log R(\phi) / \partial \hat{\phi}^3 = \pm \left\{ \left[ \partial^3 \log R(\theta) / \partial \hat{\theta}^3 \right] \left[ d\theta / d\hat{\phi} \right]^3 + 3 \left[ \partial^2 \log R(\theta) / \partial \hat{\theta}^2 \right] \left[ d\theta / d\hat{\phi} \right] \left[ d^2\theta / d\hat{\phi}^2 \right] \right\},$$

el signo + ó - se utiliza de acuerdo a si la transformación es creciente o decreciente. Para calcular  $F_3(\phi)$ , lo que procede es estandarizar la ecuación anterior (la tercera derivada) como sigue

$$F_3(\phi) = \left[ \partial^3 \log R(\phi) / \partial \hat{\phi}^3 \right] / I_\phi^{3/2} = \pm \left\{ \left[ \partial^3 \log R(\theta) / \partial \hat{\theta}^3 \right] / I_\theta^{3/2} - 3 \left[ d^2\theta / d\hat{\phi}^2 \right] / \left[ I_\theta^{1/2} (d\theta / d\hat{\phi})^2 \right] \right\}$$

y, calculando

$$d\theta / d\phi = 1 / (d\phi / d\theta); \quad d^2\theta / d\hat{\phi}^2 = -(d^2\phi / d\theta^2) / (d\phi / d\theta)^3; \quad d^3\theta / d\hat{\phi}^3 = 3(-d^2\phi / d\theta^2)^2 / (d\phi / d\theta)^5 - (d^3\phi / d\theta^3) / (d\phi / d\theta)^4,$$

sustituyendolas en  $F_3(\phi)$  se llega a

$$|F_3(\phi)| = |F_3(\theta) + 3 \left[ d^2\phi / d\hat{\phi}^2 \right] / \left[ I_\theta^{1/2} d\phi / d\hat{\phi} \right]}.$$

## APÉNDICE 2.C

Siguiendo el mismo procedimiento utilizado para calcular  $F_3(\phi)$  se puede obtener  $F_4(\phi)$ , esto es,

$$\begin{aligned} \partial^4 \log R(\phi) / \partial \hat{\phi}^4 &= \left[ \partial^4 \log R(\theta) / \partial \hat{\theta}^4 \right] \left[ d\theta / d\hat{\phi} \right]^4 + 6 \left[ \partial^3 \log R(\theta) / \partial \hat{\theta}^3 \right] \left[ d\theta / d\hat{\phi} \right]^3 \left[ d^2\theta / d\hat{\phi}^2 \right] + \\ &+ 3 \left[ \partial^2 \log R(\theta) / \partial \hat{\theta}^2 \right] \left[ d^2\theta / d\hat{\phi}^2 \right]^2 + 4 \left[ \partial^2 \log R(\theta) / \partial \hat{\theta}^2 \right] \left[ d\theta / d\hat{\phi} \right] \left[ d^3\theta / d\hat{\phi}^3 \right] \end{aligned}$$

y así

$$\begin{aligned} F_4(\phi) &= \left[ \partial^4 \log R(\phi) / \partial \hat{\phi}^4 \right] / I_\phi^2 = \left[ \partial^4 \log R(\theta) / \partial \hat{\theta}^4 \right] / I_\theta^2 + 6 \left[ \partial^3 \log R(\theta) / \partial \hat{\theta}^3 \right] \left[ -d^2\theta / d\hat{\phi}^2 \right] / I_\theta^2 \left\{ \left[ d\theta / d\hat{\phi} \right] \right\} - \\ &- 3 \left[ d^2\theta / d\hat{\phi}^2 \right]^2 / \left\{ I_\theta \left[ d\theta / d\hat{\phi} \right]^3 \right\} - 4 \left[ d^3\theta / d\hat{\phi}^3 \right] / \left\{ I_\theta \left[ d\theta / d\hat{\phi} \right]^4 \right\}. \end{aligned}$$

Pero de las relaciones de  $d\theta / d\phi$  con  $d\phi / d\theta$  expresadas en el apéndice 2.B y sustituyendolas en la ecuación anterior, se sigue directamente que

$$F_4(\phi) = F_4(\theta) - 6F_3(\theta)I_\theta^{-1/2} \ddot{\phi} / \dot{\phi} - I_\theta^{-1} \left\{ 3(\ddot{\phi} / \dot{\phi})^2 - 4\ddot{\phi} / \dot{\phi} \right\}.$$

APÉNDICE 2.D

Dado que

$$1 = \int_Y h(\underline{y}) \exp[\underline{\eta}' \underline{y} - \psi(\underline{\eta})] dy_1 \dots dy_p,$$

o sea

$$1 = \exp[-\psi(\underline{\eta})] \int_Y h(\underline{y}) \exp[\underline{\eta}' \underline{y}] dy_1 \dots dy_p \quad \Rightarrow \quad \int_Y h(\underline{y}) \exp[\underline{\eta}' \underline{y}] dy_1 \dots dy_p = \exp[\psi(\underline{\eta})];$$

de este resultado se tiene que

$$\begin{aligned} E[\exp(\underline{s}' \underline{y})] &= \int_Y h(\underline{y}) \exp[\underline{\eta}' \underline{y} - \psi(\underline{\eta})] \exp(\underline{s}' \underline{y}) dy_1 \dots dy_p = \exp[-\psi(\underline{\eta})] \int_Y h(\underline{y}) \exp[(\underline{\eta} + \underline{s})' \underline{y}] dy_1 \dots dy_p \\ &= \exp[-\psi(\underline{\eta}) + \psi(\underline{\eta} + \underline{s})] \end{aligned}$$

y; por tanto

$$K(\underline{s}) = \log E_{\underline{\eta}}[\exp(\underline{s}' \underline{y})] = \psi(\underline{\eta} + \underline{s}) - \psi(\underline{\eta}).$$

## Capítulo 3

### Algunas aproximaciones para construir Intervalos de Confianza en presencia de parámetros de ruido

En el capítulo precedente se revisaron dos propuestas para construir intervalos de confianza para el caso de un sólo parámetro. Cada una de ellas, intenta obtener una aproximación a la normal con distintos principios. Ambas propuestas establecen resultados cercanos a los exactos, en tanto que el número extra de operaciones realizadas en comparación con las de los IC estándar, reditúa en mejores aproximaciones - principio de parsimonia- y en ocasiones evita resultados inconsistentes en la construcción de los IC estándar.

La situación más realista y común en la práctica, es aquella en la que además del parámetro de interés, hay otros parámetros presentes (parámetros de ruido).

#### *3.1 Transformaciones que aproximan las funciones de verosimilitud a la normalidad en presencia de parámetros de ruido*

El trabajo de Sprott para esta situación tiene como punto clave, extender el criterio de normalidad de la función de verosimilitud relativa aplicable al caso uniparamétrico, a la normalidad de la FVR maximizada en los parámetros de ruido,  $\theta_2, \dots, \theta_p$  (la llamada verosimilitud perfil; "profile likelihood"). Kalbfleish y Sprott (1970) comentan que la anterior función de verosimilitud relativa puede resultar una

medida pobre para la incertidumbre del parámetro de interés,  $\theta_1$ , si el número de parámetros de ruido ( $p-1$ ) es grande; esto es porque la función de verosimilitud relativa supone que para una  $\theta_1$  fija,  $\theta_2, \dots, \theta_p$  son conocidas e iguales a sus estimadores máximo verosímiles y no toma en cuenta la incertidumbre debido a la falta de conocimiento de  $\theta_2, \dots, \theta_p$ ; por tanto puede acarrear problemas en precisión sobre  $\theta_1$ . Entonces, si la dimensionalidad de los parámetros de ruido no es grande, la proximidad de la FVR (maximizada en  $\theta_2, \dots, \theta_p$ ) a una función normal establecerá un mejor criterio en la construcción de los IC para  $\theta_1$  con la técnica estándar que sólo el tamaño de la muestra.

Para exhibir las cantidades pivotaes derivadas de la teoría de máxima verosimilitud (MV), se usará la siguiente notación. Sea  $\underline{\theta}=(\theta_1, \dots, \theta_p)$  el vector de parámetros de la función de densidad<sup>1</sup>  $f=f(\underline{x};\underline{\theta})$

$$I_{ij} = -\partial^2 \log f / \partial \hat{\theta}_i \partial \hat{\theta}_j, \quad I_{ijk} = \partial^3 \log f / \partial \hat{\theta}_i \partial \hat{\theta}_j \partial \hat{\theta}_k, \\ I_{ijkl} = \partial^4 \log f / \partial \hat{\theta}_i \partial \hat{\theta}_j \partial \hat{\theta}_k \partial \hat{\theta}_l, \quad i, j, k, l=1, \dots, p, \quad (1a)$$

donde  $\partial / \partial \hat{\theta}_i$  denota  $\partial / \partial \theta_i |_{\underline{\theta}=\hat{\underline{\theta}}}$ , con  $\hat{\theta}_i, \hat{\theta}_j$ , los estimadores MV irrestrictos. De manera semejante, sean  $I_{E(ij)}, I_{E(ijk)}, I_{E(ijkl)}$ , las correspondientes cantidades esperadas evaluadas en  $\hat{\underline{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_p)$ , i.e.,

$$I_{E(ij)} = -E\left(\partial^2 \log f / \partial \theta_i \partial \theta_j\right)_{\underline{\theta}=\hat{\underline{\theta}}}, \quad I_{E(ijk)} = E\left(\partial^3 \log f / \partial \theta_i \partial \theta_j \partial \theta_k\right)_{\underline{\theta}=\hat{\underline{\theta}}}, \\ I_{E(ijkl)} = E\left(\partial^4 \log f / \partial \theta_i \partial \theta_j \partial \theta_k \partial \theta_l\right)_{\underline{\theta}=\hat{\underline{\theta}}}, \quad i, j, k, l=1, \dots, p. \quad (1b)$$

Se define la función verosimilitud relativa de  $\theta_1$  como

$$R_M(\theta_1; \underline{x}) = f(\underline{x}; \underline{\theta}^*) / f(\underline{x}; \hat{\underline{\theta}}), \quad (2)$$

donde  $\underline{\theta}^* = (\theta_1, \theta_2^*, \dots, \theta_p^*)$  es el estimador de MV para  $\underline{\theta}$  en un valor específico de  $\theta_1$ , ya que en general las  $\theta_i^* \quad i=2, \dots, p$  serán función de  $\theta_1$ .

