

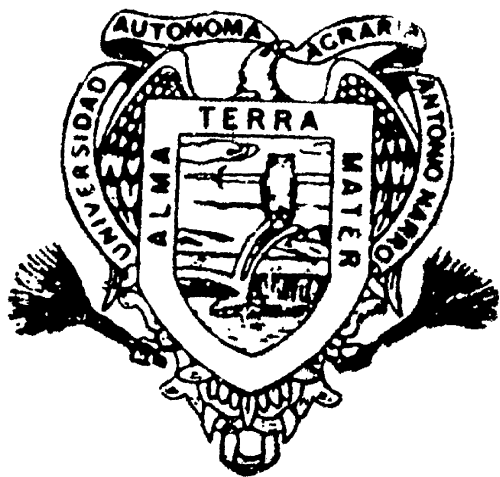
# CONSISTENCIA DEL METODO DE VEROSIMILITUD MAXIMA PARA UN MODELO NORMAL SESGADO

LUCIA MARISOL VALDES GONZALEZ

## TESIS

Presentada como Requisito Parcial para  
Obtener el Grado de:

MAESTRO EN  
ESTADISTICA APLICADA



UNIVERSIDAD AUTONOMA AGRARIA  
"ANTONIO NARRO"

PROGRAMA DE GRADUADOS

Buenavista, Saltillo, Coahuila, México  
Marzo de 2010

Universidad Autónoma Agraria Antonio Narro  
Dirección de Postgrado

CONSISTENCIA DEL MÉTODO DE VEROSIMILITUD  
MÁXIMA PARA UN MODELO NORMAL SESGADO

TESIS

Por:

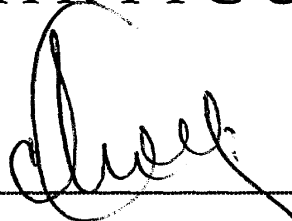
LUCÍA MARISOL VALDÉS GONZÁLEZ

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

MAESTRO EN  
ESTADÍSTICA APLICADA

COMITÉ PARTICULAR

Asesor principal:



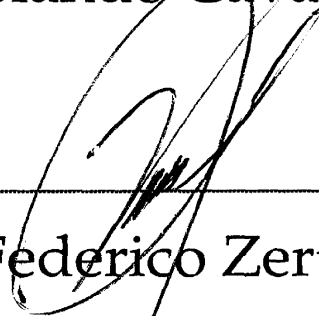
Dr. Mario Cantú Sifuentes

Asesor:



Dr. Rolando Cavazos Cadena

Asesor:



Dr. Federico Zertuche Luis



Dr. Jerónimo Landeros Flores  
Director de Postgrado

Buenvista, Saltillo, Coahuila, Marzo de 2010

# AGRADECIMIENTOS

*Dr. Mario Cantú Sifuentes*

*Dr. Rolando Cavazos Cadena*

*Dr. Federico Zertuche Luis*

*Dr. Jerónimo Landeros Flores*

# COMPENDIO

## CONSISTENCIA DEL MÉTODO DE VEROSIMILITUD MÁXIMA PARA UN MODELO NORMAL SESGADO

Por

LUCÍA MARISOL VALDÉS GONZÁLEZ

MAESTRÍA EN

ESTADÍSTICA APLICADA

UNIVERSIDAD AUTÓNOMA AGRARIA

“ANTONIO NARRO”

BUENAVISTA, SALTILLO, COAHUILA, Marzo de 2010

Dr. Mario Cantú Sifuentes –Asesor–

**Palabras clave:** Ley cero-uno de Hewitt-Savage, Ley fuerte de los grandes números, Criterio de consistencia, Acotamiento casi en toda parte.

En este trabajo se establecen la existencia y consistencia de una sucesión de estimadores de verosimilitud máxima para la familia normal sesgada unidimensional, la cual incluye parámetros de localización y escala. Se discute brevemente la importancia de este problema en el contexto de la literatura actual. El enfoque de este trabajo se basa en una reparametrización en línea de la función de verosimilitud, un mecanismo que permite obtener cotas para los maximizadores en términos de la muestra.

# ABSTRACT

## CONSISTENCY OF THE MAXIMUM LIKELIHOOD METHOD FOR A SKEW NORMAL MODEL

By

LUCÍA MARISOL VALDÉS GONZÁLEZ

MASTER

APPLIED STATISTICS

UNIVERSIDAD AUTÓNOMA AGRARIA

ANTONIO NARRO

BUENAVISTA, SALTILLO, COAHUILA, March, 2010

Dr. Mario Cantú Sifuentes –Advisor–

**Key Words:** Hewitt-Savage zero-one law, Strong law of large numbers, Consistency criterion, Boundedness almost everywhere.

For the one-dimensional skew-normal model with location and scale parameters, the existence and consistency of a sequence of maximum likelihood estimators is established, and the importance of this problem within the current literature is briefly discussed. The approach of this work is based on a reparametrization of the likelihood function, a device that allows to obtain bounds for the maximizers in terms of the sample.

# ÍNDICE DE CONTENIDO

<b>Capítulo 1</b>	<b>Introducción</b>	<b>1</b>
1.1	El Problema . . . . .	1
1.2	Los Instrumentos Básicos. . . . .	3
1.3	La Organización . . . . .	3
<b>Capítulo 2</b>	<b>Generación de una Familia de Densidades Sesgadas</b>	<b>5</b>
2.1	Introducción . . . . .	5
2.2	Construcción de una Familia Sesgada . . . . .	6
2.3	Construcción Alternativa. . . . .	11
2.4	Una Representación Alternativa. . . . .	14
2.5	Momentos de una Distribución Sesgada . . . . .	15
<b>Capítulo 3</b>	<b>El Caso Normal</b>	<b>18</b>
3.1	Introducción . . . . .	18
3.2	La Distribución Normal Sesgada . . . . .	18
3.3	Media y Varianza . . . . .	21
3.4	Relación con la Distribución Normal de Dimension Dos . . . . .	23
3.5	Importancia de la Distribución Normal Sesgada. . . . .	26
3.6	El Modelo de Localización y Escala . . . . .	26

<b>Capítulo 4</b>	<b>Consistencia de Estimadores de Verosimilitud Máxima</b>	<b>27</b>
4.1	Introducción . . . . .	27
4.2	Existencia de Estimadores de Verosimilitud Máxima . . . . .	30
4.3	Cotas para posibles maximizadores . . . . .	34
4.4	Maximización de la Función Auxiliar . . . . .	39
4.5	Demostración del Teorema de Existencia . . . . .	46
4.6	Consistencia . . . . .	48
<b>Literatura Citada</b>	. . . . .	<b>57</b>

# CAPÍTULO 1

## INTRODUCCIÓN

Este trabajo trata sobre familias de densidades sesgadas, las cuales se obtienen a partir de dos densidades simétricas. Se enfatiza el modelo normal sesgado y la aplicación del *método de estimación de verosimilitud máxima* para estimar los parámetros en un modelo sesgado de localización y escala. Dicho método es ampliamente utilizado en la práctica estadística debido a que posee propiedades de optimalidad muy importantes. Por ejemplo, bajo condiciones de regularidad adecuadas, los estimadores de verosimilitud máxima tienen *la mínima varianza asintótica* y, más aún, su distribución límite es normal (Serfling, 1988, Severini, 2001). Por otro lado, en esta tesis el análisis desarrollado se centra en un aspecto mucho más básico del esquema de verosimilitud máxima, el cual tiene que ver con el comportamiento de los estimadores conforme el tamaño de la muestra crece. La propiedad en cuestión es la *consistencia*, y el objetivo de este capítulo introductorio es describir brevemente el problema específico que se estudiará, el principal resultado que se establece, así como los instrumentos técnicos que se utilizarán.

### 1.1 EL PROBLEMA

En términos generales, el punto de partida en un problema de inferencia es un vector de datos  $Y = (Y_1, Y_2, \dots, Y_n)$ , cuya distribución no está completamente especificada. En este trabajo se supone que los datos forman una muestra de una densidad conocida como 'densidad sesgada'. Dicha densidad se genera a partir de dos densidades simétricas, mediante un proceso de condicionamiento que se describe con precisión en el Capítulo 2. El interés en este tipo de distribución se



origina en el hecho de que surge de manera natural en la práctica y en que, desde un punto de vista analítico, su análisis presenta retos importantes. Por ejemplo, aún para casos relativamente simples, como en el que la densidad sesgada se obtiene de distribuciones normales, se muestra en el Capítulo 3 que el cálculo de su valor esperado requiere de argumentos indirectos que no son necesarios en el estudio de otras distribuciones más comunes. Como ya se mencionó, el propósito central de este trabajo es analizar, en el contexto de una distribución normal sesgada, el método de verosimilitud máxima para construir estimadores, y resolver el siguiente problema:

**Problema Básico:** *Establecer la existencia y consistencia de los estimadores de verosimilitud máxima.*

En contextos más clásicos, la consistencia de los estimadores máximo verosímiles ha sido obtenida bajo las siguientes condiciones suficientes impuestas sobre las posibles densidades  $f(y, \theta)$  de  $Y$  (Newey y McFadden, 1993):

- i) Para cada  $y$ , la función  $\theta \mapsto \log(f(y, \theta))$  es continua y el espacio de parámetros  $\Theta$  es compacto;
- ii) Para cada  $y$ , la función  $\theta \mapsto \log(f(y, \theta))$  es estrictamente cóncava, el espacio de parámetros es convexo y el verdadero valor de  $\theta$  se ubica en el interior de  $\Theta$ .

Sin embargo, en el modelo normal sesgado estas condiciones no se satisfacen, de manera que es importante e interesante analizar el problema planteado anteriormente en ese contexto.

*El Resultado Principal* de este trabajo se encuentra en el Capítulo 4, donde se establece el resultado deseado de consistencia.

## 1.2 LOS INSTRUMENTOS BÁSICOS

Las herramientas estadísticas que se utilizarán para analizar la consistencia de los estimadores de verosimilitud máxima son, esencialmente, dos:

E1: La ley de los grandes números, y

E2: La desigualdad de Jensen, la cual relaciona las ideas de valor esperado y de función cóncava. Cuando se aplica esta desigualdad a la función logaritmo natural se obtiene la desigualdad de Kullback, que será muy útil en los dos últimos capítulos de este trabajo.

Por otro lado, el instrumento matemático esencial en el desarrollo subsecuente es el siguiente:

M1 La idea de conjunto compacto y propiedades básicas referentes a la noción de función continua.

## 1.3 LA ORGANIZACIÓN

*La exposición de este trabajo ha sido organizada de la siguiente manera:*

En el Capítulo 2 se introduce la clase de familias sesgadas, indicando como su construcción involucra dos densidades simétricas; en este punto, la presentación sigue de cerca el enfoque de Azzalini (1985), donde las densidades sesgadas fueron originalmente introducidas para el caso en que las densidades generadoras coinciden ambas con la densidad normal estándar. Posteriormente, se estudia otra forma de obtener familias sesgadas por medio de un argumento que involucra probabilidades condicionales, el cual se utilizará más adelante para entender la gran importancia práctica de una distribución sesgada. Todavía dentro del segundo capítulo, se describe una representación adicional de una variable aleatoria con densidad sesgada, la cual es útil para estudiar sus momentos respecto al origen. Por otro lado, en el Capítulo 3 se introduce la familia normal sesgada y, como una manera de entender el papel del parámetro de sesgo (o asimetría), se determina la probabilidad de que dicha distribución asigne a los intervalos  $(0, \infty)$  y  $(-\infty, 0)$ , así como la media y la varianza de dicha distribución. En este punto se

ilustra como la evaluación de estas cantidades requiere de una técnica especial, basada en la diferenciación. Posteriormente, se analiza la relación existente entre una distribución normal sesgada y la distribución normal bidimensional clásica, y la representación obtenida se utiliza para enfatizar el importante papel de la distribución normal sesgada en las aplicaciones. Finalmente, en el Capítulo 4 se introduce formalmente la noción de consistencia y se establece la existencia y consistencia de estimadores de verosimilitud máxima para la distribución normal sesgada. La presentación en esta parte es autocontenida y relativamente independiente del resto del material. La razón es que el enfoque es mucho mas técnico y requiere de ideas matemáticas más avanzadas que las utilizadas en los capítulos precedentes.

# CAPÍTULO 2

## GENERACIÓN DE UNA FAMILIA DE DENSIDADES SESGADAS

### 2.1 INTRODUCCIÓN

En este capítulo se introducen las familias de densidades sesgadas, las cuales permiten incluir una densidad simétrica respecto al origen, denotada por  $f$ , dentro de una clase de densidades en la que, en general,  $f$  se distingue por ser la única que es simétrica. La construcción que se presenta más adelante fue originalmente propuesta en Azzalini (1985), en donde se estudió la inclusión de la densidad normal estándar dentro de una familia de densidades. La presentación está motivada por el trabajo de Dalla Valle (2004), el cual se refiere al caso normal; sin embargo, los argumentos a continuación no requieren forma específica alguna para las densidades que generan una familia sesgada.

La exposición del capítulo ha sido organizada de la siguiente manera: En la Sección 2 se introduce la familia sesgada generada por dos densidades simétricas  $f$  y  $g$  respecto al origen, siguiendo la idea de Azzalini (1985), y se demuestra que, en general, dicha familia incluye sólo una densidad simétrica. En la Sección 3 se muestra como una familia sesgada puede generarse mediante argumentos de distribuciones condicionales, construcción que es la base para entender la importancia práctica de las familias sesgadas, como se señalará en el siguiente capítulo. Posteriormente, en la Sección 4 se representa una variable aleatoria con distribución sesgada como una suma que involucra a dos variables con distribución simétrica, y este resultado se aplica en la Sección 5 para estudiar los momentos de

una distribución sesgada.

## 2.2 CONSTRUCCIÓN DE UNA FAMILIA SESGADA

En esta sección se describe una construcción general que permite incluir una densidad simétrica  $f$  dada, en una familia de densidades  $\mathcal{A}$ , de manera que, en general,  $f$  es la única densidad simétrica en  $\mathcal{A}$ . Como punto de partida, sean  $f$  y  $g$  dos densidades simétricas (respecto al origen) definidas en la recta numérica, esto es  $f, g : \mathbb{R} \rightarrow [0, \infty)$  satisfacen

$$f(x) = f(-x) \quad \text{y} \quad g(x) = g(-x), \quad x \in \mathbb{R}, \quad (2.2.1)$$

y sea  $G(x)$  la correspondiente función de distribución asociada a  $g$ , esto es,

$$G(x) = \int_{-\infty}^x g(y) dy \quad (2.2.2)$$

la cual está sujeta a la relación

$$G(-x) = 1 - G(x), \quad x \in \mathbb{R}. \quad (2.2.3)$$

Sea ahora  $\lambda$  cualquier número real y note que la anterior ecuación implica que,

$$f(x) = f(x)G(\lambda x) + f(x)[1 - G(\lambda x)] = f(x)G(\lambda x) + f(x)G(-\lambda x), \quad x \in \mathbb{R},$$

de manera que

$$\begin{aligned} 1 &= \int_{\mathbb{R}} f(x) dx \\ &= \int_{\mathbb{R}} f(x)G(\lambda x) dx + \int_{\mathbb{R}} f(x)G(-\lambda x) dx; \end{aligned}$$

note ahora que, utilizando la simetría de  $f$ , el cambio de variable  $z = -x$  en la segunda integral arroja que

$$\int_{\mathbb{R}} f(x)G(-\lambda x) dx = \int_{\mathbb{R}} f(-z)G(\lambda z) dz = \int_{\mathbb{R}} f(z)G(\lambda z) dz$$

y combinando las dos últimas relaciones desplegadas, se desprende que

$$1 = \int_{\mathbb{R}} f(x)G(\lambda x) dx + \int_{\mathbb{R}} f(z)G(\lambda z) dz = 2 \int_{\mathbb{R}} f(x)G(\lambda x) dx;$$

este argumento se debe a Azzalini (1985), en donde se aplicó al caso en que las densidades  $f$  y  $g$  coinciden con la densidad normal estándar. La anterior igualdad establece el siguiente resultado.

