

03061

# UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

Unidad Académica de los Ciclos Profesionales y de Posgrado del CCH  
Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas



## REVISION Y DESCRIPCION DE DIVERSOS METODOS DE ESTIMACION POR INTERVALOS

# T E S I S

Que para obtener el grado de:  
Maestra en Estadística e  
Investigación de Operaciones  
Presenta la Actuarial  
*Hortensia Góngora Treviño*

**03061**

México, D.F.

TESIS CON  
FALLA DE ORIGEN

Junio 1988



Universidad Nacional  
Autónoma de México



## **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 está 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.

## INDICE

<b>Introducción</b>		<b>1</b>
<b>Capítulo 1</b>	<b>Nociones Básicas</b>	<b>3</b>
	1.1 Importancia del Problema	3
	1.2 Regiones e Intervalos Confidenciales	10
<b>Capítulo 2</b>	<b>Pivotales en la Construcción Frecuencial de Intervalos</b>	<b>14</b>
	2.1 Cantidades Pivotales	14
	2.2 Método de Construcción de Intervalos de Confianza Basado en Cantidades Pivotales	22
<b>Capítulo 3</b>	<b>Enfoques Fiducial y Bayesiano</b>	<b>43</b>
	3.1 Introducción	43
	3.2 Método Fiducial	44
	3.3 Fiducial Implícito	56
	3.4 Método Bayesiano	60
<b>Conclusiones</b>		<b>67</b>
<b>Referencias</b>		<b>69</b>

## INTRODUCCION

Dentro de la inferencia estadística, el problema de estimación por intervalos ha sido presentado ampliamente en la literatura, tanto de estadística básica como de estadística avanzada. En general se le han dado enfoques variados, siendo el más común el basado en cantidades pivotaes. El objetivo de esta tesis, consiste en presentar un panorama de los distintos métodos que existen de estimación por intervalos, así como una descripción de la filosofía en que se sustentan, por lo que éste se puede considerar como un trabajo de investigación bibliográfica.

En cada etapa de la tesis se optó por ilustrar, a través de ejemplos sencillos, la aplicación de los conceptos discutidos, así como las limitaciones y alcances de los distintos métodos asociados. Algunos de estos ejemplos, también sirvieron para ilustrar las coincidencias y discrepancias numéricas entre los diferentes enfoques.

Se espera que este trabajo, cuya profundidad se limitó al caso de un parámetro escalar desconocido, motive el estudio de problemas más complejos en esta área, que se presentan al generalizar la teoría al caso de un vector desconocido de parámetros.

La forma en la que se presenta el material es la siguiente:

A través del Capítulo 1 se pretende subrayar la relevancia del problema de estimación por intervalos, así como proporcionar la herramienta básica para introducirse al tema. En la Sección 1 se analiza lo importante que resulta el no quedarse tan sólo con un estimador puntual del parámetro, y el reflexionar acerca de las características propias de la distribución particular con que se trabaja. En la Sección 2 se definen, de

manera general y desde el punto de vista clásico, algunos conceptos fundamentales de las regiones de confianza, mencionando de manera somera su relación con la teoría de pruebas de hipótesis.

En el Capítulo 2, se aborda el tema desde el punto de vista frecuentista de la estadística clásica, haciendo una descripción de las limitaciones y alcances que poseen los distintos métodos para la construcción, mediante la utilización de las conocidas cantidades pivotaes, de intervalos de confianza. En la primera sección, se definen e ilustran las diferentes técnicas que se conocen para encontrar cantidades pivotaes exactas y asintóticas incluyéndose, como un criterio para su elección, el Principio de Verosimilitud de Fisher. Posteriormente, en la Sección 2, se presenta una extensa descripción de la utilización, en la construcción de intervalos de confianza, de las cantidades pivotaes vistas en la Sección 1 y se menciona una teoría desarrollada por Sprott, la cual, por medio de reparametrizaciones, permite utilizar resultados asintóticos con una mejor aproximación.

Finalmente, en el Capítulo 3, se aborda el tema desde otros dos puntos de vista que conducen a los métodos Fiducial y Bayesiano. En la Sección 1 se introducen, de manera breve, las diferencias filosóficas entre estos dos enfoques. En la Sección 2 se presentan, con una notación sencilla, los conceptos esenciales del método Fiducial explícito basado en la utilización de las cantidades pivotaes vistas en la Sección 1 del Capítulo 2. En la tercera sección, se plantean los conceptos fiduciales sin hacer uso explícito de cantidades pivotaes. Este planteamiento es referido como método Fiducial implícito. Por último, en la Sección 4 se expone, de manera suscita, el conocido método Bayesiano para la construcción de intervalos de probabilidad que, en cierto modo, es paralelo al problema de construcción de intervalos de confianza. También en esta sección se comenta brevemente sobre las distribuciones *a priori* Informativas y no-informativas, dando en cada caso referencias que se espera resulten útiles.

# CAPITULO 1

## NOCIONES BASICAS

### 1.1 Importancia del Problema

Cuando se desea hacer inferencias sobre una población determinada, en general se desconocen los valores de los parámetros de interés que caracterizan a dicha población. El procedimiento que se sigue en la práctica consiste en tomar una muestra al azar de la población y, a partir de ésta se hacen inferencias sobre los valores desconocidos de los parámetros. El tipo más común de inferencia consiste en estimar los valores de dichos parámetros.

Cuando se trata con el problema de estimación puntual, se obtienen como estimadores de los parámetros, funciones de la muestra observada. Los métodos desarrollados para tal fin proporcionan, para cualquier muestra dada, estimaciones únicas de los parámetros de interés, es decir que, al substituir en los estimadores los valores observados de una muestra en particular, se obtienen estimaciones únicas de los parámetros que pueden diferir de las obtenidas por otra muestra de la misma población.

No parece ser de mucho valor el contar tan sólo con estimaciones únicas que pueden diferir de los valores verdaderos de los parámetros para cualquier muestra en particular, a menos que se cuente con alguna medida del posible error cometido en la estimación. Esta incertidumbre usualmente se expresa en los libros de texto en términos de la desviación estándar muestral de los estimadores.

Intuitivamente, se dice que es "probable" que un parámetro  $\theta$ , cuyo estimador

es  $\hat{\theta}$ , se encuentre en el rango

$$\hat{\theta} \pm \sqrt{\widehat{\text{Var}}(\hat{\theta})},$$

donde  $\widehat{\text{Var}}(\hat{\theta})$  es la varianza estimada de  $\hat{\theta}$ . O bien, se dice que es "muy probable" que

$$\theta \in \left( \hat{\theta} \pm 2\sqrt{\widehat{\text{Var}}(\hat{\theta})} \right)$$

y "casi seguramente" que

$$\theta \in \left( \hat{\theta} \pm 3\sqrt{\widehat{\text{Var}}(\hat{\theta})} \right).$$

Este método de construir intervalos aproximados para un parámetro, funciona bien en situaciones prácticas, cuando la distribución de  $\hat{\theta}$  es aproximadamente *Normal*, o bien cuando posee una forma acampanada parecida a la de la *Normal* (ver Regla Empírica en Mendenhall 1978, p.56). Sin embargo, cuando la distribución del estimador es muy asimétrica o tiene colas muy pesadas, este método puede llevar a conclusiones equivocadas sobre los posibles valores que puede tomar el parámetro.

Para ilustrar lo burdo que puede resultar, el estimar puntualmente a un parámetro por medio de una función de la muestra y una medida de incertidumbre basada en más menos una o dos desviaciones estándar, se dan los siguientes tres ejemplos. En el primero de ellos, la distribución del estimador está sumamente alejada de la *Normal* y, en los siguientes dos, la distribución del estimador es muy asimétrica.

### Ejemplo 1.1

Suponga que se desea estimar la media  $\mu$  de una distribución *Uniforme* con varianza conocida, y que el tamaño de muestra  $n$  es igual a 1.

En este caso el estimador  $\hat{\mu}$  de la media (que es la observación en sí) tiene como distribución a

$$f(\hat{\mu}) = \frac{1}{2a} 1_{[\mu-a, \mu+a]}(\hat{\mu}),$$

con  $a$  conocido,  $\mu \in \mathfrak{R}$ , y  $-\infty < \mu - a < \mu + a < \infty$ .

La esperanza y varianza para este estimador son iguales a

$$E_{\mu}(\hat{\mu}) = \frac{(\mu - a) + (\mu + a)}{2} = \mu,$$

$$\text{Var}_{\mu}(\hat{\mu}) = \frac{((\mu + a) - (\mu - a))^2}{12} = \frac{4a^2}{12} = \frac{a^2}{3}.$$

Aplicando la regla empírica sin reflexión, se llega a afirmar que

$$\mu \in \left[ \hat{\mu} - \frac{2a}{\sqrt{3}}, \hat{\mu} + \frac{2a}{\sqrt{3}} \right], \quad (1)$$

aproximadamente en un 95% de las veces. Sin embargo, evaluando la probabilidad anterior como

$$p\left(\hat{\mu} - \frac{2a}{\sqrt{3}} \leq \mu \leq \hat{\mu} + \frac{2a}{\sqrt{3}}\right) = p\left(\mu - \frac{2a}{\sqrt{3}} \leq \hat{\mu} \leq \mu + \frac{2a}{\sqrt{3}}\right)$$

$$= \int_{\mu - \frac{2a}{\sqrt{3}}}^{\mu + \frac{2a}{\sqrt{3}}} 1_{[\mu - a, \mu + a]}(\hat{\mu}) \frac{d\hat{\mu}}{2a} = 1,$$

se encuentra que es igual a 1 y no a 0.95. Este hecho implica que los extremos del intervalo aleatorio en (1), caen fuera del rango  $[\mu - a, \mu + a]$ . Si se hubiese construido el intervalo con más menos una desviación estándar que, según la regla empírica tendría una confianza aproximada del 68%, un cálculo similar al anterior da una confianza real del 57.7%. Así, en el primer caso, se dio un intervalo aleatorio que incluye valores del parámetro que no son posibles y, en el segundo, la confianza real fue menor.

El ejemplo anterior pretende ilustrar que aun en el caso de contar con una distribución simétrica para el estimador, el aplicar la regla empírica sin reflexión, puede llevar a obtener intervalos confidenciales cuyo nivel de confianza real difiere del previsto, o bien, que se incluyan a valores no posibles del parámetro. Por supuesto que, al aumentar el tamaño de muestra se mejoraría la aproximación que proporciona la regla empírica. Más adelante se abunda en ésto.

## Ejemplo 1.2

Este ejemplo se tomó de un estudio realizado para probar la bondad de ajuste de la distribución *Gaussiana Inversa* (ver O'Reilly y Rueda 1986). Aquí se encontró un parámetro desconocido  $\theta$ , cuyo recorrido natural son los reales positivos, es decir, que siempre  $\theta > 0$ . El estimador puntual para este parámetro se obtuvo como un escalamiento del de máxima verosimilitud y se denotó por  $\hat{\theta}$ .

Por consideraciones de la distribución *Gaussiana Inversa*, se encontró que la esperanza y varianza para este estimador son iguales a

$$E_{\theta}(\hat{\theta}) = \theta + \frac{1}{n},$$

$$\text{Var}_{\theta}(\hat{\theta}) = \frac{2\theta^2}{n-5} + \frac{(n+1)\theta}{n(n-5)} + \frac{2n-4}{n^2(n-5)}; \quad n > 5,$$

dando como resultado un estimador consistente.

Aplicando la regla empírica a este problema, se obtendría que

$$\theta \in \left( \hat{\theta} \pm \sqrt{\widehat{\text{Var}}(\hat{\theta})} \right), \quad (2)$$

aproximadamente en un 68% de las veces, con

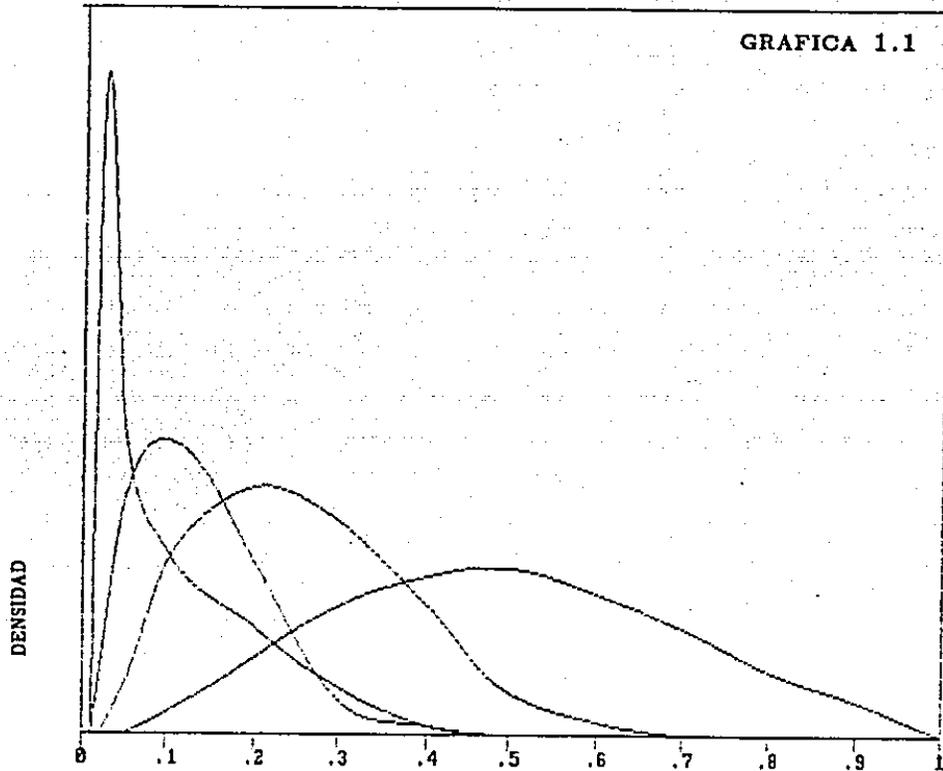
$$\widehat{\text{Var}}_{\theta}(\hat{\theta}) = \frac{2\hat{\theta}^2}{n-5} + \frac{(n+1)\hat{\theta}}{n(n-5)} + \frac{2n-4}{n^2(n-5)}; \quad n > 5. \quad (3)$$

Aquí es importante observar que para valores pequeños de  $\hat{\theta}$  (cercaos al cero), el límite inferior en (2) puede ser negativo. Por ejemplo, si  $\hat{\theta} = 0.01$ , aun con  $n = 50$ , la varianza estimada en (3) para este estimador es igual a 0.001084, y la cota inferior del supuesto intervalo para  $\theta$  en (2) resulta ser

$$\hat{\theta} - \sqrt{\widehat{\text{Var}}(\hat{\theta})} = 0.01 - 0.032931 = -0.022931.$$

La razón de este resultado equívoco, es que la distribución de  $\hat{\theta}$  es altamente asimétrica,

como se puede apreciar en la gráfica 1.1.



DISTRIBUCIONES DE  $\hat{\theta}$  PARA VALORES DE  $\theta$ : 0.062, 0.125, 0.25, 0.50.  
(Entre más pequeño es  $\theta$  más asimétrica es la distribución.)

### Ejemplo 1.3

Este ejemplo se tomó de Sprott y Viveros (1984), p 29. Sea  $\underline{X} = (X_1, X_2)$  una muestra de tamaño 2 de una distribución *Exponencial* con media  $\theta > 0$ . El estimador puntual para este parámetro desconocido  $\theta$ , obtenido por el método de máxima verosimilitud, es igual a

$$\hat{\theta} = \frac{X_1 + X_2}{2} = \frac{T}{n}. \quad (4)$$

La esperanza y varianza para este estimador son

$$E_{\theta}(\hat{\theta}) = \theta \quad ; \quad \text{Var}_{\theta}(\hat{\theta}) = \frac{\theta^2}{2}. \quad (5)$$

Aplicando la regla empírica en este problema se obtendría que

$$\theta \in \left( \hat{\theta} \pm \sqrt{\text{Var}(\hat{\theta})} \right), \quad (6)$$

aproximadamente en un 68% de las veces. Evaluando la probabilidad anterior en términos de la variable aleatoria  $T/\theta$ , se llega a que

$$\begin{aligned} p\left(\hat{\theta} - \frac{\hat{\theta}}{\sqrt{2}} \leq \theta \leq \hat{\theta} + \frac{\hat{\theta}}{\sqrt{2}}\right) &= p\left(\hat{\theta}\left(1 - \frac{1}{\sqrt{2}}\right) \leq \theta \leq \hat{\theta}\left(1 + \frac{1}{\sqrt{2}}\right)\right) \\ &= p\left(0.2929 \leq \frac{\theta}{\hat{\theta}} \leq 1.7071\right) \\ &= p\left(1.1716 \leq \frac{T}{\theta} \leq 6.8284\right), \end{aligned} \quad (7)$$

dado que  $T/\theta$  se distribuye como una *Gamma* con parámetros  $\alpha = 2$  y  $\beta = 1$ , esta probabilidad se puede evaluar exactamente como

$$\int_{1.1716}^{6.8284} \frac{1}{\Gamma(2)} \omega \exp\{-\omega\} d\omega = 0.66; \quad \text{con } \omega = \frac{T}{\theta}.$$

En este caso los niveles exacto y aproximado son muy cercanos, por lo que no parece haber problemas, ni en cuanto al nivel de confianza del intervalo confidencial obtenido, ni tampoco en cuanto a los posibles valores que puede tomar  $\theta$ . Sin embargo, al tomar una confianza mayor, digamos aproximadamente del 95%, se tendría según la regla empírica que

$$\theta \in \left( \hat{\theta} \pm 2\sqrt{\text{Var}(\hat{\theta})} \right) = \left( \hat{\theta} \left(1 \pm \frac{2}{\sqrt{2}}\right) \right),$$

es decir que el supuesto intervalo cubre valores del parámetro que no son posibles, ya que la cota inferior en (6) siempre será negativa e igual a  $-0.4142\hat{\theta}$ .

La regla empírica aplicada a este ejemplo en que el tamaño de muestra  $n$  es muy pequeño, produjo resultados aceptables para un nivel de confianza del 68%. Sin embargo, al ponerse más estrictos con un nivel de confianza del 95%, los resultados ya

fueron equívocos. Aquí al igual que en el ejemplo anterior, esto se debe a la asimetría de la distribución del estimador utilizado. Más adelante, en el capítulo 2, se retomará este mismo ejemplo con  $n = 2$ , para ilustrar cómo, a pesar de tener un tamaño de muestra tan pequeño, Sprott y Viveros (1984) obtienen intervalos confidenciales aceptables aplicando teoría asintótica al reparametrizar.

Los resultados encontrados en los tres ejemplos anteriores en donde, no se tienen distribuciones *Normales*, crean la necesidad de buscar otros métodos más confiables para inferir sobre los valores desconocidos de los parámetros de interés.

Otro hecho importante por mencionar, es el gran énfasis que se da en estimación puntual a los estimadores máximo verosímiles. Las propiedades óptimas de éstos, para muestras finitas, son el insesgamiento y varianza mínima, en la mayoría de los casos y, para muestras grandes la consistencia y la normalidad asintótica (bajo condiciones de regularidad). Debido a esto se dice que si el tamaño de muestra  $n$  crece, entonces el estimador máximo verosímil  $\hat{\theta}$  del parámetro  $\theta$  tendrá como distribución aproximada a la *Normal*, es decir que

$$\hat{\theta} \sim N\left(\theta, \frac{1}{I_E(\theta)}\right), \quad (8)$$

donde  $I_E(\theta)$  es la cantidad de información de Fisher para  $X$ , definida como

$$I_E(\theta) = -E_{\theta}\left[\frac{\partial^2 \log f(X; \theta)}{\partial \theta^2}\right]. \quad (9)$$

En la práctica, para el caso de muestras finitas, usualmente la cantidad en (9) se estima por medio de la información observada de Fisher  $I_{\theta}$ , definida como

$$I_{\theta} = -\frac{\partial^2 \log f(\underline{x}; \theta)}{\partial \hat{\theta}^2}, \quad (10)$$

y, a partir de aquí, se dice que

$$\theta \in \left(\hat{\theta} \pm z_{\alpha/2}(I_{\theta})^{-\frac{1}{2}}\right), \quad (11)$$

en un  $(1 - \alpha) \times 100\%$  de las veces, donde  $z_{\alpha/2}$  es el percentil  $\alpha/2 \times 100\%$  de la distribución *Normal*(0,1). Sin embargo, como menciona Sprott (1975), el criterio correcto para exhibir un estimador puntual del parámetro y su varianza, como un resumen de los

datos, no es el tamaño de muestra sino la normalidad de las funciones de verosimilitud que resulten. Para ilustrar este hecho, se retoma el ejemplo 1.2 de la *Gaussiana Inversa*, en el que se estimó al parámetro  $\theta$  por medio de un escalamiento del estimador máximo verosímil  $\hat{\theta}$ .