De forma semejante a lo visto para las cantidades pivotaes asintóticas normales (sección 1.2), en este caso se utilizan los estimados de MV para  $\theta_1$  que dan:

<sup>1</sup>Nótese como el caso uniparamétrico revisado en el capítulo anterior, queda incluido al hacer  $p=1$ .

$$U(\theta_1; \underline{x}) = (\hat{\theta}_1 - \theta_1) I^{1/2}, \quad (3a)$$

$$U_E(\theta_1; \hat{\theta}) = (\hat{\theta}_1 - \theta_1) I_E^{1/2}, \quad (3b)$$

$$S(\theta_1; \underline{x}) = (\partial \log f / \partial \theta_1^*) / I^{1/2}, \quad (4a)$$

$$S_E(\theta_1; \hat{\theta}) = (\partial \log f / \partial \theta_1^*) / I_E^{1/2}, \quad (4b)$$

$$\{-2 \log R_M(\theta_1; \underline{x})\}^{1/2}, \quad (5)$$

donde  $I=I_{11}$ ,  $I_E=I_{E(11)}$ ,  $\partial \log f / \partial \theta_1^* = \partial \log f(\underline{x}; \underline{\theta}) / \partial \theta_1 |_{\underline{\theta}=\underline{\theta}^*} = \partial \log f(\underline{x}; \underline{\theta}^*) / \partial \theta_1$  representa la función "score", también conocida como función de puntaje. Aunque no es la intención de este trabajo ahondar en el tema de pruebas de hipótesis, su estrecha relación teórica con la de los IC establece que un desarrollo semejante, utilizando la  $R_M(\theta_1; \underline{x})$  llevaría a mejores resultados en tales pruebas. Con este propósito se desprende (5) que es el criterio del cociente de verosimilitud para probar  $\theta_1$ .

En seguida se analiza la relación que guardan (3), (4) y (5). Para esto se desarrolla  $\log R_M(\theta_1; \underline{x}) = \log R_M$  en serie de Taylor y se supondrá que la convergencia es lo suficientemente rápida tal que

$$\log R_M \approx -\frac{1}{2} u^2(\theta_1) \left\{ 1 - \frac{1}{3} (\theta_1 - \hat{\theta}_1) \left( \partial^3 \log R_M / \partial \hat{\theta}_1^3 \right) / I - \frac{1}{12} (\theta_1 - \hat{\theta}_1)^2 \left( \partial^4 \log R_M / \partial \hat{\theta}_1^4 \right) / I \right\},$$

donde  $u(\theta_1) = U(\theta_1; \underline{x})$ . Ahora hágase

$$\log R_M \approx -\frac{1}{2} u^2(\theta_1) \left\{ 1 - \frac{1}{3} u(\theta_1) F_3(\theta_1) - \frac{1}{12} u^2(\theta_1) F_4(\theta_1) \right\}, \quad (6)$$

donde

$$F_3(\theta_1) = \left( \partial^3 \log R_M / \partial \hat{\theta}_1^3 \right) / I^{3/2} \quad (7)$$

y

$$F_4(\theta_1) = \left( \partial^4 \log R_M / \partial \hat{\theta}_1^4 \right) / I^2. \quad (8)$$

$F_3$  y  $F_4$  son, como en la interpretación del capítulo anterior, medidas estándar de la desviación de la FVR a la normalidad debido a la asimetría y la curtosis de ésta, respectivamente.

Realizando un procedimiento semejante para la función "score"

$$\partial \log f / \partial \theta_1^* \approx u(\theta_1) I^{1/2} \left\{ 1 + \frac{1}{2} (\hat{\theta}_1 - \theta_1) \left( \partial^3 \log R_M / \partial \hat{\theta}_1^3 \right) / I - \frac{1}{6} (\hat{\theta}_1 - \theta_1)^2 \left( \partial^4 \log R_M / \partial \hat{\theta}_1^4 \right) / I \right\},$$

entonces

$$\partial \log f / \partial \theta_1^* \approx u(\theta_1) I^{1/2} \left\{ 1 + \frac{1}{2} u(\theta_1) F_3(\theta_1) - \frac{1}{6} u^2(\theta_1) F_4(\theta_1) \right\}. \quad (9)$$

Escogiendo una reparametrización  $\underline{\Phi} = (\phi, \theta_2, \dots, \theta_p)$  donde  $\phi = g(\theta_1)$ , tal que  $F_3$  esté próximo a cero; sí además  $F_4$  es lo suficientemente pequeño entonces la aproximación de  $R_M(\phi; \underline{x}) = R_M(\phi)$  a la normal será aún mejor, esto es, de (6)

$$\log R_M(\phi) \approx -\frac{1}{2} u^2(\phi) \Rightarrow R_M(\phi) \approx \exp \left\{ -\frac{1}{2} u^2(\phi) \right\}. \quad (10)$$

Obsérvese que como consecuencia de (10), (5) y (3a) son casi iguales. Nótese también que bajo esta transformación, la supuesta insignificancia de  $F_4(\phi)$  y la ecuación (9), (4a) y (3a) son aproximadamente iguales y por transitividad (4a) y (5) están igualmente relacionadas. Así que, cuando  $R_M(\phi)$  se aproxime a una función normal, se obtendrán resultados muy cercanos a los exactos en la construcción de los IC.

Un análisis semejante se puede realizar con las fórmulas de las cantidades esperadas (1b), (3b) y (4b), llevando a resultados similares a los anteriores. Se denotará por  $F_{E(3)}(\theta_1)$  y  $F_{E(4)}(\theta_1)$  a los valores de asimetría y curtosis, respectivamente; utilizando valores esperados. En lo que sigue, sólo se hará referencia a las cantidades observadas  $I, F_3, F_4$  siendo el procedimiento equivalente para las cantidades esperadas  $I_E, F_{E(3)}$  y  $F_{E(4)}$ .

De forma semejante al caso uniparamétrico se pueden calcular las siguientes cantidades

$$|F_3(\phi)| = |F_3(\phi; \underline{x})| = |F_3(\theta_1) + 3[I^{1/2} \ddot{\phi} / \dot{\phi}]|, \quad (11)$$

$$F_4(\phi) = F_4(\phi; \underline{x}) = F_4(\theta_1) - 6F_3(\theta_1) \left[ \ddot{\phi} / \dot{\phi} I^{1/2} \right] \cdot \left[ 15(\ddot{\phi} / \dot{\phi})^2 - 4(\ddot{\phi} / \dot{\phi}) \right] / I, \quad (12)$$

donde  $\dot{\phi} = d\phi/d\hat{\theta}_1$ ,  $\ddot{\phi} = d^2\phi/d\hat{\theta}_1^2$  y  $\ddot{\phi} = d^3\phi/d\hat{\theta}_1^3$ . Si se reemplazan las  $I$ 's y  $F(\theta_1)$ 's por los correspondientes esperados en (11) y (12), se podrán calcular  $F_{E(3)}(\phi)$  y  $F_{E(4)}(\phi)$ .

Realizando un procedimiento semejante para la función "score"

$$\partial \log f / \partial \theta_1^* \approx u(\theta_1) I^{1/2} \left\{ 1 + \frac{1}{2} (\hat{\theta}_1 - \theta_1) \left( \partial^3 \log R_M / \partial \hat{\theta}_1^3 \right) / I - \frac{1}{6} (\hat{\theta}_1 - \theta_1)^2 \left( \partial^4 \log R_M / \partial \hat{\theta}_1^4 \right) / I \right\},$$

entonces

$$\partial \log f / \partial \theta_1^* \approx u(\theta_1) I^{1/2} \left\{ 1 + \frac{1}{2} u(\theta_1) F_3(\theta_1) - \frac{1}{6} u^2(\theta_1) F_4(\theta_1) \right\}. \quad (9)$$

Escogiendo una reparametrización  $\underline{\Phi} = (\phi, \theta_2, \dots, \theta_p)$  donde  $\phi = g(\theta_1)$ , tal que  $F_3$  esté próximo a cero; si además  $F_4$  es lo suficientemente pequeño entonces la aproximación de  $R_M(\phi; \underline{x}) = R_M(\phi)$  a la normal será aún mejor, esto es, de (6)

$$\log R_M(\phi) \approx -\frac{1}{2} u^2(\phi) \Rightarrow R_M(\phi) \approx \exp \left\{ -\frac{1}{2} u^2(\phi) \right\}. \quad (10)$$

Obsérvese que como consecuencia de (10), (5) y (3a) son casi iguales. Nótese también que bajo esta transformación, la supuesta insignificancia de  $F_4(\phi)$  y la ecuación (9), (4a) y (3a) son aproximadamente iguales y por transitividad (4a) y (5) están igualmente relacionadas. Así que, cuando  $R_M(\phi)$  se aproxime a una función normal, se obtendrán resultados muy cercanos a los exactos en la construcción de los IC.