**Lema 2.2.1.** Sean  $f, g : \mathbb{R} \rightarrow [0, \infty)$  dos densidades simétricas (vea 2.2.1). En este caso, para cada  $\lambda \in \mathbb{R}$ , la función  $\rho(\cdot, \lambda) : \mathbb{R} \rightarrow [0, \infty)$  dada por

$$\rho(x, \lambda) = 2f(x)G(\lambda x), \quad x \in \mathbb{R} \quad (2.2.4)$$

es una densidad.

Usando que  $G(0) = \frac{1}{2}$ , igualdad que se desprende de (2.2.3), se obtiene que

$$\rho(x, 0) = 2f(x)G(0) = f(x).$$

A continuación se demostrará que, en general,  $\rho(\cdot, 0) = f(\cdot)$  es la única densidad simétrica dentro de la familia de densidades  $\{\rho(\cdot, \lambda)\}_{\lambda \in \mathbb{R}}$ .

**Lema 2.2.2.** Suponga que la función de distribución  $G(x)$  es estrictamente creciente. En este caso, la densidad  $\rho(\cdot, \lambda)$  en (2.2.4) es simétrica si y sólo si  $\lambda = 0$ . Más explícitamente,

$$\rho(x, \lambda) = \rho(-x, \lambda) \quad \text{para todo } x \in \mathbb{R} \iff \lambda = 0.$$

**Demostración.** Suponga que  $\lambda = 0$ .

En este contexto, se verificó líneas arriba que  $\rho(x, 0) = f(x)$ , la cual es una densidad simétrica (vea 2.2.1). Por lo tanto,

$$\lambda = 0 \implies \rho(x, \lambda) \quad \text{es simétrica.}$$

Suponga que  $\rho(x, \lambda)$  es una densidad simétrica.

En este caso, debe demostrarse que  $\lambda = 0$ . Para alcanzar esta conclusión se procederá mediante el método de contradicción. Así, suponga que  $\lambda \neq 0$  y considere los siguientes dos casos:

**Caso 1:**  $\lambda > 0$ . En esta situación, la función  $x \mapsto G(\lambda x)$  es estrictamente creciente, de manera que

$$G(\lambda x) < G(\lambda \cdot 0) = \frac{1}{2} \quad \text{si } x < 0$$

$$G(\lambda x) > G(\lambda \cdot 0) = \frac{1}{2} \quad \text{si } x > 0$$

Considere ahora un número  $x > 0$  con la propiedad de que  $f(-x) = f(x) > 0$ , y observe que a partir de (2.2.4) se obtiene

$$\rho(x, \lambda) = 2f(x)G(\lambda x) > 2f(x)G(\lambda \cdot 0) = f(x)$$

mientras que

$$\rho(-x, \lambda) = 2f(-x)G(\lambda(-x)) < 2f(x)G(\lambda 0) = f(-x) = f(x).$$

Combinando estas últimas dos relaciones se desprende que

$$\rho(x, \lambda) > f(x) > \rho(-x, \lambda)$$

y por lo tanto  $\rho(\cdot, x)$  no es simétrica, contradiciendo el supuesto inicial. Por lo tanto,  $\lambda > 0$  no puede ocurrir.

**Caso 2:**  $\lambda < 0$ . Bajo esta condición, la función  $x \mapsto G(\lambda x)$  es estrictamente decreciente, propiedad que conduce a

$$G(\lambda x) > G(\lambda \cdot 0) = \frac{1}{2} \quad \text{si } x < 0$$

$$G(\lambda x) < G(\lambda \cdot 0) = \frac{1}{2} \quad \text{si } x > 0$$

Como en el caso precedente, considere ahora un número  $x > 0$  con la propiedad de que  $f(-x) = f(x) > 0$  y, combinando la especificación de  $\rho(x, \lambda)$  en (2.2.4) con la anterior relación desplegada, note que

$$\rho(x, \lambda) = 2f(x)G(\lambda x) < 2f(x)G(\lambda 0) = f(x)$$

mientras que

$$\rho(-x, \lambda) = 2f(-x)G(\lambda(-x)) > 2f(x)G(\lambda 0) = f(-x) = f(x).$$

de manera que

$$\rho(x, \lambda) < f(x) < \rho(-x, \lambda).$$

Estas desigualdades muestran que  $\rho(\cdot, x)$  no es simétrica, lo cual contradice el supuesto inicial. Por lo tanto,  $\lambda < 0$  no puede ocurrir.

En resumen: Se ha demostrado que si la densidad  $\rho(\cdot, \lambda)$  es simétrica, entonces las opciones  $\lambda > 0$  y  $\lambda < 0$  conducen a una contradicción, de manera que la única alternativa posible es  $\lambda = 0$ , esto es,

$$\rho(\cdot, \lambda) \text{ es simétrica} \implies \lambda = 0,$$

concluyendo la demostración. □

**Ejemplo 2.2.1.** Suponga que  $f(x)$  y  $g(x)$  son iguales a la densidad de Cauchy:

$$f(x) = g(x) = \frac{1}{\pi} \left( \frac{1}{1+x^2} \right), \quad x \in \mathbb{R}.$$

En este caso,

$$G(x) = \int_{-\infty}^x \frac{1}{\pi} \left( \frac{1}{1+y^2} \right) dy = \frac{1}{\pi} \arctan(x) + \frac{1}{2}$$

y

$$\rho(x, \lambda) = 2 \frac{1}{\pi} \left( \frac{1}{1+x^2} \right) \left[ \frac{1}{\pi} \arctan(\lambda x) + \frac{1}{2} \right]$$

Observe que  $G(0) = \frac{1}{2} + \frac{\arctan(0)}{\pi} = \frac{1}{2}$  y

$$\rho(x, 0) = 2 \frac{1}{\pi} \left( \frac{1}{1+x^2} \right) \left[ \frac{1}{\pi} \arctan(0 \cdot x) + \frac{1}{2} \right] = \frac{1}{\pi} \left( \frac{1}{1+x^2} \right) = f(x)$$

mientras que

$$\begin{aligned} \rho(x, \lambda) = \rho(-x, \lambda) &\iff \left[ \frac{1}{\pi} \arctan(\lambda x) + \frac{1}{2} \right] = \left[ \frac{1}{\pi} \arctan(-\lambda x) + \frac{1}{2} \right] \\ &\iff \arctan(\lambda x) = \arctan(-\lambda x) \\ &\iff \lambda x = -\lambda x \\ &\iff 2\lambda x = 0 \end{aligned}$$

donde la tercera equivalencia se debe a que la función  $\arctan(\cdot)$  es uno a uno. Como  $\rho(x, \lambda)$  es simétrica respecto al origen si y sólo si  $\rho(x, \lambda) = \rho(-x, \lambda)$  para



todo  $x$ , se desprende a partir de las anteriores relaciones que  $\rho(x, \lambda)$  es simétrica si y sólo si  $\lambda x = 0$  para todo  $x$ , lo que equivale a  $\lambda = 0$ , confirmando la afirmación del Lema 2.2.1.  $\square$

Para concluir esta sección, se analizará el comportamiento de las densidades  $\rho(x, \lambda)$  conforme  $\lambda$  tiende a  $+\infty$  o  $-\infty$ .

**Lema 2.2.3.** *Para cada  $\lambda \in \mathbb{R}$  sea  $\rho(x, \lambda)$  la densidad en (2.2.4), donde  $f$  y  $g$  son dos densidades simétricas respecto al origen. En este caso,*

$$\begin{aligned} \lim_{\lambda \rightarrow \infty} \rho(x, \lambda) &= 2f(x), & \text{si } x > 0 \\ &= 0 & \text{si } x < 0 \end{aligned}$$

y

$$\begin{aligned} \lim_{\lambda \rightarrow -\infty} \rho(x, \lambda) &= 2f(x), & \text{si } x < 0 \\ &= 0 & \text{si } x > 0. \end{aligned}$$

**Demostración.** Recuerde que  $\lim_{x \rightarrow \infty} G(x) = 1$  y  $\lim_{x \rightarrow -\infty} G(x) = 0$ . Usando este hecho se obtiene que,

$$\lim_{\lambda \rightarrow \infty} G(\lambda x) = 1 \quad \text{si } x > 0 \quad \text{y} \quad \lim_{\lambda \rightarrow -\infty} G(\lambda x) = 0 \quad \text{si } x < 0$$

Combinando estas convergencias con la especificación de  $\rho(x, \lambda)$  en (2.2.4) se desprende que

$$\lim_{\lambda \rightarrow \infty} \rho(x, \lambda) = 2 \lim_{\lambda \rightarrow \infty} f(x)G(\lambda x) = 2f(x) \quad \text{si } x > 0,$$

y

$$\lim_{\lambda \rightarrow \infty} \rho(x, \lambda) = 2 \lim_{\lambda \rightarrow \infty} f(x)G(\lambda x) = 2f(x) \cdot 0 = 0 \quad \text{si } x < 0,$$

estableciendo la primera afirmación del lema, mientras que la segunda puede demostrarse a través de un argumento similar.  $\square$

**Ejemplo 2.2.2.** En el contexto del Ejemplo 2.2.1,

$$\lim_{\lambda \rightarrow \infty} \rho(x, \lambda) = \frac{2}{\pi} \left( \frac{1}{1+x^2} \right) \quad \text{si } x > 0 \quad \text{y} \quad \lim_{\lambda \rightarrow \infty} \rho(x, \lambda) = 0 \quad \text{si } x < 0.$$

$\square$

de manera que

$$P[(X, Y) \in B] = \int_{\mathbb{R}} \left[ \int_{B(x)} f(x)g(y) dy \right] dx \quad (2.3.6)$$

Después de este preámbulo, ahora se construirán las anteriores densidades sesgadas  $\rho(x, \lambda)$  mediante un argumento condicional.

**Lema 2.3.1.** Sean  $X$  y  $Y$  variables aleatorias independientes con densidades simétricas  $f$  y  $g$ , respectivamente, y sea  $\lambda$  un número real arbitrario. En este contexto, la densidad condicional de  $X$  dado el evento  $Y > -\lambda X$  es  $\rho(x, \lambda)$ . Más explícitamente, para cada intervalo  $A \subset \mathbb{R}$

$$P[X \in A | Y > -\lambda X] = \int_A \rho(x, \lambda) dx,$$

donde  $\rho(x, \lambda)$  está dada por (2.2.4).

**Demostración.** Para empezar, recuerde que

$$P[X \in A | Y > -\lambda X] = \frac{P[X \in A, Y > -\lambda X]}{P[Y > \lambda X]} \quad (2.3.7)$$

Ahora se calculará el numerador de este cociente usando el procedimiento de tres pasos delineado anteriormente. Primero observe que el evento  $[X \in A, Y > -\lambda X]$  puede escribirse como  $[(X, Y) \in B]$ , donde

$$B = \{(x, y) | x \in A, y > -\lambda x\}.$$

Ahora, note para cada  $x$  fijo, el conjunto  $B(x)$  en (2.3.5) está dado por

$$\begin{aligned} B(x) = \{y | (x, y) \in B\} &= \{y | x \in A, y > -\lambda x\} = (-\lambda x, \infty) \quad \text{si } x \in A \\ &= \emptyset \quad \text{cuando } x \notin A. \end{aligned}$$

Por lo tanto, para  $x \in A$ ,

$$\begin{aligned} \int_{B(x)} f(x)g(y) dy &= \int_{y \in (-\lambda x, \infty)} f(x)g(y) dy \\ &= f(x) \int_{-\lambda x}^{\infty} g(y) dy \\ &= f(x)(1 - G(-\lambda x)) \\ &= f(x)G(\lambda x) \end{aligned}$$

donde la última igualdad se debe a (2.2.3); note, además, que, debido a que  $B(x) = \emptyset$  cuando  $x \notin A$ , se tiene que

$$\int_{B(x)} f(x)g(y) dy = 0, \quad x \notin A.$$

Por lo tanto,

$$\begin{aligned} P[X \in A, Y > -\lambda X] &= P[(X, Y) \in B] \\ &= \int_{\mathbb{R}} \left[ \int_{B(x)} f(x)g(y) dy \right] dx \quad (2.3.8) \\ &= \int_A f(x)G(\lambda x) dx. \end{aligned}$$

Finalmente, observe que la simetría e independencia de  $X$  y  $Y$  implica que las parejas  $(X, Y)$  y  $(-X, -Y)$  tienen la misma distribución, de manera que la probabilidad del evento  $[Y > -\lambda X]$  es la misma que la de  $[-Y > -\lambda(-X)] = [Y < -\lambda X]$ :

$$P[Y > -\lambda X] = P[Y < -\lambda X].$$

Como  $P[Y = -\lambda X] = 0$ , pues el vector  $(X, Y)$  tiene densidad, se sigue que

$$\begin{aligned} 1 &= P[Y > -\lambda X] + P[Y = -\lambda X] + P[Y < -\lambda X] \\ &= P[Y > -\lambda X] + P[Y < -\lambda X] \\ &= 2P[Y > -\lambda X] \end{aligned}$$

y entonces

$$P[Y > -\lambda X] = \frac{1}{2}.$$

Combinando esta relación con (2.3.7) y (2.3.8) se sigue que

$$\begin{aligned} P[X \in A | Y > -\lambda X] &= \frac{\int_A f(x)G(\lambda x) dx}{\frac{1}{2}} \\ &= \int_A 2f(x)G(\lambda x) dx \\ &= \int_A \rho(x, \lambda) dx \end{aligned}$$

concluyendo la demostración. □

## 2.4 UNA REPRESENTACIÓN ALTERNATIVA

En esta sección se representa una variable aleatoria  $W$  con densidad sesgada (2.2.4) en términos de las variables aleatorias independientes  $X$  y  $Y$  de la sección anterior. El resultado en esta dirección se presenta en el siguiente lemma, el cual permitirá establecer una fórmula interesante para los momentos de orden par de  $W$  más adelante.

**Lema 2.4.1.** Sean  $X$  y  $Y$  variables aleatorias independientes con densidades simétricas  $f(x)$  y  $g(y)$ , respectivamente, y defina la variable aleatoria  $W$  mediante

$$W = XI[Y \geq -\lambda X] - XI[Y < -\lambda X], \quad (2.4.9)$$

donde  $\lambda \in \mathbb{R}$  es arbitrario. Más explícitamente,

$$\begin{aligned} W &= X && \text{si } Y \geq \lambda X \\ W &= -X && \text{si } Y < \lambda X. \end{aligned}$$

Con esta notación,  $W$  tiene densidad  $\rho(\cdot, \lambda)$  en (2.2.4).