#### Ejemplo 1.4

Suponga que el estimador del parámetro  $\theta$  en el ejemplo 1.2 toma el valor de 0.001 y que el tamaño de muestra  $n$  se hace crecer al valor de 1000. Entonces, el estimador de la varianza de  $\hat{\theta}$  en (3) es igual a  $3.010 \times 10^{-6}$  y, haciendo uso de la normalidad asintótica de  $\hat{\theta}$ , se tendría por (11) que

$$\theta \in \left( \hat{\theta} \pm .995 \sqrt{\widehat{\text{Var}}(\hat{\theta})} \right),$$

en un 68% de las veces. Sin embargo, aquí al igual que en el ejemplo 1.2, se obtiene un límite inferior que cae fuera del rango de valores permitidos para el parámetro  $\theta$  ( $> 0$ ), siendo igual a  $-7.263 \times 10^{-4}$ . Este resultado erróneo se debe a la gran asimetría que presenta la densidad de  $\hat{\theta}$  para valores pequeños de  $\theta$ , como se pudo apreciar en la gráfica 1.1.

El resultado obtenido en este ejemplo muestra que el criterio de "n grande" llega a ser insuficiente en una situación práctica.

### 1.2 Regiones e Intervalos Confidenciales

En esta sección se pretenden definir los conceptos fundamentales asociados a las regiones e intervalos confidenciales basados en la teoría frecuentista de probabilidad, así como, introducir parte de la notación que se utilizará en el resto de este trabajo.

Sean  $X_1, \dots, X_n$  variables aleatorias, cuya función de distribución conjunta, dependiente del parámetro escalar  $\theta$ , es  $f(\underline{x}; \theta)$ , en donde  $\underline{X}$  es el vector aleatorio  $(X_1, \dots, X_n)$  y  $\underline{x}$  es el vector real  $(x_1, \dots, x_n)$ . Sea  $\Theta$  el espacio paramétrico.

En la práctica, generalmente el interés consiste en resumir toda la información que proporcionan los datos  $\underline{x}$  sobre el parámetro escalar  $\theta$ , a través de regiones que se espera resulten en la forma de un solo intervalo. En el caso en que los datos  $\underline{x}$  sean consistentes con valores parametrales contiguos, se puede resumir dicha información en un solo intervalo confidencial; en el caso contrario en que los datos  $\underline{x}$  sean consistentes con valores parametrales no contiguos, el resumen se hace a través de una región confidencial consistente en la unión de un cierto número de intervalos disjuntos.

En términos muy generales, se pueden definir las regiones e intervalos confidentiales como subconjuntos aleatorios del espacio paramétrico  $\Theta$ , en donde la probabilidad de que contengan al verdadero valor del parámetro  $\theta$  es  $1 - \alpha$ .

De manera más formal, se define una región confidencial para el parámetro  $\theta$  de nivel  $(1 - \alpha)$ , la cual se denotará como  $R_\alpha(X_1, \dots, X_n)$ , como la región aleatoria de  $\Theta$ , tal que

$$p(\theta \in R_\alpha(X_1, \dots, X_n); \theta) = 1 - \alpha \quad \text{para cada } \theta \in \Theta, \quad (12)$$

es decir, que desde el punto de vista frecuentista,  $R_\alpha(X_1, \dots, X_n)$  contiene al verdadero valor de  $\theta$  en un  $100 \times (1 - \alpha)$  % de las veces en muestras repetidas de tamaño  $n$ ; aun cuando en la práctica la interpretación de muestreo repetido pueda ser hipotética. La interpretación anterior que se da a la región confidencial en (12) es la misma para cada valor de  $\alpha$ . Así para exhibir la mayor información posible sobre el parámetro  $\theta$ , conviene considerar, al menos de manera implícita, a la familia de regiones confidentiales

$$\{R_\alpha(X_1, \dots, X_n)\}_{\alpha > 0} \quad (13)$$

que se obtiene al variar  $\alpha$  en  $(0,1)$ . Esta familia de regiones confidentiales debe poseer la propiedad básica conocida como **anidamiento**, la cual consiste en que, si  $\alpha_1$  y  $\alpha_2$  son dos valores en  $(0,1)$  tales que  $\alpha_1 > \alpha_2$ , entonces las correspondientes regiones deben satisfacer que

$$R_{\alpha_1}(X_1, \dots, X_n) \subset R_{\alpha_2}(X_1, \dots, X_n), \quad (14)$$

es decir que, entre más confianza se tenga, mayor resultará la región correspondiente y deberá contener a la construida con los mismos datos pero de menor confianza. En el caso extremo en que se consideran a todos los valores de  $\alpha$  en el intervalo  $(0,1)$  de

manera simultanea, se genera lo que se conoce como la **distribución confidencial** del parámetro  $\theta$ , la cual consiste en una sucesión infinita de regiones confidenciales cada una de las cuales está contenida en la anterior conforme el valor de  $\alpha$  decrece del 1 al 0, debido a la propiedad de anidamiento en (14).

El criterio más común para formar las regiones confidenciales, consiste en que todos los valores parametrales en  $R_\alpha(X_1, \dots, X_n)$  tengan verosimilitud más alta que los valores parametrales fuera de  $R_\alpha(X_1, \dots, X_n)$  y, en este caso, se dice que  $R_\alpha(X_1, \dots, X_n)$  es una región confidencial basada en la verosimilitud; sin embargo, pueden existir otros criterios diferentes de éste, en cuyo caso se hace necesario especificarlos claramente en cada aplicación.

Como se verá en capítulos posteriores, existen diferentes métodos para construir regiones confidenciales, por lo que se hace necesario contar con ciertos criterios que permitan comparar entre ellos y elegir alguno como óptimo para el problema en particular que nos ocupa.

Un criterio adicional comúnmente utilizado es el de **insesgamiento**. En este caso, se dice que una región confidencial de nivel  $(1 - \alpha)$  para  $\theta$  es insesgada, si se satisface que

$$p(\theta' \in R_\alpha(X_1, \dots, X_n); \theta) \leq 1 - \alpha \quad \text{para todo } \theta' \neq \theta \in \Theta, \quad (15)$$

es decir, que la confianza de que la región confidencial contenga a un valor parametral equivocado  $\theta'$ , es menor o igual que la confianza de contener al verdadero valor de  $\theta$ .

Otro criterio de optimalidad para comparar regiones confidenciales es el de escoger aquella que sea estocásticamente más pequeña. Esto es, dado algún valor pequeño de  $\alpha$  y dos regiones confidenciales distintas  $R_\alpha^{(1)}(\underline{X})$  y  $R_\alpha^{(2)}(\underline{X})$  del parámetro  $\theta$ , se dice que  $R_\alpha^{(1)}(\underline{X})$  es **preferible** a  $R_\alpha^{(2)}(\underline{X})$  si se satisface que

$$p(\theta' \in R_\alpha^{(1)}(\underline{X}); \theta) \leq p(\theta' \in R_\alpha^{(2)}(\underline{X}); \theta) \quad \text{para todo } \theta' \neq \theta \in \Theta, \quad (16)$$

es decir, que  $R_\alpha^{(1)}(\underline{X})$  es estocásticamente más pequeña que  $R_\alpha^{(2)}(\underline{X})$ , si la confianza de que  $R_\alpha^{(1)}(\underline{X})$  contenga un valor parametral equivocado  $\theta'$ , es menor o igual que la confianza de que  $R_\alpha^{(2)}(\underline{X})$  lo contenga.

Los criterios de insesgamiento y región confidencial estocásticamente más pequeña ya señalados corresponden, en teoría de pruebas de hipótesis, al insesgamiento y búsqueda de mayor potencia de pruebas, respectivamente (ver Rao 1973, pp. 449 y 454).

Como se mencionó al principio de esta sección, en ocasiones la información proporcionada por los datos  $\underline{x}$  sobre el parámetro escalar  $\theta$ , se puede resumir en un intervalo confidencial más que en una región. De manera formal se define un intervalo confidencial para el parámetro  $\theta$  de nivel  $1 - \alpha$ , el cual se denotará por  $[h_1(\underline{X}), h_2(\underline{X})]$ , como aquel intervalo aleatorio de  $\Theta$ , tal que

$$p(h_1(\underline{X}) \leq \theta \leq h_2(\underline{X}); \theta) = 1 - \alpha, \quad \text{para cada } \theta \in \Theta, \quad (17)$$

es decir que, desde el punto de vista frecuentista,  $[h_1(\underline{X}), h_2(\underline{X})]$  contiene al verdadero valor de  $\theta$  en un  $100 \times (1 - \alpha) \%$  de las veces en muestreo repetido. A  $h_1$  y  $h_2$  en (17) se les conoce, respectivamente, como los límites de confianza inferior y superior.

Dado que los intervalos confidenciales son un caso particular de las regiones confidenciales, todo lo antes mencionado para los últimos se aplica a los primeros.

Habiendo descrito brevemente los criterios generales para la conceptualización de regiones e intervalos confidenciales, se verán en los capítulos siguientes varios de los métodos desarrollados para la construcción de éstos, restringiéndonos al caso particular de intervalos confidenciales para un parámetro escalar  $\theta$ . Estos métodos en general llevan a resultados diferentes aunque, como se podrá apreciar posteriormente, en ciertas ocasiones llegan a coincidir.

## CAPITULO 2

### PIVOTALES EN LA CONSTRUCCION FRECUENCIAL DE INTERVALOS

#### 1.1 Cantidades Pivotales

Un tipo importante de métodos para construir regiones de confianza cuando el espacio paramétrico es un intervalo o todo el eje real, se basa en las llamadas **Cantidades Pivotales**. Estos métodos pretenden encontrar cantidades aleatorias que dependan de la muestra  $\underline{X}$  y del parámetro  $\theta$  y cuya distribución sea única; en este caso se tienen **pivotales exactas**. Si la distribución de dichas cantidades no es única pero al aumentar el tamaño de muestra  $n$ , tiende en el límite a una única, entonces éstas se conocen como **Cantidades Pivotales Asintóticas**. A continuación se definen estas cantidades de manera formal.

Sea  $\underline{X} = (X_1, \dots, X_n)$  una muestra aleatoria de la función de densidad  $f(x; \theta)$  dependiente de un parámetro escalar  $\theta$ , esto es, la densidad de  $\underline{x} = (x_1, \dots, x_n)$  es

$$f(\underline{x}; \theta) = \prod_{i=1}^n f(x_i; \theta). \quad (1)$$

Una **Cantidad Pivotal Exacta**  $U = U(\underline{X}; \theta)$ , se define como una función de la muestra  $\underline{X}$  y del parámetro  $\theta$  con distribución  $F_0(u)$  independiente del parámetro  $\theta$

$$p(U(\underline{X}; \theta) \leq u) = F_0(u) \quad \text{para todo } \theta \in \Theta, \quad (2)$$

es decir, que es fija para todo  $\theta \in \Theta$ .

Al ser  $F_0(u)$  independiente de  $\theta$ , ninguno de sus momentos dependen de ella, en particular los dos primeros momentos de  $F_0(u)$  siempre son cantidades fijas, es decir, que la media y varianza de  $U$  no dependen de  $\theta$ , como se puede apreciar en el siguiente ejemplo.

### Ejemplo 2.1

Sean  $X_1, \dots, X_n$  variables aleatorias i.i.d.  $N(\mu, 1)$ . Se sabe que  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ , el estimador de la media  $\mu$ , se distribuye como  $N(\mu, \frac{1}{n})$  y, por lo tanto, una cantidad pivotal exacta es

$$U = \sqrt{n}(\bar{X} - \mu), \quad (3)$$

que depende únicamente de la muestra  $\underline{X}$  y del parámetro  $\mu$ . Su correspondiente función de distribución es

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

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

Podrían haberse construido otras cantidades pivotaes basadas por ejemplo, en la mediana muestral  $X_{((n+1)/2)}$  (si  $n$  es impar), o bien, en el rango medio  $(X_{(1)} + X_{(n)})/2$ .

En los casos en que se utiliza la teoría de muestras grandes, es decir, al dejar tender  $n$  a infinito, se define una **Cantidad Pivotal Asintótica** como la función  $U_n = U_n(\underline{X}; \theta)$  de la muestra y del parámetro, cuya distribución  $F_0(u)$ , en el límite, es independiente del parámetro  $\theta$ , es decir

$$p(U_n(\underline{X}; \theta) \leq u) \rightarrow F_0(u), \quad n \rightarrow \infty. \quad (4)$$

Barnard (1973) define las cantidades pivotaes asintóticas de una manera un poco menos general, ya que pide además que sean estables en la media, es decir que

$$E_{\theta}(U_n(\underline{X}; \theta)) = 0 \quad \text{para cada } n, \quad (5)$$

y estables en media cuadrática, es decir que

$$\text{Var}_\theta(U_n(\underline{X}; \theta)) = 1 \quad \text{para cada } n. \quad (6)$$

Ver Sprott (1975).

En algunas ocasiones, se encuentran pivotaes asintóticas con sus primeros y segundos momentos exactos, es decir, independientes de  $n$  y  $\theta$  y, en otras, éstos solamente son exactos asintóticamente. Pareciera que lo ideal sería encontrar una pivotal asintótica cuya media y varianza sean cantidades fijas para  $n$  finito, pero como se verá más adelante, no siempre resultará esto en un beneficio.

El siguiente ejemplo ilustra la construcción de una cantidad pivotal asintótica.

### Ejemplo 2.2

Sean  $X_1, \dots, X_n$  variables aleatorias i.i.d. con media común  $\mu$  y varianza común igual a 1. Se sabe por el Teorema Central del Límite que la cantidad aleatoria

$$U_n = \frac{\bar{X}_n - \mu}{1/\sqrt{n}} = \sqrt{n}(\bar{X}_n - \mu), \quad (7)$$

tiene como distribución límite a la  $N(0, 1)$ .

Aquí  $U_n$  es la cantidad pivotal asintótica y  $F_0(u)$  la distribución *Normal estándar*. Esta pivotal en particular tiene sus dos primeros momentos exactos, es decir, que son cantidades fijas independientemente de los valores que tomen  $\mu$  y  $n$ .

En los dos ejemplos anteriores, fue posible encontrar cantidades pivotaes de manera bastante sencilla; sin embargo, en la práctica puede suceder que no se tenga idea de qué función utilizar para un problema determinado, es más, podría suceder que ni siquiera se sepa si existe una pivotal o no. Mood, Graybill y Boes (1974) aseguran que, siempre que la muestra provenga de una población con función de distribución continua, existirá una cantidad pivotal y proponen un método para construir dichas cantidades. Este método consiste en lo siguiente: si  $X_1, \dots, X_n$  es una muestra aleatoria de  $f(x; \theta)$  cuya función de distribución  $F(x; \theta)$  es continua en  $x$ , se sabe por la transformación

con la integral de probabilidad que  $Z_i = F(X_i; \theta)$  es una variable aleatoria con una distribución *Uniforme* sobre el intervalo  $(0,1)$ , es decir que

$$p(Z_i \leq z) = p(F(X_i; \theta) \leq z) = p(X_i \leq F^{-1}(z)) = F(F^{-1}(z)) = z, \quad (8)$$

donde  $F^{-1}(z)$  se puede definir como

$$F^{-1}(z) = \inf\{x; F(x) \geq z\},$$

o como

$$F^{-1}(z) = \sup\{x; F(x) \leq z\}. \quad (9)$$

Si ahora se toma la variable  $Y_i = -\log Z_i = -\log F(X_i; \theta)$ , esta nueva variable tiene distribución *Exponencial* con parámetro 1, es decir que

$$p(Y_i \leq y) = p(-\log Z_i \leq y) = p(\log Z_i \geq -y) = p(Z_i \geq \exp\{-y\}) = 1 - \exp\{-y\}. \quad (10)$$

También, se sabe que la suma de variables aleatorias i.i.d. *Exponenciales*, se distribuye como *Gamma*; por lo que al tomar la suma de  $Y_1, \dots, Y_n$ , la cantidad pivotal

$$U = \sum_{i=1}^n Y_i = -\sum_{i=1}^n \log Z_i = -\sum_{i=1}^n \log F(X_i; \theta) = -\log \prod_{i=1}^n F(X_i; \theta), \quad (11)$$

tiene una distribución *Gamma* con parámetros  $\alpha = n$  y  $\beta = 1$ , la cual es independiente del parámetro  $\theta$ .

También se podrían tomar como pivotaes a

$$-2 \sum_{i=1}^n \log F(X_i; \theta) \sim \text{Gamma}(\alpha = n, \beta = 1/2) = \chi_{(2n)}^2, \quad (12)$$

o bien a

$$-2 \sum_{i=1}^n \log(1 - F(X_i; \theta)) \sim \chi_{(2n)}^2,$$

ya que si,  $Z_i = F(X_i; \theta)$  se distribuye como una *Uniforme*(0,1), también  $1 - Z_i$  se distribuye *Uniforme* (0,1).

El siguiente ejemplo, presentado por Mood, Graybill y Boes (1974) p.388, ilustra la aplicación de este método de construcción de cantidades pivotaes.

### Ejemplo 2.3

Sea  $X_1, \dots, X_n$  una muestra aleatoria de la densidad

$$f(x; \theta) = \theta x^{\theta-1} 1_{(0,1)}(x), \quad \theta > 0$$

y con función de distribución

$$F(x; \theta) = x^\theta 1_{(0,1)}(x) + 1_{[1,\infty)}(x).$$

Utilizando (12), una de las cantidades pivotaes que se puede construir aquí, es

$$U = -2 \sum_{i=1}^n \log F(X_i; \theta) = -2\theta \sum_{i=1}^n \log X_i,$$

la cual se distribuye como una  $Ji - Cuadrada$  con  $2n$  grados de libertad.

En ocasiones, este método para construir cantidades pivotaes puede llevar a expresiones de  $U$  muy complicadas que, como se verá más adelante, impedirá despejar al parámetro de interés. Un caso en donde se presenta este problema se da al aplicar este método al ejemplo 2.1, como se muestra a continuación.

### Ejemplo 2.4

Retomando el ejemplo 2.1, en donde se tenía una muestra aleatoria  $X_1, \dots, X_n$  de la densidad *Normal*

$$f(x; \mu) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2}\right\},$$

con función de distribución  $N(\mu; 1)$

$$F(x; \mu) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(t-\mu)^2}{2}\right\} dt.$$

Aplicando (12), la cantidad pivotal resultante es

$$U = -2 \sum_{i=1}^n \log F(X_i; \theta) = -2 \sum_{i=1}^n \log \int_{-\infty}^{X_i} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(t-\mu)^2}{2}\right\} dt, \quad (13)$$

la cual se distribuye como una  $Ji - Cuadrada$  con  $2n$  grados de libertad. Observe sin embargo lo difícil que resultaría utilizar el conocimiento de la distribución de  $U$  para hacer una aseveración probabilística que pueda traducirse en un intervalo de confianza para  $\mu$ .

Otro camino para encontrar cantidades pivotaes en las aplicaciones, es el de buscar aquellas que sean funciones de la estadística suficiente minimal  $T(\underline{X})$ . Estos métodos, a diferencia de los dos anteriores, siempre serán consistentes con el Principio de Verosimilitud. Este principio, en esencia dice que toda la información de la muestra  $\underline{x}$  acerca del parámetro  $\theta$ , está contenida en la función de verosimilitud  $L(\theta; \underline{x})$ , por lo que toda inferencia respecto a  $\theta$  debe basarse solamente en ésta (Birnbaum 1962, Sprott 1984).

