ESTIMACIÓN DE VEROSIMILITUD MÁXIMA EN MODELOS DE TRASLACIÓN CON SOPORTE COMPACTO

JESÚS SALVADOR RUÍZ OLMOS

T E S I S

Presentada como requisito parcial
para obtener el grado de
Maestro en Ciencias
en Estadística Experimental

Universidad Autónoma Agraria
Antonio Narro

SUBDIRECCIÓN DE POSTGRADO

Buenavista, Saltillo,Coah.

Mayo de 2005

UNIVERSIDAD AUTÓNOMA AGRARIA ANTONIO NARRO

Subdirección de Postgrado

ESTIMACIÓN DE VEROSIMILITUD MÁXIMA EN MODELOS DE TRASLACIÓN CON SOPORTE COMPACTO

TESIS

Por

JESÚS SALVADOR RUÍZ OLMOS

Elaborada bajo la supervisión del comité particular de asesoría y aprobada como requisito parcial, para optar al grado de:

MAESTRO EN CIENCIAS

en Estadística Experimental

Comité Particular

Asesor principal: ______________________________

Dr. Rolando Cavazos Cadena

Asesor: ______________________________

M.C. Luis Rodríguez Gutiérrez

Asesor: ______________________________

Félix de Jesús Sánchez Pérez

____

Dr. Jerónimo Landeros Flores
Subdirector de Postgrado

Buenavista, Saltillo, Coahuila. Abril de 2005

AGRADECIMIENTOS

- A Dios, por la oportunidad de vivir y poder contemplar la belleza de su creación.

- A mis padres, que con su apoyo y ejemplo me han motivado a ser lo que ahora soy.

- Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por apoyarme en mis estudios de postgrado.

- A la Universidad Autónoma Agraria Antonio Narro, por permitir la difusión del conocimiento.

- Al Dr. Rolando Cavazos Cadena, por su paciencia y dedicación hacia sus alumnos y en especial hacia mí.

- Al M.C. Felix de J. Sánchez Pérez y al M.C. Luis Rodríguez Gutiérrez, por permitir que este trabajo viera la luz.

- A mis compañeros de la maestría: Gerardo, Jéssica y Soledad; gracias por su invaluable amistad y por compartir tantos momentos conmigo en estos dos años.

- A los maestros del Departamento de Estadística, por compartir sus consejos y conocimientos conmigo; así como también a Beatriz y a Rosario por tener siempre un espacio para mí cuando las necesitaba.

- A las familias Olmos, Ruíz y Sandoval; sé que es imposible nombrarlos a todos aquí, pero también es imposible nombrar todo lo que han hecho por mí. Gracias por su confianza y apoyo.

- Al Dr. Javier García Cantú, al MVZ Jorge Luis Macedo, al Ing. José Luis Merinos y a todos mis amigos, por su amistad sincera y desinteresada.

DEDICATORIA

- A mis hermanos, Carolina y César; por recordarme día a día que aún tengo motivos para seguirme superando.

COMPENDIO

ESTIMACIÓN DE VEROSIMILITUD MÁXIMA EN MODELOS DE TRASLACIÓN CON SOPORTE COMPACTO

POR

JESÚS SALVADOR RUÍZ OLMOS

MAESTRÍA EN CIENCIAS

en Estadística Experimental

UNIVERSIDAD AUTÓNOMA AGRARIA ANTONIO NARRO

BUENAVISTA, SALTILLO, COAHUILA. ABRIL, 2005

Dr. Rolando Cavazos Cadena -Asesor-

Palabras claves: Método de Verosimilitud Máxima (MVM), intervalos de confianza, probabilidad asintótica, densidad, modelo estadístico de traslación, estimadores invariantes, información de Fisher, distribución asintótica, consistencia, error cuadrático medio.

El presente trabajo trata sobre la problemática al utilizar el Método de Verosimilitud Máxima (MVM) para obtener intervalos de confianza con probabilidad asintótica para θ bajo un modelo estadístico de traslación; además, compara los intervalos así obtenidos con aquellos provenientes de los métodos de momentos y de cuantiles.

Se hace una revisión acerca de los resultados disponibles en la literatura para el MVM y se especifican la condiciones de regularidad necesarias para la aplicación del MVM en cualquier

modelo estadístico. Así mismo, se analiza el método de traslación propuesto y la problemática que se presenta para utilizar el MVM bajo estas condiciones.

Por otro lado, se establece la cota inferior de Cramer-Rao para el error cuadrático medio de estimadores con sesgo constante y se obtiene una forma alternativa para obtener la información de Fisher en términos de la segunda derivada de la densidad de los errores.

También se estudian los métodos de momentos y de cuantiles para construir estimadores y se utiliza su normalidad asintótica para construir intervalos de confianza con una probabilidad asintótica para el parámetro dado. El estudio concluye que los intervalos de confianza obtenidos por el MVM son asintóticamente más pequeños que los obtenidos por el método de momentos ó de cuantiles.

ABSTRACT

MAXIMUM-LIKELIHOOD ESTIMATION FOR A TRANSLATIONAL MODEL WITH COMPACT SUPPORT

BY

JESÚS SALVADOR RUÍZ OLMOS

MASTER IN SCIENCE

in Experimental Statistics

UNIVERSIDAD AUTÓNOMA AGRARIA ANTONIO NARRO

BUENAVISTA, SALTILLO, COAHUILA. APRIL, 2005

Ph.D. Rolando Cavazos Cadena -Advisor-

Key Words: Maximum-Likelihood Method (MLM),confidence interval, asymptotical probability, density, translational model, invariant estimators, Fisher information, asymptotical distribution, consistency, quadratic mean error.

This paper face the matter of using the Maximum-Likelihood Method (MLM) to obtain confidence intervals on θ with an asymptotical probability under a translational model; besides, it compares the intervals obtained with others coming from method of moments and from quantiles.

A review was made on the literature available about MLM and regularities' conditions needed were specified in order to utilize MLM. Also, the translational model proposed was analyzed along with some troubles derived of using MLM under this condition.

The Cramer-Rao lower bound was settled to invariant estimators and an alternative way to obtain the Fisher's information was established under this kind of densities.

The method of moments and quantiles were analyzed in order to obtain their estimators and their asymptotical normality was used to built confidence intervals with asymptotical probability on the given parameter.

Finally, the conclusion of this paper is that the confidence intervals provided by MLM are asymptotically smaller than others coming from method of moments and quantiles.

Contenido

Capítulo 1

PRESENTACIÓN

Introducción

Este trabajo trata sobre *el problema de estimación en un modelo estadístico de traslación.* En términos generales, en dicho modelo cada dato disponible X_i puede expresarse como $X_i = \theta + \varepsilon_i$, donde $\theta \in \mathbb{R}$ y 'el error aleatorio' ε_i tiene densidad $g(\cdot)$, de manera que la densidad de cada observación es $g(x_i - \theta)$. Además de requerimientos de diferenciabilidad sobre la densidad de los errores, *el supuesto central* de este estudio es que $g(x)$ se anula fuera de un intervalo $[a, A]$, donde a y A son finitos, condición que es bastante natural cuando se tiene la seguridad de que las observaciones X_i son acotadas, por ejemplo, cuando cada X_i representa un porcentaje, o la medición de una cantidad y el error que se comete no excede un cierto nivel. En este contexto, *el principal objetivo* es construir intervalos de confianza para θ de manera que (i) conforme el número de datos disponibles aumenta, la probabilidad de que los intervalos cubran al parámetro alcanza un nivel especificado, y (ii) que la longitud de los intervalos sea lo menor posible. *El resultado principal* en esta dirección se formula en el Capítulo 4, y establece que el método de verosimilitud máxima produce intervalos más cortos que aquéllos generados mediante las técnicas de momentos y de cuantiles. Desde luego, esta conclusión es clásica, y es oportuno enfatizar *la principal diferencia* entre los resultados disponibles en la literatura, y los obtenidos en este trabajo. Esencialmente, el análisis clásico del método de verosimilitud máxima se lleva a cabo bajo ciertas condiciones de regularidad R_1–R_3, las cuales se describen más adelante en este capítulo. Sin embargo, al menos dos de dichas condiciones—a saber, R_1 y R_3—no se aplican al modelo de traslación considerado en este estudio, así que los resultados que finalmente se obtienen no son consecuencia de aquellos disponibles en la literatura.

La organización del trabajo es la siguiente: Primeramente, en el resto de este capítulo se presentan los conceptos básicos que se utilizan más adelante, incluyendo las ideas de intervalos de confianza, convergencia y consistencia, se describen los resultados disponibles sobre el método de verosimilitud máxima, y en la parte final se describe el resultado principal del trabajo. En el Capítulo 2, se introduce el modelo estadístico de traslación, así como la clase de estadísticos invariantes, a la cual pertenecen todos los estimadores analizados posteriormente. Además, se establece la cota inferior de Cramer–Rao para el error cuadrático medio de estimadores con sesgo constante. El argumento en esta parte es similar al que se aplica bajo las condiciones de regularidad usuales, pero deben cuidarse detalles técnicos que se originan por el hecho de que la densidad de los datos tiene soporte que depende del parámetro. En el Capítulo 3 se estudian los métodos de momentos y de cuantiles para construir estimadores, y se utiliza su normalidad asintótica para construir intervalos de confianza para el parámetro con una probabilidad (asintótica) de cobertura dada. La varianza límite de estos estimadores se compara con la cota inferior de Cramer–Rao, lo cual se hace directamente para el estimador de momentos, pero es una tarea interesante para el estimador de cuantiles; el resultado de esta comparación, formulado en la Sección 5 del Capítulo 3, es una de las contribuciones técnicas del capítulo. En el Capítulo 4, referente al método de verosimilitud máxima, se ubican los principales resultados de este trabajo: se establece la consistencia de los estimadores, y se obtiene una cota para su distribución asintótica, resultado que se utiliza para construir intervalos de confianza con longitud menor o igual a la de aquéllos generados por los métodos de momentos y de cuantiles. Finalmente, la exposición concluye en el Capítulo 5 con algunos comentarios breves antes de la literatura citada.

Convergencia

En esta sección se introducen las ideas de convergencia que se utilizarán en este trabajo. La discusión es breve y tiene por objeto establecer la notación que se utilizará. Una presentación detallada puede encontrarse, por ejemplo, en Ash (1987), Billingsley (1997), Dudewicz y Mishra (1989), Dudley (2002), o Serfling (1988). Dado un vector aleatorio $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ cuya densidad $f_n(\mathbf{x}; \theta)$ depende de un parámetro desconocido $\theta \in \Theta$, el problema de estimación (puntual) para una función $h(\theta)$ consiste en construir un estadístico $T_n \equiv T_n(\mathbf{X}_n)$ cuyos valores se aproximen, en algún sentido, al valor desconocido $h(\theta)$. Entre las diversas maneras de medir la proximidad de T_n y $h(\theta)$, una sumamente útil es a través del error cuadrático medio dado por

$$E_\theta[(T_n - h(\theta))^2] = \int (T_n(\mathbf{x}) - h(\theta))^2 f(\mathbf{x}; \theta)\, d\mathbf{x}. \tag{1.2.1}$$

Si para cada n se dispone de un estimador $T_n \equiv T_n(\mathbf{X}_n)$, la sucesión $\{T_n\}$ *converge* a $h(\theta)$ *en la media cuadrática* si

$$\lim_{n\to\infty} E_\theta[(T_n - h(\theta))^2] = 0, \quad \theta \in \Theta. \tag{1.2.2}$$

Otra idea de convergencia, intuitivamente más 'clara', es la siguiente: La sucesión $\{T_n\}$ *converge en probabilidad a* $h(\theta)$ si, para cada $\varepsilon > 0$,

$$\lim_{n\to\infty} P_\theta[|T_n(\mathbf{X}_n) - h(\theta)| > \varepsilon] = 0, \quad \theta \in \Theta. \tag{1.2.3}$$

Verbalmente, esta condición expresa que, conforme el número de datos n aumenta, la probabilidad de que el estimador T_n se desvíe de $h(\theta)$ por más de una cantidad ε se aproxima a cero, lo cual significa que el esfuerzo de obtener más datos se ve recompensado con una probabilidad cada vez más alta de que el estimador se ubique cerca de $h(\theta)$, característica deseable en un 'buen' método de estimación. En adelante se escribe

$$T_n \xrightarrow{\mathrm{P}} h(\theta)$$

si $\{T_n\}$ converge en probabilidad a $h(\theta)$, esto es, si (1.2.3) ocurre para cada $\varepsilon > 0$. A partir de la desigualdad de Markov, se tiene que

$$P_\theta[|T_n(\mathbf{X}_n) - h(\theta)| > \varepsilon] \leq \frac{E_\theta[(T_n - h(\theta))^2]}{\varepsilon^2},$$

de donde se desprende que convergencia en la media cuadrática implica convergencia en probabilidad. Otra noción de convergencia es la siguiente: La sucesión de estimadores $\{T_n\}$ converge a $h(\theta)$ *casi seguramente* si,

$$P_\theta[\lim_{n\to\infty} T_n = h(\theta)] = 1, \quad \theta \in \Theta. \tag{1.2.4}$$

Esta condición significa que, a medida que n aumenta, el estimador T_n se aproxima, inexorablemente, a la cantidad desconocida $h(\theta)$. En adelante, se escribe

$$T_n \to h(\theta) \quad P_\theta\text{–c. s.}$$

para indicar que $\{T_n\}$ converge casi seguramente a $h(\theta)$. Esta idea de convergencia es más fuerte que la convergencia en probabilidad, esto es, si $T_n \to h(\theta)$ P_θ–c. s., entonces $T_n \xrightarrow{\mathrm{P}} h(\theta)$. Un instrumento básico para establecer la convergencia casi segura es el siguiente resultado, conocido en la literatura como el (segundo) lema de Borel–Cantelli.

Teorema 1.2.1. Suponga que para cada $\varepsilon > 0$ y $\theta \in \Theta$ se tiene que

$$\sum_{n=1}^{\infty} P_\theta[|T_n - h(\theta)| > \varepsilon] < \infty.$$

En este caso, $T_n \to h(\theta)$ P_θ–c. s..

Un resultado fundamental sobre convergencia casi segura es la siguiente *ley de los grandes números.*

Teorema 1.2.2. Suponga que las variables aleatorias $X_i, X_2, X_3, \ldots$ son independientes con una distribución común que depende de $\theta \in \Theta$, la cual tiene esperanza finita $\mu(\theta)$, y para cada n sea $\overline{X}_n$ la media muestral de las primeras n observaciones, esto es,

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

En este caso, $\overline{X}_n \to \mu(\theta)$ P_θ–c. s.

El método de estimación que genera los estimadores $\{T_n\}$ de $h(\theta)$—o la sucesion $\{T_n\}$ misma—se denomina *consistente* si T_n converge $h(\theta)$ en alguno de los tres sentidos discutidos anteriormente. En este ensayo, la idea de consistencia se refiere a convergencia casi segura.

Otra noción de convergencia es la idea de convergencia en distribución. Considere una sucesión $\{W_n\}$ de variables aleatorias, sea $\mathcal{D}$ una distribución de probabilidad fija, y denote mediante $F_\mathcal{D}$ a la correspondiente función de ditribución. La sucesión $\{W_n\}$ converge en distribución a $\mathcal{D}$ si para cada $x \in \mathbb{R}$ en el cual $F_\mathcal{D}$ es continua se tiene que

$$\lim_{n\to\infty} F_n(x) = F_\mathcal{D}(x), \tag{1.2.5}$$

donde $F_n(x) = P[W_n \leq x]$ es la función de distribución de W_n. Se utilizará la notación a $W_n \xrightarrow{\mathrm{d}} \mathcal{D}$ para indicar que $\{W_n\}$ converge en distribución a $\mathcal{D}$. El siguiente *teorema central de límite* es fundamental en estadística.

Teorema 1.2.3. Suponga que $X_1, X_2, X_3, \ldots,$ son variables aleatorias *independientes e idénticamente distribuidas* (iid) con esperanza μ y varianza σ^2, ambas finitas. En este caso,

$$\sqrt{n}[\overline{X}_n - \mu] \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, \sigma^2\right),$$

donde $\mathcal{N}\left(0, \sigma^2\right)$ es la distribución normal con media cero y varianza σ^2.

El siguiente resultado proporciona condiciones bajo las cuales la convergencia en (1.2.5) es uniforme.

Teorema 1.2.4. Suponga que $W_n \xrightarrow{\mathrm{d}} \mathcal{D}$ y que la función de distribución de $\mathcal{D}$, denotada por $F_{\mathcal{D}}$, es continua en todo $\mathbb{R}$. En este caso,

$$\max_{x \in \mathbb{R}} |F_n(x) - F_{\mathcal{D}}(x)| \to 0 \quad \text{conforme } n \to \infty,$$

donde F_n es la función de distribución de W_n.

Intervalos de Confianza

De acuerdo a Graybill (1985), la construcción de intervalos de confianza es el procedimiento de inferencia más útil en las aplicaciones. Dado un vector aleatorio $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ cuya distribución depende de un parámetro $\theta \in \Theta \subset \mathbb{R}$, considere un intervalo $I(\mathbf{X}_n) = [L(\mathbf{X}_n),\ U(\mathbf{X}_n)]$ cuyos extremos son estadísticos. Dado un número $\gamma \in (0, 1)$, $I(\mathbf{X}_n)$ es un intervalos de confianza para θ con nivel de cobertura (o de confianza) $1 - \gamma$ si

$$P_\theta[\theta \in I(\mathbf{X}_n)] \geq 1 - \gamma, \quad \theta \in \Theta;$$

esta condición significa que, independientemente del verdadero valor θ_0 del parámetro (el cual es desconocido para el observador), la probabilidad de que éste se encuentre dentro del intervalo $I_n(\mathbf{X}_n)$ es por lo menos $1 - \gamma$. Suponga ahora que para cada n se tiene un intervalo aleatorio $I_n(\mathbf{X}_n)$, el cual se utilizará como intervalo de confianza para θ. La sucesión $\{I_n(\mathbf{X}_n)\}$ tiene un nivel asintótico de cobertura $1-\gamma$ si

$$\lim_{n\to\infty} P_\theta[\theta \in I_n(\mathbf{X}_n)] \geq 1-\gamma, \quad \theta \in \Theta,$$

condición que significa que, conforme n se incrementa, la probabilidad de que el verdadero valor del parámetro sea cubierto por $I_n(\mathbf{X}_n)$ alcanzará, en el límite, el valor $1-\gamma$. La construcción de sucesiones de intervalos de confianza con nivel asintótico de cobertura $1-\gamma$ es el principal tema de estudio en este ensayo.

El Método de Verosimilitud Máxima

Suponga que $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ es un vector aleatorio tal que las variables X_i son iid con densidad $f(x;\theta)$ que depende de un parámetro $\theta \in \Theta$, de tal manera que $\mathbf{X}_n$ tiene densidad

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

La función de verosimilitud correspondiente está dada por

$$L_n(\theta;\mathbf{x}) = f_n(\mathbf{x};\theta);$$

el punto importante en esta igualdad es que el lado izquierdo se considera como función de θ dado que el valor $\mathbf{x}$ es fijo, mientras que en el lado derecho lo que se supone fijo es θ. Un estimador de verosimilitud máxima es un estadístico $\hat{\theta}_n \equiv \hat{\theta}_n(\mathbf{X}_n)$ con la propiedad de que, para cada $\mathbf{x} \in \mathbb{R}^n$,

$$L(\hat{\theta}_n(\mathbf{x});\mathbf{x}) \geq L(\theta;\mathbf{x}) \quad \theta \in \Theta.$$

El análisis clásico de este procedimiento de construcción de estimadores se lleva a cabo bajo las siguientes condiciones de regularidad R_1–R_3 (Greene, 2003, Griffiths y Carter-Hill, 1997, Serfling, 1988, Rao, 1988):

R_1: Las densidades $f(y;\theta)$ tienen soporte común, esto es, $\mathcal{Y}\!:=\{y\colon f(y;\theta)>0\}$ no depende de θ.

R_2: Para cada $y\in\mathcal{Y}$, $f(y;\theta)$ tiene derivadas parciales respecto a θ hasta de tercer orden.

R_3: Existen funciones $H_1, H_3\colon \mathcal{Y}\to\mathbb{R}$ tales que para cada $y\in\mathcal{Y}$ y $\theta\in\Theta$

$$\left|\frac{\partial f(y;\theta)}{\partial\theta}\right|\leq H_1(y), \qquad \left|\frac{\partial^3\log(f(y;\theta))}{\partial\theta^3}\right|\leq H_3(y)$$

donde

$$\int_{\mathcal{Y}} H_1(y)\,dy<\infty, \qquad \text{y} \qquad \int_{\mathcal{Y}} H_3(y)f(y;\theta)\,dy<\infty.$$

Más aún, para todo $\theta\in\Theta$ se tiene que

$$0<\int_{\mathcal{Y}} -\frac{\partial^2\log(f(y;\theta))}{\partial\theta^2}\,f(y;\theta)\,dy<\infty.$$

Bajo estas condiciones, se obtiene el siguiente resultado fundamental.

Teorema 1.4.1. Suponga que las anteriores condiciones de regularidad R_1–R_3 se satisfacen. En estas condiciones las siguientes afirmaciones (i)–(iii) son válidas.