**Demostración.** Se establecerá que

$$P[W \in A] = \int_A \rho(x, \lambda) dx$$

para todo intervalo  $A$ . Para lograr este objetivo, note que

$$P[W \in A] = P[W \in A, Y \geq -\lambda X] + P[W \in A, Y < -\lambda X] \quad (2.4.10)$$

Observe a continuación que la especificación de  $W$  implica que  $W = X$  cuando  $Y \geq -\lambda X$ , de manera que  $[W \in A, Y \geq -\lambda X] = [X \in A, Y \geq -\lambda X]$ . Por lo tanto,

$$P[W \in A, Y \geq -\lambda X] = P[X \in A, Y \geq -\lambda X]$$

y como se estableció en la demostración del Lema 2.3.1, la probabilidad en el lado derecho es  $\int_A f(x)G(\lambda x) dx$ , de manera que

$$P[W \in A, Y \geq -\lambda X] = \int_A f(x)G(\lambda x) dx \quad (2.4.11)$$

Por otro lado, a partir de la definición de  $W$  implica que  $W = -X$  cuando  $Y < -\lambda X$ , y entonces  $[W \in A, Y < -\lambda X] = [-X \in A, Y > -\lambda X]$ . Por lo tanto,

$$\begin{aligned} P[W \in A, Y < -\lambda X] &= P[-X \in A, Y < -\lambda X] \\ &= P[-X \in A, -Y > -\lambda(-X)] \end{aligned}$$

Recordando que las parejas  $(X, Y)$  y  $(-X, -Y)$  tienen la misma distribución, se sigue que  $P[-X \in A, -Y > -\lambda(-X)] = P[X \in A, Y > -\lambda X]$ , y entonces

$$\begin{aligned} P[W \in A, Y < -\lambda X] &= P[X \in A, Y > -\lambda X] \\ &= P[X \in A, Y \geq -\lambda X] \end{aligned}$$

donde se utilizó que  $P[Y = -\lambda X] = 0$  para establecer la última igualdad.

Combinando esta igualdad con (2.4.11) se sigue que

$$P[W \in A, Y < -\lambda X] = \int_A f(x)G(\lambda x) dx.$$

Para concluir, combine esta igualdad con (2.4.10) y (2.4.11) para obtener

$$P[W \in A] = \int_A f(x)G(\lambda x) dx + \int_A f(x)G(\lambda x) dx = \int_A 2f(x)G(\lambda x) dx,$$

es decir,  $P[W \in A] = \int_A \rho(x, \lambda) dx$ , estableciendo la conclusión deseada.  $\square$

## 2.5 MOMENTOS DE UNA DISTRIBUCIÓN SESGADA

En esta sección se analizan los momentos de la densidad  $\rho(\cdot, \lambda)$  en (2.2.4). El resultado principal establece que, si  $W$  tiene la densidad  $\rho(\cdot, \lambda)$ , entonces sus momentos absolutos coinciden con los correspondientes a  $X$ , donde  $X$  tiene densidad  $f(x)$ .

**Teorema 2.5.1.** *Suponga que  $W$  tiene densidad  $\rho(\cdot, \lambda)$ , donde  $\lambda \in \mathbb{R}$  es arbitrario, y sea  $X$  una variable aleatoria con la densidad simétrica  $f(x)$ . En este caso, las siguientes afirmaciones son válidas:*

(a) *La distribución de  $|W|$  y de  $|X|$  coinciden. Por lo tanto,*

(b)  $E[W^{2k}] = E[X^{2k}]$ ,  $k = 1, 2, \dots$  Además,

(c) Si  $\lambda \geq 0$ , entonces  $E[W^{2k+1}] \geq 0$ , mientras que  $E[W^{2k+1}] \leq 0$  para  $\lambda < 0$ .

**Demostración.** Sean  $X$  y  $Y$  variables aleatorias independientes con las densidades  $f(x)$  y  $g(y)$ , respectivamente, defina la nueva variable  $\tilde{W}$  como en la sección precedente, esto es,

$$\tilde{W} = XI[Y \geq -\lambda X] - XI[Y < -\lambda X], \quad (2.5.12)$$

En este caso,  $\tilde{W}$  tiene densidad  $\rho(\cdot, \lambda)$ , por el Lema 2.4.1, de tal suerte que

$$W \text{ y } \tilde{W} \text{ tienen la misma distribución.} \quad (2.5.13)$$

Por otro lado, a partir de (2.5.12) se sigue que  $\tilde{W} = X$  cuando  $Y \geq -\lambda X$  y  $\tilde{W} = -X$  si  $Y < -\lambda X$ , de manera que siempre se tiene que  $\tilde{W}$  es igual a  $X$  o a  $-X$ , y por lo tanto,  $\tilde{W}$  y  $X$  siempre tienen el mismo valor absoluto, es decir

$$|\tilde{W}| = |X|.$$

Combinando esta ecuación con (2.5.13) se obtiene que  $|W|$  y  $|X|$  tienen la misma distribución, estableciendo la afirmación (a). Para obtener la parte (b), note que  $|W|^m$  y  $|X|^m$  siempre tienen la misma esperanza, por la parte (a). Además, si  $m = 2k$  es un número par, se tiene que  $|W|^{2k} = W^{2k}$  y  $|X|^{2k} = X^{2k}$ , de manera que  $E[W^{2k}] = E[X^{2k}]$ , que es la conclusión deseada. Para establecer la parte (c), note que

$$\begin{aligned} E[W^{2k+1}] &= \int_{\mathbb{R}} x^{2k+1} \rho(x, \lambda) dx \\ &= \int_{\mathbb{R}} x^{2k+1} 2f(x)G(\lambda x) dx \\ &= \int_0^{\infty} x^{2k+1} 2f(x)G(\lambda x) dx + \int_{-\infty}^0 x^{2k+1} 2f(x)G(\lambda x) dx \\ &= \int_0^{\infty} x^{2k+1} 2f(x)G(\lambda x) dx + \int_0^{\infty} (-z)^{2k+1} 2f(-z)G(\lambda(-z)) dz \end{aligned}$$

donde se utilizó el cambio de variable  $x = -z$  para obtener la última expresión. Recordando que  $f(-z) = f(z)$  y que  $(-z)^{2k+1} = -z^{2k+1}$ , se desprende que

$$\begin{aligned}
E[W^{2k+1}] &= \int_0^{\infty} x^{2k+1} 2f(x)G(\lambda x) dx - \int_0^{\infty} z^{2k+1} 2f(z)G(\lambda(-z)) dz \\
&= \int_0^{\infty} x^{2k+1} 2f(x)G(\lambda x) dx - \int_0^{\infty} x^{2k+1} 2f(x)G(\lambda(-x)) dx \\
&= \int_0^{\infty} x^{2k+1} 2f(x)[G(\lambda x) - G(\lambda(-x))] dx
\end{aligned}$$

es decir,

$$E[W^{2k+1}] = \int_0^{\infty} x^{2k+1} 2f(x)[G(\lambda x) - G(-\lambda x)] dx$$

Finalmente, como  $G(-\lambda x) = 1 - G(\lambda x)$ , por (2.2.3), se obtiene la expresión

$$E[W^{2k+1}] = \int_0^{\infty} x^{2k+1} 2f(x)[2G(\lambda x) - 1] dx$$

Para concluir el argumento, considere ahora los siguientes dos casos:

**Caso 1:**  $\lambda \geq 0$ . En esta circunstancias,  $\lambda x \geq 0$  para todo  $x > 0$ , y entonces  $G(\lambda x) \geq G(0) = 1/2$ , de manera que  $2G(\lambda x) - 1 \geq 0$  para todo  $x > 0$ , y entonces (2.5.14) implica que

$$E[W^{2k+1}] = \int_0^{\infty} x^{2k+1} 2f(x)[2G(\lambda x) - 1] dx \geq 0. \quad (2.5.14)$$

**Caso 2:**  $\lambda < 0$ . En esta circunstancias,  $\lambda x < 0$  para todo  $x > 0$ , de tal manera que  $G(\lambda x) \leq G(0) = 1/2$ . Luego,  $2G(\lambda x) - 1 \leq 0$  para todo  $x > 0$ , y partir de (2.5.14) se desprende que

$$E[W^{2k+1}] = \int_0^{\infty} x^{2k+1} 2f(x)[2G(\lambda x) - 1] dx \leq 0 \quad (2.5.15)$$

concluyendo la demostración. □

# CAPÍTULO 3

## EL CASO NORMAL

### 3.1 INTRODUCCIÓN

En este capítulo se estudia el caso más importante de la construcción de familias de densidades sesgadas, a saber, aquel en que las densidades simétricas  $f$  y  $g$  coinciden ambas con la densidad normal estándar. La exposición ha sido organizada de la siguiente manera. En la Sección 2 se introduce la familia de densidades normales sesgadas y se calcula explícitamente la probabilidad de que una variable con dicha distribución tome valores positivos. La técnica empleada en el cálculo es la de derivación, mediante la cual pueden obtenerse cantidades importantes, como por ejemplo la media de la distribución, tarea que se lleva a cabo en la Sección 3. Posteriormente, en la Sección 4 se analiza la forma en que una densidad normal sesgada puede construirse a partir de una distribución normal multivariada por medio de un argumento condicional, y los resultados obtenidos se utilizan en la Sección 5 para enfatizar el importante papel de la distribución normal sesgada en las aplicaciones. La presentación finaliza en la Sección 6 introduciendo el modelo de localización y escala basado en la familia normal sesgada.

### 3.2 LA DISTRIBUCIÓN NORMAL SESGADA

En esta sección se introduce la distribución normal sesgada y se estudia la probabilidad que, bajo dicha distribución, le corresponde a los intervalos a cada lado del origen. Considere la densidad normal estándar



$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad x \in \mathbb{R},$$

la cual es claramente simétrica respecto al origen, esto es, satisface  $\varphi(x) = \varphi(-x)$ . Aplicando los resultados del capítulo precedente con  $f(x) = g(x) = \varphi(x)$ , se desprende que, para cada número real  $\lambda$ ,

$$\rho(x, \lambda) = 2\varphi(x)\Phi(\lambda x), \quad x \in \mathbb{R} \quad (3.2.1)$$

es una función de densidad, donde

$$\Phi(x) = \int_{-\infty}^x \varphi(y) dy$$

es la función de distribución acumulada de  $\varphi(x)$ , la cual se encuentra ampliamente tabulada. La densidad (3.2.1) es la densidad normal sesgada con parámetro  $\lambda$ . En adelante, la distribución correspondiente a la densidad (3.2.1) se denota mediante  $SN(\lambda)$ , esto es,

$$X \sim SN(\lambda) \iff X \text{ tiene la densidad (3.2.1).}$$

Como se estableció en el Lema 2.2.2, la densidad normal sesgada no es simétrica cuando  $\lambda \neq 0$ , y el objetivo central de esta sección es determinar la diferencia entre las probabilidades de caer en los intervalos  $(-\infty, 0)$  y  $(0, \infty)$  para una variable aleatoria con la densidad (3.2.1). Con este fin, suponga que

$$X \sim SN(\lambda). \quad (3.2.2)$$

**Lema 3.2.1.** *Sea  $X$  una variable aleatoria con la densidad normal sesgada de parámetro  $\lambda$  (vea (3.2.2)), y sea  $p(\lambda)$  la probabilidad de que  $X$  tome un valor positivo, esto es,*

$$p(\lambda) = P[X > 0] = \int_0^{\infty} 2\varphi(x)\Phi(\lambda x) dx. \quad (3.2.3)$$

Con esta notación, se tiene que

$$p(\lambda) = \frac{1}{2} + \frac{1}{\pi} \arctan(\lambda).$$

**Demostración.** Evaluar la integral que define a  $P(\lambda)$  no parece tarea sencilla, de manera que se empleará un argumento indirecto. Primero observe que

$$\frac{dp(\lambda)}{d\lambda} = \int_0^{\infty} 2\varphi(x) \frac{\partial \Phi(\lambda x)}{\partial \lambda} \quad (3.2.4)$$

Recuerde ahora la conocida regla fundamental para derivar una función de distribución: la derivada de la función de distribución acumulada es la densidad correspondiente. Aplicando este hecho conjuntamente con la regla de la cadena se sigue que

$$\frac{\partial \Phi(\lambda x)}{\partial \lambda} = x\varphi(\lambda x)$$

y entonces (3.2.4) equivale a

$$\frac{dp(\lambda)}{d\lambda} = \int_0^{\infty} 2x\varphi(x)\varphi(\lambda x) dx$$

Para evaluar la integral en esta expresión note que

$$2x\varphi(x)\varphi(\lambda x) = 2x \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{\sqrt{2\pi}} e^{-(\lambda x)^2/2} = \frac{x}{\pi} e^{-(1+\lambda^2)x^2/2}$$

de tal forma que

$$\begin{aligned} \frac{dp(\lambda)}{d\lambda} &= \frac{1}{\pi} \int_0^{\infty} x e^{-(1+\lambda^2)x^2/2} dx \\ &= \left. -\frac{1}{\pi} \frac{e^{-(1+\lambda^2)x^2/2}}{1+\lambda^2} \right|_{x=0}^{\infty} \\ &= \frac{1}{\pi} \frac{1}{1+\lambda^2} \end{aligned}$$

Integrando respecto a  $\lambda$  se obtiene

$$p(\lambda) = \frac{1}{\pi} \arctan(\lambda) + C$$

donde  $C$  es una constante. Evaluando en  $\lambda = 0$  y recordando que  $\arctan(0) = 0$  se obtiene  $p(0) = C$ ; como la densidad normal sesgada coincide con la densidad normal estándar cuando  $\lambda = 0$ , se tiene que  $C = p(0) = \frac{1}{2}$ , y entonces

$$p(\lambda) = \frac{1}{\pi} \arctan(\lambda) + \frac{1}{2}$$

que es la conclusión deseada. □

Si  $X \sim 2\varphi(x)\Phi(\lambda x)$ , el lema anterior permite concluir que

$$\begin{aligned} P[X > 0] - P[X < 0] &= P[X > 0] - (1 - P[X > 0]) \\ &= 2P[X > 0] - 1 \\ &= \frac{2}{\pi} \arctan(\lambda); \end{aligned}$$

como  $\arctan(\lambda) > 0$  equivale a  $\lambda > 0$ , se desprende que  $P[X > 0] > P[X < 0]$  si y sólo si  $\lambda$  es positivo.

### 3.3 MEDIA Y VARIANZA

El objetivo de esta sección es determinar la media y la varianza de una densidad normal sesgada, resultado que se establece a continuación.