Un análisis semejante se puede realizar con las fórmulas de las cantidades esperadas (1b), (3b) y (4b), llevando a resultados similares a los anteriores. Se denotará por  $F_{E(3)}(\theta_1)$  y  $F_{E(4)}(\theta_1)$  a los valores de asimetría y curtosis, respectivamente; utilizando valores esperados. En lo que sigue, sólo se hará referencia a las cantidades observadas  $I, F_3, F_4$  siendo el procedimiento equivalente para las cantidades esperadas  $I_E, F_{E(3)}$  y  $F_{E(4)}$ .

De forma semejante al caso uniparamétrico se pueden calcular las siguientes cantidades

$$|F_3(\phi)| = |F_3(\phi; \underline{x})| = |F_3(\theta_1) + 3[I^{1/2} \ddot{\phi} / \dot{\phi}]|, \quad (11)$$

$$F_4(\phi) = F_4(\phi; \underline{x}) = F_4(\theta_1) - 6F_3(\theta_1) \left[ \ddot{\phi} / \dot{\phi} I^{1/2} \right] - \left[ 15(\ddot{\phi} / \dot{\phi})^2 - 4(\dddot{\phi} / \dot{\phi}) \right] / I, \quad (12)$$

donde  $\dot{\phi} = d\phi/d\hat{\theta}_1$ ,  $\ddot{\phi} = d^2\phi/d\hat{\theta}_1^2$  y  $\dddot{\phi} = d^3\phi/d\hat{\theta}_1^3$ . Si se reemplazan las  $I$ 's y  $F(\theta_1)$ 's por los correspondientes esperados en (11) y (12), se podrán calcular  $F_{E(3)}(\phi)$  y  $F_{E(4)}(\phi)$ .

Sí es posible encontrar una reparametrización  $\phi=g(\theta_1)$  tal que  $F_3(\phi) \approx 0$ , entonces la FVR como función de  $\phi$ , estará más próxima a la normal. Esta reparametrización se puede encontrar resolviendo la ecuación diferencial (11). Una forma es siguiendo la propuesta de Anscombe (1964) vista en el capítulo anterior.

A continuación un ejemplo de una gamma con el parámetro  $r$  como parámetro de ruido.

Ejemplo 3.1 Suponga que una muestra de 30 datos ( $n=30$ ) de una gamma,  $f(x; \lambda, r) = \lambda^r x^{r-1} \exp(-\lambda x) / \Gamma(r)$ , arrojó una suma de 3.29008 y una suma de los logaritmos de las observaciones igual a -294.85575. Para construir un IC para  $\lambda$ , en presencia del parámetro de ruido  $r$  se hace lo siguiente (utilizando lo expuesto arriba)

$$\partial l(\lambda, r; \underline{x}) / \partial \lambda = - \sum_{i=1}^n x_i + nr/\lambda,$$

e igualando a cero el resultado anterior, se obtiene el estimador máximo verosímil de  $\lambda$

$$\hat{\lambda} = \hat{r} / \bar{x} \quad \text{donde} \quad \bar{x} = \sum_{i=1}^n x_i / n.$$

Para calcular el estimador máximo verosímil de  $r$ , se sigue de

$$\partial l(\lambda, r; \underline{x}) / \partial r = \sum_{i=1}^n \log x_i + n \log \lambda - n \frac{\partial}{\partial r} \log \Gamma(r)$$

e igualando a cero esta última derivada y sustituyendo  $\lambda$  por  $\hat{\lambda}$ , se tiene

$$n^{-1} \sum_{i=1}^n \log x_i - \log \bar{x} = \frac{\partial}{\partial r} \log \Gamma(\hat{r}) - \log \hat{r}.$$

Puesto que estimar el valor de  $r$  en la ecuación anterior es complicado, se requiere de algún software capaz de resolverla, por ejemplo el MAPLE V, o revisar de algunas tablas que los calcule, tales como las de Masuyama y Kuroiwa (1952) y Chapman (1956). Los estimaciones máximo verosímiles de  $\lambda$  y  $r$  son

$$\hat{\lambda} = 0.9656 \quad \text{y} \quad \hat{r} = 0.1059.$$

Una vez que se han estimado los valores del parámetro de ruido y del parámetro de interés, se sigue como en el capítulo anterior,

$$\partial^2 l(\lambda, r; \underline{x}) / \partial \lambda^2 = -nr/\lambda^2 \quad \text{y} \quad \partial^3 l(\lambda, r; \underline{x}) / \partial \lambda^3 = 2nr/\lambda^3.$$

Al calcular  $|F_3(\lambda)| = |2/(nr)^{1/2}|$ , se observa que la asimetría de la función no es pequeña para  $\hat{r} = 0.1059$  y  $n=30$ . Por lo tanto, una reparametrización como la propuesta por Anscombe (1964) ayudaría a eliminar el efecto de la asimetría en la FV. La transformación de Anscombe (1964) para este caso sería, como en el ejemplo 2.2

$$\phi = g(\lambda) = \lambda^{1/3}.$$

También, como en el ejemplo 2.2, la transformación  $\phi$  eliminó la asimetría de la FVR.  $|F_3(\phi)|=0$ . Así que, si se utiliza esta transformación en la serie de Taylor de la FVR el término de orden cúbico desaparece.

Obsérvese en las siguientes tablas el contraste en la forma de la FVR, alrededor del estimador de MV y para su transformada cuando el tamaño de muestra es "relativamente grande".

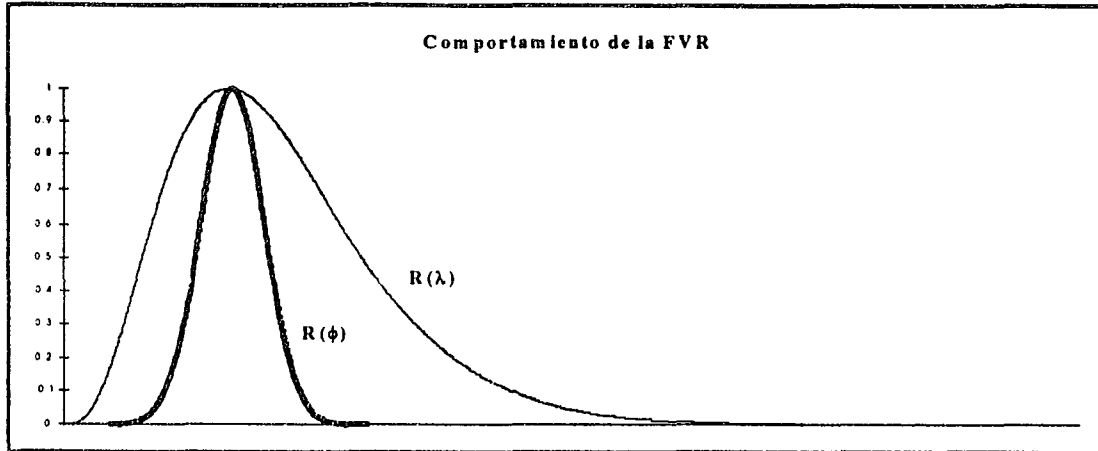
$R(\lambda; \underline{x}, \hat{r})$	0.4504	0.7700	0.9501	1.0000	0.9498	0.7696	0.4505
$\lambda$	0.4325	0.6250	0.8025	<b>0.9656</b>	1.1500	1.4125	1.8200
$ \lambda - \hat{\lambda} $	0.5331	0.3406	0.1631	0	0.1844	0.4469	0.8544

Tabla 8. Intervalos de verosimilitud relativa para el parámetro  $\lambda$  en el caso gamma con r estimada.

$R(\phi; \underline{x}, \hat{r})$	0.4993	0.8006	0.8998	1.0000	0.9000	0.7997	0.4995
$\phi$	0.7705	0.8640	0.9035	<b>0.9884</b>	1.0733	1.1120	1.2062
$ \phi - \hat{\phi} $	0.2179	0.1244	0.0849	0	0.0849	0.1236	0.2178

Tabla 9. Intervalos de verosimilitud relativa para la transformada de  $\lambda$  en el caso gamma con r estimada.

Dado que una función simétrica toma el mismo valor de la función a iguales distancias de su media. Véase en la tabla 8 que la FVR de  $\lambda$  toma a iguales o aproximadamente iguales valores de  $R(\lambda)$  diferentes desviaciones de la media, *i.e.*, no sería muy factible pensar a  $R(\lambda)$  como una función simétrica. En cambio en la tabla 9, para la FVR de la transformada de  $\lambda$ ,  $\phi$ , se observa un comportamiento más simétrico, ya que a la misma o casi la misma distancia de la media se obtienen valores de la  $R(\phi)$  aproximadamente iguales. De manera gráfica  $R(\lambda)$  y  $R(\phi)$  se observan de la siguiente forma.



Gráfica 3. Formas de la FVR para  $\theta$  y  $\phi$  con  $r$  estimada.

Se presenta abajo la confianza real de encontrar al parámetro  $\lambda$  dentro de los LC construidos con las técnicas estándar y de Sprott.