La heurística de estos métodos se basa en que la verosimilitud, que es vista siempre como función de  $\theta$  para  $\underline{x}$  fija  $L(\theta; \underline{x})$  (ver Kendall y Stuart 1979), se puede factorizar para  $T$  suficiente como el producto de dos funciones (Teorema de Factorización de Neyman)

$$L(\theta; \underline{x}) = \prod_{i=1}^n f(x_i; \theta) = g(T(\underline{x}); \theta) \cdot h(\underline{x}), \quad \forall \theta \in \Theta, \quad (14)$$

donde la función  $h$  es fija ya que únicamente depende de la muestra observada  $\underline{x}$  y no del parámetro  $\theta$ , y la función  $g$  depende de la estadística suficiente  $T(\underline{x})$  y de  $\theta$ . Así, se puede decir que, la verosimilitud es proporcional a una función de la estadística suficiente y del parámetro, esto se, puede expresar como

$$L(\theta; \underline{x}) \propto g(T(\underline{x}); \theta). \quad (15)$$

En particular, si la estadística  $T(\underline{X})$  es la suficiente minimal, entonces al basar en ella toda inferencia sobre  $\theta$ , se esta cumpliendo con el principio de verosimilitud.

Aplicando la definición general de cantidad pivotal exacta dada al principio de este capítulo, lo que se busca es, una cantidad aleatoria que sea función de la muestra, a través de la estadística suficiente minimal  $T(\underline{X})$ , y del parámetro  $\theta$ , cuya distribución sea única para cualquier valor de  $\theta \in \Theta$ .

En el ejemplo 2.1, se obtuvo la cantidad pivotal exacta en (3), función del parámetro desconocido  $\mu$  y de la estadística suficiente minimal para  $\mu$

$$T(\underline{X}) = \sum_{i=1}^n X_i. \quad (16)$$

Observe que para dos muestras  $\underline{x}$  y  $\underline{x}'$ , tales que  $T(\underline{x}) = T(\underline{x}')$ , se obtendría el mismo valor de  $U$  y, por ende, la misma inferencia para  $\mu$ . En el ejemplo 2.4, sin embargo, la pivotal que se obtuvo en (13) no es función de la estadística suficiente minimal en (16). Es obvio que si se tuviesen dos muestras  $\underline{x}$  y  $\underline{x}'$  tales que  $T(\underline{x}) = T(\underline{x}')$ , éstas no necesariamente tendrían el mismo valor de  $U$ , violandose el principio de verosimilitud, que afirma que, las inferencias obtenidas a partir de dos muestras equivalentes en verosimilitud deben ser idénticas.

Se dice que dos muestras  $\underline{x}$  y  $\underline{x}'$  son equivalentes en verosimilitud si

$$L(\theta; \underline{x}) \propto L(\theta; \underline{x}'), \quad \forall \theta \in \Theta$$

y que, la estadística suficiente  $T(\underline{X})$  es minimal si, para toda  $\underline{x}$  y  $\underline{x}'$  equivalentes en verosimilitud se tiene que

$$T(\underline{x}) = T(\underline{x}').$$

Ver Sampson y Spencer (1976).

Un método aparentemente universal que se utiliza también para construir cantidades pivotaes exactas, se basa en la transformación con la integral de probabilidad utilizando la función de distribución de la estadística suficiente minimal, si dicha función es continua. Sea  $G^*(t; \theta)$  la función de distribución continua de  $T(\underline{X})$ , a la que se aplica la transformación con la integral de probabilidad, definiendo así la cantidad pivotal exacta  $U(T(\underline{X}); \theta)$  como

$$U(T(\underline{X}); \theta) = G^*(T(\underline{X}); \theta), \quad (17)$$

que tiene una distribución *Uniforme*(0,1) independiente del parámetro  $\theta$ . Para poder apreciar mejor la aplicación de este método, se da el siguiente ejemplo.

### Ejemplo 2.5

Sean  $X_1, \dots, X_n$  variables aleatorias i.i.d. *Uniformes* sobre el intervalo  $(0, \theta)$ , con función de densidad

$$f(x; \theta) = \begin{cases} \frac{1}{\theta} & \text{si } x \in (0, \theta) \\ 0 & \text{en otro caso.} \end{cases}$$

La estadística suficiente minimal en este caso es igual a  $T(\underline{X}) = \max X_i = X_{(n)}$ , con función de distribución continua

$$\begin{aligned} G^*(t; \theta) &= p(X_{(n)} \leq t) = p(X_1 \leq t, \dots, X_n \leq t) \\ &= \prod_{i=1}^n p(X_i \leq t) = \prod_{i=1}^n F_X(t; \theta), \end{aligned}$$

donde  $F_X(t; \theta)$  es la función de distribución común de las variables  $X_1, \dots, X_n$ , definida de la siguiente manera

$$F_X(x; \theta) = \begin{cases} 0 & \text{si } x \leq 0 \\ \frac{x}{\theta} & \text{si } x \in (0, \theta) \\ 1 & \text{si } x \geq \theta. \end{cases}$$

Por lo tanto, la función de distribución continua de  $T(\underline{X})$  es igual a

$$G^*(t; \theta) = \begin{cases} 0 & \text{si } t \leq 0 \\ \left(\frac{t}{\theta}\right)^n & \text{si } t \in (0, \theta) \\ 1 & \text{si } t \geq \theta. \end{cases}$$

Aplicando (17), la cantidad pivotal exacta que resulta es

$$U(T; \theta) = G^*(T; \theta) = \left(\frac{T}{\theta}\right)^n, \quad (18)$$

con distribución *Uniforme*(0,1).

En el ejemplo anterior fue posible encontrar una expresión sencilla para la pivotal exacta  $U$ ; sin embargo, al aplicar este método a otros casos, se pueden obtener expresiones muy complicadas para  $U$ , a partir de las cuales resulte difícil despejar al parámetro de interés, para encontrar el correspondiente intervalo confidencial.

## Ejemplo 2.6

Del ejemplo 2.1 en el que se tenía una muestra aleatoria de la densidad *Normal* con media común  $\mu$  y varianza común igual a 1. La estadística suficiente minimal  $T(\underline{X}) = \sum_{i=1}^n X_i$ ; para  $\mu$ , tiene como funciones de densidad y distribución, respectivamente a

$$g^*(t; \mu) = \left(\frac{n}{2\pi}\right)^{1/2} \exp\left\{-\frac{n(t-\mu)^2}{2}\right\},$$

$$G^*(t; \mu) = \int_{-\infty}^t g^*(z; \mu) dz.$$

Dado que  $G^*(t; \mu)$  es continua, la cantidad pivotal exacta  $U(T(\underline{X}); \mu)$  con distribución *Uniforme*(0,1) que resulta de aplicar (17) es

$$U(T(\underline{X}); \mu) = \int_{-\infty}^T g^*(z; \mu) dz. \quad (19)$$

Más adelante se verá que, gracias a que se cuenta con tablas de la distribución *Normal*, resulta más o menos sencillo encontrar el intervalo confidencial para  $\mu$  utilizando esta pivotal.

## 1.2 Método de Construcción de Intervalos de Confianza basado en Cantidades Pivotalas

El método basado en cantidades pivotaes para construir intervalos de confianza para el parámetro  $\theta$ , consta de dos etapas; la primera consiste en encontrar una pivotal por alguno de los métodos ya mencionados en la sección anterior y, la segunda, no siempre posible, es la de poder "pivotear", es decir, despejar el valor del parámetro desconocido  $\theta$  de la desigualdad probabilística hecha sobre la cantidad pivotal con su propia distribución, como ya se ha venido mencionando (ej. 2.4).

Para construir un intervalo confidencial para  $\theta$ , dada una cantidad pivotal  $U = U(\underline{X}; \theta)$  con función de densidad  $f_0(u)$  y, dado cualquier nivel de confianza  $1 - \alpha$ , lo

primero es construir un intervalo de probabilidad  $1 - \alpha$  para  $U$ . Para esto, se encuentran cantidades  $a$  y  $b$  dependientes del nivel  $1 - \alpha$ , tales que

$$p(a \leq U(\underline{X}; \theta) \leq b) = F_0(b) - F_0(a) = 1 - \alpha \quad \text{para } \theta \in \Theta. \quad (20)$$

Si para cada posible valor que tome  $\underline{X}$ , la desigualdad en  $U$

$$a \leq U(\underline{X}; \theta) \leq b, \quad (21)$$

se satisface si, y sólo si,\*\*

$$h_1(\underline{X}) = U^{-1}(\underline{X}; a) \leq \theta \leq U^{-1}(\underline{X}; b) = h_2(\underline{X}), \quad (22)$$

en donde  $h_1$  y  $h_2$  son funciones que dependen de la muestra  $\underline{X}$ , pero no del parámetro  $\theta$ . Entonces, por (20), se deben satisfacer las siguientes igualdades

$$p(a \leq U(\underline{X}; \theta) \leq b) = p(h_1(\underline{X}) \leq \theta \leq h_2(\underline{X})) = 1 - \alpha, \quad (23)$$

las cuales llevan a encontrar el intervalo aleatorio  $[h_1(\underline{X}), h_2(\underline{X})]$ , conocido como el intervalo confidencial de nivel  $1 - \alpha$  para  $\theta$ , como se vio en el Capítulo 1. Al intervalo para  $\theta$  en (23) se le da la interpretación, desde el punto de vista frecuentista, de que en muestreo repetido, el valor del parámetro  $\theta$  se encuentra entre las estadísticas  $h_1$  y  $h_2$  en un  $100 \times (1 - \alpha)\%$  de las veces. A las estadísticas  $h_1(\underline{X})$  y  $h_2(\underline{X})$ , como se mencionó en el Capítulo 1, se les conoce como los límites de confianza inferior y superior, respectivamente.

Veamos cómo se puede encontrar un intervalo confidencial de nivel  $1 - \alpha$  para el parámetro  $\mu$  del ejemplo 2.1, utilizando la cantidad pivotal exacta (3).

### Ejemplo 2.7

En el ejemplo 2.1 para una muestra aleatoria  $X_1, \dots, X_n$  de una  $N(\mu, 1)$ , se encontró la cantidad pivotal exacta, función de la estadística suficiente minimal

\*\* Estas desigualdades en  $\theta$  resultan si  $U(\underline{X}; \theta)$  como función de  $\theta$  para  $\underline{X}$  fija, es una función monótona creciente; en el caso en que  $U(\underline{X}; \theta)$  sea monótona decreciente, las desigualdades en (22) se convierten en  $h_1 \geq \theta \geq h_2$  mientras que, cuando la función  $U(\underline{X}; \theta)$  no es monótona, lo que resulta es una región confidencial para  $\theta$  (definida en el Capítulo 1), es decir, una unión de intervalos disjuntos.

$T(\underline{X}) = \sum_{i=1}^n X_i$  y del parámetro  $\mu$

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

que tiene una distribución *Normal estándar* independiente del parámetro desconocido  $\mu$  y definida como

$$\Phi(u) = p(U \leq u) = \int_{-\infty}^u \phi(t) dt,$$

con  $\phi(\cdot)$  la densidad  $N(0, 1)$ .

Dado cualquier nivel de confianza  $1 - \alpha$ , se pueden encontrar valores  $a$  y  $b$ , tales que

$$p(a \leq U(T(\underline{X}); \mu) \leq b) = 1 - \alpha \quad (24)$$

es decir, tales que

$$\Phi(b) - \Phi(a) = \int_a^b \phi(t) dt = 1 - \alpha.$$

A partir de la desigualdad en probabilidad (24), el parámetro  $\mu$  se despeja de la siguiente forma

$$\begin{aligned} p(a \leq U \leq b) &= p(a \leq \sqrt{n}(\bar{X} - \mu) \leq b) \\ &= p\left(\frac{a}{\sqrt{n}} \leq (\bar{X} - \mu) \leq \frac{b}{\sqrt{n}}\right) \\ &= p\left(\bar{X} - \frac{b}{\sqrt{n}} \leq \mu \leq \bar{X} - \frac{a}{\sqrt{n}}\right) \\ &= 1 - \alpha, \end{aligned}$$

y el intervalo de confianza resultante al  $100 \times (1 - \alpha)\%$  para  $\mu$  es

$$\left(\bar{X} - \frac{b}{\sqrt{n}}, \bar{X} - \frac{a}{\sqrt{n}}\right) = (h_1, h_2).$$

Se pueden escoger muchos pares de valores  $a$  y  $b$  para cualquier nivel fijo  $1 - \alpha$ , tales que satisfagan (24). Sin embargo, como se mencionó en el Capítulo 1, en general se busca que el intervalo confidencial sea lo más corto posible; esto se logra escogiendo aquellos pares de valores  $a$  y  $b$  que hagan la longitud  $h_2 - h_1$  lo más chica posible. Si la longitud del intervalo  $h_2 - h_1$  no es aleatoria, se pueden escoger cantidades  $a$  y  $b$  tales que  $f_0(b) = f_0(a)$  ya que así, la distancia  $b - a$  será mínima para un nivel fijo  $1 - \alpha$ . Si la longitud  $h_2 - h_1$  es aleatoria, entonces se puede recurrir al intervalo más corto en

promedio, es decir, con longitud esperada más pequeña, o bien, al intervalo que resulte más corto para la muestra particular observada. Existen puntos de vista controversiales al respecto.

Para el ejemplo 2.7, en que la función  $\phi(u)$  es simétrica alrededor de  $u = 0$ , el intervalo confidencial más corto para el nivel fijo  $1 - \alpha$  se obtiene cuando  $a = -b$  en (24).

La etapa principal del método de construcción de intervalos confidenciales basado en cantidades pivotaes, consiste en poder encontrar una desigualdad doble en la que el parámetro  $\theta$  quede contenido, a partir de la desigualdad doble obtenida para la cantidad pivotal, es decir, que las desigualdades en (21), lleven a éstas en (22) para cualquier posible valor de  $\underline{X}$ . Como se observó anteriormente en el ejemplo 2.4, existen algunos casos en los cuales no es posible llevar a cabo esta etapa, debido a que no se puede despejar el parámetro de interés. En los siguientes ejemplos se trata de encontrar un intervalo de confianza para el parámetro  $\mu$  del ejemplo 2.1, a través de las pivotaes obtenidas en los ejemplos 2.4 y 2.6.

### Ejemplo 2.8

La cantidad pivotal en (13), obtenida en el ejemplo 2.4, fue

$$U = -2 \sum_{i=1}^n \log \int_{-\infty}^{X_i} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(t-\mu)^2}{2}\right\} dt, \quad (25)$$

que tiene una distribución *Ji-Cuadrada* con  $2n$  grados de libertad. Dado cualquier nivel  $(1 - \alpha)$  es posible encontrar valores  $a$  y  $b$  que, en particular, podrían ser los percentiles  $(\alpha/2) \times 100\%$  y  $(1 - \alpha/2) \times 100\%$  de una  $\chi^2_{(2n)}$ , respectivamente.

Las desigualdades probabilísticas en (21) que se obtienen para la pivotal de  $U$  en (25) son

$$P\left(\chi^2_{(2n)} \frac{\alpha}{2} \leq -2 \sum_{i=1}^n \log \int_{-\infty}^{X_i} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(t-\mu)^2}{2}\right\} dt \leq \chi^2_{(2n)} \frac{(1-\alpha/2)}{2}\right) = 1 - \alpha. \quad (26)$$

Observe que tratar de despejar al parámetro  $\mu$  de (26) para llegar a una desigualdad del tipo de (22) no parece ser nada sencillo.

### Ejemplo 2.9

La cantidad pivotal en (19) que se obtuvo en el ejemplo 2.6, fue igual a

$$U = \int_{-\infty}^T \left(\frac{n}{2\pi}\right)^{1/2} \exp\left\{-\frac{n(z-\mu)^2}{2}\right\} dz, \quad (27)$$

que se distribuye *Uniforme*(0,1). Dado un nivel de confianza  $(1 - \alpha)$ , se pueden encontrar cantidades  $a$  y  $b$  que, en particular, podrían ser los percentiles  $(\alpha/2) \times 100\%$  y  $(1 - \alpha/2) \times 100\%$  de una distribución *Uniforme*(0,1), respectivamente, tales que

$$\begin{aligned} p\left(\frac{\alpha}{2} \leq \int_{-\infty}^T \left(\frac{n}{2\pi}\right)^{1/2} \exp\left\{-\frac{n(z-\mu)^2}{2}\right\} dz \leq 1 - \frac{\alpha}{2}\right) \\ = p\left(\frac{\alpha}{2} \leq \Phi(\sqrt{n}(\bar{X} - \mu)) \leq 1 - \frac{\alpha}{2}\right) \\ = p\left(\Phi^{-1}\left(\frac{\alpha}{2}\right) \leq \sqrt{n}(\bar{X} - \mu) \leq \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right) \\ = p\left(\bar{X} - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\frac{1}{\sqrt{n}} \leq \mu \leq \bar{X} + \Phi^{-1}\left(\frac{\alpha}{2}\right)\frac{1}{\sqrt{n}}\right) \\ = 1 - \alpha. \end{aligned} \quad (28)$$

Así, el intervalo confidencial resultante al  $100 \times (1 - \alpha)\%$  para el parámetro  $\mu$  es

$$\left(\bar{X} - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\frac{1}{\sqrt{n}}, \bar{X} + \Phi^{-1}\left(\frac{\alpha}{2}\right)\frac{1}{\sqrt{n}}\right) = (h_1, h_2).$$

Como se pudo apreciar del ejemplo 2.8, en ocasiones no resulta fácil construir el intervalo confidencial de nivel  $(1 - \alpha)$  para el parámetro de interés, con el uso de ciertas pivotaes. En este ejemplo se pudo despejar el parámetro, gracias a que existen tablas de la  $N(0,1)$ , ya que los extremos del intervalo resultante requieren de la inversa de la función de distribución, o sea, los conocidos percentiles.

En el siguiente ejemplo se construye un intervalo de confianza haciendo uso de la pivotal en (17).

### Ejemplo 2.10

Retomando el ejemplo 2.5, en el que se tenían variables aleatorias i.i.d. *Uniformes* en el intervalo  $(0, \theta)$  y donde la cantidad pivotal exacta obtenida en (18) fue

igual a

$$U(T; \theta) = G^*(T; \theta) = \left(\frac{T}{\theta}\right)^n,$$

con distribución *Uniforme*(0,1). Con objeto de obtener el intervalo confidencial de nivel  $1 - \alpha$  para  $\theta$ , se parte del intervalo de probabilidad correspondiente para la pivotal anterior  $U$ , de la siguiente forma

$$\begin{aligned} p(a \leq U(T; \theta) \leq b) &= p\left(a \leq \left(\frac{T}{\theta}\right)^n \leq b\right) \\ &= p\left(\frac{T}{b^{1/n}} \leq \theta \leq \frac{T}{a^{1/n}}\right) \\ &= 1 - \alpha. \end{aligned} \tag{29}$$

Así, el intervalo confidencial resultante al  $100 \times (1 - \alpha)\%$  para el parámetro  $\theta$  es

$$(h_1(T), h_2(T)) = \left(\frac{T}{b^{1/n}}, \frac{T}{a^{1/n}}\right), \tag{30}$$

donde  $T$ , la estadística suficiente minimal, es igual al  $\max\{X_1, \dots, X_n\}$  y el par de valores  $(a, b)$  se escoge de forma tal que el área contenida entre los dos sea igual a  $1 - \alpha$ .

En general no parece haber un criterio a seguir, al utilizar la distribución *Uniforme*(0,1), para elegir las cantidades  $a$  y  $b$  en la desigualdad probabilística para la pivotal (29) del ejemplo anterior. En el caso en que la pivotal tenía una densidad *Ji - Cuadrada*, como en el ejemplo 2.8, se propusieron los percentiles  $\alpha/2 \times 100\%$  y  $(1 - \alpha/2) \times 100\%$ , pero podrían haberse propuesto aquellos que dieran el intervalo de longitud más corto para la pivotal; éstos son, los valores de  $a$  y  $b$  tales que

$$f_0(a) = f_0(b) \quad \text{sujetos a que} \quad \int_a^b f_0(t) dt = 1 - \alpha, \tag{31}$$

donde  $f_0$  es la densidad *Ji - Cuadrada*. El intervalo de probabilidad resultante para la pivotal  $U(\underline{X}; \theta)$ , contendrá los valores de  $u$  con densidad máxima. En la *Uniforme* sin embargo, todos los valores de  $u$  en el intervalo (0,1) tienen la misma densidad y por ello la ausencia aparente de un criterio para elegir los mejores valores de  $a$  y  $b$ .