**Teorema 3.3.1.** *Suponga que  $X \sim 2\varphi(x)\Phi(\lambda x)$ . En este caso,*

$$E[X] = \sqrt{\frac{2}{\pi}} \frac{\lambda}{\sqrt{1 + \lambda^2}} \quad \text{y} \quad \text{Var}[X] = 1 - \frac{2\lambda^2}{\pi(1 + \lambda^2)}.$$

**Demostración.** Sea  $\mu(\lambda)$  la esperanza de la distribución normal sesgada con parámetro  $\lambda$ , esto es,

$$\mu(\lambda) = \int_{\mathbb{R}} x 2\varphi(x)\Phi(\lambda x) dx.$$

Para evaluar esta integral recuerde que  $\frac{\partial \Phi(\lambda x)}{\partial \lambda} = x\varphi(\lambda x)$  y entonces

$$\begin{aligned}
 \mu'(\lambda) &= \frac{d}{d\lambda} \int_{\mathbb{R}} x^2 \varphi(x) \Phi(\lambda x) dx \\
 &= \int_{\mathbb{R}} x^2 \varphi(x) \frac{\partial \Phi(\lambda x)}{\partial \lambda} dx \\
 &= \int_{\mathbb{R}} x^2 \varphi(x) x \varphi(\lambda x) dx \\
 &= \int_{\mathbb{R}} 2x^2 \frac{e^{-x^2/2}}{\sqrt{2\pi}} \frac{e^{-\lambda^2 x^2/2}}{\sqrt{2\pi}} dx \\
 &= \sqrt{\frac{2}{\pi}} \int_{\mathbb{R}} x^2 \frac{e^{-x^2(1+\lambda^2)/2}}{\sqrt{2\pi}} dx.
 \end{aligned}$$

Definiendo

$$\sigma = \frac{1}{\sqrt{1 + \lambda^2}}$$

se tiene que

$$\begin{aligned}
 \mu'(\lambda) &= \sqrt{\frac{2}{\pi}} \int_{\mathbb{R}} x^2 \frac{e^{-x^2/[2\sigma^2]}}{\sqrt{2\pi}} dx \\
 &= \sqrt{\frac{2}{\pi}} \sigma \int_{\mathbb{R}} x^2 \frac{e^{-x^2/[2\sigma^2]}}{\sqrt{2\pi\sigma}} dx
 \end{aligned}$$

y puesto que la última integral es el segundo momento de una densidad normal de media cero y varianza  $\sigma^2$  se desprende que el valor de dicha integral es  $\sigma^2$ , hecho que conduce a

$$\mu'(\lambda) = \sqrt{\frac{2}{\pi}} \sigma^3 = \sqrt{\frac{2}{\pi}} \frac{1}{(1 + \lambda^2)^{3/2}}.$$

Integrando ambos lados se llega a

$$\mu(\lambda) = \sqrt{\frac{2}{\pi}} \frac{\lambda}{(1 + \lambda^2)^{1/2}} + C$$

para alguna constante  $C$ . Evaluando en  $\lambda = 0$  se obtiene  $C = \mu(0) = 0$ , donde la segunda igualdad se debe a que la densidad normal sesgada correspondiente a  $\lambda = 0$  es la densidad normal estándar, de donde se desprende que  $\mu(0) = 0$ . Por lo tanto

$$E[X] = \mu(\lambda) = \sqrt{\frac{2}{\pi}} \frac{\lambda}{(1 + \lambda^2)^{1/2}}.$$

Para concluir, note que el Teorema 2.5.1 del capítulo anterior implica que el segundo momento de  $X$  coincide con el segundo momento de la densidad normal estándar, el cual es 1. Por lo tanto,

$$\text{Var}[X] = E[X^2] - (E[X])^2 = 1 - \frac{2}{\pi} \frac{\lambda^2}{1 + \lambda^2},$$

completando la demostración. □

### 3.4 RELACIÓN CON LA DISTRIBUCIÓN NORMAL DE DIMENSION DOS

En esta sección se establece una relación entre la distribución normal sesgada y la distribución normal clásica en  $\mathbb{R}^2$ . Considere un vector aleatorio  $(X, Y)$  con distribución normal de dimensión dos:

$$(X, Y)' \sim \mathcal{N}_2(\mu, V) \tag{3.4.5}$$

donde  $\mu = (\mu_X, \mu_Y)'$  y  $V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}$  es la matriz de varianza de  $(X, Y)'$ .

**Teorema 3.4.1.** *Sea  $(X, Y)'$  un vector aleatorio con la distribución normal bidimensional como en (3.4.5), y sea*

$$\rho = \frac{V_{12}}{\sqrt{V_{11}V_{22}}}$$

el coeficiente de correlación entre  $X$  y  $Y$ . En este caso, La distribución condicional de la variable  $X$  estandarizada dado que  $Y > \mu_Y$  es normal sesgada con  $\lambda = \rho / \sqrt{1 - \rho^2}$ . Más explícitamente,

$$\frac{X - \mu_x}{\sqrt{V_{11}}} \Big|_{Y > \mu_Y} \sim SN \left( \frac{\rho}{\sqrt{1 - \rho^2}} \right)$$

**Demostración.** Defina las variables

$$U_1 = \frac{X - \mu_X}{\sqrt{V_{11}}}, \quad W = \frac{Y - \mu_Y}{\sqrt{V_{22}}} \quad (3.4.6)$$

de manera que

- (a)  $(U_1, W)$  tiene la densidad normal bidimensional con media  $(0, 0)'$ , y más aún, ambas variables tienen varianza uno y correlación  $\rho$ .

Considere ahora una variable aleatoria  $Z$  con distribución normal estándar que sea independiente de  $U_1$ , y defina

$$U_2 = \sqrt{1 - \rho^2}Z + \rho U_1.$$

Observe ahora que, como  $U_1$  y  $Z$  son independientes con distribución normal estándar y el vector  $(U_1, U_2)'$  es una transformación lineal de  $(U_1, Z)'$ , se tiene que

- (i)  $(U_1, U_2)'$  tiene distribución normal bidimensional con media  $(0, 0)'$ ;  
(ii) Como  $Z$  y  $U_1$  son independientes de varianza 1,

$$\text{Var}[U_2] = \text{Var}[\sqrt{1 - \rho^2}Z + \rho U_1] = (1 - \rho^2)\text{Var}[Z] + \rho^2\text{Var}[U_1] = 1;$$

además,

$$\begin{aligned} \text{Cov}[U_1, U_2] &= \text{Cov}[U_1, \sqrt{1 - \rho^2}Z + \rho U_1] \\ &= \sqrt{1 - \rho^2}\text{Cov}[U_1, Z] + \rho\text{Cov}[U_1, U_1] \\ &= \sqrt{1 - \rho^2} \cdot 0 + \rho \cdot 1 \\ &= \rho. \end{aligned}$$

A partir de estas propiedades (i) y (ii) se desprende,

- (b)  $(U_1, U_2)$  tiene la densidad normal bidimensional con media  $(0, 0)'$ , y más aún, ambas variables tienen varianza uno y correlación  $\rho$ .

Comparando (a) y (b) se concluye que

$(U_1, W)$  y  $(U_1, U_2)$  tienen la misma distribución.

Por lo tanto,

*La distribución condicional de  $U_1$  dado  $W > 0$  es la misma que la distribución condicional de  $U_1$  dado  $U_2 > 0$*  (3.4.7)

Note ahora que

$$\begin{aligned} U_1|_{U_2>0} &= U_1|_{\sqrt{1-\rho^2}Z+\rho U_1>0} \\ &= U_1|_{Z > -\frac{\rho}{\sqrt{1-\rho^2}}U_1} \end{aligned}$$

Como  $Z$  y  $U_1$  son independientes con la distribución normal estándar, una aplicación del Lema 2.3.1 en el capítulo precedente implica que

$$U_1|_{U_2>0} \sim SN \left( \frac{\rho}{\sqrt{1-\rho^2}} \right),$$

propiedad que combinada con (3.4.7) implica que

$$U_1|_{W>0} \sim SN \left( \frac{\rho}{\sqrt{1-\rho^2}} \right).$$

Para finalizar el argumento, observe que, via (3.4.6), la relación precedente equivale a

$$\frac{X - \mu_X}{\sqrt{V_{11}}} \Big|_{Y > \mu_Y} \sim SN \left( \frac{\rho}{\sqrt{1-\rho^2}} \right),$$

que es la conclusión deseada.

### 3.5 IMPORTANCIA DE LA DISTRIBUCIÓN NORMAL SESGADA

En esta sección se utiliza el Teorema 3.4.1 para poner de relieve la forma en que la distribución normal sesgada puede surgir en las aplicaciones. Considere dos características aleatorias  $X$  y  $Y$  asociadas con un mismo objeto, y suponga que  $Y$  se observa primero y  $X$  cierto tiempo después. Los objetos se someten a un proceso de selección, en donde la característica  $Y$  se utiliza para decidir si los objetos son aceptados o no. Suponga que solamente se aceptan los objetos que tienen un desempeño relevante medido por medio de  $Y$ , en el sentido de que un objeto es admitido sólo si el valor de  $Y$  supera al promedio, esto es,

$$\text{un objeto es aceptado si y sólo si } Y > \mu_Y$$

Dentro de la población de objetos aceptados se observa  $X$ , y si suponemos — como ocurre con frecuencia en las aplicaciones— que  $(X, Y)$  tiene distribución normal bidimensional, entonces la distribución de  $X$  dentro de dicha población es  $X|_{Y > \mu_Y}$  la cual, por el Teorema 3.4.1, es normal sesgada. Puesto que un proceso de selección como el que aquí se describe ocurre frecuentemente, se desprende que la distribución normal sesgada es un modelo susceptible de utilizarse de manera razonable en aplicaciones reales.

### 3.6 EL MODELO DE LOCALIZACIÓN Y ESCALA

Considere una variable aleatoria  $X$  con distribución  $SN(\lambda)$ . Al someter  $X$  a un cambio de localización y de escala se obtiene una nueva variable

$$W = \xi + \sigma X, \quad \xi \in \mathbb{R}, \quad \sigma > 0.$$

La densidad de  $W$  se obtiene a partir de la fórmula del cambio de variable y está dada por

$$f_W(w) = \frac{2}{\sigma} \varphi \left( \frac{w - \xi}{\sigma} \right) \Phi \left( \lambda \frac{w - \xi}{\sigma} \right)$$

En el siguiente capítulo se estudia el problema de estimación del parámetro  $\theta = (\xi, \sigma, \lambda)$  por medio del método de verosimilitud máxima y se estudia la consistencia de los estimadores.



# CAPÍTULO 4

## CONSISTENCIA DE ESTIMADORES DE VEROSIMILITUD MÁXIMA

### 4.1 INTRODUCCIÓN

Este trabajo trata con el problema de inferencia basada en la verosimilitud para la familia de densidades normal sesgada univariada, que de aquí en adelante será denotada como  $\mathcal{SN} = \{\rho(\cdot; \theta) \mid \theta \in \Theta\}$ , la cual está especificada como sigue:

Para cada  $\theta = (\sigma, \xi, \lambda) \in \Theta := (0, \infty) \times \mathbb{R}^2$ ,

$$\rho(x; \theta) := \frac{2}{\sigma} \varphi\left(\frac{x - \xi}{\sigma}\right) \Phi\left[\lambda \left(\frac{x - \xi}{\sigma}\right)\right], \quad x \in \mathbb{R}, \quad (4.1.1)$$

donde  $\varphi(x) = (2\pi)^{-\frac{1}{2}} e^{-\frac{x^2}{2}}$  es la densidad normal estándar y  $\Phi(\cdot)$  denota la correspondiente función de distribución acumulada. Cuando  $\lambda = 0$ ,  $\rho(\cdot; \theta)$  se reduce a la densidad normal con media  $\xi$  y desviación estándar  $\sigma$ , en tanto que si  $\lambda \neq 0$  entonces  $\lambda$  controla la asimetría de la distribución en torno a  $\xi$ . La familia normal sesgada fué introducida en Azzalini (1985, 1986) y durante los últimos veinte años se han estudiado intensamente, tanto modelos de regresión basados en la  $\mathcal{SN}$ , como extensiones multivariadas; vea, por ejemplo, Azzalini y Dalla Valle (1996), Azzalini y Capitanio (1999), Genton (2004), así como las referencias en esos artículos. Desde los primeros trabajos sobre el modelo  $\mathcal{SN}$ , el problema de inferencia basada en la verosimilitud ha mostrado ser el mas desafiante, debido

a la singularidad de la matriz de información de Fisher cuando  $\lambda = 0$  siendo esto la principal fuente de dificultades. Tal problema también se presenta en el análisis de la familia exponencial sesgada, la cual incluye al modelo normal sesgado con localización y escala como un caso particular. (Azzalini, 1986, DiCiccio and Monti, 2004). La singularidad de la matriz de información ha sido manejada prácticamente vía la parametrización centrada introducida en Azzalini (1985), mientras que los resultados de inferencia asintótica están basados en el reciente trabajo de Rotnitzky *et. al.* (2000). Usando las conclusiones en este último artículo y dada una secuencia consistente de estimadores de verosimilitud máxima de  $\theta$ , Chiogna (2005) obtuvo tasas de convergencia, y Sartori (2006) estudió condiciones para garantizar que el estimador de  $\lambda$  sea finito.

*El problema considerado en este trabajo es muy básico, a saber,*

- i) establecer la *existencia* de una sucesión de estimadores de verosimilitud máxima de  $\theta$ , y demostrar su *consistencia*.

Para ver *la importancia e interés* detrás de este problema, observe que los resultados generales que aseguran la existencia de una secuencia consistente de estimadores de verosimilitud máxima, se han establecido suponiendo que alguna una de las siguientes condiciones son satisfechas en el modelo estadístico subyacente:

- (i) El espacio de parámetros es compacto, o
- (ii) Con probabilidad 1, la función de verosimilitud (asociada a una muestra finita) es estrictamente cóncava; vea, por ejemplo, Lehmann y Casella (1998), Shao (1999) o, especialmente, la muy completa investigación de Newey and McFadden (1993).

Es sabido, que para el modelo normal sesgado, la función de verosimilitud  $L_n(\cdot; \mathbf{x})$  asociada a una muestra  $\mathbf{x} = (x_1, \dots, x_n)$  tiene siempre un punto crítico en  $\theta_{\mathbf{x}} = (s, \bar{x}, 0)$ , donde  $\bar{x}$  y  $s$  son, respectivamente, la media y la desviación estándar de  $\mathbf{x}$ . Sin embargo, no es difícil ver que  $\theta_{\mathbf{x}}$  no es un maximizador de  $L_n(\cdot; \mathbf{x})$  si el tercer momento central de  $\mathbf{x}$  es no nulo, de manera que, con probabilidad 1, la función de verosimilitud *no* es cóncava. Dado que el espacio de parámetros de la familia  $SN$  no es compacto, hasta donde se sabe, la existencia y consistencia de una secuencia de estimadores de verosimilitud máxima del parámetro  $\theta$  no puede ser obtenida directamente de resultados generales disponibles, y la solución del

problema básico arriba expuesto llenará un importante hueco que actualmente se presenta en la literatura del modelo normal sesgado; de hecho, una solución establecerá bases firmes en el estudio del comportamiento asintótico de los estimadores de verosimilitud máxima.

*El enfoque* usado para estudiar la existencia de una secuencia de estimadores de verosimilitud máxima está basado en una reparametrización *en línea* de la función de verosimilitud que, además de simplificar el análisis, permite obtener cotas *a priori* para posibles maximizadores.

*La organización* del material subsecuente es como sigue: En la Sección 2 se introduce la reparametrización en línea de la función de verosimilitud máxima y el resultado principal del trabajo, que establece la existencia de una secuencia de estimadores de verosimilitud máxima, así como, cotas *a priori* para los estimadores se enuncia en el Teorema 4.2.1, que se demuestra en la Sección 5 después de los preliminares técnicos necesarios presentados en las Secciones 3 y 4. Finalmente, la exposición concluye en la Sección 6 con una demostración de la consistencia del método de verosimilitud máxima.