Exacto	0.680	0.750	0.900	0.950	0.990
Estándar	0.710	0.790	0.930	0.954*	0.979*
Sprott	0.669	0.740	0.895	0.947	0.990

Tabla 10. NIVELES REALES DE CONFIANZA. Comparación de los niveles de confianza reales de los IC estándar y de Sprott de dos colas, con  $\alpha = 0.16, 0.125, 0.05, 0.025$  y  $0.005$

Como antes, los espacios de la tabla marcados con "\*" indican que el IC construido para  $\lambda$  toma valores fuera de su rango (se construyó el nivel de confianza real tomando como LCI igual a cero). Los límites de confianza inferiores negativos que fueron calculados con la técnica estándar son: -0.0962 a un nivel de confianza del 95% y -0.4294 a un nivel de confianza del 99%. Lo anterior significa, entre otras cosas, que la técnica estándar construye un IC no óptimo ya que se puede obtener otro IC más pequeños al mismo nivel de confianza tomando como límite de confianza inferior al 0. Por otro lado, los niveles de

confianza reales calculados para ambas técnicas a partir de los LC que con ellas se construyeron tienen una mayor aproximación con la técnica de Sprott que con esos construidos con la técnica estándar.

Siguiendo un desarrollo paralelo al capítulo precedente, hasta aquí, solamente se ha tomado el efecto de la asimetría como posible desviación a una aproximación normal, esto es, se ha supuesto que efectos de cuarto orden o mayores son insignificantes. En ocasiones, asumir esta postura puede llevar a resultados inconsistentes. En este sentido y con el ánimo de ser cautelosos, se puede tomar el efecto que produce la curtosis de la FVR a la desviación de la normalidad. La sugerencia de Sprott, es: si el término  $F_3(\theta_1)$  en (6) es importante entonces reparametrizar para hacer insignificante o nula su contribución en la expansión del  $\log R_M(\theta_1)$ , i.e., eliminar el efecto de la asimetría en la FVR; después, se calcula  $F_4(\theta_1)$  para examinar si una aproximación de la FVR a una distribución t de Student lleva a mejores resultados. Entonces, como en el capítulo anterior, sea  $\varpi$  igual a  $\phi$  ó  $\theta_1$  dependiendo de si hubo reparametrización o no y suponiendo que se cumple  $|F_3(\varpi)| \approx 0$  entonces

$$F_4(\varpi) = 6/(m+1) \tag{13a}$$

o la correspondiente  $F_{E(4)}$  satisfaciendo  $|F_{E(3)}(\varpi)| \approx 0$

$$F_{E(4)}(\varpi) = 6(m+2)(m+3)^2 / [m(m+1)(m+5)(m+7)] \tag{13b}$$

y

$$v\sqrt{m/(m+1)} = t_{(m)}, \tag{14}$$

donde  $v = v(\varpi; \underline{x}) = (\hat{\varpi} - \varpi) I_{\varpi}^{-1/2}$ . Luego, despejando a  $m$  de (13) se tiene que  $v$  sigue una distribución<sup>2</sup> t con  $m$  g.l.. Así que, si  $F_3(\varpi)$  esta próximo a cero, Sprott recomienda utilizar (14) como una t con  $m$  g.l. y; si  $F_4(\varpi)$  es pequeño  $m$  será grande y  $v$  será aproximadamente una  $N(0,1)$ . Sprott también recomienda tomar a  $m$  como el número entero más próximo que satisface (13), según sea el caso. Al igual que en el capítulo 2 sí

$$|F_4(\varpi)| \in [0.231, 3],$$

$$|F_{E(4)}(\varpi)| \in [5.626, 12]$$

<sup>2</sup>Sprott (1982)

sería aceptable la propuesta de Sprott de utilizar la  $t$  de Student; pero si la cantidad  $F_4(\varpi) \in (-0.231, 0.231)$  ó  $F_{L(4)}(\varpi) \in (-5.626, 5.626)$  se recomienda utilizar una aproximación normal.

**Ejemplo 3.2** Suponga que en el ejemplo 3.1 se quiere analizar el efecto que provoca la curtosis en la FVR. Para esto, primero se calcula dicho efecto con el parámetro y después con su transformación.

La función gamma con  $r$  estimada: sin reparametrización es  $F_4(\lambda) = -6/(nr)$  y para su transformación  $F_4(\phi) = -2/(9nr)$ , donde  $\phi = \lambda^{1/3}$ . Obsérvese que  $F_4(\lambda)$  es importante cuando  $n$  es 30 y  $\hat{r} = 0.1059$ ; por lo que resultaría errado asumir que su presencia es insignificante para la aproximación. En cambio, el valor de  $F_4(\phi)$  en la FVR reparametrizada es pequeño (0.0699) de lo que se concluye que, en este caso, la transformación  $\phi$  además de eliminar problemas de asimetría también reduce desviaciones de la función de verosimilitud por el efecto de la curtosis, de acuerdo al criterio expuesto arriba.

### 3.2 *Un enfoque bayesiano para la construcción de intervalos de confianza en presencia de parámetros de ruido*

Ahora se presenta el trabajo que realizó Efron (1993) para construir los IC en presencia de parámetros de ruido desde un punto de vista bayesiano. La idea de él radica, primero, en utilizar algunos elementos de ortogonalización de parámetros en el sentido de Cox y Reid (1987). Después define la "verosimilitud implicada" (implied likelihood) a la cual compara con la verosimilitud perfil y muestra su proximidad a esta última<sup>3</sup>; el hecho de compararla con la verosimilitud perfil es para sugerir que la

<sup>3</sup>Véase teorema 2 de Efron (1993).

verosimilitud implicada goza de "buenas" propiedades al igual que la verosimilitud perfil<sup>4</sup>. Finalmente, propone una función de densidad a priori para luego obtener desde un enfoque bayesiano, la función de densidad a posteriori y con ella construir los IC.

Esta presentación se divide en un breviario de la ortogonalidad de parámetros en el sentido de Cox y Reid (1987). Esto ayudará a disminuir el efecto de los parámetros de ruido en la construcción de los IC. Después se da un panorama general de la verosimilitud implicada. Se presentan algunas de sus características cuando existen o no parámetros de ruido. Una presentación semejante a la anterior se hace para la función de densidad a priori que propone Efron (1993) utilizando elementos de los métodos ABC y ABC<sub>q</sub>. Por último se muestra la función de densidad a posteriori que tiene como características principales, según Efron (1993): estar aproximado a segundo orden (SOA) y dada las particularidades que ofrece el utilizar elementos de los métodos ABC y ABC<sub>q</sub>, el que no aparezcan parámetros de ruido en la función de densidad a posteriori.

El concepto de parámetros ortogonales que se adoptará en lo que sigue es el referente al artículo de Cox y Reid (1987) y que lo definen como sigue.

Sea  $\underline{y}=(y_1, \dots, y_p)$  el vector de estadísticas suficientes,  $g_{\underline{\mu}}(\underline{y})$  la función de densidad,  $\underline{\mu}=(\mu_1, \dots, \mu_p)$  el vector de parámetros desconocidos y  $l(\underline{\mu})$  la log-verosimilitud. Si  $\underline{\mu}$  se particiona en dos vectores,  $\underline{\theta}$  y  $\underline{v}$  de dimensiones  $p_1$  y  $p_2$ , respectivamente, tal que  $p_1+p_2=p$ , se dice que  $\underline{\theta}$  es ortogonal a  $\underline{v}$  si los siguientes elementos de la matriz de información de Fisher satisfacen

$$i_{\theta, v_s} = \frac{1}{n} E_{\underline{\mu}} \left[ \frac{\partial l(\underline{\mu})}{\partial \theta_s} \frac{\partial l(\underline{\mu})}{\partial v_t} \right] = \frac{1}{n} E_{\underline{\mu}} \left[ \frac{\partial^2 l(\underline{\mu})}{\partial \theta_s \partial v_t} \right] = 0, \quad (15)$$

para  $s=1, \dots, p_1$  y  $t=p_1+1, \dots, p$ . La definición de ortogonalidad se puede extender a más de dos conjuntos de parámetros  $y$ ; en particular,  $\underline{\mu}$  es totalmente ortogonal si la matriz de información de Fisher es diagonal. Algunas consecuencias de la ortogonalidad de parámetros suponiendo  $p_1=p_2=1$  son<sup>5</sup>:

<sup>4</sup>Algunas propiedades de la verosimilitud perfil son que se equipara a la función de verosimilitud en caso de no existir parámetros de ruido  $y$ ; aún en problemas con la existencia de parámetros de ruido la verosimilitud perfil sigue el principio de verosimilitud: su valor sólo depende de la forma de la función de densidad y no de algún otro conjunto de datos que pudo haber sido observado.

<sup>5</sup>Cox y Reid (1987).

- i) Los estimadores máximo verosímiles de  $\theta$  y  $\underline{v}$  son asintóticamente independientes.
- ii) La estimación asintótica del error estándar de  $\theta$  es la misma si  $\underline{v}$  se trata como conocida o no.

Las anteriores observaciones son particularmente ciertas para el caso en que  $\theta$  y  $\underline{v}$  sean de dimensiones 1 y  $(p-1)$ , respectivamente. En el caso de que los parámetros no sean ortogonales y se desee hacer estimaciones sobre alguno de ellos (el parámetro de interés), Cox y Reid (1987) proponen hacer una reparametrización apropiada, tal que (15) se cumpla para este nuevo planteamiento y de la cual sea posible estimar el parámetro de interés.