(i) La sucesión de estimadores de verosimilitud máxima $\{\hat\theta_n\}$ es consistente, esto es,

$$\hat\theta_n\to\theta \quad P_\theta\text{–c. s.};$$

(ii) Conforme n tiende a infinito, $\sqrt{n}[\hat\theta_n-\theta]\xrightarrow{\text{d}}\mathcal{N}\left(0,1/\mathcal{I}(\theta)\right)$, donde $\mathcal{I}(\theta)$ es el número de información de Fisher dado por

$$\mathcal{I}(\theta)\!:=-\int_{\mathcal{Y}}\frac{\partial^2\log(f(y;\theta))}{\partial\theta^2}\,f(y;\theta)\,dy.$$

(iii) Dado $\gamma\in(0,1)$, para cada entero positivo n defina el intervalo de confianza $\hat I_n$ para θ mediante

$$\hat I_n=\left[\hat\theta_n-\frac{z_{\gamma/2}}{\sqrt{n\mathcal{I}(\hat\theta_n)}},\ \hat\theta_n+\frac{z_{\gamma/2}}{\sqrt{n\mathcal{I}(\hat\theta_n)}}\right]. \tag{1.4.1}$$

En este caso, la sucesión $\{\hat{I}_n\}$ tiene un nivel asintótico de cobertura de al menos $1-\gamma$, esto es, para cada $\theta \in \Theta$, $\lim_{n\to\infty} P_\theta[\theta \in \hat{I}_n(\mathbf{X}_n)] \geq 1-\gamma$.

Por otro lado, una sucesión $\{T_n \equiv T_n(\mathbf{X}_n)\}$ de estimadores de θ es asintóticamente normal si tiene la siguiente propiedad:

$$\sqrt{n}[T_n - \theta] \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, v(\theta)\right),$$

donde

$$v(\theta) = \lim_{n\to\infty} nE_\theta[(T_n-\theta)^2], \quad \theta \in \Theta$$

es una función continua. Esta sucesión puede utilizarse para construir los intervalos

$$I(T_n) = \left[T_n - \frac{z_{\gamma/2}\sqrt{v(T_n)}}{\sqrt{n}}, \;\; T_n + \frac{z_{\gamma/2}\sqrt{v(T_n)}}{\sqrt{n}}\right]$$

los cuales tienen un nivel asintótico de cobertura $1-\gamma$. La relevancia de los intervalos $\hat{I}_n$ construidos mediante el método de verosimilitud máxima, se debe a la siguiente propiedad: Si la sucesión $\{T_n\}$ es asintóticamente normal, entonces

$$\lim_{n\to\infty} \frac{|\hat{I}_n|}{|I(T_n)|} \leq 1,$$

donde $|J|$ denota la longitud del intervalo J. Esta propiedad muestra que, en el límite, los intervalos $\hat{I}_n$ son más cortos que aquellos construidos mediante los estadísticos $\{T_n\}$, y además, pone en relieve la eficiencia del método de verosimilitud máxima.

El Problema

Como ya se ha mencionado, en este trabajo se analiza un modelo estadístico de traslación, en el que cada dato X_i tiene densidad $g(x-\theta)$, donde $g(\cdot)$ es una densidad conocida que se anula fuera de un intervalo acotado, y el tema principal es el estudio de los estimadores de verosimilitud máxima de θ en este contexto. El aspecto que hace interesante al problema es que dos de las condiciones de regularidad R_1–R_3 descritas en la sección precedente no se satisfacen en el modelo estudiado en este trabajo, y por lo tanto, las conclusiones obtenidas en el Capítulo 4 no se desprenden de los resultados clásicos sobre el método de verosimilitud máxima. Para ilustrar este punto, considere el caso en que la densidad $g(x)$ es

$$g(x) = 30x^3(1-x)^3 I_{(0,1)}(x),$$

y Θ es un intervalo, digamos $\Theta = (0,1)$. En este caso $\{x\colon g(x-\theta) > 0\} = (\theta, 1+\theta)$ *si* depende de θ, y entonces la condición R_1 no se satisface. A continuación, observe que para $x \in (\theta, 1+\theta)$ se tiene

$$\log(g(x-\theta)) = 3\log(x-\theta) + 3\log(1-x+\theta) + \log(30),$$

y entonces

$$\partial_\theta^3 \log(g(x-\theta)) = -\frac{6}{(x-\theta)^3} + \frac{6}{(1-x+\theta)^3}.$$

Por lo tanto, si la funció0n $H_3(x)$ es tal que $\left|\partial_\theta^3 \log(g(x-\theta))\right| \le H_3(x)$ para todo θ y $x \in (\theta, 1+\theta)$, se tiene que

$$H_3(x) = \infty, \quad x \in \Theta \cup (\Theta + 1),$$

de manera que $\int H_3(x) g(x-\theta)\, dx = \infty$ y por lo tanto la condición R_3 no se verifica. Este ejemplo muestra que, cuando el soporte de la densidad $g(\cdot)$ es acotado, los resultados clásicos sobre el método de verosimilitud máxima no se aplican directamente, y el propósito de este ensayo es proporcionar extensiones de dichos resultados al presente marco de trabajo.

El problema específico que se estudia en este trabajo tiene tres facetas:

1. Establecer la consistencia del método de verosimilitud máxima en el modelo de traslación;

2. Utilizar los estimadores de verosimilitud máxima para construir intervalos de confianza I_n con nivel asintótico de cobertura especificado;

3. Comparar las longitudes de los intervalos I_n con la de aquéllos construidos usando los métodos de momentos y de cuantiles, y establecer que los intervalos I_n son, en el límite, más cortos.

Los resultados sobre estos problemas se presentan en el Capítulo 4.

Capítulo 2

MODELO DE TRASLACIÓN

Introducción

El propósito de este capítulo es presentar el modelo y los resultados básicos que se utilizarán en el desarrollo subsecuente. La exposición inicia introduciendo el modelo estadístico que se estudiará, el cual está caracterizado por la siguiente condición: cada componente del vector de datos, dado por $\mathbf{X} = (X_1, \ldots, X_n)$, se expresa como $X_i = \theta + \varepsilon_i$, donde las variables ε_i son independientes y tienen densidad común que se anula fuera de un intervalo acotado. Esta estructura ocasiona que el soporte de las variables X_i dependa de θ, característica que hace que el análisis del método de verosimilitud máxima presente dificultades que no pueden abordarse siguiendo el enfoque clásico; vea, por ejemplo, Serfling (1988). Los resultados que se establecen en este capítulo se utilizarán posteriormente para acotar la distribución límite del estimador de verosimilitud máxima de θ, lo que finalmente permitirá establecer intervalos de confianza para el parámetro con un nivel asintótico determinado. Además de presentar el modelo básico, este capítulo tiene dos propósitos fundamentales: (i) introducir la clase de estimadores invariantes y (ii) establecer la cota de Cramer-Rao para el modelo de traslación, resultado que establece un límite inferior para el error cuadrático medio de un estimador y, bajo condiciones de normalidad asintótica, para la longitud de intervalos de confianza cuando el tamaño de la muestra es 'grande'.

La organización del material es la siguiente: En la Sección 2 se introduce el modelo de traslación y, en el contexto de un problema de medición, se discute la conveniencia del supuesto de que la densidad de los errores tenga soporte acotado. En la Sección 3 se introduce la clase de estimadores invariantes y se demuestra que su sesgo y error estándar no dependen del parámetro. En la Sección 4 se formula la cota de Cramer-Rao, resultado que, después de establecer los preliminares técnicos necesarios, se demuestra en la Sección 5. La exposición concluye en la Sección 6 presentando una forma alternativa para calcular la cota de Cramer–Rao en términos de la segunda derivada de la densidad de los errores. Debido a la diversidad de formas (unimodales) que la densidad beta adopta al cambiar sus parámetros, en este trabajo se utilizan densidades beta para ilustrar los resultados obtenidos.

Modelo de Traslación

El supuesto básico sobre los datos disponibles, es que éstos se generan de acuerdo al siguiente modelo.

Definición 2.1. El vector aleatorio observable $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ obedece a un modelo de traslación si (i) sus componentes son independientes, y (ii) la siguiente ecuación estructural se satisface:

$$X_i = \theta + \varepsilon_i, \quad \theta \in \Theta, \quad \varepsilon_i \sim g(\cdot), \quad i = 1, 2, \ldots, n. \tag{2.2.1}$$

En esta expresión, $\Theta \subset \mathbb{R}$ es un intervalo conocido de longitud positiva, $g(\cdot)$ es una densidad en $\mathbb{R}$, y la notación $\varepsilon_i \sim g(\cdot)$ significa que la variable aleatoria ε_i tiene densidad $g(\cdot)$.

La condición de que el vector $\mathbf{X}$ sea observable significa que después de realizar el experimento aleatorio que genera al vector, el valor preciso de $\mathbf{X}$ será conocido por el analista. Por otro lado, sobre θ sólo se conoce que pertenece al espacio de parámetros Θ, el cual contiene más de un elemento, y entonces el valor preciso de $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)'$ no podrá ser determinado por el observador a partir de $\mathbf{X}$; esto significa que ε *es un vector aleatorio no observable*. Por lo tanto, (2.2.1) establece que el vector de datos $\mathbf{X}$ es la suma de un vector θ fijo pero no completamente determinado, y de un vector aleatorio no observable ε, el cual, siguiendo la convención usual, se denominará *vector de errores*.

Ejemplo 2.2.1. [Mediciones (Lehmann y Casella 2000, Lehmann, 2001).] Se trata de determinar una magnitud θ con un instrumento que produce un error de medición ε, de manera que cada una de las mediciones X que se obtienen está dada por

$$X = \theta + \varepsilon. \tag{2.2.2}$$

El espacio de parámetros Θ puede tomarse como cualquier intervalo que incluya a los posibles valores de la magnitud que se quiere determinar. En un caso extremo puede tomarse $\Theta = [0, \infty)$, o tal vez $\Theta = \mathbb{R}$, pero con frecuencia se tendrá información *a priori* que permita especificar Θ como un conjunto más reducido. Por ejemplo, si se trata de estimar la longitud θ de una varilla que cabe a lo ancho en la cabina de un auto, entonces se sabe que $\theta \leq A$, donde A es el ancho de la cabina, y si la varilla sobrepasa el ancho a del asiento, entonces $\theta \geq a$, de manera que puede tomarse $\Theta = [a, A]$. Por otro lado, para completar la especificación del modelo, es necesario estipular la densidad $g(\cdot)$ del error aleatorio ε, y se tienen multitud de posibilidades. En general, un instrumento diseñado para medir producirá un error—por defecto, o por exceso—de a lo más la mitad de la unidad que permite detectar, por lo que la condición menos restrictiva que puede suponerse sobre $g(\cdot)$ es la siguiente:

(a) El soporte de la densidad g es $[-1/2, 1/2]$.

Esta condición, conjuntamente con $\Theta = [a, A]$ y la ecuación (2.2.2) determinan un modelo no parámetrico de traslación. En el otro extremo de posibilidades sobre la especificación de la distribución de los errores, se ubica el supuesto de que ε tenga una densidad completamente determinada, por ejemplo, uniforme en $[-1/2, 1/2]$.

(b) $g(x) = I_{[-1/2,1/2]}(x)$.

En este caso, la probabilidad de obtener un error que se ubique en un intervalo J de longitud δ contenido en $[-1/2, 1/2]$ es siempre δ, de tal manera que la probabilidad no depende de la ubicación de J. En general, un buen instrumento de medición producirá errores cuya distribución es más concentrada alrededor de cero que cerca de los extremos $1/2$ y $-1/2$, y la especificación anterior viola esta condición. Como otra alternativa, suponga que ε tiene una distribución beta simétrica alrededor de cero.

(c) $g(x) = \dfrac{\Gamma(2\alpha)}{\Gamma(\alpha)\Gamma(\alpha)} \left(\dfrac{1}{2} + x\right)^{\alpha-1} \left(\dfrac{1}{2} - x\right)^{\alpha-1} I_{[-1/2,1/2]}(x)$.

En esta expresión α es un número positivo. Si $\alpha > 1$, entonces ε está más concentrada alrededor de cero que cerca de los extremos $-1/2$ y $1/2$, como debe ser para un buen instrumento de medición; el número α puede pensarse como una característica del diseño del instrumento estipulada por el fabricante. A medida que α se acerca a 1, esta densidad beta se aproxima a la densidad uniforme en el inciso (b), por lo que ésta última puede interpretarse como correspondiente a un instrumento en proceso de desgaste. Conforme α continua disminuyendo por debajo de 1, la densidad beta se concentra más alrededor de los extremos $-1/2$ y $1/2$, de modo que el caso $\alpha \in (0, 1)$ se interpreta como una característica correspondiente a un instrumento en pleno desgaste. La pericia del operador del instrumento tambien puede incorporarse en el modelo con la siguiente especificación de la densidad $g(\cdot)$.

(d) $g(x) = \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \left(\dfrac{1}{2}+x\right)^{\alpha-1} \left(\dfrac{1}{2}-x\right)^{\beta-1} I_{[-1/2,1/2]}(x).$

En este caso, α y β son dos números positivos. Si $\alpha > \beta$, entonces la media de g, dada por $\alpha/(\alpha+\beta) - 1/2$ es positiva, lo cual significa que la combinación del operario y el instrumento tienden a producir errores por exceso, mientras que si $\alpha < \beta$ los errores que se producen son, en promedio, por defecto. Finalmente, el siguiente supuesto de normalidad sobre la distribución del error ε es frecuente.

(e) $g(x) = \dfrac{1}{\sqrt{\sigma^2 2\pi}} e^{-x^2/(2\sigma^2)}.$

Bajo este supuesto el error ε puede tomar valores en todo $\mathbb{R}$, de manera que a partir de la ecuación estructural $X = \theta + \varepsilon$, es posible que X tome valores negativos o valores tan grandes que evidentemente sobrepasen a la magnitud θ que se desea determinar. Sobre este punto, el argumento que justifica el supuesto (e) tiene dos facetas. Primero, suponer una distribución normal para ε produce simplificaciones analíticas importantes y permite establecer intervalos de confianza para θ usando métodos simples y bien establecidos. Por otro lado, bajo la condición de normalidad (e) se tiene que ε toma valores en el intervalo $[-3\sigma, 3\sigma]$ con una probabilidad de alrededor de 0.99, por lo que si σ es pequeño, entonces se tiene una alta probabilidad de que los datos observados se mantengan dentro de límites razonables. □

Note que la condición estructural (2.2.1) es equivalente al enunciado de que la densidad de X_i es $g(x_i - \theta)$, por lo que la idea en la Definición 2.2.1 puede formularse equivalentemente como sigue.

Definición 2.2. El vector aleatorio n-dimensional $\mathbf{X}$ obedece a un modelo de traslación si su densidad pertenece a la familia paramétrica $\{f(\mathbf{x};\theta) \,|\, \theta \in \Theta\}$, donde el espacio de parámetros Θ es un intervalo contenido en $\mathbb{R}$,

$$f(\mathbf{x};\theta) = \prod_{i=1}^{n} g(x_i - \theta), \quad \mathbf{x} \in \mathbb{R}^n, \tag{2.2.3}$$

y $g(\cdot)$ es una densidad en $\mathbb{R}$; esta expresión para $f(\mathbf{x};\theta)$ significa que el vector $\mathbf{X}$ es una muestra de la densidad $g(x-\theta)$.

Como ya se ha mencionado, el propósito de este trabajo es estudiar el problema de inferencia estadística para el parámetro θ en el anterior modelo de traslación, y el principal objetivo es construir intervalos de confianza para θ con un nivel de confianza asintótico determinado. Este problema será abordado bajo el siguiente supuesto.

Hipótesis 2.2.1. *El vector de observaciones* $\mathbf{X}$ *obedece a un modelo de traslación en el que la densidad* g *de los errores es conocida y tiene soporte acotado, esto es,* $g(\cdot)$*, se anula fuera de un intervalo* $[a, A]$.

Estimadores Invariantes

Suponga que $\mathbf{X}$ obedece a un modelo de traslación, de manera que $\mathbf{X} = \theta\mathbb{1}+\varepsilon$, donde $\mathbb{1}$ es el vector n-dimensional con todas sus componentes iguales a uno. Sea $T(\mathbf{X})$ un estadístico que se utiliza para estimar θ, y observe que si a cada dato se le agrega una constante conocida c entonces el nuevo vector de observaciones será $\mathbf{X}+c\mathbb{1}$, de manera que el valor del estimador en esta muestra es $T(\mathbf{X}+c\mathbb{1})$. Como agregar la constante c a los datos tiene el mismo efecto que sumar c al parámetro θ, es natural requerir que

$$T(\mathbf{X}+c\mathbb{1}) = T(\mathbf{X}) + c. \tag{2.3.1}$$

Definición 3.1. El estadístico $T(\mathbf{X})$ es invariante bajo traslaciones si la igualdad (2.3.1) es válida para todo valor de $\mathbf{X}$ y $c \in \mathbb{R}$.

Ejemplos de estadísticos invariantes bajo traslación son la media y la mediana muestrales. Todos los estadísticos que se considerarán en este trabajo como estimadores de θ, incluyendo el estimador de verosimilitud máxima, son invariantes bajo traslación. La siguiente propiedad básica será de utilidad.

Lema 2.3.1. Suponga que $\mathbf{X}$ obedece a un modelo de traslación y que $T(\mathbf{X})$ es un estimador de θ invariante bajo traslación. En este caso, el sesgo de $T(\mathbf{X})$, dado por

$$E_\theta[T(\mathbf{X})] - \theta,$$

así como el correspondiente error cuadrático medio, especificado mediante

$$E_\theta[(T(\mathbf{X}) - \theta)^2],$$

no dependen de θ.

Demostración. A partir de la ecuación estructural (2.2.1) el vector $\mathbf{X}$ puede expresarse como $\mathbf{X} = \theta\mathbb{1} + \varepsilon$, de manera que $T(\mathbf{X}) = T(\varepsilon + \theta\mathbb{1}) = \theta + T(\varepsilon)$, pues $T(\mathbf{X})$ es invariante bajo traslación. Por lo tanto,

$$T(\mathbf{X}) - \theta = T(\varepsilon)$$

de donde se desprende que

$$E_\theta[T(\mathbf{X})] - \theta = E_\theta[T(\varepsilon)] \quad \text{y} \quad E_\theta[(T(\mathbf{X})] - \theta)^2] = E_\theta[T(\varepsilon)^2].$$

Como las componentes de $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)'$ son independientes e idénticamente distribuidas con densidad $g(\cdot)$, se tiene que

$$E_\theta[T(\varepsilon)] = \int_{\mathbb{R}^n} T(\varepsilon_1, \ldots, \varepsilon_n) g(\varepsilon_1) \cdots g(\varepsilon_n)\, d\varepsilon_1 \ldots d\varepsilon_n$$

y

$$E_\theta[T(\varepsilon)^2] = \int_{\mathbb{R}^n} T(\varepsilon_1, \ldots, \varepsilon_n)^2 g(\varepsilon_1) \cdots g(\varepsilon_n)\, d\varepsilon_1 \ldots d\varepsilon_n$$

no dependen de θ. $\square$

Cota de Rao-Cramer

El objetivo de esta sección es establecer una cota inferior para el error cuadrático esperado de un estimador de una función paramétrica $h(\theta)$. Este resultado se obtiene bajo el supuesto de que tanto la Hipótesis 2.2.1 como la siguiente condición de regularidad se satisfacen.

Hipótesis 2.4.1. *La densidad $g(\cdot)$ de los errores tiene primera derivada continua en cada punto $x \in \mathbb{R}$.*

Observación 2.4.1. Como la función $g(x)$ se anula fuera del intervalo $[a, A]$ se tiene que $g'(x) = 0$ para $x < a$. Por lo tanto, la Hipótesis 2.4.1 implica que $g'(a) = \lim_{x \to a-} g'(x) = 0$ y, similarmente, $g'(A) = 0$. A partir de esta observación se desprende que la densidad beta en las partes (c) y (d) del Ejemplo 2.2.1 —para las cuales $a = -1/2$ y $A = 1/2$—satisfacen la Hipóteisis 2.4.1 si y sólo si $\alpha > 1$ y $\beta > 1$.