**Notación.** En el resto de la presentación,  $X_1, X_2, X_3, \dots$ , denota una secuencia de variables aleatorias independientes e idénticamente distribuidas con densidad común perteneciente a la familia  $\mathcal{SN}$ , y la distribución de la sucesión  $X_1, X_2, \dots$  cuando  $\theta \in \Theta$  es el valor real del parámetro se denota por  $P_\theta[\cdot]$ , mientras que  $E_\theta[\cdot]$  representa el correspondiente operador de esperanza. Para cada entero positivo  $n$

$$\text{sean } \bar{X}_n := \frac{\sum_{i=1}^n X_i}{n}, S_n := \left[ \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{n} \right]^{\frac{1}{2}}, \text{ y}$$

$$X_1^n := (X_1, \dots, X_n); \quad (4.1.2)$$

similarmente, para cada  $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ ,

$$\bar{\mathbf{x}} := \frac{\sum_{i=1}^n x_i}{n} \quad \text{y} \quad s(\mathbf{x}) := \left[ \frac{\sum_{i=1}^n (x_i - \bar{\mathbf{x}})^2}{n} \right]^{\frac{1}{2}}$$

Finalmente, si  $Q$  es un conjunto finito,  $\#Q$  denota el número de elementos de  $Q$  mientras que, para un evento  $A$ ,  $I[A]$  es la variable indicadora (de Bernoulli)

correspondiente a  $A$ .

## 4.2 EXISTENCIA DE ESTIMADORES DE VEROSIMILITUD MÁXIMA

Esta sección trata sobre la existencia de una secuencia de estimadores de verosimilitud máxima, una idea que es formalmente introducida a continuación. Primero, para la densidad  $\rho(\cdot; \theta)$  en (4.1.1), sea  $\ell(\cdot; x)$  la función de verosimilitud (logarítmica) correspondiente a una sola observación  $x \in \mathbb{R}$ , donde, como es usual, constantes multiplicativas en una densidad son ignoradas:

$$\begin{aligned} \ell(\sigma, \xi, \lambda; x) = & -\log(\sigma) - \frac{(x - \xi)^2}{2\sigma^2} \\ & + \log \left[ \Phi \left( \lambda \left( \frac{x - \xi}{\sigma} \right) \right) \right], \quad \theta = (\sigma, \xi, \lambda) \in \Theta. \end{aligned} \quad (4.2.1)$$

Dado un tamaño de muestra fijo  $n \geq 1$ , para cada muestra  $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$  defina la función de verosimilitud promedio  $L_n(\cdot; \mathbf{x})$  mediante

$$L_n(\theta; \mathbf{x}) := \frac{1}{n} \sum_{k=1}^n \ell(\theta; x_k), \quad \theta \in \Theta, \quad (4.2.2)$$

y note que por medio de cálculos directos que usan (4.2.1) se obtiene que, para cada  $\mathbf{x} \in \mathbb{R}^n$ ,

$$\begin{aligned} L_n(\theta; \mathbf{x}) = & -\log(\sigma) - \frac{1}{2} \left( \frac{s(\mathbf{x})}{\sigma} \right)^2 - \frac{1}{2} \left( \frac{\bar{\mathbf{x}} - \xi}{\sigma} \right)^2 \\ & + \frac{1}{n} \sum_{i=1}^n H \left( \lambda \left( \frac{x_i - \xi}{\sigma} \right) \right), \quad \theta = (\sigma, \xi, \lambda) \in \Theta, \end{aligned} \quad (4.2.3)$$

donde

$$H(x) := \log(\Phi(x)), \quad x \in \mathbb{R}. \quad (4.2.4)$$

En este punto, es conveniente observar que  $H(\cdot) < 0$ , mientras que

$$H'(x) = \frac{\varphi(x)}{\Phi(x)} > 0 \quad \text{y} \quad H''(x) = \frac{-x\varphi(x)}{\Phi(x)} - \left( \frac{\varphi(x)}{\Phi(x)} \right)^2 < 0$$

para cada  $x \in \mathbb{R}$ , así que  $H(\cdot)$  es estrictamente creciente y también estrictamente cóncava en  $\mathbb{R}$ .

**Definición 4.2.1.** Sea  $\{\theta_n = (\sigma_n, \xi_n, \lambda_n) : \mathbb{R}^n \rightarrow \Theta\}$  una secuencia de funciones (Borel) medibles y defina

$$(\hat{\sigma}_n, \hat{\xi}_n, \hat{\lambda}_n) \equiv \hat{\theta}_n := \theta_n(X_1^n) = (\sigma_n g(X_1^n), \xi_n(X_1^n), \lambda_n(X_1^n)).$$

En este caso,  $\{\hat{\theta}_n\}$  es una sucesión de estimadores de verosimilitud máxima de  $\theta$  si

$$P_{\theta_0} \left[ \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} \left[ L_k(\hat{\theta}_k; X_1^k) \geq L_k(\theta; X_1^k) \text{ para toda } \theta \in \Theta \right] \right] = 1, \quad \theta_0 \in \Theta. \quad (4.2.5)$$

**Observación 4.2.1.** La continuidad del mapeo  $(\theta, \mathbf{x}) \mapsto L_n(\theta; \mathbf{x})$  implica que

$$\left[ L_n(\hat{\theta}_n; X_1^n) \geq L_n(\theta; X_1^n), \theta \in \Theta \right]$$

es un evento, ya que  $L_n(\hat{\theta}_n; X_1^n) \geq L_n(\theta; X_1^n)$  ocurre para cada  $\theta \in \Theta$  si y sólo si se cumple para cada  $\theta$  con componentes racionales. Por otro lado, el significado (4.2.5) es que, con probabilidad 1 y sin importar el valor real del parámetro,  $\hat{\theta}_n$  maximiza la función de verosimilitud promedio observada  $L_n(\cdot; X_1^n)$  siempre que  $n$  sea lo suficientemente grande. El evento entre corchetes en (4.2.5) es el límite inferior de la familia  $\left\{ \left[ L_k(\hat{\theta}_k; X_1^k) \geq L_k(\theta; X_1^k), \theta \in \Theta \right] \mid k = 1, 2, 3, \dots \right\}$  y, como se ha demostrado en Billingsley (1995), o en Shao (1999),  $P_{\theta_0} \left[ L_k(\hat{\theta}_k; X_1^k) \geq L_k(\theta; X_1^k) \text{ para toda } \theta \in \Theta \right] \rightarrow 1$  cuando ocurre. (4.2.5)

El resultado básico del presente trabajo es el siguiente.

**Teorema 4.2.1.**

- (i) Existe una sucesión  $\{\hat{\theta}_n\}$  de estimadores de verosimilitud máxima  $\theta$ . Mas aún,
- (ii) En el evento  $\left[ L_n(\hat{\theta}_n; X_1^n) \geq L_n(\theta; X_1^n) \text{ para toda } \theta \in \Theta \right] \cap [S_n > 0]$ , las componentes de  $\hat{\theta} = (\hat{\sigma}_n, \hat{\xi}_n, \hat{\lambda}_n)$  satisfacen las siguientes relaciones:

$$\frac{1}{2} \leq \frac{S_n}{\hat{\sigma}_n}, \quad \left( \frac{S_n}{\hat{\sigma}_n} \right)^2 + \left( \frac{\bar{X}_n - \hat{\xi}_n}{\hat{\sigma}_n} \right)^2 = 1, \quad (4.2.6)$$

y, si  $a > \sqrt{3}$ , entonces

$$\frac{2}{a - \sqrt{3}} \Phi^{-1} \left( \left( \frac{1}{2} \right)^{\frac{1}{\beta_n^+(a)}} \right) \leq \hat{\lambda}_n \leq \frac{2}{\sqrt{3} - a} \Phi^{-1} \left( \left( \frac{1}{2} \right)^{\frac{1}{\beta_n^-(a)}} \right), \quad (4.2.7)$$

donde para  $a > 0$  las variables aleatorias  $\beta_n^+(a)$  y  $\beta_n^-(a)$  están dadas por

$$\begin{aligned} \beta_n^+(a) &:= \frac{1}{n} \sum_{i=1}^n I[X_i > \bar{X}_n + aS_n], \\ \beta_n^-(a) &:= \frac{1}{n} \sum_{i=1}^n I[X_i < \bar{X}_n - aS_n]. \end{aligned} \quad (4.2.8)$$

La demostración del Teorema 4.2.1 depende fuertemente de las propiedades de una función auxiliar  $\tilde{L}_n$  introducida abajo, cuya definición está motivada por la siguiente observación: Sean  $\theta \in \Theta$  y  $\mathbf{x} \in \mathbb{R}^n$  con  $s(\mathbf{x}) > 0$  vectores arbitrarios pero fijos, y sea  $\mathbf{z} \in \mathbb{R}^n$  el vector normalizado correspondiente a  $\mathbf{x}$ , *i.e.*,

$$z_i = \frac{x_i - \bar{x}}{s(\mathbf{x})}, \quad i = 1, 2, \dots, n; \quad (4.2.9)$$

ya que  $\frac{(x_i - \xi)}{\sigma} = z_i \left( \frac{s(\mathbf{x})}{\sigma} \right) + \frac{(\bar{x} - \xi)}{\sigma}$ , de (4.2.3) se sigue que

$$\begin{aligned} L_n(\theta; \mathbf{x}) + \log(s(\mathbf{x})) &= \log \left( \frac{s(\mathbf{x})}{\sigma} \right) - \frac{1}{2} \left( \frac{s(\mathbf{x})}{\sigma} \right)^2 - \frac{1}{2} \left( \frac{\bar{x} - \xi}{\sigma} \right)^2 \\ &\quad + \frac{1}{n} \sum_{i=1}^n H \left( \lambda \left( z_i \frac{s(\mathbf{x})}{\sigma} + \frac{\bar{x} - \xi}{\sigma} \right) \right), \end{aligned} \quad (4.2.10)$$

y entonces, dado  $\mathbf{x}$ ,  $L_n(\theta; \mathbf{x}) + \log(s(\mathbf{x}))$  es una función de  $y = \frac{s(\mathbf{x})}{\sigma}$ ,  $\omega = \frac{(\bar{x} - \xi)}{\sigma}$ , y  $\lambda$ . Este hecho motiva la siguiente definición.

**Definición 4.2.2.** Sea  $n$  un entero arbitrario pero fijo.

(i)

i) Para cada  $\mathbf{z} = (z_1, \dots, z_n) \in \mathbb{R}^n$ , la función  $\tilde{L}_n(\cdot; \mathbf{z}) : \Theta \rightarrow \mathbb{R}$  está definida por

$$\begin{aligned} \tilde{L}_n(y, \omega, \lambda; \mathbf{z}) &= \log(y) - \frac{y^2 + \omega^2}{2} \\ &\quad + \frac{1}{n} \sum_{i=1}^n H \left( \lambda [z_i y + \omega] \right), \end{aligned} \quad (4.2.11)$$

para  $y > 0$ ,  $\omega, \lambda \in \mathbb{R}$

mientras que

$$m_n(\mathbf{z}) = \sup_{\theta \in \Theta} \tilde{L}_n(\theta; \mathbf{z}).$$

ii) Dado  $\mathbf{z} \in \mathbb{R}^n$ ,  $\mathcal{M}_n(\mathbf{z})$  es el conjunto de maximizadores de  $\tilde{L}_n(\cdot; \mathbf{z})$ , mientras que  $\mathcal{I}(\mathbf{z})$  consiste de la tercera coordenada de los puntos en  $\mathcal{M}_n(\mathbf{z})$ . Más explícitamente,

$$\mathcal{M}_n(\mathbf{z}) := \left\{ \theta \in \Theta \mid \tilde{L}_n(\theta; \mathbf{z}) = m_n(\mathbf{z}) \right\}$$

y

$$\mathcal{I}_n(\mathbf{z}) = \pi\left(\mathcal{M}_n(\mathbf{z})\right),$$

donde  $\pi : \mathbb{R}^3 \rightarrow \mathbb{R}$  es la proyección definida por  $\pi(x_1, x_2, x_3)' = x_3$ .

A partir de (4.2.10) y (4.2.11) se sigue que maximizar  $L_n(\cdot; \mathbf{x})$  y  $\tilde{L}_n(\cdot; \mathbf{z})$  son problemas equivalentes, un hecho que es formalmente establecido enseguida para referencia futura. Dado  $\mathbf{x} \in \mathbb{R}^n$  defina  $T_{\mathbf{x}} : \Theta \rightarrow \Theta$  como

$$T_{\mathbf{x}}(\theta) = \left( \frac{s(\mathbf{x})}{\sigma}, \frac{\bar{\mathbf{x}} - \xi}{\sigma}, \lambda \right), \quad \theta = (\sigma, \xi, \lambda) \in \Theta. \quad (4.2.12)$$

**Lema 4.2.1.** *Sea  $\mathbf{x} \in \mathbb{R}^n$  tal que la desviación estándar correspondiente  $s(\mathbf{x})$  es positiva. En este caso, las siguientes propiedades (i)–(iii) se satisfacen:*

- (i) El mapeo  $T_{\mathbf{x}} : \Theta \rightarrow \Theta$  es biyectivo;
- (ii) Para cada  $\theta \in \Theta$ ,  $L_n(\theta; \mathbf{x}) + \log(s(\mathbf{x})) = \tilde{L}_n(T_{\mathbf{x}}(\theta); \mathbf{z})$ , donde  $\mathbf{z}$  es el vector normalizado en (4.2.9);
- (iii)  $\tilde{\theta}^*$  es un maximizador de  $\tilde{L}_n(\cdot; \mathbf{z})$  si y sólo si  $\theta^* = T_{\mathbf{x}}^{-1}(\tilde{\theta}^*)$  maximiza  $L_n(\cdot; \mathbf{x})$ .

**Demostración.** (i) Puesto que  $s(\mathbf{x}) > 0$ , la parte (i) se sigue de (4.2.12); combinando esta última igualdad con (4.2.10) y con (4.2.11) se obtiene la afirmación (ii) y, usando el hecho de que  $T_{\mathbf{x}}(\cdot)$  es una biyección, se deduce (iii).  $\square$

Por este lema, el analizar  $\tilde{L}_n(\cdot; \mathbf{z})$  en vez de  $L_n(\cdot; \mathbf{x})$  es equivalente a introducir el nuevo parámetro

$$\tilde{\theta} = T_{\mathbf{x}}(\theta);$$

note, sin embargo, que esta es una reparametrización en línea de  $L_n(\cdot; \mathbf{x})$ , en el sentido de que la transformación anterior depende de los datos observados  $\mathbf{x}$ , en

contraste con las reparametrizaciones usuales del modelo, que no involucran a los datos. La estrategia usada para establecer el Teorema 4.2.1 puede describirse brevemente como: La maximización y las propiedades de los optimizadores se estudiarán para  $\tilde{L}_n$ , y el Teorema 4.2.1 se obtendrá mediante el Lema 4.2.1.

Aunque  $\tilde{L}_n(\cdot; \mathbf{z})$  se ve más simple que  $L_n(\cdot; \mathbf{x})$ , el análisis subsecuente para  $\tilde{L}_n$  es algo técnico, y es conveniente dar un esquema de las dos principales fases del argumento. Primero, dada una muestra normalizada  $\mathbf{z}$  de tamaño  $n$ , en la siguiente sección se introduce un conjunto  $\mathcal{P}_n(\mathbf{z})$  que contiene todos los posibles maximizadores de  $\tilde{L}_n(\cdot; \mathbf{z})$ , y se establecen cotas para los miembros de  $\mathcal{P}_n(\mathbf{z})$ . El segundo paso del argumento se presenta en la Sección 4, donde se identifica una clase  $\mathcal{S}_n$  de muestras normalizadas tal que  $\mathcal{P}_n(\mathbf{z})$  es compacto cuando  $\mathbf{z} \in \mathcal{S}_n$ , propiedad que conduce a la existencia de maximizadores de  $\tilde{L}_n(\cdot; \mathbf{z})$  así como a una función medible  $\theta_n$  tal que  $\theta_n(\mathbf{z})$  maximiza a  $\tilde{L}_n(\cdot; \mathbf{z})$  para cada  $\mathbf{z} \in \mathcal{S}_n$ . La sección concluye demostrando que la muestra observada normalizada está contenida en  $\mathcal{S}_n$  cuando el tamaño de muestra  $n$  es suficientemente grande. Después de estos preliminares, en la Sección 5 se da finalmente una demostración del Teorema 4.2.1.