Ahora se presenta la construcción de IC que propone Efron (1993) desde un enfoque bayesiano para el caso de la familia exponencial, esto es, los IC son construidos una vez que se conoce la función de densidad a posteriori,  $\pi(\theta)$ , esto es, de

$$\int_{h_1(\underline{x})}^{h_2(\underline{x})} \pi(\theta) d\theta = 1 - \alpha,$$

donde  $h_1$  y  $h_2$  son los límites de confianza inferior y superior, respectivamente. El cálculo de  $\pi(\theta)$  se construye siguiendo el teorema de Bayes

$$\pi(\theta) \propto L(\theta)\pi_0(\theta),$$

donde  $L(\theta)$  es la función de verosimilitud y  $\pi_0(\theta)$  la función de densidad a priori. Efron (1993) propone utilizar la verosimilitud implicada (más adelante se define) en vez de la verosimilitud estándar. Si se utiliza la verosimilitud implicada calculada con algunas constantes obtenidas con los métodos ABC y/o ABC<sub>q</sub>, ( $a$ ,  $z_0$ ,  $c_q$ ), Efron demuestra que existen similitudes entre esta metodología con respecto a las densidades y la verosimilitud y las propuestas por Welch-Peers (1963), Peers (1965) y Tibshirani (1989) dependiendo si hay parámetros de ruido presentes o no.

Para definir la verosimilitud implicada se necesita de los elementos que a continuación se presentan.

La distribución confidencial denotada por  $\pi_{\underline{x}}^*(\theta) = d\alpha_{\underline{x}}(\theta)/d\theta$  la cual conserva la idea descrita en el capítulo 1. Para entender mejor lo anterior se muestran los siguientes ejemplos para obtener  $\alpha_{\underline{x}}(\theta)$ . Para

esto, sea  $\hat{\theta}_A^{(1-\alpha)}$  el límite superior aproximado con confianza  $(1-\alpha)$  para  $\theta=t(\underline{\mu})$  el parámetro escalar de interés con segundas derivadas continuas.

En el caso de la construcción de los IC estándar

$$\hat{\theta}_A^{(1-\alpha)} = \hat{\theta} + \hat{\sigma}z_\alpha, \quad (16)$$

donde  $\hat{\theta}$  es la estimación máximo verosímil de  $\theta$ ,  $\hat{\sigma}$  es la estimación del error estándar de  $\hat{\theta}$  y  $z_\alpha$  es el cuantil de la normal estándar a nivel  $1-\alpha$ , i.e.,  $z_\alpha = \Phi^{-1}(1-\alpha)$ , con  $\Phi^{-1}(\cdot)$  como la inversa de la función de distribución normal estándar. Sea  $\alpha_x(\theta)$  el valor con nivel de confianza  $1-\alpha$  correspondiente al límite superior de  $\theta$  para el IC construido. Suponga que  $\alpha_x(\theta)$  es derivable con respecto a  $\theta$ , entonces despejando el nivel de confianza  $1-\alpha$  de (16) se tiene que  $\alpha_x(\theta) = \Phi\left[\frac{(\theta - \hat{\theta})}{\hat{\sigma}}\right]$ . Por tanto la distribución confidencial aproximada para este caso es

$$\pi_x^*(\theta) = \frac{d\alpha_x(\theta)}{d\theta} = \frac{d\Phi\left[\frac{(\theta - \hat{\theta})}{\hat{\sigma}}\right]}{d\theta} = \phi\left[\frac{(\theta - \hat{\theta})}{\hat{\sigma}}\right],$$

donde  $\phi(\cdot)$  es la función de densidad normal estándar.

Para el caso de los IC construidos con el método ABC

$$\hat{\theta}_{ABC}^{(1-\alpha)} = t\left(\hat{\underline{\mu}} + \frac{1}{\lambda}\hat{\underline{\delta}}\right)$$

y realizando algunas operaciones se llega a que  $\alpha_x(\theta) = \Phi[w(\theta) - z_0]$ , con  $w(\theta) = w$  calculado así (de la definición del método ABC)

$$\lambda = \frac{t(\hat{\theta})\left[t^{-1}(\theta) - \hat{\underline{\mu}}\right]}{\hat{\sigma}} \rightarrow w = \frac{2\lambda}{(2a\lambda + 1) + (4a\lambda + 1)^{1/2}}. \quad (17a)$$

De igual manera, para el método  $ABC_q$  se sigue de

$$\hat{\theta}_{ABC_q}^{(1-\alpha)} = \hat{\theta} + \hat{\sigma}\xi$$

y después de algunas operaciones se obtiene  $\alpha_x(\theta) = \Phi[w - z_0]$ , con  $w$  calculado siguiendo la definición del método  $ABC_q$

$$\xi = \frac{\theta - \hat{\theta}}{\hat{\sigma}} \rightarrow \lambda = \frac{2\xi}{1 + (4\xi c_q + 1)^{1/2}} \rightarrow w = \frac{2\lambda}{(2a\lambda + 1) + (4a\lambda + 1)^{1/2}}. \quad (17b)$$

Por otro lado, se ha denotado a la distribución confidencial aproximada con un subíndice el cual señala la muestra observada,  $\underline{x}$ . Suponga que se está en un esquema de muestreo repetido. Si se toman dos muestras de igual tamaño y además cada una de las muestras es igual con  $\underline{x}$ , entonces la unión de estas dos muestras será  $(\underline{x}, \underline{x})$  y su distribución confidencial aproximada se denotará como  $\pi_{\underline{xx}}^*(\theta) = d\alpha_{\underline{xx}}(\theta)/d\theta$ .

Con estos elementos Efron define la verosimilitud implicada como

$$L^*(\theta) = \frac{\pi_{\underline{xx}}^*(\theta)}{\pi_{\underline{x}}^*(\theta)}.$$

Si se calcula la verosimilitud implicada a partir de las distribuciones confidenciales obtenidas por los IC estándar se tiene

$$L^*(\theta) \propto \exp\left\{-\frac{(\theta - \hat{\theta})^2}{2\hat{\sigma}^2}\right\}.$$

Efron (1993) demostró que si se calcula la verosimilitud implicada para el parámetro  $\theta$ , a partir de la construcción de los intervalos de confianza ABC y ABC<sub>q</sub>, ésta tiene la siguiente forma

$$L^*(\theta) \propto \exp\left\{-\frac{1}{2}w^2\right\},$$

si no hay parámetros de ruido y se necesita de un ajuste si existen parámetros de ruido:

$$L^{**}(\theta) \propto \exp\left\{-\frac{1}{2}w^2\right\} \exp\left\{-\hat{\gamma}\theta/\hat{\sigma}\right\}.$$

donde  $\hat{\gamma} = \hat{b}/\hat{\sigma} - c_q$ . Se calculará  $w$  con (17a) o (17b) dependiendo el método que se haya utilizado para la construcción del IC. Como se comentó en el capítulo anterior, si la dimensión del vector canónico es  $p=1$ , o si  $\theta=t(\underline{\mu})$  es una combinación lineal del vector de medias  $\underline{\mu}$ , entonces  $\hat{\gamma} = 0$  y  $a=z_0$  (de la definición de  $\hat{\gamma} = a - z_0$ ); además  $L^{**}(\theta)$  será igual a  $L^*(\theta)$ .

En lo que sigue se supondrá que la verosimilitud implicada tomará como base para su cálculo alguno de los métodos ABC o ABCq. Ahora bien, ¿Qué tal trabajan la función de verosimilitud implicada y la función de verosimilitud implicada ajustada comparándolas con la verosimilitud perfil y la verosimilitud perfil modificada, respectivamente? En este sentido Efron (1993) demostró que la función de verosimilitud implicada es una aproximación de segundo orden a la función de verosimilitud perfil y a la función de verosimilitud perfil ajustada en caso de parámetros de ruido presentes, al menos cuando el parámetro de interés es un elemento del vector de parámetros canónicos y se está en la familia exponencial.

Es importante resaltar que para construir IC con las ideas de Efron (1993) cuando existen parámetros de ruido en la función de densidad, lo primero será ortogonalizar el parámetro de interés con respecto a los parámetros de ruido y entonces, calcular la función de verosimilitud implicada. También para la función de densidad a priori se supondrá que en caso necesario, o sea, en caso de que existan parámetros de ruido presentes se ortogonalizará antes.

Ahora bien, para dar una interpretación bayesiana es necesario una función de densidad a priori. En este contexto Efron (1993) propone las siguientes que tienen como base la construcción de IC por medio de los métodos ABC y ABC<sub>q</sub> aplicados a la distribución confidencial

$$\pi_0^*(\theta) \propto \frac{e^{z_0 w} (1 - aw)^3}{(1 + aw)(1 + 2c_q \lambda)},$$

si no hay parámetros de ruido y una modificación en caso de la existencia de dichos parámetros

$$\pi_0^{**}(\theta) \propto \frac{e^{z_0 w} (1 - aw)^3}{(1 + aw)(1 + 2c_q \lambda)} e^{\hat{\gamma} \theta / \hat{\sigma}}.$$

A  $\pi_0^*(\theta)$  y  $\pi_0^{**}(\theta)$  Efron (1993, 17-18) las llamo: la función de densidad a priori implicada (implied prior density function) y la función de densidad a priori implicada modificada (modified implied prior density function), respectivamente. Obsérvese que si  $\hat{\gamma} = 0$  entonces  $\pi_0^*(\theta)$  y  $\pi_0^{**}(\theta)$  son iguales. Efron (1993) demuestra que la aproximación de esta propuesta con la densidad a priori no informativa de Welch-Peers (1963) y la de Tibshirani (1989) es de segundo orden, dependiendo de la existencia o no de parámetros de ruido.

Una vez que se tienen la función de densidad a priori y la función de verosimilitud ¿qué tan “buena” es ésta aproximación para la construcción de IC?, esto es, ¿Qué tan buena es la aproximación en la construcción de IC a partir de la función de densidad a posteriori que Efron (1993) propone? Él demostró que su aproximación es de segundo orden, esto es,  $P[\theta < \hat{\theta}^{(1-\alpha)}; \underline{\eta}] = 1 - \alpha + O(n^{-1})$ . Su demostración toma como base la relación que existe entre sus propuestas y la teoría de Welch-Peers (1963), Peers (1965) y Tibshirani (1989), dependiendo de la presencia o no de parámetros de ruido.