Por ejemplo, suponga que en el ejemplo 2.10, se escogen los valores  $a$  y  $b$  como los percentiles  $(\alpha/2) \times 100\%$  y  $(1 - \alpha/2) \times 100\%$  respectivamente, de la distribución *Uniforme*(0,1). En tal caso, el intervalo obtenido en (30) resulta en el intervalo

$$\left(T(1 - \alpha/2)^{-1/n}, T(\alpha/2)^{-1/n}\right). \tag{32}$$

Sin embargo, también se podrían haber escogido los valores  $a$  y  $b$  como los percentiles  $(\alpha/4) \times 100\%$  y  $(1 - (3/4)\alpha) \times 100\%$  de la misma distribución, dando como resultado el intervalo confidencial

$$(T(1 - (3/4)\alpha)^{-1/n}, T(\alpha/4)^{-1/n}). \quad (33)$$

De la misma manera en que se generaron los intervalos (32) y (33), se puede generar una infinidad de intervalos para el mismo parámetro y con el mismo nivel de confianza; aquí cabe preguntarse cuál de entre todos es el mejor.

Como se ha venido mencionando, lo que se busca es que el intervalo confidencial sea lo más corto posible, por lo que podría pensarse en escoger valores de  $a$  y  $b$  tales que  $a = q\alpha$  y  $b = 1 - (1 - q)\alpha$ , para un cierto valor de  $q$  entre 0 y 1, y que minimicen la longitud del intervalo confidencial de nivel  $1 - \alpha$  para  $\theta$ .

Para  $T$  fijo, o aun para  $T$  sustituido por su esperanza, en cuyo caso se minimizaría la longitud esperada, la longitud del intervalo confidencial para  $\theta$  en (30) es

$$T(a^{-1/n} - b^{-1/n}) = T((q\alpha)^{-1/n} - (1 - (1 - q)\alpha)^{-1/n}), \quad (34)$$

y el intervalo será el más corto si, se escogen  $a$  y  $b$  tales que minimicen

$$a^{-1/n} - b^{-1/n} \quad ; \quad \text{sujetas a que } \begin{cases} 1 - a - b = 1 - \alpha \\ y \\ 0 < a + b < 1, \end{cases} \quad (35)$$

o, equivalentemente, si se escoge  $q$  tal que se minimice

$$Q(q) = (q\alpha)^{-1/n} - (1 - (1 - q)\alpha)^{-1/n}. \quad (36)$$

Resolviendo lo anterior, se tiene que la derivada parcial de la función  $Q(q)$  con respecto a  $q$  es igual a

$$\begin{aligned} \frac{\partial Q(q)}{\partial q} &= \frac{\alpha}{n} \left( (1 - (1 - q)\alpha)^{-\frac{n+1}{n}} - (q\alpha)^{-\frac{n+1}{n}} \right) \\ &= \frac{\alpha^{-1/n}}{n} \left( \left( \frac{1 - \alpha}{\alpha} + q \right)^{-\frac{n+1}{n}} - q^{-\frac{n+1}{n}} \right), \end{aligned}$$

la cual siempre es menor que cero, por lo que la función  $Q(q)$  es monótona decreciente en el intervalo  $[0, 1]$  y, por lo tanto, alcanza su mínimo cuando  $q$  es igual a 1.

Así, el intervalo más corto para el parámetro  $\theta$  en el ejemplo 2.10 se obtiene cuando  $q = 1$  o bien, cuando  $a = \alpha$  y  $b = 1$  en (30), es decir,

$$\left(T, \frac{T}{\alpha^{1/n}}\right). \quad (37)$$

Otro aspecto que en general puede inquietar respecto a este método, es el hecho de que, si bien  $U(T(\mathbf{X}); \theta)$  tiene una distribución *Uniforme*(0,1), también  $Z = 1 - U$  tendrá la misma distribución. En este caso, la nueva pivotal  $Z$  aplicada al ejemplo 2.10 es

$$Z = 1 - U = 1 - G^*(T; \theta) = 1 - \left(\frac{T}{\theta}\right)^n,$$

por lo que

$$\begin{aligned} p\left(a \leq 1 - \left(\frac{T}{\theta}\right)^n \leq b\right) &= p\left(1 - b \leq \left(\frac{T}{\theta}\right)^n \leq 1 - a\right) \\ &= p\left(\frac{T^n}{1 - a} \leq \theta^n \leq \frac{T^n}{1 - b}\right) \\ &= p\left(\frac{T}{(1 - a)^{1/n}} \leq \theta \leq \frac{T}{(1 - b)^{1/n}}\right) \\ &= 1 - \alpha, \end{aligned}$$

y el correspondiente intervalo confidencial de nivel  $1 - \alpha$  para  $\theta$ , sería

$$\left(T(1 - a)^{-1/n}, T(1 - b)^{-1/n}\right), \quad (38)$$

en lugar del anterior en (30).

Si lo que interesa es encontrar el intervalo confidencial de longitud más corta para  $T = \max\{X_1, \dots, X_n\}$  fijo, éste se obtiene, como en el caso anterior, al encontrar  $q$  tal que minimice la función

$$Q(q) = (\alpha - \alpha q)^{-1/n} - (1 - q\alpha)^{-1/n},$$

cuya parcial con respecto a  $q$  es mayor que cero, siendo la función monótona creciente en el intervalo  $[0, 1]$ , por lo que su mínimo se alcanza cuando  $q$  es igual a 0, dando ahora como resultado que el intervalo confidencial más corto de nivel  $1 - \alpha$  para el parámetro  $\theta$ , se obtiene al tomar  $a = 0$  y  $b = 1 - \alpha$  en (38). Observe que este intervalo es igual al que se obtuvo en (37), pudiéndose concluir que; siempre y cuando se minimice respecto

a  $q$ , el intervalo confidencial más corto de nivel  $(1 - \alpha)$  para  $\theta$ , resulta ser invariante a haber utilizado cualquiera de las transformaciones

$$U(T(\underline{X}); \theta) = G^*(T; \theta)$$

$$Z(T(\underline{X}); \theta) = 1 - U(T(\underline{X}); \theta) = 1 - G^*(T; \theta). \quad (39)$$

Sin embargo, cuando no se busca el intervalo más corto, la elección de  $U$  o  $Z$  en (39) lleva a los intervalos (30) o (38) respectivamente, que en general son distintos (solamente son iguales cuando  $q = 1/2$ ).

Finalmente, para ilustrar la aplicación del método de la cantidad pivotal en el caso asintótico, se hace uso de los siguientes tres ejemplos.

### Ejemplo 2.11

En el ejemplo 2.2 se encontró que, para cada posible valor de  $n$  y una muestra aleatoria  $X_1, \dots, X_n$  de una distribución con media  $\mu$  y varianza 1, la cantidad pivotal asintótica

$$U_n = \frac{\bar{X}_n - \mu}{1/\sqrt{n}}, \quad (40)$$

tenía una distribución límite  $N(0, 1)$ , la cual no depende de  $n$  ni de  $\mu$ .

Como la distribución *Normal estándar* es simétrica alrededor de cero, dado cualquier nivel fijo  $1 - \alpha$ , se pueden escoger valores  $a$  y  $b$  tales que  $\phi(a) = \phi(b)$ ; dichos valores, como se mencionó en el ejemplo 2.7, son  $a = -b$ . Así, se puede afirmar que

$$p(-b \leq U_n \leq b) \approx 1 - \alpha,$$

de donde despejando el parámetro  $\mu$ , se llega al correspondiente intervalo confidencial más corto de nivel  $1 - \alpha$

$$p\left(\bar{X}_n - \frac{b}{\sqrt{n}} \leq \mu \leq \bar{X}_n + \frac{b}{\sqrt{n}}\right) \approx 1 - \alpha. \quad (41)$$

El valor de  $b$  corresponde en este caso al percentil  $(1 - \alpha/2) \times 100\%$  de una densidad *Normal estándar*. Es importante preguntarse qué tan preciso es este intervalo, ya que

como se vio en el ejemplo 2.2, la pivotal  $U_n$  se construyó utilizando la aproximación a la densidad *Normal* cuando  $n$ , el tamaño de muestra, tendía a infinito; sin embargo, no se sabe qué tan rápida es esta convergencia a la *Normal*.

### Ejemplo 2.12

En el ejemplo 1.3, se trabajó con este problema tomado de Sprott y Viveros (1984). Recuérdese que lo que se tenía era una muestra de tamaño  $2^{**}$   $X = (X_1, X_2)$ , de una distribución *Exponencial* con media  $\theta > 0$ . Los resultados que se obtuvieron al aplicar la regla empírica para obtener un intervalo confidencial aproximado del parámetro  $\theta$ , fueron equívocos, ya que, para niveles mayores del 68% los intervalos cubrían valores incorrectos del parámetro.

Una cantidad pivotal asintótica explícita para este ejemplo, se puede construir haciendo uso de las propiedades asintóticas de los estimadores puntuales máximo verosímiles descritas en el Capítulo 1. En (1.4) se encontró que el estimador máximo verosímil de  $\theta$  era igual a

$$\hat{\theta} = \frac{X_1 + X_2}{2} = \frac{T}{n} \quad (42)$$

y de (1.10), la información observada de Fisher  $I_{\theta}$ , para este caso, es igual a

$$I_{\theta} = \frac{2}{\hat{\theta}^2}. \quad (43)$$

Obsérvese que en este ejemplo la información observada de Fisher coincide con el recíproco de la varianza estimada de  $\hat{\theta}$  obtenida en (1.5), hecho que en general no es cierto.

Por las propiedades asintóticas de  $\hat{\theta}$ , enunciadas en (1.8)-(1.11), se sabe que

$$\hat{\theta} \sim N\left(\theta; \frac{1}{I_{\theta}}\right)$$

así, una cantidad pivotal asintótica se puede construir como

$$U_n = \sqrt{I_{\theta}}(\hat{\theta} - \theta) \quad (44)$$

---

**\*\*** No parece razonable aplicar teoría asintótica a un ejemplo con un tamaño de muestra tan pequeño; sin embargo, el objetivo aquí, es precisamente ilustrar cómo a pesar de esto, se pueden obtener resultados precisos.

cuya distribución aproximada es una *Normal*(0,1). El intervalo confidencial aproximado de nivel  $1 - \alpha$  para  $\theta$  viene dado por

$$\begin{aligned} p(a \leq U_n \leq b) &= p(a \leq \sqrt{I_\theta}(\hat{\theta} - \theta) \leq b) \\ &= p\left(\hat{\theta} - \frac{b}{\sqrt{I_\theta}} \leq \theta \leq \hat{\theta} - \frac{a}{\sqrt{I_\theta}}\right) \\ &= 1 - \alpha. \end{aligned} \tag{45}$$

Si lo que se desea es el intervalo confidencial más corto, éste sería igual a

$$\left(\hat{\theta} \pm \frac{b}{\sqrt{I_\theta}}\right),$$

donde  $b$  es el percentil  $(1 - \alpha/2) \times 100\%$  de una *Normal estándar*. Sustituyendo a  $I_\theta$  y  $\hat{\theta}$  por las igualdades en (43) y (42) respectivamente, lo anterior resulta en la aseveración

$$\theta \in \frac{T}{n} \left(1 \pm \frac{b}{\sqrt{2}}\right), \tag{46}$$

aproximadamente en un  $(1 - \alpha) \times 100\%$  de las veces en muestreo repetido.

Suponga que se quieren intervalos de confianza al 75%, 90% y 95%, es decir, que  $\alpha = 0.25, 0.10$  y  $0.05$ . En este caso, se tendría por (46) que

$$\theta \in \frac{T}{n} \left(1 \pm \frac{1.15}{\sqrt{2}}\right), \quad \theta \in \frac{T}{n} \left(1 \pm \frac{1.64}{\sqrt{2}}\right) \text{ y } \theta \in \frac{T}{n} \left(1 \pm \frac{1.96}{\sqrt{2}}\right),$$

aproximadamente en un 75%, 90% y 95% de las veces. Para ver qué tan buena es esta aproximación, se puede calcular la probabilidad exacta de (46) a través de la variable  $T/\theta$ , que tiene una distribución conocida e igual a una *Gamma*( $\alpha = 2, \beta = 1$ ), de la siguiente forma

$$p\left(\frac{T}{n} \left(1 - \frac{b}{\sqrt{2}}\right) \leq \theta \leq \frac{T}{n} \left(1 + \frac{b}{\sqrt{2}}\right)\right) = p\left(\frac{c_1}{n} \leq \frac{\theta}{T} \leq \frac{c_2}{n}\right),$$

si  $c_1 \geq 0$  y  $c_2 \geq 0$ , entonces la probabilidad anterior es igual a

$$p\left(\frac{n}{c_2} \leq \frac{T}{\theta} \leq \frac{n}{c_1}\right) = \int_{\frac{n}{c_2}}^{\frac{n}{c_1}} \frac{1}{\Gamma(2)} \omega \exp\{-\omega\} d\omega,$$

pero si  $c_1 < 0$ , entonces la probabilidad es igual a

$$p\left(\frac{n}{c_2} \leq \frac{T}{\theta}\right) = 1 - p\left(\frac{T}{\theta} \leq \frac{n}{c_2}\right) = 1 - \int_0^{\frac{n}{c_2}} \frac{1}{\Gamma(2)} \omega \exp\{-\omega\} d\omega.$$

Las probabilidades exactas  $\times 100\%$  que se obtienen son: 69.76%, 76.30% y 79.50%, las cuales difieren mucho de los niveles de confianza aproximados  $\times 100\%$  del 75%, 90% y 95%. Estas diferencias en resultados, pareciera que se deben tan sólo, al hecho de que se utiliza un resultado asintótico en un ejemplo con un tamaño de muestra muy pequeño, dando como consecuencia que la pivotal utilizada en (44) no resulte ser aproximadamente *Normal*. Sin embargo, como se verá más adelante, Sprott y Viveros opinan que la diferencia también se debe al hecho de que el método utilizado es ineficiente, en el sentido de que no refleja la asimetría en la función de verosimilitud del parámetro  $\theta$ , que para este mismo ejemplo resulta ser igual a

$$L(\theta; \underline{X}) = \frac{1}{\theta^2} \exp\left\{-\frac{T}{\theta}\right\}.$$

Suponiendo que  $T = 12$ , la siguiente tabla muestra las verosimilitudes para varios valores de  $\theta$ .

$\theta$	2	4	6	8	10	20	30	35
$L(\theta; \underline{x})$	.0006	.0031	.0038	.0035	.0030	.0014	.0007	.0006

Observe lo asimétrico de esta verosimilitud, valores a la izquierda del estimador máximo verosímil  $\hat{\theta} = 6$  son mucho menos verosímiles que valores a una distancia igual a la derecha de  $\hat{\theta}$ , presentándose una asimetría a la derecha. Sprott y Viveros proponen hacer una reparametrización, reemplazando a  $\theta$  por una función uno a uno  $\eta = g(\theta)$  de modo que, aun para muestras pequeñas, se garantice que la aproximación *Normal* es razonablemente buena, como se ilustra en el siguiente ejemplo

### Ejemplo 2.13

Suponga que en el ejemplo 2.12 se utiliza el nuevo parámetro  $\eta = \theta^{-1/3}$ . El estimador máximo verosímil en este caso es

$$\hat{\eta} = \hat{\theta}^{-1/3} = \left(\frac{2}{X_1 + X_2}\right)^{1/3} = \left(\frac{n}{T}\right)^{1/3}, \quad (47)$$

y la información observada de Fisher para este nuevo parámetro resulta ser

$$I_{\eta} = -\frac{\partial^2 \log f(\underline{x}; \theta)}{\partial \hat{\eta}^2}$$

$$= I_{\theta} \cdot \left( \frac{\partial \hat{\theta}}{\partial \hat{\eta}} \right)^2 = \frac{18}{\hat{\eta}^2}. \quad (48)$$

La construcción de una cantidad pivotal asintótica similar a (44) pero ahora en términos de  $\eta$ , da como resultado

$$U_n = \sqrt{I_{\eta}}(\hat{\eta} - \eta) = \frac{3\sqrt{2}(\hat{\eta} - \eta)}{\hat{\eta}}, \quad (49)$$

cuya distribución límite también es una  $N(0,1)$ . El intervalo de confianza aproximado correspondiente para  $\eta$  es

$$p\left(a \leq \frac{3\sqrt{2}(\hat{\eta} - \eta)}{\hat{\eta}} \leq b\right) = p\left(\hat{\eta}\left(1 - \frac{b}{3\sqrt{2}}\right) \leq \eta \leq \hat{\eta}\left(1 + \frac{b}{3\sqrt{2}}\right)\right) \approx 1 - \alpha. \quad (50)$$

Este intervalo de confianza en términos del parámetro original  $\theta$ , resulta ser

$$p\left(\hat{\theta}\left(1 + \frac{b}{3\sqrt{2}}\right)^{-3} \leq \theta \leq \hat{\theta}\left(1 - \frac{b}{3\sqrt{2}}\right)^{-3}\right) \approx 1 - \alpha, \quad (51)$$

donde  $b$  sería igual al percentil  $(1 - \alpha/2) \times 100\%$  de una  $N(0,1)$ , si el intervalo de confianza que se desea es el más corto.

Seguendo un procedimiento igual al del ejemplo 2.12 para verificar la precisión del intervalo de confianza en (51), se obtiene que, para valores de  $\alpha$  de 0.25, 0.10 y 0.05, correspondiendo a valores de  $b$  de 1.15, 1.64 y 1.96 respectivamente, las probabilidades exactas  $\times 100\%$ , obtenidas de la distribución de la variable aleatoria  $T/\theta$  son: 73.4%, 89.1% y 94.7%, siendo muy cercanas a los niveles de confianza aproximados del 75%, 90% y 95% respectivamente. Comparando estos resultados con los obtenidos en el ejemplo 2.12, se puede apreciar la gran mejoría en precisión que se logró, aun con un tamaño de muestra tan pequeño como 2, al aplicar la reparametrización  $\eta = \theta^{-1/3}$ . Así mismo, la verosimilitud para este nuevo parámetro  $\eta$ , resulta ser simétrica alrededor del estimador máximo verosímil  $\hat{\eta}$ , como se puede apreciar en la siguiente tabla, que muestra las verosimilitudes para diferentes valores de  $\eta$ , suponiendo  $T = 12$  y  $\hat{\eta} = 0.5503$

$\eta$	0.1503	0.3503	0.5503	0.7503	0.9503	1.1503
$L(\eta; \underline{x})$	0.00001	0.00110	0.00376	0.00112	0.00002	$2.71 \times 10^{-8}$

Este método estudiado por Sprott para construir cantidades pivotaes asintóticas con ciertas propiedades que serán discutidas más adelante, se basa en la estimación máximo verosímil. La teoría de máxima verosimilitud trata con las propiedades frecuenciales del estimador máximo verosímil  $\hat{\theta}$ , el cual es solución de la ecuación

$$S(\underline{X}; \theta) = \frac{\partial \log L(\theta; \underline{X})}{\partial \theta} = \frac{\partial \log f(\underline{X}; \theta)}{\partial \theta} = 0, \quad (52)$$

donde  $S(\underline{X}; \theta)$  se conoce como la función de puntuación o función "score". Bajo condiciones de regularidad apropiadas, se puede demostrar que, la media y la varianza para esta función son iguales a

$$E_{\theta}(S(\underline{X}; \theta)) = 0 \quad ; \quad \text{Var}_{\theta}(S(\underline{X}; \theta)) = -E_{\theta} \left( \frac{\partial^2 \log f(\underline{X}; \theta)}{\partial \theta^2} \right) = I_{\mathbb{E}}(\theta), \quad (53)$$

donde  $I_{\mathbb{E}}(\theta)$  se definió en (1.9) como la cantidad de información de Fisher para la muestra  $\underline{X}$ . Estos resultados son exactos con muestras de cualquier tamaño, por lo que  $S(\underline{X}; \theta)$  es exactamente estable en la media.