Teorema 2.4.1. Sea $T\colon \mathbb{R}^n \to \mathbb{R}$ una función que es integrable en conjuntos acotados, esto es,

$$\int_C |T(\mathbf{x})|\, d\mathbf{x} < \infty \tag{2.4.1}$$

para cada conjunto acotado $C \subset \mathbb{R}^n$. Suponga que $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ obedece a una modelo de traslación para el cual las Hipótesis 2.2.1 y 2.4.1 se satisfacen, y sea $H(\theta)$ la esperanza del estadístico $T(X)$, esto es, para cada $\theta \in \Theta$

$$\begin{aligned} H(\theta) &= E_\theta\left[T(\mathbf{X})\right] \\ &= \int_{\mathbb{R}^n} T(\mathbf{x})\, f(\mathbf{x};\theta)\, d\mathbf{x} \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}}\cdots\int_{\mathbb{R}} T(x_1, x_2, \ldots, x_n) \prod_{i=1}^{n} g(x_i-\theta)\, dx_1\, dx_2 \ldots, dx_n. \end{aligned}$$

En este caso,

(i) La función $H(\cdot)$ es derivable y

$$H'(\theta) = \int_{\mathbb{R}^n} T(\mathbf{x})\partial_\theta f(\mathbf{x};\theta)\, d\mathbf{x}, \quad \theta \in \Theta.$$

(ii) Si el sesgo de $T(\mathbf{X})$ como estimador de la función paramétrica $h(\theta)$ es constante, esto es,

$$E_\theta[T(\mathbf{X})] - h(\theta) = H(\theta) - h(\theta) = \text{constante}, \tag{2.4.2}$$

entonces

$$\int_{\mathbb{R}^n} [T(\mathbf{x}) - h(\theta)]\partial_\theta f(\mathbf{x};\theta)\, d\mathbf{x} = h'(\theta) \tag{2.4.3}$$

y, consecuentemente,

$$E_\theta[(T(\mathbf{X}) - h(\theta)^2] \geq \frac{|h'(\theta)|^2}{n\mathcal{I}} \tag{2.4.4}$$

donde

$$\mathcal{I} = \int_{\mathbb{R}} \left(\frac{g'(x)}{g(x)}\right)^2 g(x)\, dx \tag{2.4.5}$$

es el número de información de Fisher asociado a la densidad $g(\cdot)$.

La demostración de este teorema se presenta en la siguiente sección. La relación (2.4.4) es la *desigualdad de Cramer-Rao* y desempeña un papel importante en la fase de diseño del experimento, pues permite obtener una cota inferior para el tamaño de la muestra n que se necesita para obtener intervalos de confianza con una longitud deseada ℓ. Por ejemplo, suponga que se utilizará el estadístico $T(\mathbf{X})$

para estimar $h(\theta) = \theta$, y que el intervalo de confianza para θ_0, el verdadero valor del parámetro, se obtendrá utilizando la desigualdad de Chebychev:

$$\begin{aligned} P_{\theta_0}\left[T(\mathbf{X}) - \ell/2 \leq \theta_0 \leq T(\mathbf{X}) + \ell/2\right] \\ = P_{\theta_0}\left[|T(\mathbf{X}) - \theta_0| \leq \ell/2\right] \geq \frac{4E_{\theta_0}\left[(T(X) - \theta_0)^2\right]}{\ell^2} \end{aligned}$$

Siguiendo este enfoque, para garantizar que $[T(\mathbf{X}) - \ell/2,\ T(\mathbf{X}) + \ell/2]$ tenga un nivel de confianza $1 - \gamma$ se buscará que $4E_{\theta_0}\left[(T(X) - \theta_0)^2\right]/\ell^2 = 1 - \gamma$, i.e., $E_{\theta_0}\left[(T(X) - \theta_0)^2\right] = \ell^2(1 - \gamma)/4$, de manera que (2.4.4) permite concluir que

$$\frac{(1-\gamma))\ell^2}{4} \geq \frac{|h'(\theta_0)|^2}{n\mathcal{I}} = \frac{1}{n\mathcal{I}},$$

y entonces

$$n \geq \frac{4}{\mathcal{I}(1-\gamma))\ell^2}.$$

En otros casos, el estimador 'estandarizado' $[T(\mathbf{X}) - \theta_0]/\sqrt{E_{\theta_0}\left[(T(\mathbf{X}) - \theta_0)^2\right]}$ tiene al menos aproximadamente, distribución normal estándar, de manera que, escribiendo

$$v(\theta_0) := E_{\theta_0}\left[(T(\mathbf{X}) - \theta_0)^2\right],$$

$$\begin{aligned} P_{\theta_0}\left[\theta_0 \in T(\mathbf{X}) - z_{\gamma/2}\sqrt{v(\theta_0)}\ ,\ T(\mathbf{X}) + z_{\gamma/2}\sqrt{v(\theta_0)}\ \right] \\ = P_{\theta_0}\left[\left|\frac{T(\mathbf{X}) - \theta_0}{\sqrt{E_{\theta_0}\left[(T(\mathbf{X}) - \theta_0)^2\right]}}\right| \leq z_{\gamma/2}\right] = 1 - \gamma. \end{aligned}$$

La longitud del intervalo de confianza en este enunciado es

$$\ell = 2z_{\gamma/2}\sqrt{E_{\theta_0}\left[(T(\mathbf{X}) - \theta_0)^2\right]} \geq \frac{2z_{\gamma/2}}{\sqrt{n\mathcal{I}}}$$

donde la desigualdad se debe a (2.4.4). A partir de esta relación se desprende que el tamaño de la muestra n satisface

$$n \geq \frac{4z_{\gamma/2}^2}{\ell^2\mathcal{I}}.$$

Observación 2.4.2. Es conveniente puntualizar que la desigualdad de Cramer-Rao (2.4.4) es válida siempre y cuando el sesgo de $T(\mathbf{X})$ como estimador de $h(\theta)$ sea constante (vea la condición (2.4.2)). De acuerdo al Lemma 2.3.1 este requerimiento se satisface para todos los estimadores de $h(\theta) = \theta$ que sean invariantes bajo traslación.

Demostración del Teorema 2.4.1

El argumento para demostrar la desigualdad de Cramer-Rao utiliza el siguiente resultado preliminar sobre la densidad del vector de observaciones $\mathbf{X}$.

Lema 2.5.1. Bajo las Hipótesis 2.2.1 y 2.4.1 las siguientes propiedades (i)–(v) sobre las densidades del vector $\mathbf{X}$ se satisfacen:

(i) Existe una constante $B > 0$ tal que

$$|f(\mathbf{x};\theta)| \leq B^n \quad \text{y} \quad |\partial_\theta f(\mathbf{x};\theta)| \leq nB^n, \quad \mathbf{x} \in \mathbb{R}^n, \quad \theta \in \Theta. \tag{2.5.1}$$

(ii) Si $g(x) = 0$, entonces $g'(x) = 0$.

En consecuencia,

(iii) Si $f(\mathbf{x};\theta) = 0$, entonces $\partial_\theta f(\mathbf{x};\theta) = 0$.

(iv) Adoptando la convención de que un cociente de la forma $0/0$ se define de manera arbitraria, digamos como 1, la igualdad

$$\partial_\theta f(\mathbf{x};\theta) = \left(\frac{\partial_\theta f(\mathbf{x};\theta)}{f(\mathbf{x};\theta)}\right) f(\mathbf{x};\theta) \tag{2.5.2}$$

es siempre válida. Más aún,

(v) $E_\theta\left[\left(\dfrac{\partial_\theta f(\mathbf{x};\theta)}{f(\mathbf{x};\theta)}\right)^2\right] = n\mathcal{I}$, donde el número de información de Fisher $\mathcal{I}$ está dado en (2.4.5).

Demostración. Primeramente, observe que como g se anula fuera del intervalo $[a, A]$ y tiene derivada continua en todo $\mathbb{R}$, la función $g(\cdot)$ es continua y satisface

$$g(a) = g(A) = 0 \tag{2.5.3}$$

y, más aún, existe una constante $B > 0$ tal que para todo $x \in \mathbb{R}$

$$|g(x)| \leq B \quad \text{y} \quad |g'(x)| \leq B. \tag{2.5.4}$$

(i) Usando la expresión (2.2.3) para $f(\mathbf{x}; \theta)$ se desprende que

$$|f(\mathbf{x}; \theta)| = \prod_{i=1}^{n} |g(x_i - \theta)| \leq \prod_{i=1}^{n} B = B^n, \tag{2.5.5}$$

así como la siguiente expresión para la derivada parcial de $f(\mathbf{x}; \theta)$ respecto al parámetro θ:

$$\partial_\theta f(\mathbf{x}; \theta) = \sum_{i=1}^{n} -g'(x_i - \theta) \prod_{j:1 \leq j \neq i \leq n} g(x_j - \theta). \tag{2.5.6}$$

Combinando esta expresión con (2.5.4) se desprende que

$$\begin{aligned}
|\partial_\theta f(\mathbf{x}; \theta)| &= \left| \sum_{i=1}^{n} -g'(x_i - \theta) \prod_{j:1 \leq j \neq i \leq n} g(x_j - \theta) \right| \\
&\leq \sum_{i=1}^{n} |-g'(x_i - \theta)| \prod_{j:1 \leq j \neq i \leq n} |g(x_j - \theta)| \\
&\leq \sum_{i=1}^{n} B \prod_{j:1 \leq j \neq i \leq n} B = \sum_{i=1}^{n} B \times B^{n-1}
\end{aligned}$$

y entonces $|\partial_\theta f(\mathbf{x}; \theta)| \leq nB^n$.

(ii) Suponga que $g(x) = 0$ y observe que $g(x + h) \geq 0$ para todo h, pues $g(\cdot)$ es una densidad. Por lo tanto, x es un punto de mínimo local de g, de donde se sigue que $g'(x) = 0$.

(iii) Suponga que $f(\mathbf{x}; \theta) = 0$. A partir de (2.5.5) se desprende que $g(x_r - \theta) = 0$ para algún $r \in \{1, 2, \ldots, n\}$. Observando que para $i \neq r$ el factor $g(x_r - \theta)$

aparece en $\prod_{j:1\leq j\neq i\leq n} g(x_j-\theta)$ se tiene que $\prod_{j:1\leq j\neq i\leq n} g(x_j-\theta)=0$, y entonces la expresión (2.5.6) se reduce a

$$\partial_\theta f(\mathbf{x};\theta) = -g'(x_r-\theta)\prod_{j:1\leq j\neq r\leq n} g(x_j-\theta)$$

Finalmente, observe que $g(x_r-\theta)=0$ implica que $g'(x_r-\theta)=0$, por la parte (ii), de manera que la anterior expresión implica que $\partial_\theta f(\mathbf{x};\theta)=0$.

(iv) Cuando $f(\mathbf{x};\theta)\neq 0$ la igualdad (2.5.2) es claramente válida, mientras que si $f(\mathbf{x};\theta)=0$, entonces $\partial_\theta f(\mathbf{x};\theta)=0$, por la parte (ii), y adoptando la convención mencionada en el lemma, (2.5.2) se reduce a $0=1\times 0$, igualdad que es correcta, mostrando que (2.5.2) es válida en cualquier caso.

(v) Observe que las fórmulas (2.2.3) y (2.5.6) para $f(\mathbf{x};\theta)$ y $\partial_\theta f(\mathbf{x};\theta)$ implican que

$$\left(\frac{\partial_\theta f(\mathbf{X};\theta)}{f(\mathbf{X};\theta)}\right)^2 f(\mathbf{x};\theta) = \left(\sum_{i=1}^n \frac{g'(X_i-\theta)}{g(X_i-\theta)}\right)^2 f(\mathbf{x};\theta),$$

donde se aplica la convención de que el cociente $0/0$ se define como arbitrariamente como 1. Por lo tanto

$$\begin{aligned} E_\theta\left[\left(\frac{\partial_\theta f(\mathbf{X};\theta)}{f(\mathbf{X};\theta)}\right)^2\right] &= \int_{\mathbb{R}^n}\left(\frac{\partial_\theta f(\mathbf{X};\theta)}{f(\mathbf{X};\theta)}\right)^2 f(\mathbf{x};\theta)\,d\mathbf{x} \\ &= \int_{\mathbb{R}^n}\left(\sum_{i=1}^n \frac{g'(X_i-\theta)}{g(X_i-\theta)}\right)^2 f(\mathbf{x};\theta)\,d\mathbf{x} \\ &= E_\theta\left[\left(\sum_{i=1}^n \frac{g'(X_i-\theta)}{g(X_i-\theta)}\right)^2\right] \end{aligned} \tag{2.5.7}$$

Por otro lado, el caso $n=1$ de la igualdad (2.5.2), establece que

$$\frac{g'(x-\theta)}{g(x-\theta)}g(x-\theta) = g'(x-\theta),$$

y debido a que la densidad de X_i correspondiente al parámetro θ es $g(x-\theta)$, se obtiene

$$\begin{aligned}
E_\theta\left[\frac{g'(X_i-\theta)}{g(X_i-\theta)}\right] &= \int_{\mathbb{R}} \frac{g'(x-\theta)}{g(x-\theta)} g(x-\theta)\,dx \\
&= \int_{\mathbb{R}} g'(x-\theta)\,dx \\
&= \int_a^A g'(x)\,dx \\
&= g(A)-g(a)
\end{aligned}$$

y via (2.5.3) se desprende que

$$E_\theta\left[\frac{g'(X_i-\theta)}{g(X_i-\theta)}\right] = 0. \tag{2.5.8}$$

Como las variables $g'(X_i-\theta)/g(X_i-\theta)$ son iid, esta última igualdad implica que

$$\begin{aligned}
E_\theta\left[\left(\sum_{i=1}^n \frac{g'(X_i-\theta)}{g(X_i-\theta)}\right)^2\right] &= \sum_{i=1}^n E_\theta\left[\left(\frac{g'(X_i-\theta)}{g(X_i-\theta)}\right)^2\right] \\
&= nE_\theta\left[\left(\frac{g'(X_1-\theta)}{g(X_1-\theta)}\right)^2\right] \\
&= n\int_{\mathbb{R}} \left(\frac{g'(x-\theta)}{g(x-\theta)}\right)^2 g(x-\theta)\,dx \\
&= n\int_{\mathbb{R}} \left(\frac{g'(x)}{g(x)}\right)^2 g(x)\,dx = n\mathcal{I},
\end{aligned} \tag{2.5.9}$$

y conjuntamente con (2.5.7), esto implica que $E_\theta\left[\left(\frac{\partial_\theta f(\mathbf{X};\theta)}{f(\mathbf{X};\theta)}\right)^2\right] = n\mathcal{I}$, concluyendo la demostración. □

Usando los resultados del lema precedente, a continuación se establece el Teorema 2.4.1.

Demostración del Teorema 2.4.1. (i) Sea $\theta\in\Theta$ arbitrario pero fijo, tome $\Delta\neq 0$ tal que $\theta+\rho\in\Theta$ si ρ está entre 0 y Δ, y note que para cada uno de estos valores de ρ se tiene que

$$\frac{H(\theta+\rho)-H(\theta)}{\rho} = \int_{\mathbb{R}^n} T(\mathbf{x})\frac{f(\mathbf{x};\theta+\rho)-f(\mathbf{x};\theta)}{\rho}\,d\mathbf{x}. \tag{2.5.10}$$

Por otro lado, como $g(x) = 0$ para x fuera del intervalo $[a, A]$, (2.2.3) implica que $f(\mathbf{x}; \theta) = 0$ si $x_i \notin [a+\theta, A+\theta]$ para algún i. Por lo tanto, para cada ρ entre 0 y Δ se tiene que $f(\mathbf{x}; \theta+\rho) - f(\mathbf{x}; \theta) = 0$ si $x_i \notin [a+\theta-|\Delta|, A+\theta+|\Delta|]$ para algún i, esto es,

$$\begin{aligned} f(\mathbf{x}; \theta+\rho) - f(\mathbf{x}; \theta) &= [f(\mathbf{x}; \theta+\rho) - f(\mathbf{x}; \theta)] \prod_{i=1}^{n} I_{[a+\theta-|\Delta|,\ A+\theta+|\Delta|]}(x_i) \\ &= [f(\mathbf{x}; \theta+\rho) - f(\mathbf{x}; \theta)] I_C(\mathbf{x}) \end{aligned}$$

donde

$$C = [a+\theta-|\Delta|,\ A+\theta+|\Delta|] \times \cdots \times [a+\theta-|\Delta|,\ A+\theta+|\Delta|].$$

Aplicando el teorema del valor medio, para cada $\mathbf{x}$ existe $\tilde{\rho}(\mathbf{x}; \rho)$ entre 0 y ρ tal que $f(\mathbf{x}; \theta+\rho) - f(\mathbf{x}; \theta) = \rho \partial_\theta f(\mathbf{x}; \theta + \tilde{\rho}(\mathbf{x}; \rho))$ de manera que

$$\frac{f(\mathbf{x}; \theta+\rho) - f(\mathbf{x}; \theta)}{\rho} = \partial_\theta f(\mathbf{x}; \theta + \tilde{\rho}(\mathbf{x}; \rho)) I_C(\mathbf{x}), \tag{2.5.11}$$

y (2.5.10) equivale a

$$\begin{aligned} \frac{H(\theta+\rho) - H(\theta)}{\rho} &= \int_{\mathbb{R}^n} T(\mathbf{x}) \partial_\theta f(\mathbf{x}; \theta + \tilde{\rho}(\mathbf{x}; \rho)) I_C(\mathbf{x})\, d\mathbf{x} \\ &= \int_C T(\mathbf{x}) \partial_\theta f(\mathbf{x}; \theta + \tilde{\rho}(\mathbf{x}; \rho))\, d\mathbf{x}. \end{aligned} \tag{2.5.12}$$

Note ahora que la continuidad de $\partial_\theta f(\mathbf{x}; \theta)$ y el hecho de que $\tilde{\rho}(\mathbf{x}; \rho)$ se ubique entre 0 y ρ implican que

$$\partial_\theta f(\mathbf{x}; \theta + \tilde{\rho}(\mathbf{x}; \rho)) \to \partial_\theta f(\mathbf{x}; \theta) \quad \text{conforme} \quad \rho \to 0,$$

y tomando el limite cuando $\rho \to 0$ en (2.5.11) se desprende que

$$\partial_\theta f(\mathbf{x}; \theta) = \partial_\theta f(\mathbf{x}; \theta) I_C(\mathbf{x}), \tag{2.5.13}$$

lo cual implica que $\partial_\theta f(\mathbf{x}; \theta)$ se anula si $x \notin C$, mientras que

$$\lim_{\rho \to 0} T(\mathbf{x}) \partial_\theta f(\mathbf{x}; \theta + \tilde{\rho}(\mathbf{x}; \rho)) = T(\mathbf{x}) \partial_\theta f(\mathbf{x}; \theta).$$

Además, a partir de (2.5.4) se desprende que $|T(\mathbf{x})\partial_\theta f(\mathbf{x};\theta+\tilde{\rho}(\mathbf{x};\rho))| \leq |T(\mathbf{x})|nB^n$ y, via (2.4.1), el teorema de convergencia dominada permite concluir que

$$\begin{aligned}
\lim_{\rho\to 0}\frac{H(\theta+\rho)-H(\theta)}{\rho} &= \lim_{\rho\to 0}\int_C T(\mathbf{x})\partial_\theta f(\mathbf{x};\theta+\tilde{\rho}(\mathbf{x};\rho))\,d\mathbf{x}\\
&= \int_C \lim_{\rho\to 0} T(\mathbf{x})\partial_\theta f(\mathbf{x};\theta+\tilde{\rho}(\mathbf{x};\rho))\,d\mathbf{x}\\
&= \int_C T(\mathbf{x})\partial_\theta f(\mathbf{x};\theta)\,d\mathbf{x}
\end{aligned}$$

esto es,

$$H'(\theta) = \int_C T(\mathbf{x})\partial_\theta f(\mathbf{x};\theta)\,d\mathbf{x} = \int_{\mathbb{R}^n} T(\mathbf{x})\partial_\theta f(\mathbf{x};\theta)I_C(\mathbf{x})\,d\mathbf{x};$$

combinando esta relación con (2.5.13) se obtiene

$$H'(\theta) = \int_{\mathbb{R}^n} T(\mathbf{x})\partial_\theta f(\mathbf{x};\theta)\,d\mathbf{x}.$$

(ii) Considere el estadístico $\tilde{T}(\mathbf{X}) \equiv 1$, para el cual $\tilde{H}(\theta) = E_\theta[\tilde{T}(\mathbf{X})] = 1$. Aplicando el resultado de la parte (i) con $\tilde{T}(\mathbf{X})$ en lugar de $T(\mathbf{X})$ se desprende que $0 = \tilde{H}'(\theta) = \int_{\mathbb{R}^n} \tilde{T}(\mathbf{x})\partial_\theta f(\mathbf{x};\theta)\,d\mathbf{x} = \int_{\mathbb{R}^n} \partial_\theta f(\mathbf{x};\theta)\,d\mathbf{x}$. Por lo tanto,

$$\begin{aligned}
&\int_{\mathbb{R}^n} [T(\mathbf{x})-h(\theta)]\partial_\theta f(\mathbf{x};\theta)\,d\mathbf{x}\\
&\qquad = \int_{\mathbb{R}^n} T(\mathbf{x})\partial_\theta f(\mathbf{x};\theta)\,d\mathbf{x} - h(\theta)\int_{\mathbb{R}^n} \partial_\theta f(\mathbf{x};\theta)\,d\mathbf{x}\\
&\qquad = \int_{\mathbb{R}^n} T(\mathbf{x})\partial_\theta f(\mathbf{x};\theta)\,d\mathbf{x}
\end{aligned}$$

y entonces $\int_{\mathbb{R}^n}[T(\mathbf{x}) - h(\theta)]\partial_\theta f(\mathbf{x};\theta)\,d\mathbf{x} = H'(\theta)$, por la parte (i), y como la diferencia entre $H(\theta)$ y $h(\theta)$ es constante, se tiene que

$$\int_{\mathbb{R}^n} [T(\mathbf{x})-h(\theta)]\partial_\theta f(\mathbf{x};\theta)\,d\mathbf{x} = h'(\theta),$$

estableciendo (2.4.3). A partir de este punto, usando (2.5.2) se obtiene

$$\begin{aligned}
h'(\theta) &= \int_{\mathbb{R}^n} [T(\mathbf{x})-h(\theta)]\frac{\partial_\theta f(\mathbf{x};\theta)}{f(\mathbf{x};\theta)}f(\mathbf{x};\theta)\,d\mathbf{x}\\
&= E_\theta\left[(T(\mathbf{X})-h(\theta))\left(\frac{\partial_\theta f(\mathbf{X};\theta)}{f(\mathbf{X};\theta)}\right)\right]
\end{aligned}$$

y por medio de la desigualdad de Cauchy- Schwarz se desprende que

$$|h'(\theta)|^2 \leq E_\theta\left[(T(\mathbf{X}) - h(\theta))^2\right] E_\theta\left[\left(\frac{\partial_\theta f(\mathbf{X};\theta)}{f(\mathbf{X};\theta)}\right)^2\right],$$

de manera que $|h'(\theta)|^2 \leq n\mathcal{I}E_\theta\left[(T(\mathbf{X}) - h(\theta))^2\right]$, por el Lemma 2.5.1(v), concluyendo la demostración. □