Por tanto, de todo lo anterior se tiene que independientemente de tratar con parámetros de ruido o no, la función de densidad a posteriori es

$$\pi^*(\theta) \propto L^*(\theta)\pi_0^*(\theta) = L^{**}(\theta)\pi_0^{**}(\theta)$$

y su aproximación al IC exacto es del orden  $O(n^{-1})$ .

En seguida se presenta un ejemplo aplicando esta metodología.

Ejemplo 3.3 Suponga que se tienen la misma muestra que en el ejemplo 3.1 y se quiere construir un IC para  $\phi = E[y]$ . Entonces, para calcular la función de densidad final y de ahí construir los IC se sigue de la propuesta de Efron (1993), como primer paso para calcular la función de densidad, ortogonalizar el parámetro  $\phi$  con respecto a los parámetros de ruido. En este caso  $\theta = \phi$  es ortogonal a  $r$  y esto es posible demostrarlo de la siguiente manera.

Como  $\phi = E[y] = r/\lambda$ , entonces  $\lambda = r/\phi$ . Reparametrizando  $f(u; \lambda, r)$ <sup>6</sup> como  $f(u; \phi, r)$  con  $u = \sum x_i$ ,

$$f(u; \phi, r) = \left(\frac{r}{\phi}\right)^r u^{r-1} \exp\left(-\frac{r}{\phi}u\right) / \Gamma(r)$$

y

$$l(\phi, r; u) = -\frac{r}{\phi}u + (r-1)\log u - \left[\log \Gamma(r) - r \log\left(\frac{r}{\phi}\right)\right].$$

<sup>6</sup> La suma de v.a.'s i.i.d. como gammas de parámetros  $\lambda$  y  $r$  se distribuye como otra gamma de parámetros  $\lambda$  y  $nr$ .

Siguiendo los pasos de ortogonalización de Cox-Reid (1987), se calcula el valor esperado de  $\partial^2 l / \partial \phi \partial r$ , i.e.,

$$\begin{aligned} \frac{\partial^2 l}{\partial \phi \partial r} &= \frac{u}{\phi^2} - \frac{1}{\phi}, \\ E\left[\frac{\partial^2 l}{\partial \phi \partial r}\right] &= \int_U \left[\frac{u}{\phi^2} - \frac{1}{\phi}\right] \frac{\left(\frac{u}{\phi}\right)^r u^{r-1} \exp\left(-\frac{u}{\phi}\right)}{\Gamma(r)} du = \\ &= -\frac{1}{\phi} + \frac{1}{\phi^2} \int_U \frac{\left(\frac{u}{\phi}\right)^r u^r \exp\left(-\frac{u}{\phi}\right)}{\Gamma(r)} du = -\frac{1}{\phi} + \frac{1}{\phi} \int_U \frac{\left(\frac{u}{\phi}\right)^{r+1} u^r \exp\left(-\frac{u}{\phi}\right)}{r\Gamma(r)} du = \\ &= -\frac{1}{\phi} + \frac{1}{\phi} \int_U \frac{\left(\frac{u}{\phi}\right)^{r+1} u^r \exp\left(-\frac{u}{\phi}\right)}{\Gamma(r+1)} du = 0. \end{aligned}$$

De aquí que  $\phi$  y  $r$  sean ortogonales según Cox y Reid (1987). El siguiente paso es calcular  $\hat{\phi}, \hat{\sigma}, a, z, c_q$ , entonces de la función de densidad conjunta

$$f(\underline{x}; \phi, r) = \exp\left\{-\frac{1}{\phi} \sum x_i + (r-1) \sum \log x_i - n\left[\log \Gamma(r) - r \log\left(\frac{1}{\phi}\right)\right]\right\},$$

se definen los parámetros canónicos y la función  $\psi(\cdot)$

$$\eta_1 = -\frac{r}{\phi}, \quad \eta_2 = r \quad \text{y} \quad \psi = \psi(\eta_1, \eta_2) = n\left[\log \Gamma(\eta_2) - \eta_2 \log(-\eta_1)\right].$$

Luego, se deriva  $\psi$  con respecto a los parámetros  $\eta_1$  y  $\eta_2$

$$\begin{aligned} \psi_1 &= \frac{\partial \psi}{\partial \eta_1} = \mu_1 = -\frac{n\eta_2}{\eta_1} = n\phi, & \psi_{11} &= \frac{\partial^2 \psi}{\partial \eta_1^2} = \frac{n\eta_2}{\eta_1^2} = \frac{n\phi^2}{r}, \\ \psi_{111} &= \frac{\partial^3 \psi}{\partial \eta_1^3} = -\frac{2n\eta_2}{\eta_1^3} = -\frac{2n\phi^3}{r^2}. \end{aligned}$$

Para este caso en particular las demás derivadas de  $\psi$  no son necesarias calcularlas, ya que éstas se eliminan en los productos en que aparecen. Los respectivos estimadores son

$$\hat{\psi}_1 = \hat{\mu}_1 = \sum x_i, \quad \hat{\psi}_{11} = \frac{n^3 \bar{x}}{\hat{r}} \quad \text{y} \quad \hat{\psi}_{111} = \frac{2n^4 \bar{x}}{\hat{r}^2},$$

donde  $r$  se estima como en el ejemplo 3.1. El estimador de máxima verosimilitud para  $\phi$  es

$$\hat{\phi} = \bar{x} = \frac{\sum x_i}{n}.$$

La función  $t(\mu)=\phi$  es  $t(\mu)=\mu_1/n$  y sus respectivas derivadas son

$$i_1 = \frac{1}{n}, \quad i_2 = \ddot{i}_1 = \ddot{i}_{12} = \ddot{i}_2 = 0$$

y; aplicando el algoritmo que propone Efron (1993) para calcular la función de densidad final se sigue de las cantidades estimadas

$$\hat{\sigma} = \frac{\hat{\phi}}{(nr)^{1/2}}, \quad a = z = \frac{1}{3(nr)^{1/2}} \quad \text{y} \quad c_q = 0$$

y; según sea el método que utilizado para calcular dicha función, se tiene de (17) lo siguiente.

Utilizando el método  $ABC_q$

$$\xi = \frac{\phi - \hat{\phi}}{\hat{\sigma}_\phi} = \lambda \rightarrow w = \frac{2\lambda}{(2a\lambda + 1) + (4a\lambda + 1)^{1/2}}$$

y para el método ABC

$$\lambda = \frac{\frac{1}{n}(\mu_1 - \hat{\mu}_1)}{\hat{\sigma}_\phi} = \frac{\phi - \hat{\phi}}{\hat{\sigma}_\phi} \rightarrow w = \frac{2\lambda}{(2a\lambda + 1) + (4a\lambda + 1)^{1/2}}$$

Como se puede observar para este caso, con ambos métodos se obtienen la misma función de densidad final. Dado lo anterior, se calculan  $\pi_0^{**}(\phi)$  y  $L^{**}(\phi)$  para entonces obtener la función de densidad final

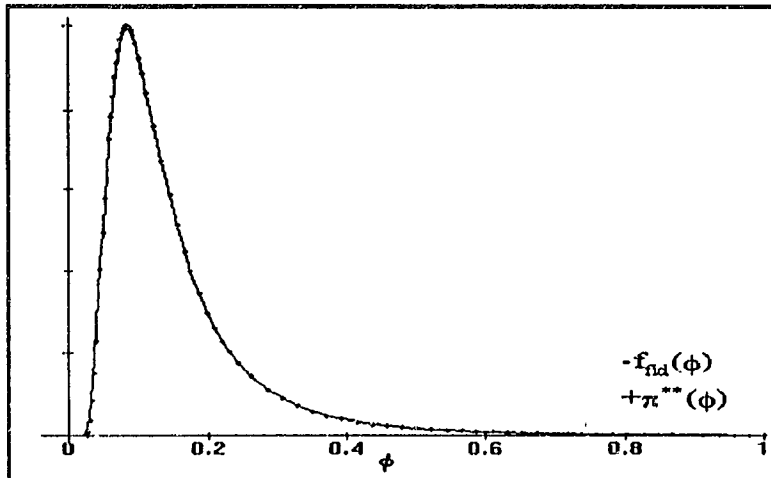
$$\pi_0^{**}(\phi) = \frac{(1 - aw)^3 \exp\left\{z_0 w - \frac{1}{2} w^2\right\}}{(1 + aw)(1 + 2c_q \lambda)}$$

Para observar la aproximación que tiene esta función respecto a la exacta, se hará una comparación gráfica entre esta función y la función de densidad fiducial<sup>7</sup> (explícita),  $f_{fid}(\phi)$ . La función de densidad fiducial para este caso es

$$f_{fid}(\phi) = \frac{\left(\frac{nr}{\phi}\right)^{nr+1} \exp\left\{-\frac{nr}{\phi}\right\}}{\Gamma(nr + 2)},$$

<sup>7</sup>Una presentación de la construcción de la función de densidad fiducial se exhibe en la tesis de Gongora (1988).

donde  $t_0 = \sum x_i$ .



Gráfica 4. Comparación de las funciones de densidad final  $\pi^{**}(\phi)$  y fiducial  $f_{nid}(\phi)$ .