La función de puntuación en (52) se puede expresar como la suma de v. a. i. i. d.

$$\begin{aligned} S(\underline{X}; \theta) &= \frac{\partial \log \prod_{i=1}^n f(X_i; \theta)}{\partial \theta} \\ &= \sum_{i=1}^n \frac{\partial}{\partial \theta} \log f(X_i; \theta) \end{aligned} \quad (54)$$

y, por el Teorema Central del Límite se implica que, asintóticamente  $S(\underline{X}; \theta)$  tiene una distribución límite  $N(0, I_{\mathbb{E}}(\theta))$ .

Por otro lado, máxima verosimilitud ha sido visto como un método para producir estimadores puntuales  $\hat{\theta}$  que son asintóticamente insesgados, es decir que

$$E_{\theta}(\hat{\theta}) \approx \theta, \quad \text{para } n \text{ grande}, \quad (55)$$

y con varianza asintótica mínima, igual a

$$\text{Var}_{\theta}(\hat{\theta}) \approx I_{\mathbb{E}}^{-1}(\theta), \quad \text{para } n \text{ grande}, \quad (56)$$

por lo que  $\hat{\theta}$  tiene como distribución aproximada a

$$\hat{\theta} \sim N(\theta, I_E^{-1}(\theta)). ** \quad (57)$$

Todos estos resultados son estrictamente asintóticos, por lo que la cantidad

$$(\theta - \hat{\theta}), \quad (58)$$

que puede verse como el error de estimación, es asintóticamente estable en la media.

La mayoría de las cantidades pivotaes que surgen de consideraciones asintóticas asociadas con máxima verosimilitud, se basan en la función de puntuación (52) y en el error de estimación (58). En Sprott (1975) aparece la siguiente lista de cantidades pivotaes asintóticas

$$S_1 = S/\sqrt{I_E(\theta)} \sim N(0,1) \quad (59)$$

$$U_1 = (\hat{\theta} - \theta)\sqrt{I_E(\theta)} \sim N(0,1) \quad (60)$$

$$S_2 = S/\sqrt{I_{\hat{\theta}}} \sim N(0,1) \quad (61)$$

$$U_2 = (\hat{\theta} - \theta)\sqrt{I_{\hat{\theta}}} \sim N(0,1) \quad (62)$$

$$S_3 = S/\sqrt{I_E(\hat{\theta})} \sim N(0,1) \quad (63)$$

$$U_3 = (\hat{\theta} - \theta)\sqrt{I_E(\hat{\theta})} \sim N(0,1), \quad (64)$$

donde  $I_E(\hat{\theta})$  y  $I_{\hat{\theta}}$  son estimadores de  $I_E(\theta)$  y se definen como

$$I_E(\hat{\theta}) = -E_{\theta} \left( \frac{\partial^2 \log f(\underline{X}; \theta)}{\partial \theta^2} \right) \Big|_{\theta=\hat{\theta}} \quad (65)$$

$$I_{\hat{\theta}} = - \frac{\partial^2 \log f(\underline{X}; \theta)}{\partial \theta^2} \Big|_{\theta=\hat{\theta}} \quad (66)$$

Las pivotaes  $S_1$  y  $U_1$  surgen de estandarizar (52) y (58) con respecto a su varianza exacta (53) y asintótica (56) respectivamente, mientras que las cantidades (61) a (64) surgen de reemplazar estas varianzas por sus estimadores dados en (65) y (66).

Bajo condiciones apropiadas de regularidad, todas las cantidades pivotaes asintóticas listadas anteriormente, son asintóticamente equivalentes, es decir que, para

\*\* Por la propiedad "BAN" ("Best Asymptotically Normal") que poseen los estimadores máximo verosímiles, vistas en el Capítulo 1.

valores grandes de  $n$ , todas tienen una distribución límite  $N(0, 1)$  por lo que, un intervalo de confianza aproximado de nivel  $1 - \alpha$  para el parámetro  $\theta$ , se puede obtener a partir del intervalo de probabilidad para la pivotal utilizada como

$$p(-b \leq U_n(\underline{X}; \theta) \leq b) \approx 1 - \alpha, \quad (67)$$

donde  $b$  es el percentil  $(1 - \alpha/2) \times 100\%$  de una  $N(0, 1)$ . Los intervalos obtenidos como la solución en  $\theta$  de  $-b \leq U_n(\underline{X}; \theta) \leq b$ , contendrán el valor verdadero de  $\theta$  con una frecuencia relativa de  $(1 - \alpha)$  aproximadamente. Sin embargo, todo esto es cierto sólo asintóticamente, pero para muestras finitas, que es lo que se encuentra en las aplicaciones, no es claro cuál de entre todas las pivotaes (59) a (64) converge más rápidamente a la distribución  $N(0, 1)$ , ni qué tan rápida sea esta convergencia.

La precisión y validez de los resultados obtenidos al aplicar teoría asintótica a muestras finitas, han sido investigadas por Sprott y Kalbfleisch (1969), Sprott (1973 y 1975) y Sprott y Viveros (1984), quienes demuestran, a través de varios ejemplos, que la precisión y validez de la aplicación está condicionada más a la normalidad de la función de verosimilitud del parámetro  $\theta$ , que al tamaño de muestra  $n$ .

El efecto que tiene la forma de la función de verosimilitud del parámetro  $\theta$  en las inferencias, ha sido estudiado por varios autores; en particular Sprott (1973 y 1984) ha encontrado, que cuando dicha función se aleja de la forma *Normal*, ya sea porque es muy asimétrica o con colas muy pesadas, entonces afirmaciones de la forma  $\theta = \hat{\theta} \pm \sigma$  en términos de un estimador puntual y una "varianza", son engañosas ya que esconden dichas "anomalías". Se propone así, como un primer paso para resumir la información relevante en los datos respecto a un parámetro, graficar la verosimilitud estandarizada o relativa del parámetro,  $R(\theta; \underline{X})$ , o dar un conjunto de rangos anidados que converjan a  $\hat{\theta}$ , estos últimos son conocidos como rangos o intervalos de verosimilitud.

La función de verosimilitud relativa se define como

$$R(\theta; \underline{X}) = \frac{L(\theta; \underline{X})}{\sup_{\theta} L(\theta; \underline{X})} = \frac{L(\theta; \underline{X})}{L(\hat{\theta}; \underline{X})}. \quad (68)$$

Esta función mide la verosimilitud de todos los valores específicos de  $\theta$  relativa al valor más verosímil de  $\theta$ , que es el máximo verosímil  $\hat{\theta}$ . De esta forma,  $R(\theta; \underline{X})$  da una escala

de verosimilitud, que varía de cero a uno para todos los valores de  $\theta$ . Esta función también posee la propiedad de invarianza funcional ante reparametrizaciones.

La forma de esta función queda determinada por los primeros términos en su expansión de Taylor. Sea  $\theta$  un parámetro escalar con función de verosimilitud relativa  $R(\theta; \underline{X})$  como en (68) y sea  $\hat{\theta}$  el estimador máximo verosímil que satisface (52). Observe que de (52) y (68) se obtienen las siguientes igualdades

$$S(\underline{X}; \theta) = \frac{\partial \log L(\theta; \underline{X})}{\partial \theta} = \frac{\partial \log R(\theta; \underline{X})}{\partial \theta} = 0. \quad (69)$$

Entonces, si  $R(\theta; \underline{X})$  satisface condiciones apropiadas de regularidad, la expansión de Taylor del logaritmo de  $R$  alrededor de  $\hat{\theta}$  hasta el término cúbico, da la aproximación siguiente (Sprott y Viveros 1984)

$$\begin{aligned} \log R(\theta; \underline{X}) &\doteq -\frac{(\theta - \hat{\theta})^2}{2} I_{\theta} + \frac{\partial^3 \log L(\theta; \underline{X})}{\partial \theta^3} \Big|_{\theta=\hat{\theta}} \frac{(\theta - \hat{\theta})^3}{6} + \dots \\ &\doteq -\frac{1}{2} U_2^2 \left[ 1 - \left( \frac{U_2}{3} \right) F_3(\hat{\theta}) \right], \end{aligned} \quad (70)$$

donde la pivotal  $U_2$  se definió en (62), la información observada de Fisher  $I_{\theta}$  se definió en (66), y

$$F_3(\hat{\theta}) = \left[ \frac{\partial^3 \log L(\theta; \underline{X})}{\partial \theta^3} \Big|_{\theta=\hat{\theta}} \right] I_{\theta}^{-3/2}. \quad (71)$$

Si  $F_3(\hat{\theta})$  en (70) es cercano a cero, entonces

$$\begin{aligned} \log R(\theta; \underline{X}) &\doteq -\frac{1}{2} U_2^2 \doteq -\frac{1}{2} (\theta - \hat{\theta})^2 I_{\theta} \\ R(\theta; \underline{X}) &\doteq \exp \left\{ -\frac{1}{2} (\theta - \hat{\theta})^2 I_{\theta} \right\}. \end{aligned} \quad (72)$$

Es decir que  $\theta$  tiene aproximadamente una verosimilitud *Normal*, centrada en  $\hat{\theta}$  y con varianza  $I_{\theta}^{-1}$ . Si éste es el caso, la verosimilitud de  $\theta$  se podrá representar por el estimador puntual  $\hat{\theta}$  y por su precisión  $I_{\theta}$ , la cual determina la precisión de  $\hat{\theta}$  como un estimador de  $\theta$  en la función de verosimilitud aproximadamente *Normal* de  $\theta$ .

Si  $F_3(\hat{\theta})$  en (70) no es pequeña, entonces la verosimilitud presenta asimetría apreciable, como fue el caso del ejemplo 2.12, por lo que ésta no podrá resumirse adecuadamente por  $(\hat{\theta}, I_{\theta})$  y las inferencias tendrán que hacerse, no a partir de  $\theta$ , sino

de un nuevo parámetro  $\eta$ , que sea función uno a uno de  $\theta$  y cuya verosimilitud sea aproximadamente *Normal*, como se hizo en el ejemplo 2.13. Esta función  $\eta = g(\theta)$ , existirá siempre que la verosimilitud sea una función "bien comportada".

Para encontrar la reparametrización apropiada, se escoge una función  $g(\theta)$  que remueva completamente la componente principal de asimetría, es decir, tal que  $F_3(\hat{\eta}) = 0$ , donde  $\hat{\eta} = g(\hat{\theta})$ . Anscombe (1964) encontró que estas funciones son soluciones de la ecuación diferencial

$$g'(\theta) = \frac{d\hat{\eta}}{d\hat{\theta}} = C \exp \left\{ \frac{1}{3} \int \frac{\partial^3 \log R(\theta; \underline{X}) / \partial \hat{\theta}^3}{\partial^2 \log R(\theta; \underline{X}) / \partial \hat{\theta}^2} d\hat{\theta} \right\}. \quad (73)$$

Recuerde que en el ejemplo 2.13 se reparametrizó con  $\eta = g(\theta) = \theta^{-1/3}$ ; este nuevo parámetro surge como resultado de aplicar (73), ya que

$$\frac{\partial^3 \log R(\theta; \underline{X}) / \partial \hat{\theta}^3}{\partial^2 \log R(\theta; \underline{X}) / \partial \hat{\theta}^2} = \left( \frac{4T}{\hat{\theta}^4} \right) \left( -\frac{\hat{\theta}^3}{T} \right) = -\frac{4}{\hat{\theta}},$$

entonces

$$g'(\theta) = C \exp \left\{ -\frac{4}{3} \log t \right\} = C t^{-4/3}$$

y

$$g(\theta) = \eta = C \int_0^\theta t^{-4/3} dt = -3C\theta^{-1/3},$$

escogiendo por simplicidad,  $C = -1/3$ .

En general cuando se desea examinar el efecto que tiene cualquier transformación del tipo  $\eta = g(\theta)$ , en producir verosimilitudes aproximadamente normales, se puede hacer uso de la igualdad siguiente obtenida por Sprott (1973)

$$|F_3(\hat{\eta})| = \left| F_3(\hat{\theta}) + \frac{3(d^2\hat{\eta}/d\hat{\theta}^2)}{(d\hat{\eta}/d\hat{\theta})I_{\hat{\theta}}^{1/2}} \right|. \quad (74)$$

Se puede verificar fácilmente que, para el ejemplo 2.13, se obtiene

$$|F_3(\hat{\eta})| = \left| \frac{4}{\sqrt{n}} - \frac{4}{3}\hat{\theta}^{-7/3} \cdot \frac{3}{\hat{\theta}^{-4/3}} \cdot \frac{\hat{\theta}}{\sqrt{n}} \right| = 0.$$

Este procedimiento de obtener verosimilitudes aproximadamente normales, evidencia el hecho de que, en algunos problemas prácticos es posible lograr una

verosimilitud aproximadamente *Normal*, sin tener necesariamente un tamaño de muestra grande, mejorando así la precisión de las inferencias al aplicar resultados asintóticos a muestras finitas, como fue el caso del ejemplo 2.13 en que, aun con  $n = 2$ , se pudieron obtener intervalos de confianza bastante precisos.

Las ventajas de tener una función de verosimilitud del parámetro  $\theta$ , o bien, de un nuevo parámetro función uno a uno del original, que sea aproximadamente *Normal*, también se ven reflejadas en las cantidades pivotaes (59) a (64). Observe que de (69) y (70) surge la siguiente aproximación para la función de puntuación  $S$

$$S(\underline{X}; \theta) \doteq U_2 \sqrt{I_\theta} \left[ 1 - \left( \frac{U_2}{2} \right) F_3(\hat{\theta}) \right] \quad (75)$$

y, suponiendo que  $F_3(\hat{\theta})$  se pueda hacer igual a cero, se tiene que

$$S(\underline{X}; \theta) \doteq U_2 \sqrt{I_\theta}, \quad (76)$$

por lo que las pivotaes en (61) y (62) resultan ser aproximadamente iguales, es decir

$$S_2 = \frac{S}{\sqrt{I_\theta}} \doteq U_2 = (\hat{\theta} - \theta) \sqrt{I_\theta}, \quad (77)$$

lográndose también con ésto que (61) sea aproximadamente lineal en  $\theta$ . Esta linealidad en las pivotaes nos permitirá despejar el parámetro  $\theta$  para obtener su intervalo de confianza.

En ocasiones, la información observada de Fisher  $I_\theta$  en (66) no se puede expresar como una función de  $\hat{\theta}$ , por lo que resulta más conveniente utilizar el estimador  $I_E(\hat{\theta})$  en (65) y la expansión de Taylor de  $E\{\log R(\theta; \underline{X})\}$ , la cual fue obtenida por Sprott (1973) y resulta ser igual a

$$E\{\log R(\theta; \underline{X})\} \doteq -\frac{1}{2} U_3^2 \left\{ 1 + \left( \frac{U_3}{3} \right) E \left\{ \frac{\partial^3 \log R(\theta; \underline{X})}{\partial \theta^3} \right\} I_E^{-3/2}(\hat{\theta}) \right\}. \quad (78)$$

En este caso, Sprott (1973) también ha demostrado que al lograr una buena aproximación *Normal* de  $R$  como en (72), se logra que la distribución de la pivotal  $U_3$  en (64) sea más cercana a la distribución *Normal*, así también la de  $S_3$  en (63), mejorándose con ésto la precisión de las inferencias basadas en ellas.

Resumiendo, se puede decir que, cuando se obtiene una buena aproximación *Normal* de  $R$ , existe muy poca diferencia en cuanto al uso de las pivotaes (61) a (64) ya que éstas resultan ser aproximadamente iguales.

En cuanto a las pivotaes  $S_1$  y  $U_1$  en (59) y (60), las cuales, como se mencionó anteriormente, surgen de estandarizar la función de puntuación  $S$  y el error de estimación  $(\theta - \hat{\theta})$ , con su varianza exacta y asintótica, respectivamente; pareciera que estas pivotaes son preferibles a las pivotaes (61) a (64), debido a que no se presenta el problema de pérdida de información como resultado de reemplazar la varianza muestral por la varianza poblacional; sin embargo, Sprott y Viveros (1984) y Sprott (1975), demuestran que esta aparente desventaja en las pivotaes (61) a (64) puede compensarse por la posibilidad de incrementar la precisión de la aproximación *Normal* y la linealidad de estas pivotaes, como se puede apreciar en el siguiente ejemplo.

#### Ejemplo 2.14

Aquí se continúa con lo visto en el ejemplo 2.13, para el que Sprott y Viveros (1984) obtienen las siguientes gráficas, donde la línea continua (—) corresponde a la densidad *Normal*, la línea -- corresponde a la densidad

$$f(u) = 2\sqrt{2} \left(1 - \frac{u}{3\sqrt{2}}\right)^5 \exp \left[-2 \left(1 - \frac{u}{3\sqrt{2}}\right)^3\right],$$

de la pivotal

$$U = U_\phi = (\hat{\phi} - \phi) \sqrt{I_\phi},$$

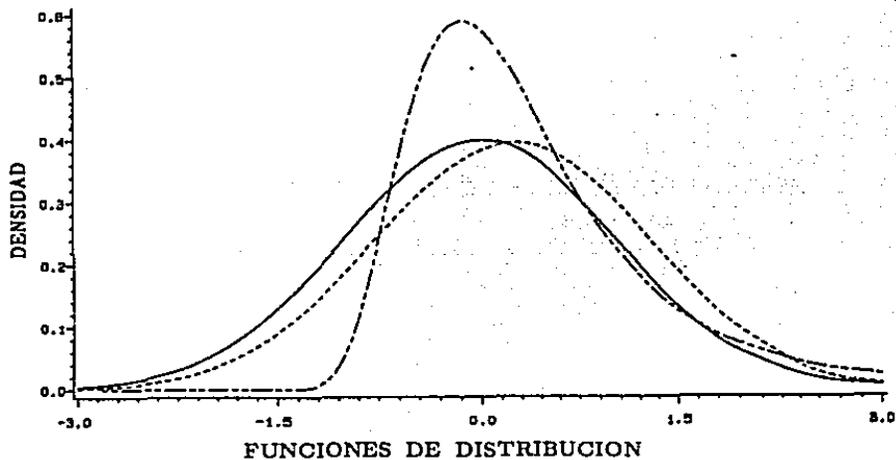
en que se utiliza un estimador de la varianza asintótica y la línea --- --- --- corresponde a la densidad

$$f(v) = 12k(1 + kv)^{-7} \exp[-2(1 + kv)^{-3}],$$

de la pivotal

$$V_\phi = \frac{(\hat{\phi} - \phi)}{\text{Var}(\hat{\phi})},$$

en la que se reemplazó  $I_{\phi}^{-1}$  en  $U_{\phi}$  por  $\text{Var}(\hat{\phi}) = (k\phi)^2$ ,  $k = 0.3520$ .



Como se puede apreciar en esta gráfica, se logra una mejor aproximación a la *Normal* cuando se utiliza el estimador de la varianza del estimador, que cuando se utiliza la varianza del estimador, además de que la pivotal  $U_{\phi}$  resulta ser lineal en  $\phi$ .

## CAPITULO 3

### ENFOQUES FIDUCIAL Y BAYESIANO

#### 3.1 Introducción

En este capítulo se hace una exposición de tres métodos, el Fiducial, Fiducial Implícito y Bayesiano, a partir de los cuales es posible hacer inferencias acerca del valor de un parámetro escalar desconocido  $\theta$ . Los dos primeros tratan a  $\theta$  como constante, mientras que el último como una variable aleatoria y, en los tres casos se pueden construir intervalos para  $\theta$ : para los primeros dos métodos se llaman fiduciales y, para el último de probabilidad (*a-posteriori*).

En la primera sección se presenta el procedimiento de construcción de intervalos fiduciales, basado en la distribución fiducial del parámetro  $\theta$ . El objetivo de esta sección no es juzgar la validez de este método, que ha sido sumamente controversial desde su aparición en 1930 sino, más bien, exponer de una manera sencilla y clara el procedimiento de construcción y aplicación del mismo, sin dejar de comentar algunos de los puntos que han sido más discutidos y criticados por muchos autores: Barnard (1963a,b, 1987), Dempster (1963), Fraser (1961a,b), Lindley (1958), Neyman (1942), Pederzen (1978), Sprott (1963, 1966), Sprott y Kalbfleisch (1967), Tukey (1957), por citar algunos.