Fórmula de la Segunda Derivada

El cálculo del número de información de Fisher frecuentemente se facilita con el uso de una fórmula alternativa que se presenta en el siguiente teorema. Note que si $g(x) \neq 0$ se tiene que

$$\frac{d\log(g(x))}{dx} = \frac{g'(x)}{g(x)},$$

y entonces

$$\frac{d^2\log(g(x))}{dx^2} = \frac{d}{dx}\left(\frac{g'(x)}{g(x)}\right). \tag{2.6.1}$$

Teorema 2.6.1. Suponga que las Hipótesis 2.2.1 y 2.4.1 son válidas y, adicionalmente, que $g(\cdot)$ tiene segunda derivada. Defina $\dfrac{d^2\log(g(x))}{dx^2}$ arbitrariamente como cero en los puntos en que $g(x) = 0$, y suponga que esta función es integrable. En este caso, el número de información de Fisher en (2.4.5) satisface

$$\mathcal{I} = -\int_{\mathbb{R}} \frac{d^2\log(g(x))}{dx^2} g(x)\,dx. \tag{2.6.2}$$

Demostración. Primeramente, note que $\mathcal{D} = \{x\colon g(x) \neq 0\} \subset (a, A)$ es una unión numerable de intervalos,

$$\mathcal{D} = \bigcup_j (a_j, A_j), \tag{2.6.3}$$

donde $g(a_j) = g(A_j) = 0$, de manera que el Lemma 2.5.1(ii) implica que

$$g'(a_j) = g'(A_j) = 0. \tag{2.6.4}$$

Considere ahora un intervalo (a_j, A_j) fijo, y note que para cada $\delta > 0$ que satisfaga $2\delta < A_j - a_j$. La fórmula de integración por partes permite concluir que

$$\begin{aligned}
&\int_{a_j+\delta}^{A_j-\delta} \frac{d^2 \log(g(x))}{dx^2} g(x)\, dx \\
&\qquad = \frac{d \log(g(x))}{dx} g(x) \Big|_{x=a_j+\delta}^{x=A_j-\delta} - \int_{a_j+\delta}^{A_j-\delta} \frac{d \log(g(x))}{dx} g'(x)\, dx \\
&\qquad = \frac{g'(x)}{g(x)} g(x) \Big|_{x=a_j+\delta}^{x=A_j-\delta} - \int_{a_j+\delta}^{A_j-\delta} \frac{g'(x)}{g(x)} g'(x)\, dx \\
&\qquad = g'(A_j - \delta) - g'(a_j + \delta) - \int_{a_j+\delta}^{A_j-\delta} \left(\frac{g'(x)}{g(x)}\right)^2 g(x)\, dx
\end{aligned}$$

Por otro lado, la continuidad de $g'(\cdot)$ y (2.6.4) implican que

$$\lim_{\delta \to 0} g'(A_j - \delta) - g'(a_j + \delta) = g'(A_j) - g'(a_j) = 0,$$

mientras que la integrabilidad de $[g'(x)/g(x)]^2$ permite concluir que

$$\lim_{\delta \to 0} \int_{a_j+\delta}^{A_j-\delta} \left(\frac{g'(x)}{g(x)}\right)^2 g(x)\, dx = \int_{a_j}^{A_j} \left(\frac{g'(x)}{g(x)}\right)^2 g(x)\, dx.$$

Combinando las últimas tres relaciones desplegadas, y usando el hecho de que $\dfrac{d^2 \log(g(x))}{dx^2}$ es integrable, se desprende que

$$\begin{aligned}
\int_{a_j}^{A_j} \frac{d^2 \log(g(x))}{dx^2} g(x)\, dx &= \lim_{\delta \to 0} \int_{a_j+\delta}^{A_j-\delta} \frac{d^2 \log(g(x))}{dx^2} g(x)\, dx \\
&= -\int_{a_j}^{A_j} \left(\frac{g'(x)}{g(x)}\right)^2 g(x)\, dx,
\end{aligned}$$

y via (2.6.3) se desprende que

$$\begin{aligned}
\int_{\mathcal{D}} \frac{d^2 \log(g(x))}{dx^2} g(x)\, dx &= \sum_j \int_{a_j}^{A_j} \frac{d^2 \log(g(x))}{dx^2} g(x)\, dx \\
&= \sum_j -\int_{a_j}^{A_j} \left(\frac{g'(x)}{g(x)}\right)^2 g(x)\, dx \\
&= -\int_{\mathcal{D}} \left(\frac{g'(x)}{g(x)}\right)^2 g(x)\, dx.
\end{aligned}$$

Puesto que $g(\cdot)$ se anula fuera del conjunto $\mathcal{D}$, esta relación equivale a

$$\int_{\mathbb{R}} \frac{d^2 \log(g(x))}{dx^2} g(x)\, dx = -\int_{\mathbb{R}} \left(\frac{g'(x)}{g(x)}\right)^2 g(x)\, dx = -\mathcal{I},$$

concluyendo la demostración. $\square$

Ejemplo 2.6.1. Considere la densidad beta con parámetros $\alpha, \beta > 0$ trasladada media unidad a la izquierda, esto es, por

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}(x+1/2)^{\alpha-1}(1-(x+1/2))^{\beta-1} I_{(0,1)}(x+1/2);$$

vea las partes (c) y (d) del Ejemplo 2.2.1. A continuación el número de información de Fisher se calculará usando ambas fórmulas (2.4.5) y (2.6.2). Aplicando una traslación de media unidad a la derecha la anterior densidad se transforma en

$$g(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} I_{(0,1)}(x)$$

y, por el Lemma 2.5.1(v) aplicado con $n = 1$, el número de información no se altera. En este caso $g(x)$ es no nula si $x \in (0,1)$, y $\log(g(x)) = (\alpha-1)\log(x) + (\beta-1)\log(1-x) + \text{constante}$. Por lo tanto

$$\frac{d\log(g(x))}{dx} = \frac{g'(x)}{g(x)} = \frac{\alpha-1}{x} - \frac{\beta-1}{1-x};$$

$$\frac{d^2\log(g(x))}{dx^2} = -\frac{\alpha-1}{x^2} - \frac{\beta-1}{(1-x)^2};$$

A partir de resultados establecidos, por ejemplo, en Dudewicz y Mishra (1989) o en Casella y Berger (2001), se tiene que para $\alpha, \beta > 2$,

$$\begin{aligned} \int_0^1 \frac{1}{x^2} g(x)\, dx &= \frac{C_{\alpha,\beta}}{(\alpha-1)(\alpha-2)}, \\ \int_0^1 \frac{1}{(1-x)^2} g(x)\, dx &= \frac{C_{\alpha,\beta}}{(\beta-1)(\beta-2)}, \\ \int_0^1 \frac{1}{x(1-x)} g(x)\, dx &= \frac{C_{\alpha,\beta}}{(\alpha-1)(\beta-1)}, \end{aligned} \qquad (2.6.5)$$

donde

$$C_{\alpha,\beta} = (\alpha+\beta-1)(\alpha+\beta-2). \tag{2.6.6}$$

Usando (2.4.5), $\mathcal{I}$ se calcula como sigue:

$$\begin{aligned}
\mathcal{I} &= \int_0^1 \left(\frac{g'(x)}{g(x)}\right)^2 g(x)\,dx \\
&= \int_0^1 \left(\frac{\alpha-1}{x} - \frac{\beta-1}{1-x}\right)^2 g(x)\,dx \\
&= \int_0^1 \left(\frac{(\alpha-1)^2}{x^2} + \frac{(\beta-1)^2}{(1-x)^2} - 2\frac{(\alpha-1)(\beta-1)}{x(1-x)}\right) g(x)\,dx \\
&= \left[\frac{\alpha-1}{\alpha-2} + \frac{\beta-1}{\beta-2} - 2\right] C_{\alpha,\beta}
\end{aligned}$$

y via (2.6.6) se obtiene

$$\mathcal{I} = \left[\frac{1}{\alpha-2} + \frac{1}{\beta-2}\right] C_{\alpha,\beta} = \frac{(\alpha+\beta-4)(\alpha+\beta-1)(\alpha+\beta-2)}{(\alpha-2)(\beta-2)}. \tag{2.6.7}$$

Por otro lado, via (2.6.5), $\mathcal{I}$ se calcula por medio de (2.6.2) como sigue:

$$\begin{aligned}
\mathcal{I} &= -\int_0^1 \frac{d^2\log(g(x))}{dx^2} g(x)\,dx \\
&= -\int_0^1 \left[-\frac{\alpha-1}{x^2} - \frac{\beta-1}{(1-x)^2}\right] g(x)\,dx \\
&= \int_0^1 \left[\frac{\alpha-1}{x^2} + \frac{\beta-1}{(1-x)^2}\right] g(x)\,dx \\
&= \left[\frac{\alpha-1}{(\alpha-1)(\alpha-2)} + \frac{\beta-1}{(\beta-1)(\beta-2)}\right] C_{\alpha,\beta} \\
&= \left[\frac{1}{\alpha-2} + \frac{1}{\beta-2}\right] C_{\alpha,\beta}
\end{aligned}$$

expresión que coincide con (2.6.7). $\square$

Capítulo 3

DOS PROCEDIMIENTOS INVARIANTES

Introducción

En este capítulo se estudian dos métodos de estimación del parámetro θ, a saber, las técnicas de momentos y de cuantiles, las cuales generan estimadores invariantes que, consecuentemente, satisfacen la cota de Cramer–Rao. El propósito fundamental es disponer de estimadores que permitan obtener intervalos de confianza para θ con un nivel asintótico determinado, los que más adelante serán comparados con los intervalos que se generarán mediante el método de verosimilitud máxima. Para llevar a cabo esta comparación, es necesario estudiar la relación entre el error cuadrático medio del estimador para muestras finitas, y la varianza de su distribución límite. Esta tarea es directa para el estimador de momentos, pero en el caso del estimador de cuantiles requiere de un análisis técnico particular, el cual involucra un estudio detallado de la relación entre estadísticos de orden y la distribución binomial.

La organización del material es la siguiente: En la Sección 2 se introducen los estimadores de momentos, se utiliza su normalidad asintótica para construir intervalos de confianza para θ, y la validez de la cota de Cramer–Rao para la varianza de la distribución límite se ilustra en el caso de densidades beta. En la Sección

3 se estudia la estimación por cuantiles, utilizándolos para obtener intervalos de confianza para el parámetro. Sin embargo, establecer la cota de Cramer-Rao para la varianza de la distribución límite de estos estimadores no es una tarea directa, y la verificación de este resultado se lleva a cabo en las Secciones 4–6, siendo el Teorema 3.4.1 la principal contribución técnica del capítulo.

El Método de Momentos

Dada una muestra $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ de una distribución con densidad $h(\cdot;\theta)$, el método de momentos para estimar θ consiste de tres pasos (Graybill, 1985, Casella y Berger, 2001). Primero, es necesario expresar al parámetro θ en términos de los momentos $\mu_j(\theta)$ de la densidad $h(\cdot;\theta)$, obteniendo una fórmula como

$$\theta = H(\mu_1(\theta), \ldots, \mu_r(\theta)). \tag{3.2.1}$$

Posteriormente, cada momento $\mu_j(\theta)$ se estima mediante el correspondiente momento muestral $\hat{\mu}_j$ dado por

$$\hat{\mu}_j = \frac{1}{n}\sum_{i=1}^{n} X_i^j,$$

y el último paso consiste en sustituir cada momento poblacional $\mu_j(\theta)$ en (3.2.1) por su estimador $\hat{\mu}_j$ para obtener el estadístico

$$H(\hat{\mu}_1, \ldots, \hat{\mu}_r), \tag{3.2.2}$$

el cual es un estimador de momentos de θ. A continuación se aplicará este método para estimar θ en el modelo de traslación, en el cual el vector de observaciones $\mathbf{X}$ es una muestra de la distribución con densidad $g(x-\theta)$. El punto de partida es determinar fórmulas para los momentos de la densidad $g(x-\theta)$ hasta que sea posible expresar a θ en términos de esos momentos. Note que

$$\mu_1(\theta) = \int_{\mathbb{R}} x g(x-\theta)\,dx = \int_{\mathbb{R}} (x+\theta) g(x)\,dx = \mu_g + \theta, \tag{3.2.3}$$

donde $\mu_g = \int_{\mathbb{R}} x\, g(x)\, dx$ es la esperanza de la densidad $g(\cdot)$; debido a que $g(\cdot)$ es una densidad conocida, la constante μ_g también lo es. A partir de la expresión anterior ya es posible expresar a θ en términos del primer momento de $g(x-\theta)$:

$$\theta = \mu_1(\theta) - \mu_g = H(\mu_1(\theta)). \tag{3.2.4}$$

Como el primer momento muestral es

$$\hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i =: \overline{X}_n$$

el estimador de momentos buscado se obtiene sustituyendo $\mu_1(\theta)$ por $\overline{X}_n$ en (3.2.4) para obtener el estadístico

$$T_{1,\,n}(\mathbf{X}) = \overline{X}_n - \mu_g. \tag{3.2.5}$$

Debido a que la media muestral de $\mathbf{X} + c\mathbb{1}$ es $\overline{X}_n + c$, se desprende que $T_{1,\,n}(\mathbf{X})$ es invariante bajo traslación; vea la Definición 2.3.1.

Teorema 3.2.1. El estimador de momentos $T_{1,\,n}(\mathbf{X})$ en (3.2.5) satisface las siguientes propiedades (i) y (ii):

(i) Para cada $\theta \in \Theta$

$$E_\theta[T_{1,\,n}(\mathbf{X})] = \theta \qquad \text{y} \qquad \mathrm{Var}_\theta[T_{1,\,n}] = \frac{\sigma_g^2}{n},$$

donde

$$\sigma_g^2 = \int_R (x-\mu_g)^2 g(x)\, dx$$

es la varianza de la densidad $g(\cdot)$.

Además

$$\sigma_g^2 \geq \frac{1}{\mathcal{I}}. \tag{3.2.6}$$

(ii) Denote por θ_0 al verdadero valor del parámetro. En este caso

$$\sqrt{n}[T_{1,\,n} - \theta_0] \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, \sigma_g^2\right) \quad \text{conforme } n \to \infty. \tag{3.2.7}$$

Demostración. Note que la varianza de la densidad $g(x-\theta)$ es

$$\begin{aligned}\sigma_g^2(\theta) &= \int_{\mathbb{R}} (x-\mu_1(\theta))^2 g(x-\theta)\,dx \\ &= \int_{\mathbb{R}} (x-[\theta+\mu_g])^2 g(x-\theta)\,dx \\ &= \int_{\mathbb{R}} (y-\mu_g)^2 g(y)\,dy\end{aligned}$$

donde la segunda igualdad se debe a (3.2.3), y se utilizó el cambio de variable $y = x-\theta$ en la última etapa. Por lo tanto $\sigma_g^2(\theta) = \sigma_g^2$ para todo $\theta \in \Theta$. Puesto que $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ es una muestra de la distribución con densidad $g(x-\theta)$, se tiene que

$$E_\theta[\overline{X}_n] = \mu_1(\theta) = \theta + \mu_g \quad \text{y} \quad \mathrm{Var}_\theta[\overline{X}_n] = \frac{1}{n}\sigma_g^2(\theta) = \frac{1}{n}\sigma_g^2.$$

Por lo tanto, $E_\theta[T_{1,n}] = E_\theta[\overline{X}_n - \mu_g] = \theta$, y $\mathrm{Var}_\theta[T_{1,n}] = \mathrm{Var}_\theta[\overline{X}_n - \mu_g] = \mathrm{Var}_\theta[\overline{X}_n] = \frac{1}{n}\sigma_g^2(\theta) = \frac{1}{n}\sigma_g^2$; como (3.2.6) se obtiene de la desigualdad de Cramer-Rao (2.4.4) con $h(\theta) = \theta$, esto establece la parte (i). Para concluir, note que

$$\sqrt{n}[T_{1,n} - \theta_0] = \sqrt{n}[(\overline{X}_n - \mu_g) - \theta_0] = \sqrt{n}[\overline{X}_n - \mu_1(\theta_0)],$$

por lo que la parte (ii) se desprende del teorema central de límite. □

La convergencia (3.2.7) permite construir intervalos de confianza para el verdadero valor del parámetro θ_0 con un nivel asintótico especificado. Dado $\gamma \in (0,1)$, sea $z_{\gamma/2} \in \mathbb{R}$ tal que $P[Z > z_{\gamma/2}] = \gamma/2$, donde Z tiene distribución normal estándar. Con esta notación, (3.2.7) implica que

$$\lim_{n\to\infty} P_{\theta_0}\left[\frac{\sqrt{n}}{\sigma_g}|T_{1,n} - \theta_0| \le z_{\gamma/2}\right] = 1-\gamma$$

Puesto que la desigualdad $\dfrac{\sqrt{n}}{\sigma_g}|T_{1,n} - \theta_0| \le z_{\gamma/2}$ equivale a $\theta_0 \in I_n$, donde

$$I_n = \left[T_{1,n} - \sigma_g z_{\gamma/2}/\sqrt{n},\ T_{1,n} + \sigma_g z_{\gamma/2}/\sqrt{n}\right] \tag{3.2.8}$$

se obtiene la siguiente conclusión.

Corolario 3.2.1. (i) Conforme $n \to \infty$,

$$P_{\theta_0}[\theta_0 \in I_n] \to 1 - \gamma,$$

esto es, como intervalos de confianza para θ_0, los intervalos I_n en (3.2.8) tiene un nivel de confianza asintótico $1 - \gamma$. Más aún,

(ii) La longitud $|I_n|$ del intervalo en (3.2.8) es $|I_n| = 2\sigma_g z_{\gamma/2}/\sqrt{n}$.

Observación 3.2.1. Considere la densidad $g(x)$ en el Ejemplo 2.2.1(c), de manera que $g(x)$ es simétrica respecto al origen. En este caso $\mu_g = 0$ y $T_{1,n}(\mathbf{X}) = \overline{X}_n$. Como ya se ha mencionado, con la finalidad de utilizar las herramientas disponibles para un modelo con errores normales, por simplicidad puede suponerse que la densidad $g(\cdot)$ en el modelo de traslación es normal con media cero y, más aún, que su varianza σ^2 es desconocida. En este caso, los mejores estimadores de θ y de σ^2 son $\overline{X}_n = T_{1,n}(\mathbf{X})$, y $s_n^2 = \sum_{i=1}^n (X_i - \overline{X}_n)^2/(n-1)$ (Graybill, 1985, Lehmann, 2001). Bajo este supuesto de normalidad—erróneo, pero conveniente para el analista—el intervalo de confianza para θ con nivel $1 - \gamma$ es

$$I'_n = \left[\overline{X}_n - t_{n-1,\gamma/2}\frac{s_n}{\sqrt{n}}, \overline{X}_n + t_{n-1,\gamma/2}\frac{s_n}{\sqrt{n}}\right],$$

y la longitud de I'_n es

$$|I'_n| = 2t_{n-1,\gamma/2}\frac{s}{\sqrt{n}}.$$

Note ahora que, a pesar de que (erróneamente) se suponga normalidad de los errores, la varianza muestral s_n^2 converge a la varianza poblacional σ_g^2, esto es, $s_n/\sigma_g \to 1$ con probabilidad uno. Por otro lado, usando que $\lim_{n\to\infty} t_{n-1,\gamma/2} = z_{\gamma/2}$, se desprende que

$$\frac{|I_n|}{|I'_n|} = \frac{2z_{\gamma/2}\sigma_g/\sqrt{n}}{2t_{n-1,\gamma/2}s/\sqrt{n}} = \left(\frac{z_{\gamma/2}}{t_{n-1,\gamma/2}}\right)\left(\frac{\sigma_g}{s_n}\right) \to 1 \qquad P_{\theta_0}\text{–c. s.};$$

en este sentido, los intervalos I_n e I'_n son asintóticamente equivalentes.