### 4.3 COTAS PARA POSIBLES MAXIMIZADORES

Para comenzar con el análisis del problema de maximización de la función  $\tilde{L}_n$ , en esta sección se introduce un conjunto que contiene todos los maximizadores de  $\tilde{L}_n(\cdot; \mathbf{z})$ , y se establecen cotas para sus miembros. Para cada  $n = 1, 2, 3, \dots$  y  $\mathbf{z} \in \mathbb{R}^n$ , defina

$$\mathcal{P}_n(\mathbf{z}) := \left\{ (y, \omega, \lambda) \in \Theta \mid y^2 + \omega^2 = 1, \tilde{L}_n(y, \omega, \lambda; \mathbf{z}) \geq \tilde{L}_n(1, 0, 0; \mathbf{z}) \right\}, \quad (4.3.1)$$

el cual es un subconjunto cerrado de  $\Theta$ , por la continuidad de  $\tilde{L}_n(\cdot; \mathbf{z})$ .

**Lema 4.3.1.** *Sean el entero positivo  $n$  y  $\mathbf{z} \in \mathbb{R}^n$  arbitrarios.*

(i) *Dado  $\theta = (y, \omega, \lambda) \in \Theta$ , sea*

$$\theta_1 = (y_1, \omega_1, \lambda_1) := \left( \frac{y}{\sqrt{y^2 + \omega^2}}, \frac{\omega}{\sqrt{y^2 + \omega^2}}, \lambda \sqrt{y^2 + \omega^2} \right).$$

*Con esta notación,*

$$\tilde{L}_n(\theta; \mathbf{z}) \leq \tilde{L}_n(\theta_1; \mathbf{z}), \quad (4.3.2)$$



y la desigualdad es estricta si  $\theta \neq \theta_1$ , i.e., si  $y^2 + \omega^2 \neq 1$ .

En consecuencia,

(ii)  $\mathcal{M}_n(\mathbf{z}) \subset \mathcal{P}_n(\mathbf{z})$  para toda  $\mathbf{z} \in \mathbb{R}^n$ .

### Demostración.

(i) Considere la función  $g : (0, \infty) \rightarrow \mathbb{R}$  dada por

$$g(t) = \tilde{L}_n\left(ty, t\omega, \frac{\lambda}{t}; \mathbf{z}\right) = -t^2 \frac{y^2 + \omega^2}{2} + \log(ty) + \frac{1}{n} \sum_{i=1}^n H\left(\lambda[z_i y + \omega]\right),$$

y note que

$$g'(t) = -t(y^2 + \omega^2) + \frac{1}{t} \quad \text{y} \quad g''(t) = -(y^2 + \omega^2) - \frac{1}{t^2} < 0.$$

Entonces,  $g(\cdot)$  es estrictamente cóncava y alcanza su máximo en el único punto  $t_1 = (y^2 + \omega^2)^{-\frac{1}{2}}$ , así que  $\tilde{L}_n(\theta; \mathbf{z}) = g(1) \leq g(t_1) = L_n(\theta_1; \mathbf{z})$ , donde la desigualdad es estricta si  $t_1 \neq 1$ .

(ii) Si  $\theta = (y, \omega, \lambda) \in \mathcal{M}_n(\mathbf{z})$ , entonces  $\tilde{L}_n(\theta; \mathbf{z}) \geq \tilde{L}_n(1, 0, 0; \mathbf{z})$ , y  $\theta$  maximiza la función  $\tilde{L}_n(\cdot; \mathbf{z})$ , de modo que  $y^2 + \omega^2 = 1$ , por la parte (i). Entonces,  $(y, \omega, \lambda) \in \mathcal{P}_n(\mathbf{z})$ .  $\square$

En seguida, se establecerán cotas para los elementos del conjunto  $\mathcal{P}_n(\mathbf{z})$  de posibles maximizadores. Para un subconjunto  $Q$  de  $\{1, 2, \dots, n\}$  y  $\mathbf{z} \in \mathbb{R}^n$ , escriba

$$f_n(Q) := \frac{\#Q}{n}, \quad \text{y} \quad \bar{\mathbf{z}}_Q := \frac{1}{\#Q} \sum_{i \in Q} z_i \quad \text{si } Q \text{ es no vacío o.} \quad (4.3.3)$$

**Lema 4.3.2.** Para un entero positivo  $n$  y  $\mathbf{z} \in \mathbb{R}^n$ , las siguientes afirmaciones (i) y (ii) se cumplen para cada  $(y, \omega, \lambda) \in \mathcal{P}_n(\mathbf{z})$ :

(i)  $y \geq \frac{1}{2}$  y  $\lambda[\bar{\mathbf{z}}_Q y + \omega] \geq \Phi^{-1}\left(\left(\frac{1}{2}\right)^{\frac{1}{f_n(Q)}}\right)$  si  $\emptyset \neq Q \subset \{1, 2, \dots, n\}$ ;

(ii) Para cada  $a > \sqrt{3}$

$$\frac{2}{a - \sqrt{3}} \Phi^{-1} \left( \left( \frac{1}{2} \right)^{\frac{1}{f_n(Q_a^+)}} \right) \leq \lambda \leq \frac{2}{\text{sqrt}3 - a} \Phi^{-1} \left( \left( \frac{1}{2} \right)^{\frac{1}{f_n(Q_a^-)}} \right), \quad (4.3.4)$$

donde

$$Q_a^- := \{i \mid z_i \leq -a, 1 \leq i \leq n\}, \quad Q_a^+ := \{i \mid z_i \geq a, 1 \leq i \leq n\}, \quad (4.3.5)$$

y se adopta la siguiente convención:  $\Phi^{-1} \left( \left( \frac{1}{2} \right)^{\frac{1}{0}} \right) := -\infty$ .

### **Demostración.**

(i) A partir de (4.2.11) y (4.3.1) se obtiene

$$\begin{aligned} (y, \omega, \lambda) \in \mathcal{P}_n(\mathbf{z}) &\Rightarrow \tilde{L}_n(y, \omega, \lambda; \mathbf{z}) \geq \tilde{L}_n(1, 0, 0; \mathbf{z}) \quad \text{y} \quad y^2 + \omega^2 = 1 \\ &\Rightarrow \frac{1}{2} + \log(y) + \frac{1}{n} \sum_{i=1}^n H(\lambda[z_i y + \omega]) \geq \frac{1}{2} + H(0) \end{aligned} \quad (4.3.6)$$

$$\begin{aligned} &\Rightarrow \log(y) + \frac{1}{n} \sum_{i=1}^n H(\lambda[z_i y + \omega]) \geq H(0) \\ &= \log \left( \frac{1}{2} \right). \end{aligned} \quad (4.3.7)$$

Por otra parte, usando el que  $H(\cdot)$  es no positiva y cóncava, se sigue que para cada conjunto no vacío  $Q \subset \{1, 2, \dots, n\}$ ,

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n H(\lambda[z_i y + \omega]) &\leq \frac{1}{n} \sum_{i \in Q} H(\lambda[z_i y + \omega]) \\
&= \frac{\#Q}{n} \frac{1}{\#Q} \sum_{i \in Q} H(\lambda[z_i y + \omega]) \\
&\leq \frac{\#Q}{n} H\left(\lambda \left[ y \frac{1}{\#Q} \sum_{i \in Q} z_i + \omega \right]\right) \\
&= f_n(Q) H(\lambda [\bar{z}_Q y + \omega]) \\
&= f_n(Q) \log \left( \Phi(\lambda [\bar{z}_Q y + \omega]) \right) \\
&= \log \left( \Phi(\lambda [\bar{z}_Q y + \omega])^{f_n(Q)} \right).
\end{aligned}$$

Combinando esta relación con (4.3.7) se sigue que para cada  $(y, \omega, \lambda) \in \mathcal{P}_n(\mathbf{z})$ ,

$$\begin{aligned}
\log \left( \frac{1}{2} \right) &\leq \log(y) + \log \left( \Phi(\lambda [\bar{z}_Q y + \omega])^{f_n(Q)} \right) \\
&= \log \left( y \Phi(\lambda [\bar{z}_Q y + \omega])^{f_n(Q)} \right),
\end{aligned}$$

*i.e.*,

$$\frac{1}{2} \leq y \Phi(\lambda [\bar{z}_Q y + \omega])^{f_n(Q)},$$

y usando el que  $\Phi(\cdot) < 1$  y que  $y \leq 1$  cuando  $(y, \omega, \lambda) \in \mathcal{P}_n(\mathbf{z})$ , la parte (i) se desprende de esta última relación.

(ii) Sean  $a > \sqrt{3}$  y  $(y, \omega, \lambda) \in \mathcal{P}_n(\mathbf{z})$  arbitrarios pero fijos. En este caso  $y^2 + \omega^2 = 1$  y la primera desigualdad en la parte (i) conducen a

$$\frac{1}{2} \leq y \leq 1, \quad -\frac{\sqrt{3}}{2} \leq \omega \leq \frac{\sqrt{3}}{2}. \tag{4.3.8}$$

Considere ahora los siguientes casos exhaustivos:

**Caso 1:**  $\lambda \geq 0$ . En este caso sólo la cota superior en (4.3.4) requiere de una demostración. Suponga que  $Q_a^- \neq \emptyset$ , así que  $\bar{z}_{Q_a^-} \leq -a$  (vea (4.3.3) y (4.3.5)), y note que la parte (i) y (4.3.8) juntas conducen a

$$\lambda \frac{(-a + \sqrt{3})}{2} \geq \lambda[-ay + \omega] \geq \lambda[\bar{z}_{Q_a^-}y + \omega] \geq \Phi^{-1} \left( \left( \frac{1}{2} \right)^{1/f_n(Q_a^-)} \right)$$

y ya que  $\sqrt{3} - a < 0$ ,

$$\lambda \leq \frac{2}{\sqrt{3} - a} \Phi^{-1} \left( \left( \frac{1}{2} \right)^{1/f_n(Q_a^-)} \right),$$

desigualdad que también se cumple cuando  $Q_a^-$  es vacío, dado que en este caso el lado derecho es  $+\infty$ .

**Caso 2:**  $\lambda < 0$ . en este contexto la acotación superior para  $\lambda$  se cumple con certeza, mientras que el siguiente argumento, paralelo al usado en el caso previo, establece la cota inferior: suponga que  $Q_a^+ \neq \emptyset$ , y note que (4.3.3) y (4.3.5) implican que  $\bar{z}_{Q_a^+} \geq a$ , de tal manera que, mediante (4.3.8),  $\bar{z}_{Q_a^+}y + \omega \geq \frac{\bar{z}_{Q_a^+}}{2} - \frac{\sqrt{3}}{2} \geq \frac{(a-\sqrt{3})}{2}$ . Entonces, ya que  $\lambda < 0$ ,  $\lambda[\bar{z}_{Q_a^+}y + \omega] \leq \frac{\lambda(a-\sqrt{3})}{2}$ , se tiene que

$$\Phi^{-1} \left( \left( \frac{1}{2} \right)^{\frac{1}{f_n(Q_a^+)}} \right) \leq \lambda[\bar{z}_{Q_a^+}y + \omega] \leq \lambda \frac{(a - \sqrt{3})}{2},$$

por la parte (i); y usando el que  $a - \sqrt{3} > 0$  se sigue que

$$\lambda \geq \frac{2}{a - \sqrt{3}} \Phi^{-1} \left( \left( \frac{1}{2} \right)^{\frac{1}{f_n(Q_a^+)}} \right),$$

una relación que es además válida aún cuando  $Q_a^+$  es vacío, dado que en este caso el lado derecho es  $-\infty$ .  $\square$

## 4.4 MAXIMIZACIÓN DE LA FUNCIÓN AUXILIAR

Esta sección contiene la segunda parte de los resultados preliminares que serán usados para establecer el Teorema 4.2.1. Primero, se introduce una clase  $\mathcal{S}_{a,n}$  de muestras normalizadas con la propiedad de que  $\tilde{L}_n(\cdot; \mathbf{z})$  tiene un maximizador si  $\mathbf{z} \in \mathcal{S}_{a,n}$ ; enseguida, se construye un selector medible para la familia  $\{\mathcal{M}_n(\mathbf{z}) \mid \mathbf{z} \in \mathcal{S}_{a,n}\}$ , y la sección concluye demostrando que, con probabilidad 1, la muestra normalizada observada pertenece al conjunto  $\mathcal{S}_{a,n}$  cuando el tamaño de muestra  $n$  es suficientemente grande. Para iniciar el argumento, dado  $a > 0$ , sea  $\mathcal{S}_{a,n}$  la clase de todos los vectores  $n$ -dimensionales que tienen componentes a ambos lados del intervalo  $[-a, a]$ ; formalmente,

$$\mathcal{S}_{a,n} = \{\mathbf{z} \in \mathbb{R}^n \mid z_i < -a, z_j > a \text{ para alguna } i \neq j \text{ con } 1 \leq i, j \leq n\}, \quad (4.4.1)$$

el cual es un subconjunto abierto de  $\mathbb{R}^n$ .