Se puede observar que  $\pi^{**}(\phi)$  es prácticamente igual a la  $f_{nid}(\phi)$ . De lo anterior se puede concluir que para  $\theta = \phi$ , en la gamma, la aproximación de Efron (1993) es cercana a la exacta, de la cual se puede suponer que los IC construidos con  $\pi^{**}(\phi)$  llevarán a resultados próximos a los exactos.

## Conclusiones

La construcción de intervalos de confianza cercanos a los exactos es importante, ya que ellos son o deberían ser en muchas ocasiones el resultado final de un análisis estadístico.

Al paso de las páginas algo que se puntualizó fue la necesidad de tener un mayor cuidado al momento de construir los intervalos de confianza, ya que por simplicidad algorítmica, costumbre, desconocimiento de los supuestos que sustentan a la técnica estándar, etc., en ocasiones se pueden construir intervalos de confianza que no caractericen debidamente al parámetro bajo estudio y lleve a conclusiones poco próximas a las verdaderas.

Usualmente se construyen los intervalos de confianza suponiendo que con un tamaño de muestra razonablemente grande existirá normalidad, *i.e.*, se construyen intervalos de confianza a partir de la regla de la media  $\pm 2$  desviaciones estándar (para un intervalo de confianza de dos colas a un nivel de confianza de 95%).

Algunas veces tomar el supuesto de normalidad tan sólo por el tamaño de la muestra puede llevar, entre otras cosas, a poca precisión en los resultados; como por ejemplo: una región más pequeña o más grande de la supuestamente construida; o bien, resultados inconsistentes como lo es obtener valores parametrales fuera de su rango.

Por tanto, la revisión de otras alternativas que puedan reducir tales inconsistencias es una idea que debe permanecer latente. En este trabajo se presentaron algunas propuestas como son: la construcción de intervalos de confianza con y sin parámetros de ruido de Sprott (1973, 1980, 1982), las técnicas ABC y ABC<sub>q</sub> de Diccio y Efron (1992) para la familia exponencial y la técnica de Efron (1993) en el caso de

parámetros de ruido presentes. En general, todas estas técnicas resultan en buenas aproximaciones a los resultados exactos.

Sprott (1973, 1980, 1982) fija su atención para construir los intervalos de confianza, en la proximidad de la función de verosimilitud a una normal, en vez de tan sólo esperar que el tamaño de la muestra sea lo suficientemente importante como para suponer normalidad. Las ideas de Sprott (1973, 1980, 1982) estriban en que si la función de verosimilitud para el parámetro bajo estudio  $\theta$  no se aproxima a una normal, entonces se busca una transformación de  $\theta$ ,  $\phi=g(\theta)$ , tal que la función de verosimilitud en esta reparametrización se acerque más a la de una normal. Sprott (1973, 1980, 1982) corrige el efecto de asimetría con esta transformación y propone una aproximación a la distribución t de Student en caso de problemas de curtosis. Como aportación de este trabajo a las ideas de Sprott, se propone un criterio para decidir sobre la posible presencia o no de curtosis de la función de verosimilitud.

Cuando existen parámetros de ruido, Sprott (1980) sugiere utilizar la verosimilitud perfil como primer paso para eliminar los parámetros de ruido y luego emplear lo antes mencionado para el caso uniparamétrico. Un problema que se puede presentar al utilizar la verosimilitud perfil es que se pueden obtener estimaciones inconsistentes o ineficientes. Quizás una mejor aproximación de este resultado sería: primero, ortogonalizar en el sentido de Cox y Reid (1987) el parámetro de interés con respecto a los de ruido  $y$ ; después calcular la función de verosimilitud perfil o alguna otra versión de la verosimilitud. Lo anterior podría disminuir el efecto de los parámetros de ruido sobre el comportamiento del parámetro de interés en la función de verosimilitud.

Por otra parte, los intervalos de confianza construidos por Diccio y Efron (1992) se basan en hacer algunos ajustes de orden  $O(n^{-1})$  al cuantil exacto a partir del cuantil de una normal estándar. Estas técnicas toman algunos elementos del remuestreo para mejorar las aproximaciones y facilitar los cálculos. Además, se sugiere hacer un ajuste al método  $ABC_q$  en caso de que la muestra sea pequeña, llamado  $ABC_{2q}$ .

Efron (1993) se enfoca en el problema de construir intervalos de confianza en presencia de parámetros de ruido, al cual le da una solución bayesiana. Propone utilizar las ideas de ortogonalización de Cox y Reid (1987) para reducir el efecto de los parámetros de ruido. Luego, construye las funciones de verosimilitud implicada y de densidad a priori, que llevan a una función de densidad a posteriori libre de parámetros de ruido y cercana a la distribución fiducial del parámetro de interés. Esta técnica tiene algunas cuestiones que no son del todo obvias, por ejemplo: dado que, la función de densidad a priori está en función de la muestra a través de  $a$ ,  $z_0$  y  $c_q$ , entonces ¿qué tan bayesiana es esta técnica?. Otro punto a discutir es acerca del supuesto de obtener dos muestras independientes e idénticas para calcular la llamada verosimilitud implicada, ya que este supuesto es difícil que suceda en muchos experimentos.

El pago de todas estas aproximaciones recae en el número adicional de operaciones que se necesitan calcular con respecto a la técnica estándar; sin embargo, generalmente se obtendrán mejores resultados. Así que, el uso de técnicas más refinadas para la construcción de intervalos de confianza se puede ponderar a través del principio de parsimonia.

Los principales defectos de las técnicas presentadas se refieren a la dificultad de cálculo de algunos de sus elementos que conforman sus respectivos algoritmos, por ejemplo: calcular la transformación  $\phi$  del parámetro de interés,  $\theta$ , en el caso de la técnica de Sprott; o la ortogonalización de los parámetros de ruido y de interés en la técnica de Efron (1993).

Al revisar estas técnicas surgieron algunas dudas tales como: en el caso de que no sea posible obtener una reparametrización como la propuesta por Anscombe (1964) ¿qué aplicar?. En el caso de parámetros de ruido y trabajando con la propuesta de Sprott (1980), ¿cuántos parámetros de ruido se pueden admitir para que la aproximación siga siendo “buena”? o ¿cómo se deben comportar los parámetros de ruido para que la aproximación siga siendo “buena”? ¿Se puede extender los métodos ABC y ABC<sub>q</sub> a otras familias además de la exponencial?. ¿Cómo definir la mejor reparametrización ortogonal que permita hacer los cálculos más sencillos?. Se espera que esas cuestiones abran nuevos campos de investigación e interese más al lector sobre la construcción de intervalos de confianza.

## Anexo

### Construcción de Intervalos de Confianza basada en el análisis de Sprott

La estructura del análisis de Sprott (1973, 1980, 1982) para interpretar la forma de la función de verosimilitud y construir los intervalos de confianza para el caso uniparamétrico, se puede representar con el siguiente algoritmo:

- 1 Revisar la forma de la función de verosimilitud (FV) del parámetro de interés,  $\theta$ .
- 2 Calcular  $F_3(\theta)$  y  $F_4(\theta)$  y
  - 2.1 Sí  $F_3(\theta)$  es pequeño y
    - 2.1.1 Sí  $F_4(\theta) \in (-0.231, 0.231)$ , entonces  $u_0 = (\hat{\theta} - \theta)I_{\hat{\theta}}$  es aproximadamente una v.a. normal estándar, *i.e.*, sí se construye el IC estándar, éste estará muy próximo del IC exacto; o
    - 2.1.2 Sí  $|F_4(\theta)| \in [0.231, 3]$ , entonces  $u_0$  sigue aproximadamente una distribución t de Student con  $m$  g.l., donde  $m$  se calcula de (2.9a); o
  - 2.2 Sí  $F_3(\theta)$  no es pequeño, entonces reparametrizar la FV tal que  $F_3(\phi) \approx 0$  y<sup>1</sup>
    - 2.2.1 Sí  $F_4(\phi) \in (-0.231, 0.231)$ , entonces  $u_\phi$  es aproximadamente una v.a. normal; o
    - 2.2.2 Sí  $|F_4(\phi)| \in [0.231, 3]$ , entonces  $u_\phi$  sigue aproximadamente una distribución t de Student con  $m$  g.l., donde  $m$  se calcula de (2.9a).
- 3 Construir el IC de acuerdo a las especificaciones del problema a analizar:
  - 3.1 Dos colas  $P[LI \leq u \leq LS] = 1 - \alpha$ ; o
  - 3.2 Cola a la derecha  $P[LI \leq u] = 1 - \alpha$ ; o
  - 3.3 Cola a la izquierda  $P[u \leq LS] = 1 - \alpha$ .

<sup>1</sup>El resultado de Anscombe (1964) puede ser una opción para conseguir esta reparametrización.

Donde LI y LS (los límites de confianza inferior y superior) son los cuantiles de la distribución normal estándar o de la t de Student, según sea el caso y  $u$  será  $u_\phi$  ó  $u_0$  dependiendo si hubo reparametrización o no.

Una interpretación del algoritmo hace posible dividirlo en 3 etapas, cada uno indicando lo siguiente.

El primer paso da una interpretación visual de la forma de la FV, esto es, una gráfica de la FV permite observar si hay indicios de posibles desviaciones de su forma con respecto a la de una función normal. En esta etapa no se explica matemáticamente tal representación, por lo que se recomienda observarla pero no basar las decisiones en esta etapa - a menos que el analista tenga la suficiente experiencia como para interpretarla.