Ejemplo 3.2.1. Considere el modelo de traslación en el que la densidad $g(\cdot)$ es como en el Ejemplo 2.2.1 (d), i.e.,

$$g(z) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \left(x + \frac{1}{2}\right)^{\alpha-1} \left(\frac{1}{2} - x\right)^{\beta-1} I_{[-1/2,\, 1/2]}(x), \tag{3.2.9}$$

donde α y β son mayores que dos.

La media y la varianza de g son

$$\mu_g = \frac{\alpha}{\alpha+\beta} - \frac{1}{2} \qquad \text{y} \qquad \sigma_g^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}, \tag{3.2.10}$$

respectivamente. El estimador de momentos en (3.2.5) es $T_{1,\,n} = \overline{X}_n - \mu_g$, el error cuadrático medio de $T_{1,\,n}$ es

$$E_\theta[(T_{1,\,n} - \theta)^2] = \frac{\sigma_g^2}{n},$$

y los intervalos de confianza I_n en (3.2.8) tienen longitud

$$|I_n| = \frac{2z_{\gamma/2}}{\sqrt{n}}\sigma_g = \frac{2z_{\gamma/2}}{\sqrt{n}} \sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}}.$$

□

En el contexto del ejemplo anterior, la desigualdad de Cramer–Rao dada por $E_\theta[(T_{1,\,n} - \theta)^2] \geq 1/[n\mathcal{I}]$ equivale a

$$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \geq \frac{(\alpha-2)(\beta-2)}{(\alpha+\beta-1)(\alpha+\beta-2)(\alpha+\beta-4)}; \tag{3.2.11}$$

vea (3.2.10) y el Ejemplo 2.6.7 del Capítulo precedente. En esta relación, el numerador y denominador en el lado derecho son menores a los correspondientes al lado izquierdo, y no es inmediatamente claro que (3.2.11) sea válida. Aunque el Teorema

2.4.1 establece que esta relación es cierta, es interesante verificarla directamente, lo cual se hace a continuación.

Proposición 3.2.1. Si α y β son mayores que 2, entonces (3.2.11) es válida.

Demostración. Escriba $s = \alpha + \beta$, de manera que (3.2.11) expresa que

$$\frac{\alpha\beta}{s^2(s+1)} \geq \frac{\alpha\beta - 2s + 4}{(s-1)(s-2)(s-4)}, \quad \alpha, \beta > 2, \quad s = \alpha + \beta.$$

Simplificaciones sucesivas de esta relación permiten establecer las siguientes desigualdades equivalentes:

$$\begin{aligned}
\alpha\beta(s-1)(s-2)(s-4) &\geq (\alpha\beta - 2s + 4)s^2(s+1) \\
2(s-2)s^2(s+1) &\geq \alpha\beta[s^2(s+1) - (s-1)(s-2)(s-4)] \\
2(s-2)s^2(s+1) &\geq \alpha\beta[2s^2 + s^2(s-1) - (s-1)(s-2)(s-4)] \\
2(s-2)s^2(s+1) &\geq \alpha\beta[2s^2 + (s-1)(s^2 - (s-2)(s-4))] \\
2(s-2)s^2(s+1) &\geq \alpha\beta[2s^2 + (s-1)(6s-8)] \\
2(s-2)s^2(s+1) &\geq \alpha\beta[8s^2 - 14s + 8]
\end{aligned}$$

Por lo tanto, lo que se debe demostrar es que

$$2(s-2)s^2(s+1) \geq \alpha\beta[8s^2 - 14s + 8], \quad \alpha, \beta > 2, \quad s = \alpha + \beta.$$

Como el valor máximo del producto $\alpha\beta$ sujeto a $\alpha + \beta = s$ es $s^2/4$, la anterior relación desplegada equivale a que, para todo $s > 4$, la siguiente desigualdad sea válida:

$$2(s-2)s^2(s+1) \geq \frac{s^2}{4}[8s^2 - 14s + 8]. \tag{3.2.12}$$

Simplificaciones sucesivas de esta desigualdad permiten obtener la siguiente cadena de enunciados equivalentes cuando $s \neq 0$:

$$8(s-2)s^2(s+1) \geq s^2[8s^2 - 14s + 8]$$

$$8(s-2)(s+1) \geq [8s^2 - 14s + 8]$$

$$8[s^2 - s - 2] \geq [8s^2 - 14s + 8]$$

$$-8s - 16 \geq -14s + 8$$

$$6s \geq 24$$

Esto muestra que la condición $s \geq 4$ es equivalente a (3.2.12), mostrando que, efectivamente, esta última desigualdad ocurre estrictamente cuando $s > 4$. $\square$

Estimación por Cuantiles

En esta sección se introduce un método adicional para generar estimadores de θ que son invariantes bajo traslación. Dada una densidad $m(x)$ en $\mathbb{R}$ y $p \in (0,1)$, un número ξ_p se denomina *cuantil de orden p* de la densidad $m(\cdot)$, si

$$\int_{-\infty}^{\xi_p} m(x)\, dx = p,$$

de tal manera que la probabilidad acumulada desde $-\infty$ hasta ξ_p es exactamente p. Una condición suficiente para que un cuantil de orden p sea único, es que la densidad $m(\cdot)$ sea continua y positiva en ξ_p. Si se dispone de una muestra $X_1, X_2, \ldots, X_n$ de la densidad $m(x)$ y se desea aproximar el valor de ξ_p, un procedimiento natural es el siguiente: Determine los estadísticos de orden $Y_1, Y_2, \ldots, Y_n$ de la muestra, y aproxime a ξ_p mediante $Y_{[np]}$. Esta idea es intuitivamente atractiva pues, dentro de la muestra, la proporción de valores desde $-\infty$ hasta $Y_{[np]}$ es $[np]/n \approx p$, de manera que la posición de $Y_{[np]}$ dentro de la muestra es similar a la ubicación de ξ_p dentro de la población. Suponga ahora que $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ es una muestra de una densidad $h(x;\theta)$ que depende de un parámetro θ. Como el método de momentos, la técnica de cuantiles para estimar θ consiste de tres pasos (Dudewicz y Mishra 1989, Casella y Berger 2001): Primero, dado $p \in (0,1)$, es necesario calcular el cuantil $\xi_p(\theta)$ de la densidad $h(x;\theta)$. El siguiente paso es expresar el parámetro θ en términos de $\xi_p(\theta)$ para obtener

$$\theta = H(\xi_p(\theta)).$$

Finalmente, en la expresión anterior se sustituye el percentil $\xi_p(\theta)$ por el estimador $Y_{[np]}$ obteniendo el estadístico

$$H(Y_{[np]})$$

el cual es el estimador de cuantiles de θ. A continuación, se aplicará este método para determinar el estimador de cuantiles en el modelo de traslación. Como punto de partida, denote por ξ_p^g al cuantil de orden $p \in (0,1)$ de la densidad $g(\cdot)$, el cual satisface,

$$\int_{-\infty}^{\xi_p^g} g(x)\,dx = p.$$

Usando el cambio de variable $z = x + \theta$, se desprende que

$$p = \int_{-\infty}^{\xi_p^g} g(x)\,dx = \int_{\infty}^{\theta+\xi_p^g} g(z-\theta)\,dz,$$

y en consecuencia el cuantil de orden p de $g(x-\theta)$ es

$$\xi_p(\theta) = \theta + \xi_p^g, \tag{3.3.1}$$

de manera que $\theta = \xi_p(\theta) - \xi_p^g$. Sustituyendo $\xi_p(\theta)$ por su estimador $Y_{[np]}$, se obtiene el estimador de cuantiles de θ:

$$T_{2,n}(\mathbf{X}) = Y_{[np]} - \xi_p^g. \tag{3.3.2}$$

Note que si Y_k es el k-ésimo estadístico de orden asociado al vector $\mathbf{X}$, entonces el k-ésimo estadístico de orden asociado a $\mathbf{X} + c\mathbb{1}$ es $Y_k + c$, de donde se desprende que $T_{2,n}(\mathbf{X})$ es invariante bajo traslación. La distribución asintótica del estimador de cuantiles de θ está dada en el siguiente resultado.

Teorema 3.3.1. Sea $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ una muestra de la densidad $g(x - \theta)$. Si $g(\cdot)$ es continua y no se anula en el cuantil ξ_p^g, entonces

$$\sqrt{n}\,[T_{2,n}(\mathbf{X}) - \theta] = \sqrt{n}\,[Y_{[np]} - \xi_p(\theta)] \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, \frac{p(1-p)}{g(\xi_p^g)^2}\right). \tag{3.3.3}$$

Observando que

$$T_{2,n}(\mathbf{X}) - \theta = Y_{[np]} - \xi_p^g - \theta = Y_{[np]} - \xi_p(\theta)$$

y $g(\xi_p(\theta) - \theta) = g(\xi_p^g)$, una demostración puede verse, por ejemplo, en Casella y Berger (2001), Dudewicz y Mishra (1989), o en Sefling (1988). Dado $\gamma \in (0,1)$, este resultado implica que

$$\lim_{n\to\infty} P_\theta \left[\sqrt{n}|T_{2,n}(\mathbf{X}) - \theta| \leq z_{\gamma/2} \sqrt{\frac{p(1-p)}{g(\xi_p^g)^2}} \right] = 1 - \gamma;$$

puesto que la desigualdad en esta relación equivale a $\theta \in J_n$, donde

$$J_n = \left[T_{2,n}(\mathbf{X}) - \frac{z_{\gamma/2}}{\sqrt{n}} \sqrt{\frac{p(1-p)}{g(\xi_p^g)^2}},\ T_{2,n}(\mathbf{X}) + \frac{z_{\gamma/2}}{\sqrt{n}} \sqrt{\frac{p(1-p)}{g(\xi_p^g)^2}} \right], \tag{3.3.4}$$

se desprende el siguiente resultado.

Corolario 3.3.1. (i) Conforme $n \to \infty$,

$$P_\theta[\theta \in J_n] \to 1 - \gamma,$$

es decir, como intervalos de confianza para θ, los intervalos J_n en (3.3.4) tiene un nivel de confianza asintótico $1 - \gamma$. Más aún,

(ii) La longitud $|J_n|$ del intervalo en (3.3.4) es

$$|J_n| = 2\frac{z_{\gamma/2}}{\sqrt{n}} \sqrt{\frac{p(1-p)}{g(\xi_p^g)^2}}.$$

Ejemplo 3.3.1. En el modelo de traslación suponga que $g(x)$ es la densidad beta en (3.2.9), donde $\alpha = \beta > 2$. Seleccionando $p = 1/2$ se tiene que $\xi_p^g = 0$ es la mediana de la densidad $g(\cdot)$, y $T_{2,\,n}(\mathbf{X}) = Y_{[n/2]}$ es la mediana muestral. Observando que

$$\sqrt{\frac{p(1-p)}{g(\xi_p^g)^2}} = \frac{1}{2g(0)} = \frac{2^{2\alpha-3}\Gamma(\alpha)^2}{\Gamma(2\alpha)},$$

en este caso el Teorema 3.3.1 establece que

$$\sqrt{n}\,[T_{2,\,n}(\mathbf{X}) - \theta] \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, \left[\frac{2^{2\alpha-3}\Gamma(\alpha)^2}{\Gamma(2\alpha)}\right]^2\right)$$

y la longitud del intervalo J_n en (3.3.4) es

$$|J_n| = \frac{2z_{\gamma/2}}{\sqrt{n}} \frac{2^{2\alpha-3}\Gamma(\alpha)^2}{\Gamma(2\alpha)}.$$

□

Error Cuadrático Medio

En esta sección se estudia el comportamiento límite del segundo momento de $[T_{2,\,n} - \theta]$, y el principal objetivo es establecer el siguiente resultado.

Teorema 3.4.1. Dado $p \in (0, 1)$, sea $T_{2,\,n}(\mathbf{X})$ el estimador en (3.3.2). Suponga que las Hipótesis 2.2.1 y 2.4.1 son válidas, y que $g(\xi_p^g) \neq 0$. En estas circunstancias

$$\lim_{n\to\infty} nE_\theta[(T_{2,\,n}(\mathbf{X}) - \theta)^2] = \frac{p(1-p)}{g(\xi_p^g)^2}.$$

La importancia de este resultado se debe a que permite comparar la varianza asintótica de $\sqrt{n}(T_{2,\,n}(\mathbf{X}) - \theta)$—dada por el lado derecho de la anterior ecuación—con el recíproco $1/\mathcal{I}$ del número de información de Fisher en el modelo de traslación. En efecto, como $T_{2,\,n}(\mathbf{X})$ es invariante bajo traslación, se tiene que

$$E_\theta[(T_{2,\,n}(\mathbf{X}) - \theta)^2] \geq \frac{1}{n\mathcal{I}},$$

por el Teorema 2.4.1(ii), de manera que el siguiente resultado se desprende de inmediato del Teorema 3.4.1.

Corolario 3.4.1. Bajo las condiciones del Teorema 3.4.1,

$$\frac{p(1-p)}{g(\xi_p^g)^2} \geq \frac{1}{\mathcal{I}}.$$

Antes de proporcionar una demostración del Teorema 3.4.1, es conveniente discutir la razón por la cual este resultado es interesante. Primero, escriba

$$W_n = \sqrt{n}\,[T_{2,\,n}(\mathbf{X}) - \theta], \quad \text{y} \quad v = \frac{p(1-p)}{g(\xi_p^g)^2}. \tag{3.4.1}$$

Con esta notación, se estableció en el Teorema 3.3.1 que

$$W_n \xrightarrow{\mathrm{d}} \mathcal{N}\,(0, v)\,, \tag{3.4.2}$$

y lo que el Teorema 3.4.1 asegura es que

$$\lim_{n\to\infty} E[W_n^2] = v. \tag{3.4.3}$$

A partir de estas dos últimas convergencias, la siguiente pregunta surge naturalmente:

¿Es posible obtener la convergencia (3.4.3) a partir de la normalidad asintótica en (3.4.2)?

El siguiente ejemplo muestra que la respuesta es negativa, y este hecho hace que el resultado del Teorema 3.4.1 sea interesante, pues se requiere un análisis especial para establecer (3.4.3) en el caso en que $W_n = \sqrt{n}[T_{2,\,n}(\mathbf{X}) - \theta]$.

Ejemplo 3.4.1. Considere una variable aleatoria X_0 con distribución normal estandar y para cada n, suponga que X_n es independiente de X_0 con la siguiente distribución:

$$P[X_n = n] = P[X_n = -n] = \frac{1}{2n^2}, \quad P[X_n = 0] = 1 - \frac{1}{n^2}.$$

En este caso se tiene que $E[X_n] = 0$ y $\mathrm{Var}\,(X_n) = 1$. Más aún, para cada $n \geq 1$, $P[|X_n| > 0] = P[X_n = n] + P[X_n = -n] = 1/n^2$, de manera que

$$X_n \xrightarrow{\mathrm{P}} 0. \tag{3.4.4}$$

Defina ahora

$$W_n = \frac{X_0 + X_n}{\sqrt{2}},$$

y observe que

$$E[W_n^2] = \mathrm{Var}\,(W_n) = 1. \tag{3.4.5}$$

Por otro lado, como $X_0 \sim \mathcal{N}\,(0,1)$, se tiene que $X_0/\sqrt{2} \sim \mathcal{N}\,(0,1/2)$, enunciado que permite escribir

$$\frac{X_0}{\sqrt{2}} \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, \frac{1}{2}\right).$$

Combinando esta convergencia con (3.4.4), el teorema de Slutsky implica que

$$W_n = \frac{X_0 + X_n}{\sqrt{2}} \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, \frac{1}{2}\right),$$

esto es, (3.4.2) ocurre en este ejemplo con $v = 1/2$. Sin embargo, a partir de (3.4.5) se tiene que $\lim_{n\to\infty} E[W_n^2] = 1 > v = 1/2$, mostrando que, en general, la convergencia (3.4.3) no se desprende de la normalidad asintótica (3.4.2). □

Este ejemplo muestra que la validez del Teorema 3.4.1 depende de propiedades especiales del estimador $T_{2,\,n} = Y_{[np]} - \xi_p^g$, cuyo estudio se inicia en la siguiente sección.

Conexión con la Distribución Binomial

En esta sección se establecen las propiedades del estimador $T_{2,\,n}(\mathbf{X}) = Y_{[np]} - \xi_p^g$ que posteriormente se utilizarán para demostrar el Teorema 3.4.1. El análisis subsecuente se refiere a las probabilidades de los eventos

$$[|T_{2,\,n}(\mathbf{X}) - \theta| > x] = [|Y_{[np]} - \xi_p^g - \theta| > x] = [|Y_{[np]} - \xi_p(\theta)| > x], \qquad (3.5.1)$$

donde $x > 0$, y el principal objetivo es establecer el siguiente resultado.

Teorema 3.5.1. Bajo las condiciones del Teorema 3.4.1, existe una constante $C > 0$ tal que

$$P_\theta[n(T_{2,\,n} - \xi_p(\theta))^2 > x] \leq \frac{C}{x^2}, \quad x > 0, \quad \theta \in \Theta. \qquad (3.5.2)$$

El argumento para establecer este resultado depende de los instrumentos preliminares establecidos en los cuatro siguientes lemas, los cuales se refieren a la relación entre los estadísticos de orden y la distribución binomial. Como punto de partida, observe que $Y_{[np]} > z$ significa que el número de datos observados menores o iguales a z es menor a $[np]$, esto es

$$Y_{[np]} > z \iff \sum_{i=1}^{n} I[X_i \leq z] < [np] \qquad (3.5.3)$$

mientras que $Y_{[np]} \leq z$ ocurre si y sólo si el número de datos observados menores o iguales a z es, por lo menos $[np]$, de manera que

$$Y_{[np]} \leq z \iff \sum_{i=1}^{n} I[X_i \leq z] \geq [np]. \qquad (3.5.4)$$

Para cada z, defina la variable aleatoria $N_n(z)$ mediante

$$N_n(z) = \sum_{i=1}^{n} I[X_i \leq z]. \qquad (3.5.5)$$

Recordando que $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ es una muestra de la densidad $g(x - \theta)$ se tiene que las variables aleatorias $I[X_i \leq z]$ son independientes y que su distribución común es $Ber(G(z - \theta))$, donde $G(\cdot)$ es la función de distribución de $g(x)$. Por lo tanto,

$$N_n(z) \sim \mathcal{B}(n,\ G(z - \theta)). \qquad (3.5.6)$$

Las relaciones (3.5.3)–(3.5.6) muestran la estrecha conexión entre la distribución de $Y_{[np]}$ y las probabilidades binomiales. La siguiente consecuencia de esta discusión será útil.

Lema 3.5.1. Para cada $w > 0$, las siguientes afirmaciones (i)–(iii) son válidas.

(i) $[Y_{[np]} - \xi_p(\theta) > w] = [N_n(\xi_p(\theta) + w) < [np]]$;

(ii) $[Y_{[np]} - \xi_p(\theta) \leq -w] = [N_n(\xi_p(\theta) - w) \geq [np]]$.

(iii) $N_n(\xi_p(\theta) + w) \sim \mathcal{B}(n,\ G(\xi_p^g + w))$ y $N_n(\xi_p(\theta) - w) \sim \mathcal{B}(n,\ G(\xi_p^g - w))$.

Demostración. Observando que $[Y_{[np]} - \xi_p(\theta) > w] = [Y_{[np]} > \xi_p(\theta) + w]$, la parte (i) se desprende de (3.5.3) y (3.5.5) con $\xi_p(\theta) + w$ en lugar de z. Similarmente, puesto que $[Y_{[np]} - \xi_p(\theta) \leq -w] = [Y_{[np]} \leq \xi_p(\theta) - w]$, la parte (ii) se obtiene de (3.5.4) y (3.5.5) sustituyendo z por $\xi_p(\theta) - w$.

(iii) Via (3.5.6) se tiene que $N_n(\xi_p(\theta) + w) \sim \mathcal{B}(n,\ G(\xi_p(\theta) + w - \theta))$ y $N_n(\xi_p(\theta) - w) \sim \mathcal{B}(n,\ G(\xi_p(\theta) - w - \theta))$, y la conclusión deseada se obtiene recordando que $\xi_p(\theta) = \theta + \xi_p^g$; vea (3.3.1). □

A continuación se obtiene una cota para el cuarto momento de una distribución binomial.

Lema 3.5.2. Suponga que $N \sim \mathcal{B}(n,\ p)$. En este caso las siguientes desigualdades son válidas:

$$E[(N - np)^4] \leq 3n^2, \tag{3.5.7}$$

y para cada $\delta \in (0, 1)$

$$E[(\delta + N - np)^4] \leq 22n^2. \tag{3.5.8}$$

Demostración. Sin pérdida de generalidad, escriba $N = \sum_{i=1}^{n} B_i$, donde $B_1, \ldots, B_n$ son independientes con distribución común $Ber(p)$. Por lo tanto

$$N - np = \sum_{i=1}^{n} (B_i - p),$$

de manera que

$$(N - np)^4 = \sum_{1 \leq i_1, i_2, i_3, i_4 \leq n} (B_{i_1} - p)(B_{i_2} - p)(B_{i_3} - p)(B_{i_4} - p).$$

Observe ahora que el valor esperado de $(B_{i_1} - p)(B_{i_2} - p)(B_{i_3} - p)(B_{i_4} - p)$ es cero si uno de los índices no se repite; esto se debe a que las variables B_i son independientes y a que su media es p. Por lo tanto, los únicos productos cuya esperanza no se anula son de la forma

(a) $(B_i - p)^4$, de los cuales hay n y tienen esperanza menor o igual a uno, y

(b) $(B_i - p)^2(B_j - p)^2$ con $i \neq j$, de los cuales hay $3n(n - 1)$ y su esperanza es menor o igual a uno.