En la siguiente sección se presenta otra versión del método fiducial, para el caso en que la distribución fiducial no se exhibe explícitamente en la construcción de los intervalos fiduciales, sino que queda subyacente en un procedimiento aquí referido como el método Fiducial Implícito.

Finalmente, en la tercera sección, se presenta el método Bayesiano de construcción de intervalos de probabilidad (*a-posteriori*) de densidad máxima, basado en el Teorema de Bayes, comentando algunas de las diferentes técnicas que se han desarrollado en torno a la búsqueda de la distribución inicial (*a-priori*) del parámetro  $\theta$ .

### 3.2 Método Fiducial

Fisher en 1930 introdujo por primera vez un procedimiento alternativo al de Bayes, conocido como argumento fiducial, para poder hacer inferencias sobre el valor de un parámetro escalar  $\theta$ . Este argumento permite obtener una distribución fiducial para  $\theta$ , basándose tan sólo en las observaciones disponibles  $\underline{x}$ , y su aplicación es válida solamente en los casos donde no se tiene un conocimiento inicial (*a-priori*) sobre el parámetro.

Existen en la literatura diferentes planteamientos del argumento fiducial; por ejemplo, Pederzen (1978), utiliza los conjuntos de referencia de Fisher y Fraser (1961), modelos de transformación de parámetros. El planteamiento del argumento fiducial elegido aquí, requiere de conceptos mencionados en el Capítulo 2 el de cantidad pivotal exacta y el de estadística suficiente minimal.

Recuerde que dos de los métodos presentados para la construcción de cantidades pivotaes exactas, se basaban en la estadística suficiente minimal, siendo así consistentes con el principio de verosimilitud. El primero de ellos consistía en encontrar una cantidad pivotal exacta  $U(T; \theta)$ , que fuese función de la estadística suficiente minimal uni-dimensional  $T$ , y del parámetro  $\theta$ , cuya distribución fuese conocida y única para cualquier valor de  $\theta \in \Theta$ , es decir que

$$p(U(T; \theta) \leq u; \theta) = F_0(u), \quad \forall \theta \in \Theta. \quad (1)$$

El segundo método consistía en encontrar la estadística suficiente minimal uni-dimensional  $T$ , con función de distribución  $F(t; \theta)$  que se supone continua en  $t$ , para después construir la cantidad pivotal exacta como la transformación con la integral de

probabilidad utilizando a  $F(t; \theta)$ , es decir que

$$U(T; \theta) = F(T; \theta) \sim \text{Uniforme}(0, 1),$$

$$p(U(T; \theta) \leq u; \theta) = p(F(T; \theta) \leq u; \theta) = u, \quad \forall \theta \in \Theta. \quad (2)$$

La utilización hecha en el Capítulo 2 de la cantidad pivotal exacta  $U(T; \theta)$ , consistía en construir primero un intervalo de probabilidad  $\gamma = 1 - \alpha$  para  $U$ , para después ver si era posible "pivotear", es decir, construir una desigualdad en la que el parámetro  $\theta$  quedase contenido, con la misma confianza  $\gamma$ . Para lograr lo anterior fue necesario pedir que la pivotal utilizada, vista como función de  $\theta$ , fuese una función monótona para cada valor posible de  $t$ .

En el primer método se llegó a que si la cantidad pivotal  $U(t, \theta)$  era una función creciente en  $\theta$ , para cada  $t$  fija, entonces se cumplía

$$U(T, \theta) \leq u \iff \theta \leq U^{-1}(T, u). \quad (3)$$

Así por (1) y para cada  $\gamma$  en  $(0, 1)$ , se tenían las siguientes igualdades en probabilidad

$$p(U(T, \theta) \leq u_\gamma) = p(\theta \leq U^{-1}(T; u_\gamma)) = F_0(u_\gamma) = \gamma, \quad (4)$$

donde  $u_\gamma$  es el percentil  $\gamma$  de  $F_0(u)$ . En el caso en que  $U(T, \theta)$  fuera decreciente, entonces la desigualdad en (3) simplemente se invertiría.

Para el segundo método se tiene que las propiedades pedidas a la pivotal  $U(t; \theta)$  resultan ser por (2), las mismas que las de la distribución  $F(t; \theta)$  por lo que, si  $F(t, \theta)$  era una función creciente en  $\theta$  para cada  $t$  fija, entonces se cumplía

$$U(T, \theta) \leq \gamma \iff F(T, \theta) \leq \gamma \iff \theta \leq F^{-1}(T, \gamma). \quad (5)$$

Así por (2) y para cada  $\gamma$  en  $(0, 1)$ , se obtenían las siguientes igualdades en probabilidad

$$p(F(T, \theta) \leq \gamma) = p(\theta \leq F^{-1}(T, \gamma)) = \gamma, \quad \forall \theta \in \Theta, \quad (6)$$

donde obviamente  $\gamma$  es el percentil de una distribución  $\text{Uniforme}(0, 1)$ . Si  $F(t, \theta)$  fuera una función decreciente, entonces la desigualdad en (5) se invertiría.

A las cantidades aleatorias  $U^{-1}(T, u_\gamma)$  en (4) y  $F^{-1}(T, \gamma)$  en (6), se les conoce como los límites de confianza superiores para  $\theta$  de nivel  $\gamma$  y, en caso de trabajar con funciones monótonas decrecientes, lo que se obtendría serían límites de confianza inferiores para  $\theta$  de nivel  $\gamma$ .

Desde el punto de vista frecuentista, las aseveraciones probabilísticas en (4) y (6) son válidas, ya que antes de observar la muestra  $\underline{x}$ , los límites de confianza superiores  $U^{-1}(T, u_\gamma)$  y  $F^{-1}(T, \gamma)$  estarán por arriba del valor de  $\theta$ , en un  $100 \times \gamma\%$  de las veces en un número grande de muestras. Aquí  $\theta$  es una constante desconocida y los límites de confianza superiores son variables aleatorias. En particular, los límites superiores que resultan al haber observado  $T = t_o$ ,  $U^{-1}(t_o, u_\gamma)$  y  $F^{-1}(t_o, \gamma)$ , también se les identifica como límites de confianza superiores para  $\theta$  de nivel  $\gamma$ .

Hasta aquí no se ha introducido ningún nuevo tipo de razonamiento, de hecho, todo lo expuesto anteriormente se utilizó en el Capítulo 2 para plantear la teoría de Intervalos de Confianza de Neyman. Justo en este momento del planteamiento, es cuando el argumento fiducial de Fisher introduce un nuevo tipo de razonamiento en el que  $\theta$  juega el papel lógico de una variable aleatoria, sin ser variable aleatoria, y  $T$  llega a jugar el papel lógico de un parámetro que define a la distribución fiducial de  $\theta$ .

Para hacer la exposición un poco más clara, se obtendrá la densidad fiducial para el primer método, en el que se utiliza una cantidad pivotal función de la estadística suficiente minimal, así como sus intervalos fiduciales y, posteriormente, se obtendrá para el segundo método, en el que la pivotal es justo la función de la estadística suficiente minimal.

El argumento fiducial supone que la muestra  $\underline{x}$  ya ha sido observada y, en consecuencia,  $T = t_o$ , por lo que la desigualdad en (3) se convierte en

$$\theta \leq U^{-1}(t_o, u), \quad (7)$$

y, a ésta, se le asocia una "probabilidad fiducial" similar a (4)

$$p_{Fid}(\theta \leq U^{-1}(t_o, u_\gamma)) = F_0(u_\gamma) = \gamma. \quad (8)$$

Al número  $U^{-1}(t_o, u_\gamma)$  se le llamará, por simplicidad de notación,  $\theta^*$ . Este obviamente depende de la probabilidad  $\gamma$  y del valor  $t_o$  que tomó la estadística suficiente

minimal  $T$ , como se podrá apreciar de las siguientes dos ecuaciones

$$\theta^* = U^{-1}(t_o, u_\gamma) \iff U(t_o, \theta^*) = u_\gamma \quad (9)$$

y, reescribiendo (8) en términos de  $\theta^*$ , se define la función de distribución fiducial como

$$F_{Fid}(\theta^*) = p_{Fid}(\theta \leq \theta^*) = F_0(U(t_o, \theta^*)) = \gamma. \quad (10)$$

Para poder obtener la densidad fiducial, se debe suponer que  $F_0$ , la distribución de la pivotal  $U$ , tiene una derivada continua y que  $U(t_o, \cdot)$  es diferenciable con respecto a  $\theta$ , de donde se obtiene que

$$f_{Fid}(\theta^*) = \frac{dF_{Fid}(\theta^*)}{d\theta^*} = \frac{dF_0(U(t_o, \theta^*))}{d\theta^*} = f_0(U(t_o, \theta^*)) \cdot \frac{\partial U(t_o, \theta^*)}{\partial \theta^*}, \quad (11)$$

donde  $f_0$  es la función de densidad de la pivotal  $U$ .

En el caso en que la pivotal  $U$  sea una función monótona decreciente en  $\theta$ , para cada  $t$ , (10) se escribiría como

$$F_{Fid}(\theta^*) = 1 - p_{Fid}(\theta \leq \theta^*) = 1 - F_0(U(t_o, \theta^*)) = \gamma, \quad (12)$$

y la densidad fiducial entonces resulta ser

$$f_{Fid}(\theta^*) = -f_0(U(t_o, \theta^*)) \cdot \frac{\partial U(t_o, \theta^*)}{\partial \theta^*}. \quad (13)$$

Tomando en cuenta (11) y (13), es usual que en la literatura aparezca la función de densidad fiducial para  $\theta$  como

$$f_{Fid}(\theta^*) = f_0(U(t_o, \theta^*)) \cdot \left| \frac{\partial U(t_o, \theta^*)}{\partial \theta^*} \right|, \quad (14)$$

con  $\theta^*$  relacionado con  $\gamma$  y  $t_o$  (que está fijo), a través de

$$\gamma = F_0(U(t_o, \theta^*)) = F_0(u_\gamma), \quad (15)$$

y la función de distribución de  $\theta$ , simplemente como

$$F_{Fid}(\theta^*) = \begin{cases} F_0(U(t_o, \theta^*)) & \text{si } U \text{ es creciente en } \theta \\ 1 - F_0(U(t_o, \theta^*)) & \text{si } U \text{ es decreciente en } \theta. \end{cases} \quad (16)$$

Con objeto de ilustrar la aplicación del método anterior se da el siguiente ejemplo:

### Ejemplo 3.1

Sea  $X_1, \dots, X_n$  una muestra aleatoria de la densidad *Exponencial* ( $\theta > 0$ )

$$f(x; \theta) = \theta \exp(-x\theta) 1_{(0, \infty)}(x).$$

La estadística suficiente minimal en este caso es

$$T(\underline{X}) = \sum_{i=1}^n X_i,$$

con función de distribución continua y monótona creciente en  $\theta$ , de acuerdo a una *Gamma* ( $\alpha = n, \beta = \frac{1}{\theta}$ )

$$F(t, \theta) = \int_0^t \frac{1}{\Gamma(n)} \theta^n \exp(-\theta z) \cdot z^{n-1} dz.$$

Si se elige como cantidad pivotal a

$$U(T, \theta) = \theta T \sim \text{Gamma}(\alpha = n, \beta = 1),$$

entonces

$$F_0(u) = \int_0^{u\theta} \frac{1}{\Gamma(n)} \exp(-z) z^{n-1} dz.$$

La distribución fiducial en este caso, aplicando (16), es igual a

$$f_{Fid}(\theta^*) = \int_0^{\theta^* t_0} \frac{1}{\Gamma(n)} \exp(-z) z^{n-1} dz$$

y la densidad fiducial, aplicando (11), resulta ser

$$f_{Fid}(\theta^*) = \frac{1}{\Gamma(n)} \exp(-\theta^* t_0) \cdot (\theta^* t_0)^{n-1} \cdot t_0,$$

que corresponde a decir que, dado el valor observado  $T = t_0$ ,  $\theta$  tiene como distribución fiducial a una *Gamma* ( $\alpha = n, \beta = \frac{1}{t_0}$ ).

Suponga que se desea obtener un intervalo fiducial central para  $\theta$  (es decir, con límites inferior y superior), con probabilidad fiducial  $\gamma$  y basado en el valor observado  $T = t_0$  como parámetro. Este intervalo se puede obtener encontrando cantidades  $\theta_1^*$  y  $\theta_2^*$  tales que

$$P_{Fid}(\theta_1^* \leq \theta \leq \theta_2^*) = F_{Fid}(\theta_2^*) - F_{Fid}(\theta_1^*) = \gamma. \quad (17)$$

Si adicionalmente se desea que este intervalo sea el más pequeño, entonces, además de (17),  $\theta_1^*$  y  $\theta_2^*$  deben satisfacer que

$$f_{Fid}(\theta_1^*) = f_{Fid}(\theta_2^*). \quad (18)$$

Para el ejemplo 3.1,  $\theta_1^*$  y  $\theta_2^*$  serían tales que

$$\int_{\theta_1^* t_0}^{\theta_2^* t_0} \frac{1}{\Gamma(n)} \exp(-z) z^{n-1} dz = \gamma$$

y

$$\exp(-\theta_1^* t_0) \cdot \theta_1^{*n-1} = \exp(-\theta_2^* t_0) \cdot \theta_2^{*n-1}.$$

La interpretación que se da a los intervalos fiduciales en (17) es que, para la muestra particular que se observó y que produjo el valor  $t_0$ , el valor desconocido del parámetro  $\theta$  se encontrará entre las cantidades  $\theta_1^*$  y  $\theta_2^*$  con una probabilidad fiducial  $\gamma$ .

### Ejemplo 3.2

Sea  $\underline{X} = (X_1, \dots, X_n)$  una muestra aleatoria de una  $N(\mu, 1)$ . La estadística suficiente minimal en este caso es

$$T(\underline{X}) = \frac{1}{n} \sum_{i=1}^n X_i$$

con función de distribución continua y monótona decreciente en  $\mu$ , de acuerdo a una  $N(\mu, \frac{1}{n})$

$$F(t; \mu) = \int_{-\infty}^t \frac{\sqrt{n}}{\sqrt{2\pi}} \exp\left(-\frac{(z-\mu)^2}{2/n}\right) dz.$$

Si se elige como pivotal a la cantidad

$$U(T; \mu) = (T - \mu)\sqrt{n} \sim N(0, 1),$$

entonces

$$F_0(u) = \int_{-\infty}^u \frac{1}{\sqrt{2\pi}} \exp(-z^2/2) dz.$$

Una vez observado el valor  $T = t_o$ , la distribución fiducial, aplicando (16), es igual a

$$\begin{aligned} F_{Fid}(\mu^*) &= 1 - F_0(U(t_o, \mu^*)) = 1 - \int_{-\infty}^{\sqrt{n}(t_o - \mu^*)} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz \\ &= \int_{\sqrt{n}(t_o - \mu^*)}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz, \end{aligned}$$

y la densidad fiducial aplicando (13) resulta ser

$$f_{Fid}(\mu^*) = -f_0(U(t_o, \mu^*)) \cdot \frac{\partial U(t_o, \mu^*)}{\partial \mu^*} = \frac{\sqrt{n}}{\sqrt{2\pi}} \exp\left(-\frac{(\mu - t_o)^2}{2/n}\right),$$

que equivale a decir que, dado el valor observado  $T = t_o$ ,  $\mu$  tiene como distribución fiducial una  $N(t_o, \frac{1}{n})$ .

El intervalo fiducial más pequeño para  $\mu$  de probabilidad fiducial  $\gamma$ , dada la muestra observada que produjo el valor  $T = t_o$ , se obtendrá, encontrando cantidades  $\mu_1^*$  y  $\mu_2^*$  tales que

$$P_{Fid}(\mu_1^* \leq \mu \leq \mu_2^*) = \int_{\sqrt{n}(t_o - \mu_1^*)}^{\sqrt{n}(t_o - \mu_2^*)} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz = \gamma$$

y

$$f_{Fid}(\mu_1^*) = f_{Fid}(\mu_2^*) \Rightarrow \exp\left(-\frac{(\mu_1^* - t_o)^2}{2/n}\right) = \exp\left(-\frac{(\mu_2^* - t_o)^2}{2/n}\right).$$

Como se podrá apreciar a continuación, el planteamiento del argumento fiducial a través del segundo método mencionado al inicio de esta sección, resulta ser un caso particular del método anterior, ya que se utiliza como pivotal a la función de distribución de la estadística suficiente minimal  $T$ .

Una vez observado el valor que tomó la estadística suficiente minimal  $T = t_o$ , la desigualdad en (5) se convierte en

$$\theta \leq F^{-1}(t_o, \gamma), \quad (19)$$

con probabilidad fiducial asociada igual a

$$p_{Fid}(\theta \leq F^{-1}(t_o, \gamma)) = \gamma. \quad (20)$$

Igual que se hizo en el caso anterior, el número  $F^{-1}(t_o, \gamma)$  se denotará por  $\theta^*$ , sin olvidar que éste está relacionado con la probabilidad  $\gamma$  en una *Uniforme* (0,1) y el valor fijo  $t_o$ , como se puede apreciar en la siguiente ecuación

$$\theta^* = F^{-1}(t_o, \gamma) \iff F(t_o, \theta^*) = \gamma. \quad (21)$$

Así, reescribiendo (20) en términos de  $\theta^*$  y por (21), se define la función de distribución fiducial como

$$F_{Fid}(\theta^*) = p_{Fid}(\theta \leq \theta^*) = F(t_o, \theta^*) = \gamma. \quad (22)$$

Para obtener la función de densidad fiducial es necesario pedir adicionalmente que  $F(t, \cdot)$  sea una función diferenciable con respecto a  $\theta$ , para cada  $t$ ; así se obtiene

$$f_{Fid}(\theta^*) = \frac{dF_{Fid}(\theta^*)}{d\theta^*} = \frac{\partial F(t_o, \theta^*)}{\partial \theta^*}, \quad (23)$$

donde  $F$  en este caso es la función de distribución continua de la estadística suficiente minimal.

Suponiendo que  $F(t, \theta)$  fuese una función continua monótona decreciente en  $\theta$ , para cada  $t$ , entonces (22) se escribirá como

$$F_{Fid}(\theta^*) = 1 - p_{Fid}(\theta \leq \theta^*) = 1 - F(t_o, \theta^*) = \gamma \quad (24)$$

y así, la densidad fiducial resulta ser

$$f_{Fid}(\theta^*) = \frac{dF_{Fid}(\theta^*)}{d\theta^*} = -\frac{\partial F(t_o, \theta^*)}{\partial \theta^*}. \quad (25)$$

Por (23) y (25), es común encontrar en la literatura las funciones de densidad y distribución fiducial como

$$f_{Fid}(\theta^*) = \left| \frac{\partial F(t_o, \theta^*)}{\partial \theta^*} \right| \quad (26)$$

y

$$F_{Fid}(\theta^*) = \begin{cases} F(t_o, \theta^*) & \text{si } F \text{ es creciente en } \theta \\ 1 - F(t_o, \theta^*) & \text{si } F \text{ es decreciente en } \theta. \end{cases} \quad (27)$$

Para ilustrar la aplicación del método anterior, se dan los siguientes dos ejemplos; el primero es el ejemplo 3.1 y el segundo el 2.5

### Ejemplo 3.3

Suponga que en lugar de haber elegido como cantidad pivotal a  $\theta T$ , como se hizo en el ejemplo 3.1, se elige a

$$U(T, \theta) = F(t, \theta) \sim \text{Uniforme}(0, 1),$$

donde la función de distribución  $F(t, \theta)$  de  $U$ , es una *Gamma* ( $\alpha = n, \beta = \frac{1}{\theta}$ ).

Una vez observada la muestra, es decir que  $T = t_o$ , la función de distribución fiducial para  $\theta$  aplicando (22), es igual a

$$F_{Fid}(\theta^*) = p(\theta \leq \theta^*) = F(t_o, \theta^*) = \gamma,$$

donde

$$F(t_o, \theta^*) = \int_0^{t_o} \frac{1}{\Gamma(n)} \theta^{*n} \exp(-\theta^* z) z^{n-1} dz,$$

y la densidad fiducial aplicando (23) resulta ser

$$f_{Fid}(\theta^*) = \frac{\partial F(t_o, \theta^*)}{\partial \theta^*} = \frac{\partial}{\partial \theta} \int_0^{t_o} \frac{\theta^{*n}}{\Gamma(n)} \exp(-\theta^* z) z^{n-1} dz.$$

Para derivar  $F(t_o, \theta^*)$  se puede hacer el cambio de variable  $v = \theta^* z$  en la integral, reduciendo el cálculo de la función de densidad fiducial al del ejemplo 3.1. Alternativamente, se puede intercambiar el orden de integración y diferenciación. Por

cualquiera de los dos caminos, se llega al mismo resultado del ejemplo 3.1, es decir que, dado  $T = t_0$ ,  $\theta$  tiene una distribución fiducial *Gamma* ( $\alpha = n, \beta = \frac{1}{t_0}$ ).