**Lema 4.4.1.** *Sean el entero  $n \geq 2$  y  $a > \sqrt{3}$  arbitrarios pero fijos. En este caso, las propiedades (i)–(iv) se satisfacen.*

- (i) *Para cada  $\mathbf{z} \in \mathcal{S}_{a,n}$ ,  $\mathcal{M}_n(\mathbf{z})$  es no vacío y compacto;*
- (ii) *Suponga que las sucesiones convergentes  $\{\mathbf{z}_k\} \subset \mathcal{S}_{a,n}$  y  $\{\nu_k\} \subset \Theta$  son tales que*
  - (a)  *$\lim_{k \rightarrow \infty} \mathbf{z}_k = \mathbf{z} \in \mathcal{S}_{a,n}$ , y*
  - (b)  *$\nu_k \in \mathcal{M}_n(\mathbf{z}_k)$  para  $k = 1, 2, \dots$ , y  $\lim_{k \rightarrow \infty} \nu_k = \nu$ .**En esta situación,  $\nu \in \mathcal{M}_n(\mathbf{z})$ .*
- (iii) *Para cada  $\mathbf{z} \in \mathcal{S}_{a,n}$ , el conjunto  $\mathcal{I}_n(\mathbf{z})$  de terceras coordenadas de  $\mathcal{M}_n(\mathbf{z})$  es no vacío y compacto; vea Definición 4.2.2(ii).*
- (iv) *Defina el mapeo  $\lambda_n^+ : \mathcal{S}_{a,n} \rightarrow \mathbb{R}$  mediante*

$$\lambda_n^+(\mathbf{z}) = \sup \mathcal{I}_n(\mathbf{z}), \quad \mathbf{z} \in \mathcal{S}_{a,n}. \quad (4.4.2)$$

*En este caso,  $\lambda_n^+(\mathbf{z}) \in \mathcal{I}_n(\mathbf{z})$  para cada  $\mathbf{z} \in \mathcal{S}_{a,n}$  y  $\lambda_n^+(\cdot)$  es una función semi-continua superior.*

**Demostración.** Note que para  $\mathbf{z} \in \mathcal{S}_{a,n}$  los conjuntos  $Q_a^+$  y  $Q_a^-$  en (4.3.5) son no vacíos, así que  $f_n(Q_a^+) \geq \frac{1}{n}$  y  $f_n(Q_a^-) \geq \frac{1}{n}$ . Por lo tanto, el Lema 4.3.1(ii) y el Lema 4.3.2 conjuntamente conducen a que, para cada  $\mathbf{z} \in \mathcal{S}_{a,n}$ ,

$$\begin{aligned} \mathcal{M}_n(\mathbf{z}) \subset \mathcal{P}_n(\mathbf{z}) \subset & \left[ \frac{1}{2}, 1 \right] \times \left[ \frac{-\sqrt{3}}{2}, \frac{\sqrt{3}}{2} \right] \\ & \times \frac{2}{a - \sqrt{3}} \left[ \Phi^{-1} \left( \frac{1}{2^n} \right), -\Phi^{-1} \left( \frac{1}{2^n} \right) \right] =: K_n. \end{aligned} \quad (4.4.3)$$

(i) Sea  $\mathbf{z} \in \mathcal{S}_{a,n}$  arbitrario pero fijo. Como ya se había notado,  $\mathcal{M}_n(\mathbf{z})$  y  $\mathcal{P}_n(\mathbf{z})$  son siempre conjuntos cerrados y entonces las anteriores inclusiones implican que estos conjuntos son, además, compactos, puesto que están contenidos en  $K_n$ . Por otra parte,  $\mathcal{P}_n(\mathbf{z})$  es no vacío dado que  $(1, 0, 0) \in \mathcal{P}_n(\mathbf{z})$  (vea (4.3.1)), y entonces existe un  $\theta^* \in \mathcal{P}_n(\mathbf{z})$  que satisface

$$\tilde{L}_n(\theta^*; \mathbf{z}) = \sup_{\theta \in \mathcal{P}_n(\mathbf{z})} \tilde{L}_n(\theta; \mathbf{z}) \geq \tilde{L}_n(1, 0, 0; \mathbf{z}), \quad (4.4.4)$$

debido a la continuidad de  $\tilde{L}_n(\cdot; \mathbf{z})$ . Se demostrará que  $\theta^*$  es un maximizador de  $\tilde{L}_n(\cdot; \mathbf{z})$  sobre el espacio de parámetros completo  $\Theta$ . Para alcanzar esta meta, sea  $\theta \in \Theta$  arbitrario pero fijo, y sea  $\theta_1$  como en el Lema 4.3.1(i) así que

$$\tilde{L}_n(\theta; \mathbf{z}) \leq \tilde{L}_n(\theta_1; \mathbf{z}).$$

Por lo tanto, si  $\tilde{L}_n(\theta_1; \mathbf{z}) \leq \tilde{L}_n(1, 0, 0; \mathbf{z})$ , entonces

$$\tilde{L}_n(\theta_1; \mathbf{z}) \leq \tilde{L}_n(\theta^*; \mathbf{z}),$$

por (4.4.4). Recordando que los dos primeros componentes de  $\theta_1$  determinan un punto en el círculo unitario, si  $\tilde{L}_n(\theta_1; \mathbf{z}) \geq \tilde{L}_n(1, 0, 0; \mathbf{z})$  entonces  $\theta_1 \in \mathcal{P}_n(\mathbf{z})$  y, nuevamente, la relación anterior se cumple. Entonces,  $\tilde{L}_n(\theta; \mathbf{z}) \leq \tilde{L}_n(\theta^*; \mathbf{z})$ , y por lo tanto  $\tilde{L}_n(\theta^*; \mathbf{z}) = \sup_{\theta \in \Theta} \tilde{L}_n(\theta; \mathbf{z})$ , esto es,  $\theta^* \in \mathcal{M}_n(\mathbf{z})$ .

- (ii) Note primero que  $\nu_k \in \mathcal{M}_n(\mathbf{z}_k) \subset K_n \subset \Theta$ , implica que  $\nu = \lim_k \nu_k \in K_n \subset \Theta$ , puesto que  $K_n$  es compacto. Por otra parte, la inclusión  $\nu_k \in \mathcal{M}_n(\mathbf{z}_k)$  conduce a  $\tilde{L}_n(\nu_k; \mathbf{z}_k) \geq \tilde{L}_n(\theta; \mathbf{z}_k)$  para toda  $k$  y para toda  $\theta \in \Theta$  y, después de tomar el límite cuando  $k$  tiende a  $\infty$ , la continuidad de  $\tilde{L}_n$  implica que  $\tilde{L}_n(\nu; \mathbf{z}) \geq \tilde{L}_n(\theta; \mathbf{z})$  para cada  $\theta \in \Theta$ , esto es,  $\nu \in \mathcal{M}_n(\mathbf{z})$ .
- (iii) La conclusión se desprende combinando la continuidad de la proyección  $\pi$  en la Definición 4.2.2 y la parte (i).
- (iv) Dado que  $\mathcal{I}_n(\mathbf{z})$  es compacto para cada  $\mathbf{z} \in \mathcal{S}_{a,n}$ , la inclusión  $\lambda_n^+(\mathbf{z}) \in \mathcal{I}_n(\mathbf{z})$  se sigue de (4.4.2). Para demostrar que  $\lambda_n^+(\cdot)$  es semicontinua por arriba, sea  $\mathbf{z} \in \mathcal{S}_{a,n}$  arbitrario pero fijo, y seleccione una secuencia  $\{\mathbf{z}_k\} \subset \mathcal{S}_{a,n}$  tal que

$$\mathbf{z}_k \rightarrow \mathbf{z} \quad \text{y} \quad \lambda_n^+(\mathbf{z}_k) \rightarrow \limsup_{\mathbf{y} \rightarrow \mathbf{z}} \lambda_n^+(\mathbf{y}) \quad \text{cuando } k \rightarrow \infty.$$

Enseguida, seleccione una pareja  $(y_k, \omega_k) \in (0, \infty) \times \mathbb{R}$  tal que

$$\nu_k := (y_k, \omega_k, \lambda_n^+(\mathbf{z}_k)) \in \mathcal{M}_n(\mathbf{z}_k),$$

lo cual es posible, dado que  $\lambda_n^+(\mathbf{z}_k) \in \mathcal{I}_n(\mathbf{z}_k)$ . Por (4.4.3),  $\{\nu_k\} \subset K_n$ , el cual es un conjunto compacto. Tomando, de ser necesario, una subsecuencia, puede suponerse que

$$\nu_k = (y_k, \omega_k, \lambda_n^+(\mathbf{z}_k)) \rightarrow \nu^* = (y^*, \omega^*, \lambda^*) \in K_n \subset \Theta.$$

En este caso, la parte (ii) implica que  $\nu^* \in \mathcal{M}_n(\mathbf{z})$ , y entonces  $\lambda^* \leq \lambda^+(\mathbf{z})$ , por (4.4.2); puesto que  $\lambda^* = \limsup_{\mathbf{y} \rightarrow \mathbf{z}} \lambda_n^+(\mathbf{y})$ , por las dos últimas ecuaciones, se sigue que

$$\limsup_{\mathbf{y} \rightarrow \mathbf{z}} \lambda_n^+(\mathbf{y}) \leq \lambda_n^+(\mathbf{z}),$$

y entonces  $\lambda_n^+(\cdot)$  es semicontinua superior. □

Ahora, se construirá un selector medible para la familia  $\{\mathcal{M}_n(\mathbf{z}) \mid \mathbf{z} \in \mathcal{S}_{a,n}\}$ .

**Lema 4.4.2.**

(i) Para cada  $\lambda \in \mathbb{R}$  y para cada  $\mathbf{z} \in \mathbb{R}^n$ , el mapeo

$$(y, \omega) \mapsto \tilde{L}_n(y, \omega, \lambda; \mathbf{z}), \quad y > 0, \quad \omega \in \mathbb{R},$$

es estrictamente cóncavo y tiene un maximizador único  $(\tilde{y}(\lambda, \mathbf{z}), \tilde{\omega}(\lambda, \mathbf{z})) \in (0, \infty) \times \mathbb{R}$  el cual es una función diferenciable de  $(\lambda, \mathbf{z})$ .

(ii) Dados un entero positivo  $n$  y  $a > \sqrt{3}$ , defina una función  $\theta_n^+(\cdot)$  en  $\mathcal{S}_{a,n}$  mediante

$$\theta_n^+(\mathbf{z}) = (\tilde{y}(\lambda_n^+(\mathbf{z}), \mathbf{z}), \tilde{\omega}(\lambda_n^+(\mathbf{z}), \mathbf{z}), \lambda_n^+(\mathbf{z})), \quad \mathbf{z} \in \mathcal{S}_{a,n}. \quad (4.4.5)$$

en este caso  $\theta_n^+(\cdot)$  es medible y satisface  $\theta_n^+(\mathbf{z}) \in \mathcal{M}_n(\mathbf{z})$  si  $\mathbf{z} \in \mathcal{S}_{a,n}$ .

**Demostración.**

(i) Dados  $\lambda \in \mathbb{R}$  y  $\mathbf{z} \in \mathbb{R}^n$  arbitrarios, note que la matriz Hessiana del mapeo  $(y, \omega) \mapsto \tilde{L}_n(y, \omega, \lambda; \mathbf{z})$  esta dada por

$$\begin{aligned} \mathcal{H}(y, \omega; \lambda, \mathbf{z}) &= \begin{bmatrix} \partial_y^2 \tilde{L}_n & \partial_\omega \partial_y \tilde{L}_n \\ \partial_y \partial_\omega \tilde{L}_n & \partial_\omega^2 \tilde{L}_n \end{bmatrix} \\ &= \begin{bmatrix} -1 - \frac{1}{y^2} & 0 \\ 0 & -1 \end{bmatrix} + \frac{\lambda^2}{n} \sum_{i=1}^n H''(\lambda[z_i y + \omega]) \begin{bmatrix} z_i \\ 1 \end{bmatrix} [z_i, 1]; \end{aligned}$$

vea (4.2.11). Recordando que  $H''(\cdot) < 0$ , se sigue que  $\mathcal{H}$  es siempre una matriz negativa definida, con lo que  $\tilde{L}_n(\cdot, \cdot; \lambda, \mathbf{z})$  es estrictamente cóncava. Observe ahora que  $\tilde{L}_n(y, \omega; \lambda, \mathbf{z}) \rightarrow -\infty$  cuando  $y \searrow 0$  o  $y + |\omega| \rightarrow \infty$ , de manera que  $\tilde{L}_n(\cdot, \cdot; \lambda, \mathbf{z})$  tiene un maximizador  $(\tilde{y}(\lambda, \mathbf{z}), \tilde{\omega}(\lambda, \mathbf{z}))$  — que es único debido a la concavidad estricta— el cual está caracterizado como la solución para  $(y, \omega)$  de las ecuaciones  $\partial_y \tilde{L}_n(y, \omega, \lambda; \mathbf{z}) = 0$  y  $\partial_\omega \tilde{L}_n(y, \omega, \lambda; \mathbf{z}) = 0$ ; dado que la matriz Hessiana  $\mathcal{H}$  es negativa definida, la suavidad del mapeo  $(\lambda, \mathbf{z}) \mapsto (\tilde{y}(\lambda, \mathbf{z}), \tilde{\omega}(\lambda, \mathbf{z}))$  se obtiene a partir del teorema de la función implícita.



(ii) Dado que  $\tilde{y}$  y  $\tilde{\omega}$  son funciones suaves, la semicontinuidad superior de  $\lambda_n^+$  conduce a que  $\theta_n^+(\cdot)$  es medible. Para concluir, sea  $\mathbf{z} \in \mathcal{S}_{a,n}$  arbitrario pero fijo y, usando que  $\lambda_n^+(\mathbf{z}) \in \mathcal{I}_n(\mathbf{z})$ , por el Lema 4.4.1(iv), seleccione  $(y^*, \omega^*)$  tal que  $(y^*, \omega^*, \lambda_n^+(\mathbf{z})) \in \mathcal{M}_n(\mathbf{z})$  i.e.,

$$\sup_{\theta \in \Theta} \tilde{L}_n(\theta; \mathbf{z}) = \tilde{L}_n(y^*, \omega^*, \lambda_n^+(\mathbf{z}); \mathbf{z}).$$

Se sigue que  $(\tilde{y}(\lambda_n^+(\mathbf{z})), \mathbf{z}), \tilde{\omega}(\lambda_n^+(\mathbf{z})), \mathbf{z}) = (y^*, \omega^*)$ , puesto que de otra forma,

$$\tilde{L}_n(y^*, \omega^*, \lambda_n^+(\mathbf{z}); \mathbf{z}) < \tilde{L}_n(\tilde{y}(\lambda_n^+(\mathbf{z})), \mathbf{z}), \tilde{\omega}(\lambda_n^+(\mathbf{z})), \mathbf{z}), \lambda_n^+(\mathbf{z}); \mathbf{z}),$$

por la parte (i), lo que contradice la formulación previa. Luego,

$$\theta_n^+(\mathbf{z}) = (\tilde{y}(\lambda_n^+(\mathbf{z})), \mathbf{z}), \tilde{\omega}(\lambda_n^+(\mathbf{z})), \mathbf{z}), \lambda_n^+(\mathbf{z})) = (y^*, \omega^*, \lambda_n^+(\mathbf{z})) \in \mathcal{M}_n(\mathbf{z}),$$

completando la demostración.  $\square$

El paso final antes de la demostración del 4.2.1 consiste en demostrar que, con probabilidad 1, la muestra normalizada pertenecerá al conjunto  $\mathcal{S}_{a,n}$  siempre que  $n$  sea lo suficientemente grande. Dada una muestra  $X_1^n = (X_1, \dots, X_n)$ , sea

$$\mathbf{Z}_n := (Z_{1,n}, \dots, Z_{n,n}) \tag{4.4.6}$$

la correspondiente muestra estandarizada, donde,

$$\begin{aligned} \text{si } S_n > 0, \quad Z_{i,n} &:= \frac{X_i - \bar{X}_n}{S_n}, \quad i = 1, 2, \dots, n, \\ \text{si } S_n = 0, \quad Z_{i,n} &:= 0 \quad i = 1, 2, \dots, n. \end{aligned} \tag{4.4.7}$$

**Lema 4.4.3.** Para cada  $\theta_0 \in \theta$  y para cada  $a > 0$

$$P_{\theta_0} \left[ \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} [\mathbf{Z}_k \in \mathcal{S}_{a,k}] \right] = 1, \quad \theta_0 \in \Theta;$$

vea (4.4.1), (4.4.6) y (4.4.7).