Por lo tanto, $E[(N - np)^4] \leq n + 3n(n - 1) \leq 3n^2$, estableciendo (3.5.7). Para concluir observe que

$$(\delta + N - np)^4 = \delta^4 + 4\delta^3(N - np) + 6(N - np)^2\delta^2 + 4(N - np)^3\delta + (N - np)^4.$$

Recuerde ahora que $E[N - np] = 0$, $E[(N - np)^2] = np(1 - p)$, mientras que $E[(N - np)^3] \leq E[|N - np|^4]^{3/4}$, por la desigualdad de Hölder. Tomando esperanza en ambos lados de la anterior igualdad desplegada y usando que δ se ubica entre cero y uno se desprende que

$$\begin{aligned} E[(\delta + N - np)^4] &\leq 1 + 6np(1 - p) + 4E[(N - np)^4]^{3/4} + E[(N - np)^4] \\ &\leq 1 + 6n + 4[3n^2]^{3/4} + 3n^2 \\ &\leq n^2 + 6n^2 + 4[3n^2] + 3n^2 \end{aligned}$$

y por lo tanto $E[(\delta + N - np)^4] \leq 22n^2$, concluyendo la demostración. $\square$

Combinando los dos lemas precedentes, a continuación se obtendrá una cota para la probabilidad de que $Y_{[np]}$ supere a $\xi_p(\theta)$ por más de w unidades.

Lema 3.5.3. Suponga que las condiciones del Teorema 3.4.1 son válidas.

(i) Defina

$$Q(w) = \begin{cases} [G(\xi_p^g + w) - p]/w, & \text{si } 0 < w \leq A - \xi_p^g \\ g(\xi_p^g), & \text{si } w = 0. \end{cases}$$

Con esta notación,

$$\min_{w \in [0, A - \xi_p^g]} Q(w) =: C_0 > 0.$$

(ii) Para todo $w > 0$,

$$P_\theta[Y_{[np]} - \xi_p(\theta) > w] \leq \frac{3}{C_0^4 n^2 w^4}. \tag{3.5.9}$$

Demostración. (i) Note que

$$\begin{aligned} \lim_{w \to 0+} Q(w) &= \lim_{w \to 0+} \frac{G(\xi_p^g + w) - p}{w} \\ &= \lim_{w \to 0+} \frac{G(\xi_p^g + w) - G(\xi_p^g)}{w} \\ &= G'(\xi_p^g) \end{aligned}$$

y entonces $\lim_{w \to 0+} Q(w) = g(\xi_p^g) = Q(0)$, de manera que $Q(\cdot)$ es continua y positiva en cero. Por otro lado, la función Q es claramente continua en $(0, A - \xi_p^g]$, y como el cuantil ξ_p^g de $g(x)$ es único, se tiene que $G(\xi_p^g + w) > p$ para todo $w > 0$. Por lo tanto, $Q(w)$ es continua y positiva en todo su dominio, y entonces su valor mínimo C_0 es positivo (Fulks, 1980).

(ii) Observe que si $w > A - \xi_p^g$, entonces $Y_{[np]} > \xi_p(\theta) + w \Longrightarrow Y_{[np]} > \xi_p(\theta) + A - \xi_p^g = A + \theta$ (vea (3.3.1)), y entonces $P_\theta[Y_{[np]} > \xi_p(\theta) + w] \leq P_\theta[Y_{[np]} > A + \theta] = 0$, pues $P_\theta[Y_{[np]} \leq A + \theta] = 1$; recuerde que la densidad $g(x - \theta)$ se anula cuando $x > A + \theta$. Por lo tanto, (3.5.9) es válida si $w > A - \xi_p^g$, y es suficiente establecer la desigualdad suponiendo que $w \in (0, A - \xi_p^g]$. En este caso, note que combinando (3.5.3) y (3.5.5) con $z = \xi_p(\theta) + w$ se tiene que

$$\begin{aligned} P_\theta[Y_{[np]} - \xi_p(\theta) > w] \\ &= P_\theta[Y_{[np]} > \xi_p(\theta) + w] \\ &= P_\theta[N_n(\xi_p(\theta) + w) < [np]] \\ &= P_\theta[N_n(\xi_p(\theta) + w) - nG(\xi_p^g + w) < [np] - nG(\xi_p^g + w)] \end{aligned}$$

Note ahora que $G(\xi_p^g + w) > G(\xi_p^g) = p$ para $w > 0$, de donde se desprende que $[np] - nG(\xi_p^g + w) \le np - nG(\xi_p^g + w) < 0$. Por lo tanto, la anterior relación desplegada implica que

$$\begin{aligned} P_\theta[Y_{[np]} - \xi_p(\theta) > w] \\ &= P_\theta[|N_n(\xi_p(\theta) + w) - nG(\xi_p^g + w)| > nG(\xi_p^g + w) - [np]] \\ &\le P_\theta[|N_n(\xi_p(\theta) + w) - nG(\xi_p^g + w)| > nG(\xi_p^g + w) - np] \end{aligned}$$

de donde, aplicando la desigualdad de Markov, se obtiene que

$$P_\theta[Y_{[np]} - \xi_p(\theta) > w] \le \frac{E_\theta[(N_n(\xi_p(\theta) + w) - nG(\xi_p^g + w))^4]}{(nG(\xi_p^g + w) - np)^4}.$$

Usando que $N_n(\xi_p(\theta) + w) \sim \mathcal{B}(n,\ G(\xi_p^g + w))$, por el Lemma 3.5.1(ii), el Lemma 3.5.2 permite concluir que

$$\begin{aligned} P_\theta[Y_{[np]} - \xi_p(\theta) > w] &\le \frac{3n^2}{(nG(\xi_p^g + w) - np)^4} \\ &\le \frac{3n^2}{n^4 w^4([G(\xi_p^g + w) - p]/w)^4} \\ &\le \frac{3n^2}{n^4 w^4 (Q(w))^4} \\ &\le \frac{3n^2}{n^4 w^4 C_0^4} \end{aligned}$$

donde la definición de $Q(w)$ se utilizó para escribir la tercera desigualdad, y la última etapa se debe a la parte (i). Por lo tanto $P_\theta[Y_{[np]} - \xi_p(\theta) > w] \le 3/[n^2 C_0^4 w^4]$.

□

El último resultado preliminar antes de la demostración del Teorema 3.5.1 es la siguiente cota para la probabilidad de que $\xi_p(\theta)$ supere a $Y_{[np]}$ por al menos w unidades.

Lema 3.5.4. Suponga que las condiciones del Teorema 3.4.1 son válidas.

(i) Defina

$$\tilde{Q}(w) = \begin{cases} [p - G(\xi_p^g - w)]/w, & \text{si } 0 < w \leq \xi_p^g - a \\ g(\xi_p^g), & \text{si } w = 0 \end{cases}$$

Con esta notación,

$$\min_{w \in [0, \xi_p^g - a]} \tilde{Q}(w) =: \tilde{C}_0 > 0.$$

(ii) Para todo $w > 0$

$$P_\theta[Y_{[np]} - \xi_p(\theta) \leq -w] \leq \frac{22}{\tilde{C}_0^4 n^2 w^4}. \tag{3.5.10}$$

Demostración. (i) La demostración de esta parte se similar a la del Lemma 3.5.3(i). (ii) Primeramente, note que si $w > \xi_p^g - a$, entonces $Y_{[np]} - \xi_p(\theta) < -w \Longrightarrow Y_{[np]} < \xi_p(\theta) + a - \xi_p^g = \theta + a$, y entonces $P_\theta[Y_{[np]} - \xi_p(\theta) < -w] \leq P_\theta[Y_{[np]} < \theta + a] = 0$, pues $P_\theta[Y_{[np]} \geq \theta + a] = 1$, de manera que (3.5.10) es claramente válida si $w > \xi_p^g - a$. Por lo tanto, a continuación se supone que $w \in (0, \xi_p^g - a]$. En este caso, combinando (3.5.4) y (3.5.5) con $z = \xi_p(\theta) - w$ se desprende que

$$\begin{aligned} P_\theta[Y_{[np]} - \xi_p(\theta) \leq -w] & \\ &= P_\theta[Y_{[np]} \leq \xi_p(\theta) - w] \\ &= P_\theta[N_n(\xi_p(\theta) - w) \geq [np]] \\ &= P_\theta[N_n(\xi_p(\theta) - w) - nG(\xi_p^g - w) \geq [np] - nG(\xi_p^g - w)], \end{aligned}$$

y como $[np] = np - \delta$ donde $\delta \in (0, 1)$, se tiene que

$$P_\theta[Y_{[np]} - \xi_p(\theta) \leq -w] = P_\theta[\delta + N_n(\xi_p(\theta) - w) - nG(\xi_p^g - w) \geq np - nG(\xi_p^g - w)].$$

Note ahora que $G(\xi_p^g - w) < G(\xi_p^g) = p$ para $w > 0$, de donde se desprende que $np - nG(\xi_p^g - w) > 0$. Por lo tanto, al anterior relación desplegada y la desigualdad de Markov implican que

$$P_\theta[Y_{[np]} - \xi_p(\theta) \leq -w] \leq \frac{E_\theta[(\delta + N_n(\xi_p(\theta) + w) - nG(\xi_p^g + w))^4]}{(np - nG(\xi_p^g - w))^4}.$$

Usando ahora que $N_n(\xi_p(\theta) - w) \sim \mathcal{B}(n,\, G(\xi_p^g - w))$, por el Lemma 3.5.1(ii), el Lemma 3.5.2 permite concluir que

$$\begin{aligned}
P_\theta[Y_{[np]} - \xi_p(\theta) \leq -w] &\leq \frac{22n^2}{(np - nG(\xi_p^g - w))^4} \\
&\leq \frac{22n^2}{n^4 w^4([p - G(\xi_p^g - w)]/w)^4} \\
&\leq \frac{22n^2}{n^4 w^4(\tilde{Q}(w))^4} \\
&\leq \frac{22n^2}{n^4 w^4 \tilde{C}_0^4}
\end{aligned}$$

donde la definición de $\tilde{Q}(w)$ se utilizó para escribir la tercera desigualdad, y la última etapa se debe a la parte (i). Por lo tanto $P_\theta[Y_{[np]} - \xi_p(\theta) \leq -w] \leq 22/[n^2\tilde{C}_0^4 w^4]$. □

Demostración del Teorema 3.5.1. Sea $\theta \in \Theta$ fijo y observe que *para todo* $w > 0$

$$\begin{aligned}
P_\theta[|Y_{[np]} - \xi_p(\theta)| > w] &= P_\theta[Y_{[np]} - \xi_p(\theta) > w] + P_\theta[Y_{[np]} - \xi_p(\theta) < -w] \\
&\leq \frac{3}{n^2 C_0 w^4} + \frac{22}{n^2 \tilde{C}_0 w^4} \\
&= \frac{C}{n^2 w^4},
\end{aligned}$$

donde C_0 y $\tilde{C}_0$ son como en los Lemmas 3.5.3 y 3.5.4, respectivamente, y la constante C está dada por $C := 2\max\{3/C_0,\ 22/\tilde{C}_0\}$. Por lo tanto, para cada $x > 0$

$$\begin{aligned}
P_\theta[n|T_{2,\,n}(\mathbf{X}) - \xi_p(\theta)|^2 > x] &= P_\theta[n|Y_{[np]} - \xi_p(\theta)|^2 > x] \\
&= P_\theta[Y_{[np]} - \xi_p(\theta) > \sqrt{x/n}] \\
&\leq \frac{C}{n^2(\sqrt{x/n})^4}
\end{aligned}$$

y entonces $P_\theta[n|T_{2,\,n}(\mathbf{X}) - \xi_p(\theta)|^2 > x] \le C/x^2$. $\square$

Demostración del Teorema 3.4.1

Los resultados preliminares en la sección anterior se utilizarán ahora para establecer el Teorema 3.4.1. Primeramente, sea $\mathbf{X} = (X_1, \ldots, X_n)$ una muestra de la densidad $g(x - \theta)$, donde $\theta \in \Theta$ arbitrario pero fijo, y defina

$$v\colon = \sqrt{\frac{p(1-p)}{g(\xi_p^g)^2}}. \tag{3.6.1}$$

Observe ahora que a partir del Teorema 3.3.1 se tiene que

$$\frac{\sqrt{n}[T_{2,\,n}(\mathbf{X}) - \theta]}{v} \to \mathcal{N}\left(0, 1\right),$$

y entonces el Teorema de Slutsky (Billingsley 1997, Dudewicz y Mishra, 1989) implica que

$$\frac{n[T_{2,\,n}(\mathbf{X}) - \theta]^2}{v^2} \to \mathcal{X}_1^2.$$

Por lo tanto, si Z es una variable aleatoria tal que

$$Z \sim \mathcal{X}_1^2, \tag{3.6.2}$$

se tiene que para cada $t > 0$

$$P_\theta\left[\frac{n[T_{2,\,n}(\mathbf{X}) - \theta]^2}{v^2} > t\right] \to P[Z > t].$$

Más aún, debido a que la distribución $\mathcal{X}_1^2$ es continua, la anterior convergencia es uniforme en $[0, \infty)$, esto es,

$$\Delta_n\colon = \max_{t \ge 0}\left|P_\theta\left[\frac{n[T_{2,\,n}(\mathbf{X}) - \theta]^2}{v^2} > t\right] - P[Z > t]\right| \to 0 \quad \text{conforme } n \to \infty; \tag{3.6.3}$$

vea el Teorema 1.2.4. Por otro lado, recuerde que si W es una variable aleatoria no negativa, se tiene la relación $E[W] = \int_0^\infty P[W > t]\,dt$; vea, por ejemplo, Mood *et. al.* (1987). A partir de esta hecho se desprenden las dos siguientes relaciones:

$$E_\theta\left[\frac{n[T_{2,\,n}(\mathbf{X}) - \theta]^2}{v^2}\right] = \int_0^\infty P_\theta\left[\frac{n[T_{2,\,n}(\mathbf{X}) - \theta]^2}{v^2} > t\right]\,dt; \tag{3.6.4}$$

$$E\left[Z\right] = \int_0^\infty P\left[Z > t\right]\,dt = 1, \tag{3.6.5}$$

donde la segunda igualdad se debe a que el valor esperado de la distribución $\mathcal{X}_1^2$ es uno. A continuación, seleccione un número positivo $T > 0$, y note que

$$\begin{aligned}
&\left|\int_0^T P_\theta\left[\frac{n[T_{2,\,n}(\mathbf{X}) - \theta]^2}{v^2} > t\right]\,dt - \int_0^T P[Z > t]\,dt\right| \\
&\qquad\le \int_0^T \left|P_\theta\left[\frac{n[T_{2,\,n}(\mathbf{X}) - \theta]^2}{v^2} > t\right]\,dt - \int_0^T P[Z > t]\right|\,dt \\
&\qquad\le \int_0^T \Delta_n\,dt \\
&\qquad= T\Delta_n
\end{aligned}$$

de manera que (3.6.3) implica

$$\lim_{n\to\infty}\left|\int_0^T P_\theta\left[\frac{n[T_{2,\,n}(\mathbf{X}) - \theta]^2}{v^2} > t\right]\,dt - \int_0^T P[Z > t]\,dt\right| = 0. \tag{3.6.6}$$

Por otro lado, via el Teorema 3.5.1, se tiene que

$$\int_T^\infty P_\theta\left[\frac{n[T_{2,\,n}(\mathbf{X}) - \theta]^2}{v^2} > t\right]\,dt \le \int_T^\infty \frac{C}{t^2}dt = \frac{C}{T},$$

mientras que

$$
\begin{aligned}
&\left|\int_0^\infty P_\theta\left[\frac{n[T_{2,\,n}(\mathbf{X})-\theta]^2}{v^2}>t\right]dt-\int_0^\infty P[Z>t]\,dt\right|\\
&\qquad\le\left|\int_0^T P_\theta\left[\frac{n[T_{2,\,n}(\mathbf{X})-\theta]^2}{v^2}>t\right]dt-\int_0^T P[Z>t]\,dt\right|\\
&\qquad\qquad+\left|\int_T^\infty P_\theta\left[\frac{n[T_{2,\,n}(\mathbf{X})-\theta]^2}{v^2}>t\right]dt-\int_T^\infty P[Z>t]\,dt\right|\\
&\qquad\le\left|\int_0^T P_\theta\left[\frac{n[T_{2,\,n}(\mathbf{X})-\theta]^2}{v^2}>t\right]dt-\int_0^T P[Z>t]\,dt\right|\\
&\qquad\qquad+\int_T^\infty P_\theta\left[\frac{n[T_{2,\,n}(\mathbf{X})-\theta]^2}{v^2}>t\right]dt+\int_T^\infty P[Z>t]\,dt.\\
&\qquad\le\left|\int_0^T P_\theta\left[\frac{n[T_{2,\,n}(\mathbf{X})-\theta]^2}{v^2}>t\right]dt-\int_0^T P[Z>t]\,dt\right|\\
&\qquad\qquad\qquad+\frac{C}{T}+\int_T^\infty P[Z>t]\,dt.
\end{aligned}
$$

Por lo tanto, (3.6.6) implica que, para todo $T>0$,

$$
\begin{aligned}
\lim_{n\to\infty}&\left|\int_0^\infty P_\theta\left[\frac{n[T_{2,\,n}(\mathbf{X})-\theta]^2}{v^2}>t\right]dt-\int_0^\infty P[Z>t]\,dt\right|\\
&\qquad\qquad\qquad\le\frac{C}{T}+\int_T^\infty P[Z>t]\,dt.
\end{aligned}
$$

Para concluir, note que (3.6.5) implica que $\lim_{T\to\infty}\int_T^\infty P[Z>t]\,dt=0$, de manera que tomando limite cuando T tiende a ∞ en la anterior desigualdad desplegada se obtiene que

$$
\lim_{n\to\infty}\left|\int_0^\infty P_\theta\left[\frac{n[T_{2,\,n}(\mathbf{X})-\theta]^2}{v^2}>t\right]dt-\int_0^\infty P[Z>t]\,dt\right|=0,
$$

lo que equivale a $\lim_{n\to\infty} E_\theta\left[\frac{n[T_{2,\,n}(\mathbf{X})-\theta]^2}{v^2}\right]=E[Z]=1$, por (3.6.4) y (3.6.5), de manera que $\lim_{n\to\infty} E_\theta\left[n[T_{2,\,n}(\mathbf{X})-\theta]^2\right]=v^2=\frac{p(1-p)}{g(\xi_p^g)^2}$, estableciendo la conclusión del Teorema 3.4.1.

Capítulo 4

MÉTODO DE VEROSIMILITUD MÁXIMA

Introducción

En este capítulo se estudia el método de verosimilitud máxima para el modelo de traslación introducido en la Definición 2.2.1. El principal objetivo es mostrar que los intervalos de confianza que el método genera son eficientes, en el sentido de que su longitud es menor o igual a la de aquellos construidos mediante las técnicas de momentos o cuantiles. El instrumento que permite alcanzar el propósito deseado, es una cota para la distribución límite del estimador de verosimilitud máxima $\hat{\theta}_n$ establecida en el Teorema 4.3.1, resultado que es una extensión de las conclusiones clásicas sobre la normalidad asintótica de $\hat{\theta}_n$ establecidas en el Teorema 1.4.1 y es el principal resultado de este trabajo.

La presentación está organizada de la siguiente manera: En la Sección 2 se inicia el estudio de las propiedades del estimador de verosimilitud máxima $\hat{\theta}_n$ estableciendo su consistencia. El argumento muestra que el supuesto de que la densidad de los errores tiene soporte acotado es suficiente para obtener el resultado, pues en esta parte no se requieren condiciones de diferenciabilidad. En la Sección 3 se establece la ecuación de verosimilitud que $\hat{\theta}_n$ debe satisfacer cuando la densidad de los errores es derivable, mientras que en la Sección 4 se obtiene una cota para la distribución límite de las desviaciones de $\hat{\theta}_n$ alrededor del parámetro θ, resultado que es finalmente usado en la Sección 5 para establecer intervalos de confianza con nivel asintótico de cobertura especificado, y se muestra que su longitud no excede a la de los intervalos generados por los métodos de momentos y cuantiles.

Consistencia

El objetivo de esta sección es establecer el siguiente resultado básico.

Teorema 4.2.1. Suponga que $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ obedece a un modelo de traslación que satisface la Hipótesis 2.2.1, y sea $\hat{\theta}_n \equiv \hat{\theta}_n(\mathbf{X})$ un maximizador de la función de verosimilitud correspondiente a $\mathbf{X}$. En este caso, para cada $\theta_0 \in \Theta$,

$$\lim_{n\to\infty} \hat{\theta}_n = \theta_0 \qquad P_{\theta_0}\text{–c. s.}$$

La demostración de este teorema se basa en los tres lemas que se establecen más adelante. Primeramente, es conveniente introducir la siguiente definición.