### Ejemplo 3.4

Sean  $X_1, \dots, X_n$  variables aleatorias i. i. d. *Uniformes*( $0, \theta$ ). En el Capítulo 2, se obtuvo para este problema, el intervalo de confianza más pequeño para  $\theta$  de nivel  $1 - \alpha = \gamma$  que resultó ser igual a (ver 2.37)

$$\theta \in \left( T, \frac{T}{\alpha^{1/n}} \right).$$

Aquí se desea encontrar el intervalo fiducial para  $\theta$  aplicando el planteamiento anterior. La estadística suficiente minimal en este caso es igual a

$$T(\underline{X}) = \max X_i = X_{(n)},$$

con función de distribución continua y monótona decreciente en  $\theta$  para cada  $t$ ,

$$F(t, \theta) = \begin{cases} 0 & \text{si } t < 0 \\ \left(\frac{t}{\theta}\right)^n & \text{si } 0 \leq t \leq \theta \\ 1 & \text{si } t > \theta \end{cases}$$

y función de densidad

$$f(t, \theta) = \begin{cases} n \left(\frac{t}{\theta}\right)^{n-1} \cdot \frac{1}{\theta} & \text{si } 0 \leq t < \theta \\ 0 & \text{en otro caso.} \end{cases}$$

La cantidad pivotal aplicando (2) sería igual a

$$U(T, \theta) = F(T, \theta) = \left(\frac{T}{\theta}\right)^n \sim \text{Uniforme}(0, 1).$$

Una vez observado el valor  $T = t_0$ , la distribución fiducial para  $\theta$ , aplicando (24), es

$$F_{Fid}(\theta^*) = 1 - F(t_0, \theta^*) = \begin{cases} 0 & \text{si } \theta \leq t_0 \\ 1 - \left(\frac{t_0}{\theta^*}\right)^n & \text{si } t_0 < \theta^* \end{cases}$$

y la densidad fiducial para  $\theta$ , por (25), resulta ser

$$f_{Fid}(\theta^*) = \frac{dF_{Fid}(\theta^*)}{d\theta^*} = -\frac{\partial F(t_0, \theta^*)}{\partial \theta^*} = \begin{cases} 0 & \text{si } \theta^* \in (0, t_0] \\ \frac{n}{\theta^*} \left(\frac{t_0}{\theta^*}\right)^n & \text{si } \theta^* > t_0. \end{cases}$$

Para obtener el intervalo fiducial para  $\theta$  de probabilidad fiducial igual a  $\gamma$  se necesita encontrar  $\theta_1^*$ , tal que

$$P_{Fid}(\theta \leq \theta_1^*) = F_{Fid}(\theta_1^*) = \gamma = 1 - \left(\frac{t_o}{\theta_1^*}\right)^n,$$

de donde

$$\frac{t_o^n}{\theta_1^{*n}} = 1 - \gamma = \alpha \implies \theta_1^* = \frac{t_o}{\alpha^{1/n}},$$

por lo que se dirá que, para la muestra observada, que produjo el valor  $T = t_o$ ,

$$\theta \in \left(t_o, \frac{t_o}{\alpha^{1/n}}\right),$$

con una probabilidad fiducial  $\gamma = 1 - \alpha$ .

Para este ejemplo coinciden el intervalo de confianza más pequeño de Neyman y el intervalo fiducial para  $\theta$ . Obsérvese que este último es más rápido de obtener, ya que no requiere minimizar la longitud del intervalo de probabilidad de la pivotal.

En los ejemplos anteriores, fue siempre posible resumir toda la información proporcionada por la muestra completa sobre el parámetro  $\theta$  en una estadística suficiente minimal  $T$  de dimensión igual a la del parámetro; sin embargo, no siempre es posible encontrar una estadística tal, aunque puede suceder que dicha estadística esté en correspondencia 1:1 con  $(R, S)$ , donde  $R$ , una estadística uni-dimensional y  $S$ , una estadística ancilaria (i.e., que su distribución no depende de  $\theta$ ), son tales que la función de verosimilitud se puede expresar como

$$L(\theta; \underline{x}) = f(\underline{x}; \theta) = f(R, S; \theta) f(\underline{x}/R, S), \quad (28)$$

donde la distribución condicional de  $\underline{x}$  dado  $R$  y  $S$  no depende de  $\theta$ , por lo que

$$L(\theta; \underline{x}) \propto f(R, S; \theta) = f(R; \theta | S) \cdot f(S). \quad (29)$$

Como  $S$  es ancilaria, la verosimilitud resulta proporcional a

$$L(\theta; \underline{x}) \propto f(R; \theta | S) \quad (30)$$

y así, uno puede sin pérdida de información, basar las inferencias sobre  $\theta$  en la distribución condicional de  $R$  dado  $S$ , en este caso se dice que  $R$  es una estadística exhaustiva o condicionalmente suficiente.

Si la distribución condicional  $F(R; \theta | S)$  de  $R$ , es continua en  $R$  y monótona en  $\theta$  para cada posible valor de  $R$  y el valor fijo y observado de  $S$ , entonces, se puede encontrar la distribución fiducial de  $\theta$  como en (27) si, adicionalmente  $F(R; \theta | S)$  es una función diferenciable con respecto a  $\theta$  para cada  $R$  y el valor fijo de  $S$ , entonces la función de densidad fiducial se puede encontrar como en (26) (ver Sprott 1967, Pederzen 1978).

Otro punto importante por mencionar es que, una vez que se ha obtenido la función de distribución fiducial de un parámetro  $\theta$ , ésta se puede utilizar para obtener la función de distribución fiducial de cualquier función 1:1 de  $\theta$ ,  $\eta = g(\theta)$ , por lo que ambas  $\theta$  y  $\eta$  juegan el papel lógico de variables aleatorias. Sin embargo, cuando  $\eta$  no es una función monótona de  $\theta$  esto no se puede hacer como sucede en el caso clásico, es más, Barnard y Sprott (1983), demuestran que la función de distribución fiducial de  $\eta$  no existe, cuando  $\theta$  varía en el eje real. Así, si  $\eta = \theta^2$ , entonces desde el punto de vista fiducial,  $\theta^2$  no puede jugar el papel lógico de variable aleatoria. Este hecho contradice la definición de Kolmogorov de que cualquier función medible de una variable aleatoria es variable aleatoria, por lo que las llamadas probabilidades fiduciales no se pueden ver como probabilidades en el sentido de Kolmogorov. Pederzen (1978) demuestra que no se puede obtener la distribución fiducial de  $\mu^2$  a partir de la de  $\mu$  en una  $N(\mu, 1)$  y Barnard (1987) lo demuestra para un caso discutido por Laplace (Stigler, 1986).

Como se pudo observar de lo expuesto anteriormente, en el argumento fiducial  $\theta$  juega el papel lógico de una variable aleatoria, sin ser variable aleatoria y  $T$  llega a jugar el papel lógico de un parámetro que define a la distribución fiducial. Este tipo de razonamiento iniciado por Fisher ha creado grandes controversias entre los estadísticos clásicos y Bayesianos, ya que los primeros, niegan el que se puedan hacer aseveraciones probabilísticas sobre una constante  $\theta$  del tipo (10), a menos que dicha probabilidad sea 1 ó 0 de que  $\theta^*$  sea mayor o igual a  $\theta$ ; los segundos por otro lado, aseguran que siempre es posible obtener una distribución *a-priori* de  $\theta$ , por lo que resulta más conveniente utilizar el procedimiento de Bayes.

### 3.3 Fiducial Implícito

En la sección anterior quedó establecido que para obtener los intervalos fiduciales de un parámetro escalar  $\theta$ , era necesario encontrar de manera explícita la función de distribución fiducial de  $\theta$ . Esto se logró, en general, con una cantidad pivotal exacta, función de la estadística suficiente minimal cuya distribución fuese conocida, o bien, cuando no se tenía a la mano una cantidad pivotal exacta, con la estadística suficiente minimal  $T$  y su función de distribución  $F(t; \theta)$ , la cual se supone continua. Desafortunadamente, cuando la distribución de la estadística suficiente minimal posee una expresión analítica complicada, no se puede obtener de manera sencilla la distribución fiducial de  $\theta$ .

En esta sección se presenta un método al que se refiere como **fiducial implícito**, que viene a ser otra versión del fiducial, y que permite obtener intervalos fiduciales para el parámetro  $\theta$  sin necesidad de exhibir de manera explícita la función de distribución fiducial de  $\theta$  (aunque esta última también se podría obtener).

La aplicación de este método resulta muy conveniente en los casos en que la función de distribución de la estadística suficiente minimal, posea una expresión analítica muy complicada, evitándose el algebra necesaria al derivar dicha función para encontrar la densidad fiducial de  $\theta$  y con ella, los intervalos fiduciales para  $\theta$ .

Para la construcción del método fiducial implícito se necesita de una estadística suficiente minimal y uni-dimensional  $T$  con función de distribución  $F(t; \theta)$ , que sea continua en  $t$  y monótona en  $\theta$ ; con éllo y para cada  $\gamma = 1 - \alpha$  en el intervalo  $(0,1)$  existe  $t_\gamma(\theta)$  tal que

$$p(T \leq t_\gamma(\theta); \theta) = F(t_\gamma(\theta), \theta) = \gamma, \quad (31)$$

donde  $t_\gamma(\theta)$  es el percentil  $\gamma$  de  $F(t, \theta)$ .

Lo primero es construir un conjunto de valores de  $\theta$ , para lo cual se supone por el momento, que  $F(t, \cdot)$  es creciente en  $\theta$  para cada posible valor de  $t$ , y que la muestra ya ha sido observada, es decir que  $T = t_o$ . Este conjunto de valores de  $\theta$  se define como

$$A(t_o) = \{\theta; t_o \leq t_\gamma(\theta)\} \quad (32)$$

y suponga que

$$\theta_1 \in A(t_0), \quad \text{o sea que} \quad t_0 \leq t_\gamma(\theta_1), \quad (33)$$

donde  $\theta_1$  es un posible valor del parámetro  $\theta$ , por lo que

$$F(t_0, \theta_1) \leq F(t_\gamma(\theta_1), \theta_1) = \gamma. \quad (34)$$

Sea  $\theta_2 < \theta_1$ ; como  $F(t, \theta)$  es creciente en  $\theta$  se tiene que

$$F(t_0, \theta_2) \leq F(t_0, \theta_1), \quad (35)$$

y de (34) y (35) se implica que

$$F(t_0, \theta_2) \leq F(t_\gamma(\theta_2), \theta_2) = \gamma. \quad (36)$$

Esta última desigualdad implica que

$$t_0 \leq t_\gamma(\theta_2) \quad \text{y por lo tanto} \quad \theta_2 \in A(t_0). \quad (37)$$

Así el conjunto  $A(t_0)$  de valores de  $\theta$  en (32), tiene la forma del intervalo

$$(-\infty, \theta^*], \quad (38)$$

donde  $\theta^*$  es tal que

$$t_0 = t_\gamma(\theta^*). \quad (39)$$

El conjunto  $A(t_0)$  en (32) que tiene la forma en (38) es el intervalo fiducial de  $\theta$  de probabilidad fiducial  $\gamma$ , para  $T = t_0$ , o sea que

$$p_{Fid}(\theta \in (-\infty, \theta^*]) = \gamma. \quad (40)$$

Para obtener la distribución fiducial de  $\theta$  obsérvese que

$$\theta \in A(t_0) \iff \theta \in (-\infty, \theta^*] \iff t_0 \in (-\infty, t_\gamma(\theta^*)], \quad (41)$$

y que en la obtención del valor  $\theta^*$ ,  $t_0$  está fijo y  $\gamma$  fue elegido en  $(0,1)$ , así para cada valor de  $\gamma$  existirá  $\theta^*(\gamma)$  que satisfaga (40). Al observar  $\gamma$  como función de  $\theta^*$  a través de (39) se puede escribir a (40) como

$$p_{Fid}(\theta \in (-\infty, \theta^*]) = \gamma(\theta^*), \quad (42)$$

y al hacer dicha descripción variando  $\theta^*$  los valores correspondientes de  $\gamma(\theta^*)$  serán los de la función de distribución fiducial de  $\theta$ , es decir

$$F_{Fid}(\theta^*) = \gamma(\theta^*). \quad (43)$$

En caso de que  $F(t, \theta)$  sea decreciente en  $\theta$  para cada valor de  $t$ , el conjunto de valores  $A(t_0)$  de  $\theta$  en (32) dará como resultado un intervalo fiducial para  $\theta$  de la forma

$$[\theta^*, \infty), \quad (44)$$

con  $\theta^*$  satisfaciendo (39).

Se enfatiza que este método implícito permite obtener intervalos fiduciales para el parámetro  $\theta$ , sin que sea necesario obtener la distribución fiducial de  $\theta$  explícitamente, como era el caso en la sección anterior.

Aquí se ha ilustrado el método para obtener intervalos de la forma (38) y (44), la obtención de intervalos centrales de la forma  $(\theta_1^*, \theta_2^*)$  es inmediata.

### Ejemplo 3.5

Sean  $X_1, \dots, X_n$  variables aleatorias *i.i.d.* Uniformes  $(0, \theta)$ . En el ejemplo 3.4 se obtuvo que la estadística suficiente minimal  $T$  era igual a

$$T(\underline{X}) = X_{(n)},$$

con funciones de densidad y distribución, respectivamente

$$f(t, \theta) = \begin{cases} \frac{n}{\theta^n} t^{n-1} & \text{si } 0 < t \leq \theta \\ 0 & \text{en otro caso,} \end{cases}$$

$$F(t, \theta) = \begin{cases} 0 & \text{si } t < 0 \\ \left(\frac{t}{\theta}\right)^n & \text{si } 0 \leq t \leq \theta \\ 1 & \text{si } t > \theta. \end{cases}$$

Con  $f(t, \theta)$  una función creciente en  $t$  para cada posible valor de  $\theta$  y  $F(t, \theta)$  una función continua de  $t$  y decreciente en  $\theta$  para cada posible de valor en  $t$ .

con  $c(n)$ , una constante y  $K_{n/2}(\cdot)$ , una función de *Bessel* de "segundo tipo" de orden  $n/2$ .

La evaluación de la densidad resultó factible solamente desde el punto de vista numérico y no fue posible obtener una cantidad pivotal exacta; por lo que se utilizó el procedimiento referido como fiducial implícito, para obtener intervalos de la forma

$$[\theta^*, \infty),$$

de probabilidad fiducial  $\gamma$ .

### 3.4 Método Bayesiano

Dentro de este enfoque el parámetro escalar desconocido  $\theta$  es tratado como una variable aleatoria, por lo que se supone que, antes de realizar un experimento para obtener información a través de una muestra  $\underline{x}$ , éste posee una distribución inicial  $\pi(\theta)$ , comúnmente conocida como distribución *a priori*. Esta distribución juega un papel importante dentro del análisis Bayesiano, y representa el estado de conocimiento sobre el parámetro desconocido  $\theta$ , antes de que se observe a  $\underline{x}$ . Para los estadísticos Bayesianos esta probabilidad inicial siempre existe, ya que se rigen bajo el supuesto de que la probabilidad es una expresión matemática de nuestro grado de credibilidad con respecto a cierta proposición (ver Ramsey 1926, De Finetti 1937, Savage 1954, 1961a, b, 1962 y Edwards, Lindman y Savage 1963).

Dado el valor de  $\theta$ , se considera a la densidad condicional de  $\underline{x}$  como

$$f(\underline{x} | \theta) = f(\underline{x}; \theta) \text{ (en la notación clásica),} \quad (45)$$

y una vez observado  $\underline{x}$ , la distribución condicional de  $\theta$  aplicandó el teorema de Bayes, resulta ser

$$h(\theta | \underline{x}) = C(\underline{x}) f(\underline{x} | \theta) \pi(\theta). \quad (46)$$

donde

$$C^{-1}(\underline{x}) = \begin{cases} \sum f(\underline{x} | \theta) \pi(\theta) & \text{si } \theta \text{ discreta,} \\ \int f(\underline{x} | \theta) \pi(\theta) d\theta & \text{si } \theta \text{ continua,} \end{cases} \quad (47)$$

es una constante normalizadora que asegura que la probabilidad total en (46) sea igual a 1.

A la densidad en (46) se le conoce como la distribución final o *posterior* de  $\theta$ , y representa el estado de conocimiento sobre  $\theta$  después de que se incorpora al conocimiento inicial la información proporcionada por la muestra  $\underline{x}$ .

Dado  $\underline{x}$ ,  $f(\underline{x} | \theta)$  en (46), visto como función de  $\theta$ , es proporcional a la función de verosimilitud de  $\theta$ ,  $l(\theta | \underline{x})$ . Esta función juega un papel muy importante en la fórmula de Bayes, ya que a través de ella la muestra  $\underline{x}$  puede modificar el conocimiento inicial sobre  $\theta$  representado a través de  $\pi(\theta)$ . Combinando ambas fuentes de información sobre el parámetro desconocido se describe la distribución *posterior* de  $\theta$  como

$$h(\theta | \underline{x}) \propto l(\theta | \underline{x}) \pi(\theta). \quad (48)$$

Resumiendo, aplicando el Teorema de Bayes cuando  $\pi(\theta)$  es conocida se encuentra una distribución de probabilidad para  $\theta$ , en la que se incorpora la información proporcionada por  $\underline{x}$ , y desde el punto de vista Bayesiano todas las inferencias sobre  $\theta$  se deben extraer de esta distribución final o *posterior*.

En el contexto Bayesiano no existe el concepto de intervalos de confianza para  $\theta$  de nivel  $1 - \alpha$ , sino más bien, se maneja el concepto de intervalos de probabilidad *posterior*  $1 - \alpha$ , los cuales pueden poseer ciertas propiedades óptimas como:

a) que la densidad *posterior* para cada valor de  $\theta$  dentro del intervalo sea al menos tan grande como la de los valores fuera del intervalo.

b) que para una probabilidad dada  $\alpha$  el intervalo contenga el menor número posible de valores de  $\theta$ , dicho de otra manera, que el intervalo sea lo más pequeño posible.

Un poco de reflexión hace obvio el hecho de que ambas propiedades son equivalentes. Así, si el intervalo de probabilidad *posterior* para  $\theta$  posee alguna de las propiedades anteriores entonces a éste se le conoce como el intervalo de máxima densidad *posterior* comúnmente abreviado como hpd "highest posterior density".

De manera un poco más formal, se puede decir que dada la distribución *posterior*  $h(\theta | \underline{x})$  de  $\theta$ , un intervalo en el espacio paramétrico  $\Theta$  se llamará de máxima densidad

posterior y contenido  $(1 - \alpha)$  si

$$\int_{\theta_1}^{\theta_2} h(\theta | \underline{x}) d\theta = 1 - \alpha, \quad (49)$$

y para  $\theta'$  dentro del intervalo y  $\theta''$  fuera del intervalo, se tiene que

$$h(\theta' | \underline{x}) \geq h(\theta'' | \underline{x}). \quad (50)$$

La interpretación que se les da a estos intervalos de probabilidad es que  $\theta$  se encuentra entre  $\theta_1$  y  $\theta_2$  con una probabilidad (en el sentido de Kolmogorov) igual a  $1 - \alpha$ .

Hasta ahora se ha supuesto que el conocimiento inicial sobre  $\theta$  se puede expresar a través de una distribución inicial  $\pi(\theta)$ . Este se pudo haber obtenido por experiencia previa y expresado en términos de frecuencias (ver Fisher 1959), o bien pudo haber sido expresado basando los argumentos en la teoría de decisión Bayesiana (ver Schlaifer 1959, Raiffa y Schlaifer 1961). En la mayoría de las aplicaciones, sin embargo, se desconoce cómo expresar este conocimiento inicial sobre  $\theta$  y es precisamente éste el punto crítico del método Bayesiano, donde el uso del teorema de Bayes ha llegado a ser controversial.