En el segundo paso se calculan las cantidades  $F_3(\theta)$  y  $F_4(\theta)$ . Estas cantidades (representando medidas estándar de asimetría y curtosis de la FV) darán las condiciones para interpretar la forma de la FV, esto es, qué decidir sobre la forma de la FV si las cantidades son pequeñas o no. Se complementa este análisis de Sprott con un criterio para la cantidad  $F_4$ , la cual permite determinar sobre la aproximación de la FV a una función normal o a una t de Student por el efecto de la curtosis.

El último paso expresa la construcción de los IC.

Si se está trabajando con valores esperados, solamente es necesario hacer algunos pequeños cambios a este algoritmo, tales cambios son:  $u_\sigma$  por  $u_{E(\sigma)} = (\hat{\omega} - \varpi) I_{E(\hat{\omega})}$ ,  $F_3(\theta)$  por  $F_{E(3)}(\varpi)$  y el criterio de  $F_4(\varpi)$  por su esperado  $F_{E(4)}(\varpi)$  para poder establecer si se aproxima mejor a una función normal o a una t de Student por el efecto de la curtosis.

En caso de que hallan parámetros de ruido en la FV, el algoritmo se puede extender simplemente utilizando la FV maximizada en los parámetros de ruido para algún valor dado del parámetro de interés (verosimilitud perfil). En esta situación es necesario mencionar que la aproximación al resultado exacto depende, entre otras cosas, del número de parámetros de ruido y las variaciones de éstas. Una alternativa

sería el uso de la ortogonalización de parámetros (según Cox y Reid, 1987), la cual podría eliminar o reducir el efecto de los parámetros de ruido en la función de verosimilitud.

## Bibliografía

- Anscombe, F. J. (1964) "Normal likelihood functions". *Ann. Inst. Statist. Math.*, 16, pp. 1-19.
- Barndorff-Nielsen, O. E. (1983) "On a formula for the distribution of a maximum likelihood estimator". *Biometrika*, 70, pp. 343-65.
- Barndorff-Nielsen, O. E. y Cox, D. R. (1979) "Edgeworth and saddle-point approximations with statistical applications". *JRSS*, B, 41, 3, pp. 279-312.
- Barndorff-Nielsen, O. E. y Cox, D. R. (1984) "Bartlett adjustments to the likelihood ratio statistics and the distribution of the maximum likelihood estimator". *JRSS*, B, 46, pp. 483-95.
- Bartlett, M. S. (1953) "Approximate confidence intervals". *Biometrika*, 40, pp. 12-19.
- Bernardo, J. M. y Smith, A. F. M. (1994) "*Bayesian theory*". Ed. J. Wiley, Chichester, 586 p.
- Box, G. E. y Cox, D. R. (1964) "An analysis of transformation". *JRSS*, B, 26, 2, pp. 211-252.
- Chapman, D. G. (1956) "Estimating the parameters of a truncated gamma distribution". *Annals of Mathematical Statistics*, 27, pp. 498-506.
- Cox, D. R. (1975) "Partial likelihood". *Biometrika*, 62, 2, pp. 269-276.
- Cox, D. R. y Reid, N. (1987) "Parameter orthogonality and approximate conditional inference". *JRSS*, B, 49, 1, pp. 1-39.
- Dawid, A. P., Stone, M. y Zideck, J. (1973) "Marginalization paradoxes in Bayesian and structural inference (with discussion)". *JRSS*, B, 35, pp. 189-233.
- Diciccio, T. y Efron, B. (1992) "More accurate confidence intervals in exponential families". *Biometrika*, 79, 2, pp. 231-245.
- Diciccio, T. y Martin, M. A. (1992) "Fast and accurate approximate double bootstrap confidence intervals". *Biometrika*, 79, 2, pp. 285-295.
- Diciccio, T. y Tibshirani, R. (1987) "Bootstrap confidence intervals and bootstrap approximations". *JASA*, 87, 397, pp. 163-170

- Efron, B. (1978) "The geometry of exponential families". *The Annals of Statistics*, 6, 2, pp. 362-376.
- Efron, B. (1981) "Nonparametric standar errors and confidence intervals". *Canadian Journal of Statistist*, 9, pp. 139-172.
- Efron, B. (1982) "Transformation theory: How normal is an one parameter familiy of distributions?". *The Annals of Statistics*, 10, pp. 323-339.
- Efron, B. (1984) "Comparing non-nested linear models". *JASA*, 79, pp. 791-803.
- Efron, B. (1985) "Bootstrap confidence intervals for a class of parametric problems". *Biometrika*, 72, 1, pp. 45-58.
- Efron, B. (1987) " Better bootstrap confidence intervals". *JASA*, 87, 397, pp. 171-185.
- Efron, B. (1993) "Bayes and likelihood calculation from confidence intervals". *Biometrika*, 80, 1, pp. 3-26.
- Efron, B. y Hinkley, D. V. (1978) "Assesing accuracy of maximum likelihood estimator: observed versus expected Fisher information". *Biometrika*, 65, pp. 457-487.
- Fisher, R. D. y Cornish, E. A. (1960) "The percentile of distributions having known cumulants". *Technometrics*, 2, 2, pp. 209-225.
- Gongora, H. (1988) "*Revisión y descripción de diversos métodos de estimación de intervalos*". Tesis de Maestría, CCH-IIMAS. UNAM.
- Gradshteyn, Y. S. y Ryzhik, Y. M. (1980) "*Table of integrals, series and products*". Academic Press, Inc., USA.
- Greenwood, J. A. y Durand, D. (1960) "Aids for fitting the gamma distribution by maximum likelihood". *Technometrics*, 2, pp. 55-65.
- Hall, P. (1988) "Theoretical comparasion of bootstrap confidence intervals". *The Annals of Statistics*, 16, 3, pp. 927-953.
- Johnson, N. L. y Kotz, S. (1970a) "*Continuous univariate distributions*". Vol. 1, New York: Wiley.
- Johnson, N. L. y Kotz, S. (1970b) "*Continuous univariate distributions*". Vol 2, New York: Wiley.
- Johnson, N. L., Kotz, S. y Balakrishnan, N. (1994) "*Continuous univariate distributions*". Second Edition, Vol 1, New York: Wiley.
- Kalbfleish, J. D. y Sprott, D. A. (1970) "Application of likelihood methods to models involving large numbers of parameters". *JRSS, B*, 32, pp. 175-208.

- Kendall, M. y Stuart, A. (1979a) "*The advanced theory of statistics*". Fourth Edition, Vol. 1, Griffin, London.
- Kendall, M. y Stuart, A. (1979b) "*The advanced theory of statistics*". Fourth Edition, Vol. 2, Griffin, London.
- Kendall, M., Stuart, A. y Ord, J. K. (1987) "*Kendall's advanced theory of statistics*". Fifth Edition, Vol. 1, Griffin, London.
- Lehmann, E. L. (1991) "*Theory of point estimation*". Wadsworth, Inc., USA.
- Liang, K-Y. y Tsou, D. (1992) "Empirical Bayes and conditional inference with many nuisance parameters". *Biometrika*, 79, 2, pp. 261-270.
- Lindsay, B. (1982) "Conditional score functions: some optimality results". *Biometrika*, 69, pp. 503-512.
- Monahan, J. F. y Boss, D. D. (1992) "Proper likelihoods for bayesian analysis". *Biometrika*, 79, 2, pp. 271-278.
- Mood, A. M., Graybill, F. A. y Boes, D. C. (1974) "*Introduction to the theory of statistics*". Third Edition, Mc Graw-Hill, Singapur.
- Natrella, M. G. (1963) "*Experimental statistics*". National Bureau of Standars, Handbook 91, United States Depatment of Commerce.
- Peers, H. W. (1965) "On confidence point and bayesian probability points in the case of several parameters". *JRSS. B*, 27, pp. 457-465.
- Peers, H. W. (1978) "Second-order sufficiency and statistical invariants". *Biometrika*, 65, 3, pp. 489-496.
- Shenton, R. L. (1959) "The distribution of moment estimators". *Biometrika*, 46, pp. 296-305.
- Shenton, R. L. y Bowman, K. (1963) "Higher moments of a maximum-likelihood estimate". *JRSS, B*, 25, pp. 317.
- Sprott, D. A. (1973). "Normal likelihoods and their relation to large sample theory of estimation". *Biometrika*, 60, 3, pp. 457-465.
- Sprott, D. A. (1980) "Maximum likelihood in small samples: Estimation in the presence of nuisance parameters". *Biometrika*, 67, 3, pp. 515-523.
- Sprott, D. A. (1982) "Robutness and likelihood estimation". *Commun. Statist.-Theor. Meth.*, 11(22), pp. 2513-2529.

- Sprott, D. A. y Kalbfleish, J. D.** (1969). "Examples of likelihoods and comparison with point estimates and large sample approximations". *JASA*, 64, pp. 468-484.
- Sprott, D. A. y Viveros, R.** (1984) "The interpretation of maximum-likelihood estimation". *The Canadian Journal of Statistics*, 12, 1, pp. 27-38.
- Tibshirani, R.** (1989) "Noninformative priors for one parameter of many". *Biometrika*, 76, pp. 604-608.
- Wang, S.** (1995) "One-step saddlepoint approximations for quantiles". *Computational Statistics and Data Analysis*, 20, pp. 65-74.
- Welch, B. L.** (1965) "On comparisons between confidence point procedures in the case of a single parameter". *JRSS, B*, 27, pp. 1-8.
- Welch, B. L. y Peers, H. W.** (1963) "On formulae for confidence points based on integrals of weighted likelihoods". *JRSS, B*, 25, pp. 318-329.
- Withers, C. S.** (1984) "Asymptotic expansions for distributions and quantiles with power series cumulants". *JRSS, B*, 46, 3, pp. 389-396.