Definición 2.1. Sea $g(\cdot)$ una densidad que se anula fuera de un intervalo acotado, y denote mediante $G(\cdot)$ a la correspondiente función de distribución. El mínimo esencial de $g(\cdot)$ se define como

$$b = \max\{a : G(a) = 0\}, \tag{4.2.1}$$

mientras que el máximo esencial de la densidad $g(\cdot)$ está dado por

$$B = \min\{A : G(A) = 1\}. \tag{4.2.2}$$

Verbalmente, el mínimo esencial b es el punto en que la función de distribución empieza a crecer hacia uno, mientras que B es el punto en que la función de distribución alcanza el valor uno por primera vez. Note que como $G(\cdot)$ es continua,

$$G(b) = 0 \quad \text{y} \quad G(b+\delta) > 0, \quad \delta > 0, \tag{4.2.3}$$

mientras que

$$G(B) = 1 \quad \text{y} \quad G(B-\delta) < 1, \quad \delta > 0. \tag{4.2.4}$$

Observe que la densidad de $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ está dada por

$$f_n(\mathbf{X};\theta) = \prod_{i=1}^{n} g(X_i - \theta), \quad \mathbf{X} \in \mathbb{R}^n, \tag{4.2.5}$$

donde $g(\cdot)$ es la densidad de los errores, y note que como $G(b) = 0$ y $G(B) = 1$, sin pérdida de generalidad puede suponerse que $g(x) = 0$ si $x < b$ o $x > B$, esto es,

$$g(x) = g(x) I_{[b,B]}(x), \quad x \in \mathbb{R}. \tag{4.2.6}$$

Lema 4.2.1. La función de verosimilitud está dada por

$$L_n(\theta;\mathbf{X}) = \prod_{i=1}^{n} g(X_i - \theta) I_{[X_{(n)}-B,\, X_{(1)}-b]}(\theta), \tag{4.2.7}$$

donde $X_{(n)} = \max\{X_1, X_2, \ldots, X_n\}$ y $X_{(1)} = \min\{X_1, X_2, \ldots, X_n\}$. Por lo tanto, con probabilidad uno, un maximizador $\hat{\theta}_n$ de $L_n(\theta;\mathbf{X})$ satisface

$$\hat{\theta}_n(\mathbf{X}) \equiv \hat{\theta}_n \in [X_{(n)} - B,\, X_{(1)} - b] \quad P_{\theta_0}\text{–c. s.} \tag{4.2.8}$$

Demostración. Usando la igualdad $L_n(\theta;\mathbf{X}) = f_n(\mathbf{X};\theta)$, (4.2.5) y (4.2.6) implican que

$$L_n(\theta;\mathbf{X}) = \prod_{i=1}^{n} \left[g(X_i - \theta) I_{[b,B]}(X_i - \theta) \right] = \prod_{i=1}^{n} g(X_i - \theta) \prod_{i=1}^{n} I_{[b,B]}(X_i - \theta). \tag{4.2.9}$$

Por otro lado, observando que

$$\begin{aligned} I_{[b,B]}(X_i-\theta)=1 &\iff X_i-\theta\in[b,B] \\ &\iff \theta-X_i\in[-B,-b] \\ &\iff \theta\in[X_i-B,X_i-b] \\ &\iff I_{[X_i-B,X_i-b]}(\theta)=1, \end{aligned}$$

se obtiene

$$\prod_{i=1}^{n} I_{[b,B]}(X_i-\theta)=\prod_{i=1}^{n} I_{[X_i-B,X_i-b]}(\theta)=I_J(\theta),$$

donde

$$\begin{aligned} J &= \bigcap_{i=1}^{n}[X_i-B,X_i-b] \\ &= [\max_{i=1,2\ldots,n}(X_i-B),\ \min_{i=1,2\ldots,n}(X_i-B)]=[X_{(n)}-B,X_{(1)}-b]. \end{aligned}$$

Luego, $\prod_{i=1}^{n} I_{[b,B]}(X_i-\theta) = I_{[X_{(n)}-B,X_{(1)}-b]}(\theta)$ y conjuntamente con (4.2.9), esta igualdad implica (4.2.7), mientras que (4.2.8) se obtiene notando que cuando $\theta\notin[X_{(n)}-B,\ X_{(1)}-b]$ la verosimilitud es cero. Puesto que, con probabilidad uno, $\mathbf{X}$ asume valores en la zona en que $f_n(\mathbf{X};\theta)=L_n(\theta;\mathbf{X})>0$, un maximizador de $L_n(\theta;\mathbf{X})$ debe pertenecer a $[X_{(n)}-B,\ X_{(1)}-b]$ casi seguramente. □

A continuación, se demostrará que θ_0, el verdadero valor del parámetro, está incluido en el intervalo (4.2.8).

Lema 4.2.2. Suponga que $\mathbf{X}=(X_1,X_2,\ldots,X_n)$ es una muestra de $g(x-\theta_0)$. Bajo la Hipótesis 2.2.1 la siguiente afirmación es válida:

$$\theta_0\in[X_{(n)}-B,\ X_{(1)}-b]\quad P_{\theta_0}\text{–c. s..}$$

Demostración. De acuerdo al modelo de traslación,

$$X_i=\theta_0+\varepsilon_i,\quad i=1,2,\ldots,n\quad P_{\theta_0}\text{–c. s.} \tag{4.2.10}$$

donde $\varepsilon_i \sim g(\cdot)$, y por lo tanto $P_{\theta_0}[b \le \varepsilon_i \le B] = 1$. Luego,

$$b \le \min_{i=1,2,\ldots,n}\{\varepsilon_i\} = \varepsilon_{(1)} \le \varepsilon_{(n)} = \max_{i=1,2,\ldots,n}\{\varepsilon_i\} \le B, \quad P_{\theta_0}\text{–c. s.;}$$

combinando esta relación con $\varepsilon_{(1)} = X_{(1)} - \theta_0$ y $\varepsilon_{(n)} = X_{(n)} - \theta_0$, se desprende que

$$b \le X_{(1)} - \theta_0 \le X_{(n)} - \theta_0 \le B \quad P_{\theta_0}\text{–c. s.}$$

La primera de estas desigualdades equivale a $\theta_0 \le X_{(1)} - b$, mientras que la segunda establece que $X_{(n)} - B \le \theta_0$, de manera que $\theta_0 \in [X_{(n)} - B,\ X_{(1)} - b]$ P_{θ_0}–c. s. □

El siguiente resultado es la última etapa antes de la demostración del Teorema 4.2.1.

Lema 4.2.3. Denote por θ_0 al verdadero valor del parámetro. Conforme $n \to \infty$

(a) $X_{(n)} \to B + \theta_0 \quad P_{\theta_0}$–c. s., y

(b) $X_{(1)} \to b + \theta_0 \quad P_{\theta_0}$–c. s..

Demostración. (a) Dado $\delta > 0$, considere el evento

$$[X_{(n)} \le B + \theta_0 - \delta] = \bigcap_{i=1}^{n}[X_i \le B + \theta_0 - \delta].$$

La independencia de las variables X_i permite calcular

$$\begin{aligned} P_{\theta_0}[X_{(n)} \le B + \theta_0 - \delta] &= \prod_{i=1}^{n} P_{\theta_0}[X_i \le B + \theta_0 - \delta] \\ &= \prod_{i=1}^{n} P_{\theta_0}[\varepsilon_i \le B - \delta] = \prod_{i=1}^{n}\int_b^{B-\delta} g(x)\,dx \end{aligned}$$

Por otro lado, debido a que B es el máximo esencial de $g(\cdot)$, se tiene que

$$\beta(\delta) := \int_b^{B-\delta} g(x)\,dx = G(B-\delta) - G(b) = G(B-\delta) < 1;$$

vea (4.2.3) y (4.2.4). Por lo tanto,

$$P_{\theta_0}[X_{(n)} \leq B + \theta_0 - \delta] = \beta(\delta)^n \tag{4.2.11}$$

Observe ahora que $[|X_{(n)} - (B+\theta_0)| > \delta]$ es la unión de los eventos $[X_{(n)} < (B+\theta_0)-\delta]$ y $[X_{(n)} > (B+\theta_0)+\delta]$; además, debido a que el soporte de las variables ε_i está contenido en $[b, B]$, se tiene que $X_i = \varepsilon_i + \theta_0 \in [b + \theta_0,\ B + \theta_0]$ P_{θ_0}–c. s., y entonces $X_{(n)} = \max\{X_i\} \in [b + \theta_0,\ B + \theta_0]$ P_{θ_0}–c. s., inclusión que implica $P_{\theta_0}[X_{(n)} > B + \theta_0 + \delta] = 0$, y por lo tanto $P_{\theta_0}[|X_{(n)} - (B+\theta_0)| > \delta] = P_{\theta_0}[X_{(n)} \leq B + \theta_0 - \delta]$, de tal manera que (4.2.11) permite concluir que

$$P_{\theta_0}[|X_{(n)} - (B + \theta_0)| > \delta] = \beta(\delta)^n.$$

Debido a que $\beta(\delta) < 1$, se tiene que

$\sum_{n=1}^{\infty} P_{\theta_0}[|X_{(n)} - (B + \theta_0)| > \delta] \leq \sum_{n=1}^{\infty} \beta(\delta)^n = \beta(\delta)/(1 - \beta(\delta)) < \infty$, y como esto ocurre para todo $\delta > 0$, a partir del Teorema 1.2.1 se desprende que $X_{(n)} \to B + \theta_0$ P_{θ_0}–c. s., completando la demostración de la parte (a), mientras que la parte (b) puede establecerse de forma similar. □

Los resultados preliminares en los tres lemas precedentes permiten demostrar el resultado principal de esta sección.

Demostración del Teorema 4.2.1. A partir de las inclusiones

$$\hat{\theta}_n \in [X_{(n)} - B, X_{(1)} - b] \quad \text{y} \quad \theta_0 \in [X_{(n)} - B, X_{(1)} - b] \quad P_{\theta_0}\text{–c. s.}$$

establecidas en los Lemas 4.2.1 y 4.2.2, respectivamente, se desprende que, con probabilidad uno respecto a P_{θ_0}

$$\begin{aligned} |\hat{\theta}_n - \theta_0| &\leq X_{(1)} - b - (X_{(n)} - B) \\ &= [X_{(1)} - (b + \theta_0)] - [X_{(n)} - (B + \theta_0)], \end{aligned}$$

y tomando límite conforme $n \to \infty$ se tiene que, con probabilidad uno respecto a P_{θ_0},

$$\lim_{n\to\infty} |\hat{\theta}_n - \theta_0| \leq \lim_{n\to\infty}[X_{(1)} - (b+\theta_0)] - \lim_{n\to\infty}[X_{(n)} - (B+\theta_0)]$$

y entonces el Lemma 4.2.3 implica que $\lim_{n\to\infty} |\hat{\theta}_n - \theta_0| = 0$ P_{θ_0}–c. s.. □

Ecuación de Verosimilitud

En esta sección se establece la ecuación que el estimador de verosimilitud máxima debe satisfacer cuando la densidad $g(\cdot)$ de los errores es derivable. Defina

$$\mathcal{L}_n(\theta;\mathbf{X}) := \sum_{i=1}^{n} \log(g(X_i - \theta)) = \log(f_n(\mathbf{X};\theta)) = \log(L_n(\theta;\mathbf{X})), \tag{4.3.1}$$

y observe que

$$\partial_\theta \mathcal{L}_n(\theta;\mathbf{X}) := -\sum_{i=1}^{n} \frac{g'(X_i - \theta)}{g(X_i - \theta)} \quad \text{si } L_n(\theta;\mathbf{X}) \neq 0. \tag{4.3.2}$$

Teorema 4.3.1. Sea $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ una muestra de $g(x-\theta)$ y suponga que las Hipótesis 2.2.1 y 2.4.1 se satisfacen. En este caso, con probabilidad uno las siguientes afirmaciones (i) y (ii) son válidas:

(i) La función de verosimilitud se maximiza en un punto interior de $[X_{(n)} - B, b - X_{(1)}]$.

(ii) Cualquier maximizador de $L_n(\theta;\mathbf{X})$ satisface la ecuación de verosimilitud dada por

$$\partial_\theta \mathcal{L}_n(\theta;\mathbf{X}) = 0. \tag{4.3.3}$$

Demostración. (i) Se estableció en el Lemma 4.2.1 que

$$L_n(\theta;\mathbf{X}) = \prod_{i=1}^{n} g(X_i - \theta) I_{[X_{(n)}-B,\, X_{(1)}-b]}(\theta); \tag{4.3.4}$$

de manera que $L_n(\theta;\mathbf{X})$ se anula si $\theta < X_{(n)} - B$ o $\theta > X_{(1)} - b$. Se mostrará ahora que Ł$_n(\theta;\mathbf{X})$ también se anula cuando $\theta = X_{(n)} - B$ o $\theta = b - X_{(1)}$. Con este fin, observe que $X_{(n)} = X_k$ para algún k, de manera que $g(X_k - (X_{(n)} - B)) = g(X_k - (X_k - B)) = g(B) = 0$, donde la última igualdad se debe a (4.2.6). Puesto que el factor $g(X_k - (X_{(n)} - B))$ aparece en la fórmula para $L_n((X_{(n)} - B);\mathbf{X})$, se desprende que $L_n((X_{(n)} - B);\mathbf{X}) = 0$, mientras que un argumento semejante permite establecer que $L_n((X_{(1)} - b);\mathbf{X}) = 0$. Por lo tanto, $L(\theta;\mathbf{X})$ se anula fuera del intervalo $(X_{(n)} - B,\ X_{(1)} - b)$, y por lo tanto, como función de θ, la verosimilitud $L_n(\theta;\mathbf{X})$ asume su valor máximo en $(X_{(n)} - B,\ X_{(1)} - b)$.

(ii) Denote por θ_0 el verdadero valor del parámetro, y observe que, con probabilidad 1, $\mathbf{X}$ toma valores en la región $\{\mathbf{x}\colon f_n(\mathbf{x};\theta_0) > 0\} = \{\mathbf{x}\colon L_n(\theta_0;\mathbf{x}) > 0\}$ y, por lo tanto, con probabilidad uno el valor máximo de $L_n(\theta;\mathbf{X})$ es positivo. Si θ^* denota a un maximizador, entonces

$$L_n(\theta^*;\mathbf{X}) > 0 \quad \text{y} \quad \partial_\theta L_n(\theta^*;\mathbf{X}) = 0$$

donde la segunda ecuación se debe a que $L_n(\theta^*;\mathbf{X})$ es derivable respecto a θ. Note ahora que

$$\begin{aligned}\partial_\theta L_n(\theta;\mathbf{X}) &= \partial_\theta\left[\prod_{i=1}^n g(X_i - \theta)\right]\\ &= -\sum_{j=1}^n g'(X_j - \theta) \prod_{i:1\le i\ne j\le n} g(X_i - \theta),\end{aligned}$$

de manera que

$$\begin{aligned}\partial_\theta L_n(\theta^*;\mathbf{X}) = 0 &\iff \frac{\partial_\theta L_n(\theta^*;\mathbf{X})}{L_n(\theta^*;\mathbf{X})} = 0\\ &\iff -\sum_{j=1}^n g'(X_j - \theta^*)\frac{\prod_{i:1\le i\ne j\le n} g(X_i - \theta^*)}{L_n(\theta^*,\mathbf{X})},\\ &\iff -\sum_{j=1}^n \frac{g'(X_j - \theta^*)}{g(X_i - \theta^*)}\\ &\iff \partial_\theta \mathcal{L}_n(\theta^*;\mathbf{X}) = 0\end{aligned}$$

completando la demostración. □

De acuerdo a este resultado, cualquier maximizador de $L_n(\theta; \mathbf{X})$ satisface (4.3.3). El argumento que se utilizará para construir intervalos de confianza en base al estimador de verosimilitud máxima requiere comparar $\partial_\theta \mathcal{L}(\hat{\theta}_n; \mathbf{X})$ con la derivada parcial evaluada en el verdadero valor del parámetro θ_0, esto es, con $\partial_\theta \mathcal{L}(\theta_0; \mathbf{X})$. El siguiente teorema establece la distribuición límite de esta última variable aleatoria.

Teorema 4.3.2. Suponga que las Hipótesis 2.2.1 y 2.4.1 se satisfacen, y que el número de información de Fisher $\mathcal{I}$ es finito. SI θ_0 es el verdadero valor del parámetro, entonces

$$\sqrt{n}\frac{\partial_\theta \mathcal{L}(\theta_0; \mathbf{X})}{n} \xrightarrow{\mathrm{d}} \mathcal{N}(0, \mathcal{I}).$$

Demostración. Como $\mathbf{X}$ asume valores en la región $\{\mathbf{x}\colon L_n(\theta_0, \mathbf{x}) > 0\}$ P_{θ_0}–c. s., (4.3.2) implica que

$$\partial_\theta \mathcal{L}_n(\theta_0; \mathbf{X}) = -\sum_{i=1}^{n} \frac{g'(X_i - \theta_0)}{g(X_i - \theta_0)} \quad P_{\theta_0}\text{–c. s..}$$

Observando que las variables aleatorias $\dfrac{g'(X_i - \theta_0)}{g(X_i - \theta_0)}$ son iid y que su esperanza y varianza son cero e $\mathcal{I}$, respectivamente (vea (2.5.8) y (2.5.9)), la conclusión se desprende del teorema central de límite. □

Cota Para la Distribución Asintótica

En esta sección se obtiene una cota para la distribución límite de

$$\sqrt{n}|\hat{\theta}_n - \theta_0|,$$

donde $\hat{\theta}_n$ denota al estimador de verosimilitud máxima. El siguiente teorema es el principal resultado de este trabajo y en la siguiente sección se utilizará para construir intervalos de confianza para el valor desconocido del parámetro θ.

Teorema 4.4.1. Suponga que las siguientes condiciones (i)–(iii) se satisfacen:

(i) Las Hipótesis 2.2.1 y 2.4.1 son válidas;

(ii) La densidad $g(\cdot)$ de los errores tiene segunda derivada continua y $g(\cdot)$ no se anula en el intervalo (b, B), donde b y B son como en (4.2.1) y (4.2.2), respectivamente. Más aún,

$$\frac{d^2 \log(g(x))}{dx^2} < 0, \quad x \in (b, B). \tag{4.4.1}$$

(iii) $\displaystyle\int_{\mathbb{R}} \frac{d^2 \log(g(x)}{dx^2}\, g(x)\, dx$ es finita.

En este caso, para cada $x > 0$ y $\theta_0 \in \Theta$,

$$\lim_{n\to\infty} P_{\theta_0}[\sqrt{n}|\hat{\theta}_n - \theta_0| \le x] \ge \Phi(x\sqrt{\mathcal{I}}) - \Phi(-x\sqrt{\mathcal{I}}) \tag{4.4.2}$$

donde $\mathcal{I}$ es el número de información de Fisher y $\Phi(\cdot)$ es la función de distribución de la densidad normal estándar.

La demostración de este resultado se basa en el lema que se presenta a continuación. Primero, observe que bajo las condiciones del teorema, a partir de (4.3.2) se desprende que

$$\partial_\theta^2 \mathcal{L}_n(\theta; \mathbf{X}) = -\sum_{i=1}^n \partial_\theta \left[\frac{g'(X_i - \theta)}{g(X_i - \theta)}\right] = \sum_{i=1}^n \partial_\theta^2 \log(g(X_i - \theta)) < 0 \tag{4.4.3}$$

para $\theta \in (X_{(n)} - B,\ b - X_{(1)})$, de manera que la ecuación de verosimilitud (4.3.3) tiene a lo más una solución, y entonces el estimador de verosimilitud máxima es único.