Siendo el objetivo de este trabajo el dar un panorama global de los distintos enfoques para resolver el problema de estimación por intervalos y, dado que resultaría imposible presentar una discusión de las distintas formas que pueden existir para expresar el conocimiento inicial a través de una distribución *a priori*, se hace a continuación una mención breve de algunos de los métodos para encontrarla.

Las llamadas distribuciones *a priori* conjugadas surgen como una respuesta de los estadísticos Bayesianos al caso en que  $\pi(\theta)$  y  $l(\underline{x} | \theta)$  poseen expresiones matemáticas complicadas, haciendo difícil la aplicación del teorema de Bayes para obtener la distribución *posterior* del parámetro  $\theta$ . Así, para cualquier modelo y función de verosimilitud en particular, se intenta determinar una familia conjugada de distribuciones *a priori*, cada una de las cuales puede ser combinada con  $l(\underline{x} | \theta)$  sin dificultad.

En general se pide que la familia conjugada de distribuciones, posea propiedades tales como que sea matemáticamente tratable, en el sentido descrito anteriormente; que sea rica, en el sentido de que incluya, lo más que sea posible,

distribuciones con diferentes localizaciones, varianzas y formas para que puedan representar diferentes estados iniciales de información, y por último, que sean fáciles de interpretar.

Las familias conjugadas dan una amplia variedad de modelos de  $\pi(\theta)$ , en los que se espera exista uno que describa el conocimiento inicial sobre  $\theta$ . Existen ocasiones, sin embargo, en las que no se quiere expresar dicho conocimiento inicial a través de alguno de estos modelos, o bien no se tiene un conocimiento inicial muy definido sobre  $\theta$ , para lo cual, los estadísticos Bayesianos han desarrollado las llamadas distribuciones *a priori* difusas, no-informativas o de referencia, en donde, el problema es seleccionar una  $\pi(\theta)$  que proporcione poca información relativa a la información que se espera de la muestra  $\underline{x}$  sobre el parámetro  $\theta$ .

Los trabajos desarrollados en este campo se pueden dividir, a grosso modo, en tres grupos. Existen quienes buscan una transformación uno a uno  $\phi(\theta)$  del parámetro  $\theta$ , cuya verosimilitud sea aproximadamente conocida *a priori* excepto por su localización, la cual será determinada por la muestra  $\underline{x}$ . Comúnmente se le llama a esta verosimilitud "trasladada por los datos". Esta misma idea es equivalente a pedir que la distribución *a priori* para  $\phi(\theta)$  sea localmente *Uniforme* (ver Jeffreys 1961 y Box y Tiao 1973). Dentro de este mismo grupo, están los que además encuentran a  $\pi(\theta)$  basados en principios de invarianza, es decir, encontrar una *a priori* que sea invariante ante transformaciones 1:1 del parámetro (ver Hartigan 1964, Jeffreys 1946, Perks 1947 y Villegas 1971, 1977a, b).

Otros enfoques se basan en el uso de formas límite de familias conjugadas (ver DeGroot 1970, Novick y Hall 1965 y Novick 1969).

Finalmente, existe el grupo de Bayesianos quienes modelan el conocimiento inicial mediante una *a priori* seleccionada de acuerdo a ciertas medidas de información (ver Bernardo 1979, Jaynes 1968, Lindley 1961 y Zellner 1977).

Es importante mencionar, que en ocasiones cuando se aplican algunos de los enfoques antes mencionados para obtener la distribución inicial  $\pi(\theta)$ , ésta resulta ser *impropia*, en el sentido que

$$\begin{aligned} \sum \pi(\theta) &= \infty && \text{si } \theta \text{ discreta,} \\ \int \pi(\theta) d\theta &= \infty && \text{si } \theta \text{ continua.} \end{aligned} \quad (51)$$

Sin embargo, siempre que la distribución *posterior*  $h(\theta | \underline{x})$  resulte propia no habrá ningún problema en utilizar distribuciones *a priori* impropias.

Los intervalos de máxima densidad *posterior* y contenido  $1 - \alpha$  obtenidos a partir de estas distribuciones *a priori* no-informativas, en muchas ocasiones coinciden numéricamente con los intervalos de confianza más pequeños para  $\theta$  de nivel  $1 - \alpha$  (ver Lindley 1958, Welch y Peers 1963).

### Ejemplo 3.7

Retomando el ejemplo 2.5, en el que se tenía una muestra aleatoria  $X_1, \dots, X_n$  de una distribución *Uniforme*(0,  $\theta$ ). Suponga que se da la distribución *a priori* propia para  $\theta$

$$\pi_1(\theta) = 2\theta, \quad \text{con } \theta \in (0, 1).$$

Dado que la verosimilitud de  $\theta$  es proporcional a la función de densidad de la estadística suficiente minimal, se puede obtener la distribución *posterior* de  $\theta$  como

$$h(\theta | \underline{x}) = C(\underline{x})g^*(t | \theta)\pi_1(\theta),$$

donde  $g^*(t | \theta)$  es la densidad de la estadística suficiente minimal  $T(\underline{X}) = X_{(n)}$ . Del ejemplo 2.10, en que se obtuvo la función de distribución de esta estadística, se tiene que

$$g^*(t | \theta) = \begin{cases} \frac{nt^{n-1}}{\theta^n} & \text{si } t \in (0, \theta) \\ 0 & \text{en otro caso,} \end{cases}$$

por lo que la distribución *posterior* de  $\theta$  sería proporcional a

$$g^*(t | \theta)\pi_1(\theta) = \begin{cases} \frac{2nt^{n-1}}{\theta^{n-1}} & \text{si } \theta \in (t, 1) \\ 0 & \text{en otro caso,} \end{cases}$$

La constante normalizadora  $C(\underline{x})$ , se encuentra por (47) como

$$\begin{aligned} C^{-1}(\underline{x}) &= \int_t^1 g^*(t | \theta)\pi_1(\theta) = \int_t^1 \frac{2nt^{n-1}}{\theta^{n-1}} d\theta \\ &= \frac{2nt}{n-2} (1 - t^{n-2}). \end{aligned}$$

Por lo que la densidad *posterior* de  $\theta$  es igual a

$$h(\theta | t = x_{(n)}) = \begin{cases} \frac{(n-2)t^{n-2}}{(1-t^{n-2})\theta^{n-1}} & \text{si } \theta \in (t, 1) \\ 0 & \text{en otro caso.} \end{cases}$$

Para encontrar el intervalo de máxima densidad *posterior* y contenido  $1 - \alpha$ , es necesario encontrar la constante  $k_\alpha$ , tal que

$$\int_t^{k_\alpha} h(\theta | t = x_{(n)}) d\theta = 1 - \alpha.$$

Esta constante resulta ser igual a

$$k_\alpha = t \left( t^{n-2}(1 - \alpha) + \alpha \right)^{-\frac{1}{n-2}},$$

por lo que el intervalo hpd de contenido  $1 - \alpha$  para  $\theta$  es

$$\left( t \leq \theta \leq t \left( t^{n-2}(1 - \alpha) + \alpha \right)^{-\frac{1}{n-2}} \right).$$

Suponga ahora, que se da la distribución *a priori* no-informativa impropia

$$\pi_2(\theta) = \frac{1}{\theta}, \quad \text{con } \theta \in (0, \infty).$$

Con esta *a priori* la densidad *posterior* de  $\theta$  resulta ser proporcional a

$$g^*(t | \theta)\pi_2(\theta) = \begin{cases} \frac{nt^{n-1}}{\theta^{n+1}} & \text{si } \theta \in (t, \infty) \\ 0 & \text{en otro caso,} \end{cases}$$

y la constante normalizadora  $C(\underline{x})$  en este caso, es

$$C^{-1}(\underline{x}) = \int_t^\infty g^*(t | \theta)\pi_2(\theta) d\theta = \int_t^\infty \frac{nt^{n-1}}{\theta^{n+1}} d\theta = \frac{1}{t},$$

por lo que la densidad *posterior* de  $\theta$  es igual a

$$h(\theta | t = x_{(n)}) = \begin{cases} \frac{nt^n}{\theta^{n+1}} & \text{si } \theta \in (t, \infty) \\ 0 & \text{en otro caso.} \end{cases}$$

En este caso la constante  $k_\alpha$  necesaria para obtener el intervalo de máxima densidad *posterior* y contenido  $1 - \alpha$  para  $\theta$ , es igual a

$$k_\alpha = \frac{t}{\alpha^{1/n}},$$

por lo que el intervalo resulta ser

$$\left( t \leq \theta \leq \frac{t}{\alpha^{1/n}} \right).$$

Observe que este intervalo es numéricamente igual al intervalo de confianza más pequeño de nivel  $1 - \alpha$  basado en  $T(\underline{X}) = X_{(n)}$ , en el ejemplo 2.10, así como a los intervalos fiduciales obtenidos en los ejemplos 3.4 y 3.5. Aunque aquí la interpretación que se le da, es que el parámetro  $\theta$  se encuentra entre  $t$  y  $t\alpha^{-1/n}$  con una probabilidad  $1 - \alpha$ .

## CONCLUSIONES

Resulta sorprendente constatar que en la actualidad, aún existan controversias respecto a la forma en que debe ser abordado el problema de hacer inferencias sobre un parámetro escalar. Alrededor del método frecuentista tradicional se ha suscitado una serie de críticas por parte de los estadísticos Bayesianos y fiduciales, quienes opinan que en ocasiones se llega a resultados erróneos para la muestra en particular que se observó. El método Bayesiano también ha sido criticado por los estadísticos fiduciales, por el hecho de requerir de una distribución inicial del parámetro, ya que opinan que no siempre es posible obtenerla, así como por los estadísticos clásicos quienes opinan que no es posible dar una distribución de probabilidad a un parámetro que es fijo. Finalmente, el más criticado ha sido el método fiducial, no sólo por los frecuentistas, quienes opinan que no se puede dar una distribución de probabilidad a un parámetro fijo, y por los Bayesianos, quienes opinan que no es posible dar una distribución final del parámetro sin haber utilizado la información proporcionada por la distribución inicial, sino también dentro de los mismos fiduciales existen diferencias para descifrar, esclarecer y formalizar lo que Fisher ya no pudo decir respecto a su argumento fiducial.

A pesar de lo mencionado en el párrafo anterior, resulta reconfortante observar que, al aplicar la filosofía correcta en cada uno de los enfoques vistos a través de este trabajo, se llega en ocasiones a resultados numéricos similares para un mismo problema, aunque las interpretaciones que se les da, marquen la diferencia.

Finalmente, creemos necesario mencionar una serie de dudas que surgieron a lo largo de este trabajo y que se espera puedan motivar futuros estudios. Dentro del

enfoque frecuentista, ¿qué hacer cuando se aplica teoría asintótica a un problema donde no es factible reparametrizar para obtener resultados más precisos?. Dentro del método fiducial explícito, ¿cómo se puede garantizar la unicidad de la distribución fiducial cuando se escogen diferentes cantidades pivotaes, todas funciones de la estadística suficiente minimal?. Dentro del método fiducial implícito, ¿cómo se pueden obtener los intervalos fiduciales de transformaciones 1:1 del parámetro, a partir de los intervalos del parámetro?. Y finalmente, ¿cuál sería la generalización de todos estos métodos en la presencia de parámetros de ruido?.

Como ya se mencionó en la introducción, se espera que las dudas anteriores y muchas otras alrededor del problema tratado, se exploren en trabajos futuros, y que este trabajo haya cumplido sus objetivos de informar y motivar para que exista un mayor interés en el área.

## REFERENCIAS

1. ANSCOMBE, F. J. (1964). *Normal Likelihood Functions*. Ann. Inst. Statist. Math. (Tokyo) 16, 1-19.
2. BARNARD, G. A. (1963a). *Some Logical Aspects of the Fiducial Argument*. J. R. Statist. Soc. B 25, 111-114.
3. BARNARD, G. A. (1963b). *Logical Aspects of the Fiducial Argument*. Bull. Int. Statist. Inst. 40 (2), 870-883.
4. BARNARD, G. A. (1973). *Maximum Likelihood y Nuisance Parameters*. Sankhya, A 35, 133-138.
5. BARNARD, G. A. y SPROTT, D. A. (1983). *The Generalised Problem of the Nile: Robust Confidence Sets for Parametric Functions*. Ann. Statist. 11, 104-113.
6. BARNARD, G. A. (1987). *R. A. Fisher — a true Bayesian?* Int. Statist. Rev. 55, 2, 183-189.
7. BERNARDO, J. M. (1979). *Reference Posterior Distributions for Bayesian Inference*. J. Roy. Statist. Soc. Ser. B 41, 113-147.
8. BIRNBAUM, A. (1962). *On the Foundations of Statistical Inference*. J. Amer. Statist. Assoc. 57, 269-326.
9. BOX, G. E. P. y TIAO, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Addison-Wesley.
10. COX, D. R. y HINKLEY, D. V. (1974). *Theoretical Statistics*. New York, Wiley.
11. DE FINETTI, B. (1937). *“La Prévicion: Ses Lois Logiques, Ses Sources Subjectives”*. Ann. Inst. H. Poincaré, 7, 1-68. Traducido al inglés en H. E. KYBURG, JR. y H. G. SMOKLER EDS. (1964). *Studies in Subjective Probability*. New York, Wiley.
12. DEGROOT, M. H. (1970). *Optimal Statistical Decisions*. New York, Mc Graw Hill.
13. DEMPSTER, A. P. (1963). *On Direct Probabilities*. J. R. Statist. Soc. B 25, 100-110.
14. EDWARDS, W., LINDMAN, H. y SAVAGE, L. J. (1963). *Bayesian Statistical Inference for Psychological Research*. Psychological Rev. 70, 193-242.

15. FISHER, R. A. (1959). *Statistical Methods and Scientific Inference*. Second Edition, London, Oliver and Boyd.
16. FRASER, D. A. S. (1961a). *On Fiducial Inference*. Ann. Math. Statist. 32, 661-676.
17. FRASER, D. A. S. (1961b). *The Fiducial Method and Invariance*. Biometrika 48, 261-281.
18. HARTIGAN, J. A. (1964). *Invariant Prior Distributions*. Ann. Math. Statist. 35, 836.
19. JAYNES, E. T. (1968). "Prior Probabilities". IEEE Transactions on Systems Science and Cybernetics. SSC 4, 227.
20. JEFFREYS, H. (1946). *An Invariant Form for the Prior Probability in Estimation Problems*. Proc. R. Soc., London A 186, 453-461.
21. JEFFREYS, H. (1961). *Theory of Probability*. Third Edition, Oxford, Clarendon Press.
22. KENDALL, M. y STUART, A. (1979). *The Advanced Theory of Statistics*. Fourth Edition, Vol. 2, London, Griffin.
23. LINDLEY, D. V. (1958). *Fiducial Distributions and Bayes's Theorem*. J. R. Statist. Soc. B 20, 102-107.
24. LINDLEY, D. V. (1961). *The use of Prior Probability Distributions in Statistical Inference and Decisions*. Proc. Fourth Berkeley Symp. 1, 453-468.
25. MENDENHALL, W. y REINMUTH, J. E. (1978). *Statistics for Management and Economics*. Third Edition, Massachusetts, Duxbury Press.
26. MOOD, A. M., GRAYBILL, F. A. y BOES, D. C. (1974). *Introduction to the Theory of Statistics*. Third Edition, Mc Graw-Hill.
27. NEYMAN, J. (1942). *Fiducial Argument and the Theory of Confidence Intervals*. Biometrika, Vol. 82, 128-150.
28. NOVICK, M. R. y HALL, W. J. (1965). *A Bayesian Indifference Procedure*. J. Amer. Statist. Assoc. 60, 1104-1117.
29. NOVICK, M. R. (1969). *Multiparameter Bayesian Indifference Procedure*. J. Roy. Statist. Soc. B, 31, 29-64.
30. O'REILLY, F. y RUEDA, R. (1986). *Highest Posterior Density and Confidence Intervals for the Ratio of the Parameters of the Inverse Gaussian*. Presentado en 33<sup>d</sup> NBER-NSF Seminar on Bayesian Inference in Econometrics. Riverside, California.
31. PEDERZEN, J. G. (1978). *Fiducial Inference*. Int. Statist. Rev. 46, 147-170.
32. PERKS, F. J. A. (1947). *Some Observations on Inverse Probability, Including a New Indifference Rule*. J. Inst. Actuaries 73, 285.
33. RAMSEY, F. P. (1926). *Truth and Probability*. En "The Foundations of Mathematics and other Logical Essays". London: Regan, Paul, Trench, Trubner and Co. LTD. Reimpreso en *Studies in Subjective Probability*. H. E. Kyburg Jr. and H. E. Smokler Eds. New York, Wiley (1964).
34. RAO, C. R. (1973). *Linear Statistical Inference and Its Applications*. Second Edition. New York, Wiley.

35. SAMPSON, A. y SPENCER, B. (1976). *Sufficiency, Minimal Sufficiency, and the Lack Thereof*. Amer. Statistic. Vol. 30, 1, 7-8.
36. SAVAGE, L.J. (1954). *The Foundation of Statistics*. New York, Wiley.
37. SAVAGE, L. J. (1961a). *The Subject Basis of Statistical Practice*. Un manuscrito no publicado de la Universidad de Michigan.
38. SAVAGE, L. J. (1961b). *The Foundation of Statistics Reconsidered*. Proc. 4<sup>th</sup> Berkeley Symp. 1, 575.
39. SAVAGE, L. J. (1962). *The Foundation of Statistical Inference*. London, Methuen.
40. SPROTT, D. A. (1963). *A Transformation Model for the Investigation of Fiducial Distributions*. Bull. Int. Statist. Inst. 40, 2, 856-869.
41. SPROTT, D. A. (1966). *Statistical Estimation — Some Approaches and Controversies*. Statistische Hefte 6, 97-111.
42. SPROTT, D. A. y KALBFLEISCH, J. G. (1967). *Fiducial Probability*. Statistische Hefte 2, 99-109.
43. SPROTT, D. A. y KALBFLEISCH, J. G. (1969). *Examples of Likelihoods and Comparisons with Point Estimates and Large Sample Approximations*. J. Amer. Statist. Assoc. 64, 468-484.
44. SPROTT, D. A. (1973). *Normal Likelihoods and their Relation to Large Sample Theory of Estimation*. Biometrika 60, 457-465.
45. SPROTT, D. A. (1975). *Application of Maximum Likelihood to Finite Samples*. Sankhya Ser. B 37, 259-270.
46. SPROTT, D. A. (1984). *Verosimilitud y Estimación de Máxima Verosimilitud*. Versión revisada de un artículo publicado en inglés en Mathematical Reports of the Academy of Science, Royal Society of Canada. (1984). Vol. 6, 225-242.
47. SPROTT, D. A. y VIVEROS, R. (1984). *The Interpretation of Maximum Likelihood Estimation*. Canad. J. Statist. 12, 27-38.
48. STIGLER, G. J. (1986). *Laplace's 1774 Memoir on Inverse Probability*. Statist. Sci. 1, 150-164.
49. TUKEY, J. W. (1957). *Some Examples with Fiducial Relevance*. Ann. Math. Statist. 28, 687-695.
50. VILLEGAS, C. (1971). *On Haar Priors*. Foundation of Statistical Inference (Godambe and Sprott Eds.). 409-414. Toronto: Holt, Rine Hart and Winston.
51. VILLEGAS, C. (1977a). *Inner Statistical Inference*. J. Amer. Statist. Assoc. 72, 453-458.
52. VILLEGAS, C. (1977b). *On the Representation of Ignorance*. J. Amer. Statist. Assoc. 72, 651-654.
53. WELCH, B. L. y PEERS, H. W. (1963). *On Formulae for Confidence Points Based on Integrals of Weighted Likelihood*. J. Roy. Statist. Soc. B 25, 318.
54. ZELLNER, A. (1977). *Maximal Data Information Prior Distributions*. New Developments in the Applications of Bayesian Methods. (A. Aykac and C. Brumat Eds.).