Lema 4.4.1. Bajo las condiciones del Teorema 4.4.1, suponga que el vector aleatorio $\mathbf{X} = (X_1, \ldots, X_n)$ es una muestra de $g(x - \theta_0)$, donde $\theta_0 \in \Theta$ es arbitrario, pero fijo, y que $\tilde{\theta}_n(\mathbf{X})$ es una variable aleatoria tal que, con probabilidad uno,

$$\tilde{\theta}_n(X) \equiv \tilde{\theta}_n \in (X_{(n)} - B,\ X_{(1)} - b). \tag{4.4.4}$$

En este caso,

$$\lim_{n\to\infty} \frac{1}{n} \left| \partial_\theta \mathcal{L}(\tilde{\theta}_n; \mathbf{X}) \right| \geq \mathcal{I} \quad P_{\theta_0}\text{–c. s..}$$

Demostración. Dado $\rho \in (0, B - b)$ y $\delta > 0$, defina

$$\Delta(\rho, \delta) := \max_{x \in [b+\rho,\ B-\rho]} \left| \frac{d^2 \log(g(x))}{dx^2} - \frac{d^2 \log(g(x-\delta))}{dx^2} \right|. \tag{4.4.5}$$

Note ahora que como $g(\cdot)$ no se anula en (b, B) y tiene segunda derivada continua, la función $\dfrac{d^2 \log(g(x))}{dx^2}$ es continua en (b, B), y entonces

$$\lim_{\delta \to 0} \Delta(\rho, \delta) = 0; \tag{4.4.6}$$

vea, por ejemplo, Fulks (1980). Defina

$$\tilde{\delta}_n = \tilde{\theta}_n - \theta_0$$

y note que la inclusión (4.4.4) y el Lema 4.2.2 implican que

$$|\tilde{\delta}_n| \leq X_{(1)} - b - (X_{(n)} - B) = X_{(1)} - (b + \theta_0) - (X_{(n)} - (B + \theta_0))$$

y entonces

$$\lim_{n\to\infty} |\tilde{\delta}_n| = 0 \quad P_{\theta_0}\text{–c. s.,} \tag{4.4.7}$$

por el Lemma 4.2.3. A continuación, note que el supuesto de negatividad (4.4.1) y (4.4.3) permiten obtener

$$\begin{aligned}
\left|\partial_\theta^2 \mathcal{L}_n(\tilde{\theta}_n; \mathbf{X})\right| &\geq -\sum_{i=1}^n \partial_\theta^2 \log(g(X_i - \tilde{\theta}_n)) \\
&\geq -\sum_{i=1}^n \partial_\theta^2 \log(g(X_i - \tilde{\theta}_n)) I_{[b+\theta_0+\rho,\ B+\theta_0-\rho]}(X_i) \\
&\geq -\sum_{i=1}^n \partial_\theta^2 \log(g(X_i - \theta_0 - \tilde{\delta}_n)) I_{[b+\rho,\ B-\rho]}(X_i - \theta_0) \\
&\geq -\sum_{i=1}^n \left[\partial_\theta^2 \log(g(X_i - \theta_0) I_{[b+\rho,\ B-\rho]}(X_i - \theta_0) - \Delta_n(\rho, \tilde{\delta}_n)\right]
\end{aligned}$$

y entonces

$$\frac{1}{n}\left|\partial_\theta^2 \mathcal{L}_n(\tilde{\theta}_n; \mathbf{X})\right| \geq -\frac{1}{n}\sum_{i=1}^n \partial_\theta^2 \log(g(X_i - \theta_0) I_{[b+\rho,\ B-\rho]}(X_i - \theta_0) - \Delta_n(\rho, \tilde{\delta}_n)$$

Observe ahora que las variables aleatorias $-\partial_\theta^2 \log(g(X_i - \theta_0)\ I_{[b+\rho,\ B-\rho]}$ $(X_i - \theta_0)$ son iid y que su esperanza común es

$$\begin{aligned}
-\int_{\mathbb{R}} \partial_\theta^2 \log(g(x - \theta_0) I_{[b+\rho,\ B-\rho]}(x - \theta_0) g(x - \theta_0)\, dx& \\
&= -\int_{b+\theta_0+\rho}^{B+\theta_0-\rho} \partial_\theta^2 \log(g(x - \theta_0) g(x - \theta_0)\, dx \\
&= -\int_{b+\theta_0+\rho}^{B+\theta_0-\rho} \partial_{\mathbf{x}}^2 \log(g(x - \theta_0) g(x - \theta_0)\, dx \\
&= -\int_{b+\rho}^{B-\rho} \frac{d^2 \log(g(x)}{dx^2} g(x)\, dx
\end{aligned}$$

Combinando este hecho con la anterior desigualdad desplegada, la ley de los grandes números, (4.4.6) y (4.4.7) implican que, con probabilidad uno,

$$\lim_{n\to\infty} \frac{1}{n}\left|\partial_\theta^2 \mathcal{L}_n(\tilde{\theta}_n; \mathbf{X})\right| \geq -\int_{b+\rho}^{B-\rho} \frac{d^2 \log(g(x)}{dx^2} g(x)\, dx$$

y puesto que esta desigualdad es válida para todo $\rho \in (0, B - b)$, tomando límite conforme ρ tiende a cero por la derecha se desprende que

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n}\left|\partial_\theta^2 \mathcal{L}_n(\tilde{\theta}_n; \mathbf{X})\right| &\geq -\int_b^B \frac{d^2 \log(g(x)}{dx^2} g(x)\, dx \\
&= -\int_{\mathbb{R}} \frac{d^2 \log(g(x)}{dx^2} g(x)\, dx = \mathcal{I}
\end{aligned}$$

donde la segunda igualdad se debe al Teorema 2.6.1. $\square$

Demostración del Teorema 4.4.1. Recuerde que el EVM $\hat{\theta}_n$ satisface $\partial_\theta \mathcal{L}_n(\hat{\theta}_n; \mathbf{X}) = 0$, por el Teorema 4.3.1. Combinando este hecho con el teorema de valor medio (Fulks, 1980) se desprende que existe $\tilde{\theta}_n$ entre θ_0 y $\hat{\theta}_n$ tal que

$$\partial_\theta \mathcal{L}_n(\theta_0) = \partial_\theta \mathcal{L}_n(\theta_0) - \partial_\theta \mathcal{L}_n(\hat{\theta}) = \partial_\theta^2 \mathcal{L}_n(\tilde{\theta})(\theta_0 - \hat{\theta}_n)$$

y por lo tanto

$$\sqrt{n}\left|\frac{\partial_\theta \mathcal{L}_n(\theta_0)}{n}\right| = \left|\frac{\partial_\theta^2 \mathcal{L}_n(\tilde{\theta})}{n}\right| \; |\sqrt{n}[\theta_0 - \hat{\theta}_n]|. \tag{4.4.8}$$

Como $\hat{\theta}_n$ y θ_0 pertenecen a $[X_{(n)} - B,\ b - X_{(1)}]$, se tiene que $\tilde{\theta}_n \in (X_{(n)} - B,\ b - X_{(1)})$, y entonces

$$\lim_{n\to\infty}\left|\frac{\partial_\theta^2 \mathcal{L}_n(\tilde{\theta})}{n}\right| \geq \mathcal{I}, \tag{4.4.9}$$

por el Lema 4.4.1. A continuación, note que para cada $x > 0$ y $\rho \in (0, \mathcal{I})$

$$\begin{aligned}
&\left[\sqrt{n}\left|\frac{\partial_\theta \mathcal{L}_n(\theta_0)}{n}\right| \leq x\right] \\
&\qquad = \left[\left|\frac{\partial_\theta^2 \mathcal{L}_n(\tilde{\theta})}{n}\right| \; |\sqrt{n}\,[\theta_0 - \hat{\theta}_n]| \leq x\right] \\
&\qquad = \left[\left|\frac{\partial_\theta^2 \mathcal{L}_n(\tilde{\theta})}{n}\right| \; |\sqrt{n}\,[\theta_0 - \hat{\theta}_n]| \leq x,\ \left|\frac{\partial_\theta^2 \mathcal{L}_n(\tilde{\theta})}{n}\right| \geq \mathcal{I} - \rho\right] \\
&\qquad\qquad \cup \left[\left|\frac{\partial_\theta^2 \mathcal{L}_n(\tilde{\theta})}{n}\right| \; |\sqrt{n}\,[\theta_0 - \hat{\theta}_n]| \leq x,\ \left|\frac{\partial_\theta^2 \mathcal{L}_n(\tilde{\theta})}{n}\right| < \mathcal{I} - \rho\right] \\
&\qquad \subset \left[(\mathcal{I} - \rho)\ |\sqrt{n}\,[\theta_0 - \hat{\theta}_n]| \leq x,\ \left|\frac{\partial_\theta^2 \mathcal{L}_n(\tilde{\theta})}{n}\right| \geq \mathcal{I} - \rho\right] \\
&\qquad\qquad \cup \left[\left|\frac{\partial_\theta^2 \mathcal{L}_n(\tilde{\theta})}{n}\right| < \mathcal{I} - \rho\right] \\
&\qquad \subset \left[(\mathcal{I} - \rho)\ |\sqrt{n}\,[\theta_0 - \hat{\theta}_n]| \leq x\ \right] \cup \left[\left|\frac{\partial_\theta^2 \mathcal{L}_n(\tilde{\theta})}{n}\right| < \mathcal{I} - \rho\right] \\
&\qquad = \left[|\sqrt{n}\,[\theta_0 - \hat{\theta}_n]| \leq \frac{x}{\mathcal{I} - \rho}\ \right] \cup \left[\left|\frac{\partial_\theta^2 \mathcal{L}_n(\tilde{\theta})}{n}\right| < \mathcal{I} - \rho\right]
\end{aligned}$$

y por lo tanto,

$$\begin{aligned} &P_{\theta_0}\left[\sqrt{n}\left|\frac{\partial_\theta \mathcal{L}_n(\theta_0)}{n}\right| \leq x\right] \\ &\leq P_{\theta_0}\left[|\sqrt{n}\,[\theta_0 - \hat{\theta}_n]| \leq \frac{x}{\mathcal{I}-\rho}\right] + P_{\theta_0}\left[\left|\frac{\partial_\theta^2 \mathcal{L}_n(\tilde{\theta})}{n}\right| < \mathcal{I}-\rho\right] \end{aligned} \tag{4.4.10}$$

Por otro lado, observe que

$$\lim_{n\to\infty} P_{\theta_0}\left[\sqrt{n}\left|\frac{\partial_\theta \mathcal{L}_n(\theta_0)}{n}\right| \leq x\right] = \Phi(x/\sqrt{\mathcal{I}}) - \Phi(-x/\sqrt{\mathcal{I}}),$$

por el Teorema 4.3.2, mientras que (4.4.9) implica que

$$\lim_{n\to\infty} P_{\theta_0}\left[\left|\frac{\partial_\theta^2 \mathcal{L}_n(\tilde{\theta})}{n}\right| < \mathcal{I}-\rho\right] = 0.$$

Luego, tomando límite cuando $n \to \infty$ en ambos lados de (4.4.10) se desprende que

$$\Phi(x/\sqrt{\mathcal{I}}\,) - \Phi(-x/\sqrt{\mathcal{I}}\,) \leq \lim_{n\to\infty} P_{\theta_0}\left[|\sqrt{n}\,[\theta_0 - \hat{\theta}_n]| \leq \frac{x}{\mathcal{I}-\rho}\right].$$

Sustituyendo x por $x(\mathcal{I}-\rho)$ se obtiene

$$\Phi(x(\mathcal{I}-\rho)/\sqrt{\mathcal{I}}\,) - \Phi(-x(\mathcal{I}-\rho)/\sqrt{\mathcal{I}}\,) \leq \lim_{n\to\infty} P_{\theta_0}\left[|\sqrt{n}\,[\theta_0 - \hat{\theta}_n]| \leq x\right].$$

Esta relación es válida para todo $\rho \in (0, \mathcal{I})$, de manera que dejando que ρ tienda a cero por la derecha, se obtiene

$$\Phi(x\sqrt{\mathcal{I}}\,) - \Phi(-x\sqrt{\mathcal{I}}\,) \leq \lim_{n\to\infty} P_{\theta_0}\left[|\sqrt{n}\,[\theta_0 - \hat{\theta}_n]| \leq x\right]$$

concluyendo la demostración. □

Observación 4.4.1. El Teorema 4.4.1 es una extensión débil del resultado clásico de convergencia para la distribución del estimador de verosimilitud máxima enunciado en el Teorema 1.4.1(ii):

$$\sqrt{n}\,[\hat{\theta}_n - \theta_0] \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, 1/\mathcal{I}\right), \tag{4.4.11}$$

el cual es válido bajo las condiciones de regularidad R_1–R_3 formuladas en la Sección 1.4. Cuando esta convergencia ocurre, se tiene que para cada $x > 0$

$$P_{\theta_0}\left[\sqrt{n}\,[\hat{\theta}_n - \theta_0] \le x\right] \to \Phi(x\sqrt{\mathcal{I}}\,) - \Phi(-x\sqrt{\mathcal{I}}\,)$$

conclusión que es más fuerte que (4.4.2). Como se mencionó en el Capítulo 1, la razón por la que, en el contexto de este trabajo, el argumento clásico no puede utilizarse para demostrar (4.4.11) es que la condición de acotamiento R_3 generalmente no se satisface cuando la densidad $g(\cdot)$ de los errores se anula fuera de un intervalo acotado. En efecto, bajo R_3, la tercera derivada $\partial_\theta^3 \mathcal{L}_n(\theta; \mathbf{X})$ debe tener esperanza finita. Sin embargo, si $g(\cdot)$ es la densidad beta en el Ejemplo 2.2.1(d), se tiene que

$$\partial_\theta^3 \mathcal{L}_n(\theta; \mathbf{X}) = \sum_{i=1}^{n} [-\alpha/(X_i - \theta + 1/2)^3 + \beta/(1/2 - X_i + \theta)^3] I_{[-1/2,\,1/2]}(X_i - \theta)$$

la cual tiene esperanza finita si y sólo si $\alpha, \beta > 3$, mientras que las condiciones del Teorema 4.4.1 se satisfacen cuando $\alpha, \beta > 2$. Más aún, es posible verificar que no existe función $H(\mathbf{X})$ tal que su esperanza sea finita respecto a cada P_θ y que satisfaga $|\partial_\theta^3 \mathcal{L}_n(\theta; \mathbf{X})| \le H(\mathbf{X})$, pues a partir de la fórmula anterior, $\partial_\theta^3 \mathcal{L}_n(\theta; \mathbf{X}) \to \infty$ conforme alguna variable X_i se aproxima a $\theta \pm 1/2$. □

Intervalos de Confianza

En esta sección se utiliza el Teorema 4.4.1 para establecer intervalos de confianza para θ con un coeficiente asintótico determinado, y sus longitudes se comparan con aquellas de los intervalos obtenidos mediante los métodos de momentos y de cuantiles. Con este fin, suponga que las condiciones del Teorema 4.4.1 se satisfacen, y considere el intervalo

$$K_n := \left[\hat{\theta}_n - \frac{z_{\gamma/2}}{\sqrt{n\mathcal{I}}},\ \ \hat{\theta}_n + \frac{z_{\gamma/2}}{\sqrt{n\mathcal{I}}}\right]. \tag{4.5.1}$$

Observando que la inclusión $\theta \in K_n$ equivale a $\sqrt{n}|\hat{\theta}_n - \theta| \leq \dfrac{z_{\gamma/2}}{\sqrt{\mathcal{I}}}$, se tiene que

$$\begin{aligned}
\lim_{n\to\infty} P_\theta[\theta \in K_n] &= \lim_{n\to\infty} P_\theta \left[\sqrt{n}|\hat{\theta}_n - \theta| \leq \frac{z_{\gamma/2}}{\sqrt{\mathcal{I}}}\right] \\
&\geq \Phi\left(\frac{z_{\gamma/2}}{\sqrt{\mathcal{I}}}\sqrt{\mathcal{I}}\right) - \Phi\left(-\frac{z_{\gamma/2}}{\sqrt{\mathcal{I}}}\sqrt{\mathcal{I}}\right) \\
&= \Phi\left(z_{\gamma/2}\right) - \Phi\left(-z_{\gamma/2}\right) \\
&= 1 - \gamma
\end{aligned}$$

estableciendo la parte (i) del siguiente resultado.

Teorema 4.5.1. Bajo las condiciones del Teorema 4.4.1 las afirmaciones (i) y (ii) son válidas:

(i) Como intervalos de confianza para θ, los intervalos K_n en (4.5.1) tienen un nivel asintótico de confianza $1 - \gamma$, esto es, $\lim_{n\to\infty} P_\theta[\theta \in K_n] \geq 1 - \gamma$.

(ii) Sean I_n y J_n los intervalos de confianza para θ obtenidos por el método de los momentos y de cuantiles, respectivamente; vea (3.2.8) y (3.3.4). En este caso,

$$\frac{|K_n|}{|I_n|} \leq 1 \quad \text{y} \quad \frac{|K_n|}{|J_n|} \leq 1;$$

en este sentido, los intervalos de confianza obtenidos mediante el método de verosimilitud máxima son más eficientes que aquellos generados por las técnicas de cuantiles y de momentos.

Demostración. (ii) Las longitudes de I_n, J_n y K_n son

$$|I_n| = 2\frac{z_{\gamma/2}}{\sqrt{n}}\sigma_g, \quad |J_n| = 2\frac{z_{\gamma/2}}{\sqrt{n}}\sqrt{\frac{p(1-p)}{g(\xi_p^g)^2}}, \quad |K_n| = 2\frac{z_{\gamma/2}}{\sqrt{n}}\frac{1}{\sqrt{\mathcal{I}}}.$$

Por lo tanto,

$$\frac{|K_n|}{|I_n|} = \frac{1}{\sigma_g\sqrt{\mathcal{I}}} \qquad \text{y} \qquad \frac{|K_n|}{|J_n|} = \sqrt{\frac{g(\xi_p^g)^2}{p(1-p)}}\frac{1}{\sqrt{\mathcal{I}}}$$

y, por (3.2.6) y el Corolario 3.4.1, ambos cocientes son menores o iguales a uno. □

Ejemplo 4.5.1. Consider el modelo de traslación en el que $g(\cdot)$ es la densidad beta

$$g(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \left(\frac{1}{2}+x\right)^{\alpha-1} \left(\frac{1}{2}-x\right)^{\beta-1} I_{[-1/2,1/2]}(x),$$

donde $\alpha, \beta > 2$. En este caso, se verificó directamente en la Proposición 3.2.0 que $\sigma_f^2 \mathcal{I} > 1$ y por lo tanto $|K_n|/|I_n| = 1/(\sigma_g\sqrt{\mathcal{I}}) < 1$. De hecho

$$\begin{aligned} \sigma_g^2 \mathcal{I} &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \frac{(\alpha+\beta-4)(\alpha+\beta-1)(\alpha+\beta-2)}{(\alpha-2)(\beta-2)} \\ &\geq \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \frac{4(\alpha+\beta-1)(\alpha+\beta-2)}{(\alpha+\beta-4)} \end{aligned}$$

de manera que $\sigma_g^2 \mathcal{I} \to \infty$ conforme $\alpha, \beta \searrow 2$, y entonces $1/(\sigma_g\sqrt{\mathcal{I}}) \searrow 0$ si $\alpha, \beta \searrow 2$, lo cual significa que K_n es sustancialmente más corto que I_n si α y β son cercanos a 2. Un resultado semejante es válido para los intervalos J_n construidos por el método de cuantiles. □

Capítulo 5

CONCLUSIONES

En este trabajo se tomó el problema de obtener intervalos de confianza con probabilidad asintótica determinada en un modelo estadístico de traslación. Al buscar entre los métodos de estimación con que contamos para construir estos intervalos resalta, por sus características, el método de verosimilitud máxima; sin embargo, este método requiere que el modelo en análisis cumpla con ciertas condiciones de regularidad. En este caso, el modelo no cumple con esas condiciones de regularidad por lo que no contamos con el respaldo teórico que nos permita utilizar el MVM como método de estimación para construir los intervalos de confianza que queremos generar.

Así, después de una revisión del modelo en análisis y de los métodos de estimación de momentos y de cuantiles, se demostró la consistencia y se obtuvo una cota para la distribución asintótica del EVM bajo condiciones más débiles que las condiciones de regularidad usuales. En el ejemplo visto, la condición de que la familia de distribuciones $\{f(\mathbf{x};\theta)|\ \theta \in \Theta\}$ tuvieran el mismo soporte sin importar el valor que tomara θ no se satisface; así también, la existencia de una función que sirviera de cota para la tercera derivada del logaritmo de $f(\mathbf{x};\theta)$ y cuya integral fuera finita tampoco fue cumplida. Se optó por utilizar la distribución beta como distribución representativa del modelo en cuestión debido a la variedad de formas que toma al cambiar el valor de sus parámetros.

Como resultado final del trabajo realizado, se construyeron intervalos de confianza con probabilidad asintótica determinada de contener a θ, obtenidos a través del MVM; despues de esto, se demostró que los intervalos que produce el MVM son en el límite más cortos que aquellos producidos por los métodos de momentos y de cuantiles.

LITERATURA CITADA

Ash, R. (1987), *Real Analysis and Probability*, *Academic Press*, New York.

Billingsley, P. (1997), *Probability & Measure*, *John Wiley & Sons*, New York.

Casella, G. y R. Berger (2001), *Statistical Inference*, *Duxbury Press*, New York.

Dudewicz, E. y N. Mishra (1989), *Modern Mathematical Statistics*, *John WIley & Sons*, New York.

Dudley, R.M. (2002), *Real Analysis and Probability*, *Cambridge University Press*, Boston.

Fulks, W. (1980), *Cálculo Avanzado*, *Limusa*, México D.F.

Greene, W.H. (2003), *Econometric Analysis*, *Prentice–Hall*, New York.

Graybill, F.A. (1985), *Theory and Applications of the Linear Model*, *Duxbury Press*, Boston.

Griffiths, W.E., R. Carter-Hill y George G. Judge (1997), *Learning and Practicing Econometrics*, *John Wiley & Sons*, New York.

Lehmann, E.L. y G. Casella (2000), *Point Estimation*, *John Willey & Sons*, New York.

Lehmann, E.L. (2001), *Testing Statistical Hypotheses*, *John Willey & Sons*, New York.

Mood, A.M., F. A. Graybill y D. C. Boes (1987), *Introduction to the Theory of Statistics*, *McGraw-Hill*, New York

Rao, C.R. (1988), *Linear Statistical Inference and Its Application*, *John Wiley & Sons*, New York.

Serfling, R.J. (1988), *Approximation Theorems of Mathematical Statistics*, *John Wiley & Sons*, New York.

Severini, T.A. (2001), *Likelihood Methods in Statistics*, *Oxford University Press*, London.