**Demostración.** Dados  $a > 0$  y  $\theta_0 \in \Theta$ , sean  $\mu(\theta_0)$  y  $v(\theta_0)$  la media y la desviación estándar de  $X_1$  con respecto a la densidad  $\rho(\cdot; \theta_0)$  en (4.1.1), y defina

$$\Omega_* := \left[ \lim_{n \rightarrow \infty} \bar{X}_n = \mu(\theta_0), \lim_{n \rightarrow \infty} S_n = v(\theta_0) \right]. \quad (4.4.8)$$

Entonces, por la ley fuerte de los grandes números

$$P_{\theta_0}[\Omega_*] = 1. \quad (4.4.9)$$

Enseguida, note que ambas probabilidades  $P_{\theta_0}[X_1 \leq b]$  y  $P_{\theta_0}[X_1 \geq b]$  son menores que 1 para cada  $b$ , puesto que la densidad  $\rho(\cdot; \theta_0)$  es positiva en la línea real. Entonces,

$$\begin{aligned} P_{\theta_0}[X_i \leq b \text{ para toda } i] &= \lim_{n \rightarrow \infty} P_{\theta_0}[X_i \leq b, 1 \leq i \leq n] \\ &= \lim_{n \rightarrow \infty} P_{\theta_0}[X_1 \geq b]^n = 0 \end{aligned}$$

y, similarmente,  $P_{\theta_0}[X_i \geq b \text{ para toda } i] = 0$ . Luego, definiendo

$$\Omega_0 := [X_i \leq \mu(\theta_0) - av(\theta_0) \text{ para toda } i] \cup [X_i \geq \mu(\theta_0) + av(\theta_0) \text{ para toda } i]$$

se sigue que

$$P_{\theta_0}[\Omega_0] = 0.$$

Se demostrará que

$$\Omega_* \cap \left[ \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} [\mathbf{Z}_k \notin \mathcal{S}_{a,k}] \right] \subset \Omega_0 \quad (4.4.10)$$

Con este fin, sea  $X_1, X_2, X_3, \dots$  una trayectoria muestral arbitraria pero fija tal que el evento  $\Omega_* \cap \left[ \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} [\mathbf{Z}_k \notin \mathcal{S}_{a,k}] \right]$  ocurre, y observe que los siguientes hechos (a)–(c) se cumplen a lo largo de esta trayectoria:

(a) Dado que la trayectoria  $X_1, X_2, X_3, \dots$ , es tal que  $\Omega_*$  ocurre, existe un entero  $M$  tal que

$$S_n > \frac{v(\theta_0)}{2} > 0, \quad n > M, \quad \bar{X}_n \pm aS_n \rightarrow \mu(\theta_0) \pm av(\theta_0) \quad \text{cuando } n \rightarrow \infty.$$

(b) Usando que  $\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} [\mathbf{Z}_k \notin \mathcal{S}_{a,k}]$  ocurre sobre la trayectoria  $X_1, X_2, X_3, \dots$  es observada, existe una sucesión  $\{n_r\}$  tal que

$$n_r \nearrow \infty, \quad \mathbf{Z}_{n_r} \notin \mathcal{S}_{a,n_r} \quad \text{y} \quad n_r > M \quad \text{para toda } r.$$

Puesto que  $S_n$  es positiva cuando  $n > M$ , por (a), un vistazo a (4.4.1), (4.4.6) y (4.4.7) conducen a que, para cada  $r$ , uno de las siguientes afirmaciones se cumple:

$$X_i \leq \bar{X}_{n_r} - aS_{n_r} \quad \text{para toda } i = 1, 2, \dots, n_r, \quad (4.4.11)$$

o

$$X_i \geq \bar{X}_{n_r} + aS_{n_r} \quad \text{para toda } i = 1, 2, \dots, n_r. \quad (4.4.12)$$

(c) Por (b), una de (4.4.11) o (4.4.12) ocurre para un número infinito de enteros  $r$ . Si la primer afirmación es válida con una frecuencia infinita, entonces tomando el límite cuando  $r$  tiene a  $\infty$ , la aseveración (a) conduce a que  $X_i \leq \mu(\theta_0) - av(\theta_0)$  para todos los enteros positivos  $i$ , así que  $\Omega_0$  ocurre. Similarmente, si (4.4.12) se cumple para infinitas  $r$ 's, entonces (a) conduce a  $X_i \geq \mu(\theta_0) + av(\theta_0)$  para toda  $i \geq 1$  y, nuevamente,  $\Omega_0$  ocurre.

Resumiendo, se ha demostrado que el evento  $\Omega_0$  ocurre a lo largo de una trayectoria muestral  $\{X_i\}$  tal que se observa  $\Omega_* \cap \left[ \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} [\mathbf{Z}_k \notin \mathcal{S}_{a,k}] \right]$ , estableciendo (4.4.10). Por lo tanto, dado que  $P_{\theta_0}[\Omega_0] = 0$  y  $P_{\theta_0}[\Omega_*] = 1$ ,

$$P_{\theta_0} \left[ \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} [\mathbf{Z}_k \notin \mathcal{S}_{a,k}] \right] = 0,$$

que es equivalente a la conclusión deseada. □

## 4.5 DEMOSTRACIÓN DEL TEOREMA DE EXISTENCIA

Después del trabajo previo, en esta sección se establece, finalmente, la existencia de una secuencia de estimadores de verosimilitud máxima.

### Demostración del Teorema 4.2.1.

- (i) Sea  $a > \sqrt{3}$  fija y, para un entero positivo  $n$ , sea  $\mathbf{Z}_n$  la muestra normalizada asociada a  $X_1^n = (X_1, X_2, \dots, X_n)$ ; vea (4.4.6) y (4.4.7). A continuación, sea  $\theta_n : \mathbb{R}^n \rightarrow \Theta$  la función especificada mediante

$$\theta_n(\mathbf{z}) := \begin{cases} \theta_n^+(\mathbf{z}), & \text{si } \mathbf{z} \in \mathcal{S}_{a,n} \\ \theta^*, & \text{de otra forma,} \end{cases}$$

donde  $\theta_n^+$  es el selector medible en el Lema 4.4.2, y el punto  $\theta^* \in \Theta$  es arbitrario pero fijo; ya que  $\mathcal{S}_{a,n}$  es un conjunto abierto y  $\theta_n^+(\cdot)$  es un mapeo medible, se desprende que  $\theta_n(\cdot)$  también lo es. Ahora, defina el vector aleatorio  $\tilde{\theta}_n$  como

$$\tilde{\theta}_n := \theta_n(\mathbf{Z}_n)$$

y note que el Lema 4.4.2(ii) implica que

$$\begin{aligned} [\mathbf{Z}_n \in \mathcal{S}_{a,n}] \subset [\tilde{\theta}_n \in \mathcal{M}_n(\mathbf{Z}_n)] &= [\tilde{L}_n(\tilde{\theta}_n; \mathbf{Z}_n) \\ &\geq \tilde{L}_n(\theta_n; \mathbf{Z}_n), \quad \theta \in \Theta]. \end{aligned} \quad (4.5.1)$$

Ahora defina

$$\hat{\theta}_n := \begin{cases} T_{X_1^n}^{-1}(\tilde{\theta}_n), & \text{si } S_n > 0, \\ \theta^*, & \text{de otra forma.} \end{cases} \quad (4.5.2)$$

Ya que la desviación estándar muestral  $S_n$  es positiva en el evento  $[\mathbf{Z}_n \in \mathcal{S}_{a,n}]$ , usando (4.5.1) se desprende que la definición anterior de  $\hat{\theta}_n$  y el Lema 4.2.1(iii) conjuntamente conducen a que

$$[\mathbf{Z}_n \in \mathcal{S}_{a,n}] \subset [L_n(\hat{\theta}_n; X_1^n) \geq L_n(\theta; X_1^n), \quad \theta \in \Theta]$$

y a través del Lema 4.4.3 se obtiene

$$P_{\theta_0} \left[ \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} [L_k(\hat{\theta}_k; X_1^k) \geq L_k(\theta; X_1^k), \theta \in \Theta] \right] \\ \geq P_{\theta_0} \left[ \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} [\mathbf{Z}_k \in \mathcal{S}_{a,k}] \right] = 1, \quad \theta_0 \in \Theta.$$

Por lo tanto, se sigue que  $\{\hat{\theta}_n\}$  es una sucesión de estimadores de verosimilitud máxima; vea la Definición 4.2.1.

(ii) Sea  $\{\hat{\theta}_n = (\hat{\sigma}_n, \hat{\xi}_n, \hat{\lambda}_n)\}$  una secuencia de estimadores de verosimilitud máxima y, en el evento  $[S_n > 0]$ , defina

$$\tilde{\theta}_n := T_{X_1^n}(\hat{\theta}_n) = \left( \frac{S_n}{\hat{\sigma}_n}, \frac{\bar{X}_n - \hat{\xi}_n}{\hat{\sigma}_n}, \hat{\lambda}_n \right).$$

Observe ahora que, en el evento

$$[L_n(\hat{\theta}_n; X_1^n) \geq L_n(\theta; X_1^n), \theta \in \Theta] \cap [S_n > 0]$$

los hechos (a) y (b) abajo listados se cumplen:

(a) Por el Lema 4.2.1(iii),  $\tilde{\theta}_n$  maximiza  $\tilde{L}_n(\cdot; \mathbf{Z}_n)$ , esto es,

$$\tilde{\theta}_n \in \mathcal{M}_n(\mathbf{Z}_n) \subset \mathcal{P}_n(\mathbf{Z}_n),$$

y entonces  $\left(\frac{S_n}{\hat{\sigma}_n}\right)^2 + \left(\frac{\bar{X}_n - \hat{\xi}_n}{\hat{\sigma}_n}\right)^2 = 1$ , por (4.3.1), siempre que  $\left(\frac{S_n}{\hat{\sigma}_n}\right) \geq \frac{1}{2}$ , por el Lema 4.3.2(i); esto establece (4.2.6).

(b) Recordando que  $a > \sqrt{3}$ , la parte (a) y el Lema 4.3.2(ii) conduce a

$$\frac{2}{a - \sqrt{3}} \Phi^{-1} \left( \left( \frac{1}{2} \right)^{\frac{1}{f_n(Q_a^+)}} \right) \leq \hat{\lambda}_n \leq \frac{2}{\sqrt{3} - a} \Phi^{-1} \left( \left( \frac{1}{2} \right)^{\frac{1}{f_n(Q_a^-)}} \right),$$

donde  $Q^- = \{i \mid Z_{i,n} \leq -a\}$  y  $Q^+ = \{i \mid Z_{i,n} \geq -a\}$ ; a partir de aquí, después de una comparación con (4.2.8), (4.3.3) y (4.4.7), no es difícil ver que  $f_n(Q_n^+) = \beta_n^+(a)$  y  $f_n(Q_n^-) = \beta_n^-(a)$ , y (4.2.7) se sigue de esas igualdades y de las desigualdades anteriores, completándose la demostración.  $\square$

## 4.6 CONSISTENCIA

En esta sección se demuestra que una sucesión  $\{\hat{\theta}_n\}$  de estimadores de verosimilitud máxima es (fuertemente) consistente, *i.e.*, que converge al valor real del parámetro con probabilidad 1.

**Teorema 4.6.1.** *Sea  $\{\hat{\theta}_n\}$  una secuencia de estimadores de verosimilitud máxima de  $\theta$ . En este caso  $\{\hat{\theta}_n\}$  es consistente, esto es,*

$$P_{\theta_0} \left[ \lim_{n \rightarrow \infty} \hat{\theta}_n = \theta_0 \right] = 1, \quad \theta_0 \in \Theta.$$

El núcleo de la demostración de este resultado consiste en demostrar que, con probabilidad 1,  $\{\hat{\theta}_n\}$  es una secuencia acotada en  $\Theta$ , esto es, que la sucesión está contenida en un subconjunto compacto de  $\Theta$ . Esta propiedad se establece formalmente en la segunda parte del siguiente lema.

**Lema 4.6.1.** *Sea  $\theta_0 \in \Theta$  arbitrario pero fijo.*

(a) *Para cada  $a > 0$ , sean  $\beta_n^+(a)$  y  $\beta_n^-(a)$  como en (4.2.8). En este caso, las siguientes convergencias se cumplen casi seguramente con respecto a  $P_{\theta_0}$ :*

$$\lim_{n \rightarrow \infty} \beta_n^+(a) = P_{\theta_0}[X_1 > \mu(\theta_0) + av(\theta_0)],$$

y

$$\lim_{n \rightarrow \infty} \beta_n^-(a) = P_{\theta_0}[X_1 < \mu(\theta_0) - av(\theta_0)],$$

donde, como antes,  $\mu(\theta_0)$  y  $v(\theta_0)$  son, respectivamente, la media y la desviación estándar de  $\rho(\cdot; \theta_0)$ .

(b) *Dada una secuencia  $\{\hat{\theta}_n\}$  de estimadores de verosimilitud máxima, existe un conjunto compacto  $K_{\theta_0} \subset \Theta$  tal que, cuando  $\theta_0$  es el valor real del parámetro,  $\hat{\theta}_n$  cae en  $K_{\theta_0}$  para  $n$  suficientemente grande; más precisamente,*

$$P_{\theta_0} \left[ \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} [\hat{\theta}_k \in K_{\theta_0}] \right] = 1.$$

**Demostración.** Sea  $\mathcal{Q}$  el conjunto de los números racionales, y defina el evento  $\Omega_{**}$  por

$$\Omega_{**} = \left[ \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n I[X_i < r] = P_{\theta_0}[X_1 < r] \quad \text{para toda } r \in \mathcal{Q} \right].$$

Entonces, por la ley fuerte de los grandes números  $P_{\theta_0}[\Omega_{**}] = 1$ . Luego, si el evento  $\Omega_*$  es como en (4.4.8), se tiene que

$$P_{\theta_0}[\Omega_* \cap \Omega_{**}] = 1; \quad (4.6.1)$$

vea (4.4.9).

(i) Sea  $X_1, X_2, \dots$  una trayectoria arbitraria pero fija tal que  $\Omega_{**} \cap \Omega_*$  ocurre, y observe que los siguientes hechos (a)–(c) se cumplen a lo largo de esta trayectoria:

(a) Ya que  $\Omega_*$  ocurre cuando  $X_1, X_2, \dots$  es observada, se sigue que  $\bar{X}_n + aS_n \rightarrow a\mu(\theta_0) + av(\theta_0)$ . Por lo tanto, si  $r, s \in \mathcal{Q}$  satisfacen

$$r < a\mu(\theta_0) + av(\theta_0) < s, \quad (4.6.2)$$

entonces existe un entero positivo  $M_{r,s}$  tal que

$$r < \bar{X}_n + aS_n < s, \quad n \geq M_{r,s}.$$

(b) Por (a), a lo largo de la trayectoria  $X_1, X_2, \dots$ , si  $n > M_{r,s}$  y  $i \leq n$  entonces

$$X_i > \bar{X}_n + aS_n \Rightarrow X_i > r, \quad \text{y} \quad X_i > s \Rightarrow X_i > \bar{X}_n + aS_n,$$

y se sigue que

$$\sum_{i=M_{r,s}}^n I[X_i > s] \leq \sum_{i=M_{r,s}}^n I[X_i > \bar{X}_n + aS_n] \leq \sum_{i=M_{r,s}}^n I[X_i > r],$$

*i.e.*, Para  $n > M_{r,s}$

$$\frac{1}{n} \sum_{i=M_{r,s}}^n I[X_i > s] \leq \frac{1}{n} \sum_{i=M_{r,s}}^n I[X_i > \bar{X}_n + aS_n] \leq \frac{1}{n} \sum_{i=M_{r,s}}^n I[X_i > r].$$

(c) Antes de continuar, es conveniente recordar la definición de  $\beta_n^+(a)$  y de  $\beta_n^-(a)$  en (4.2.8). Usando que  $\Omega_{**}$  ocurre en la trayectoria  $X_1, X_2, \dots$ , dejando que  $n$  crezca al infinito en las desigualdades anteriores, se sigue

que  $P_{\theta_0}[X_1 > s] \leq \liminf_{n \rightarrow \infty} \beta_n^+(a) \leq \limsup_{n \rightarrow \infty} \beta_n^+(a) \leq P_{\theta_0}[X_1 > r]$ , y entonces  $\lim_{n \rightarrow \infty} \beta_n^+(a) = P_{\theta_0}[X_1 > \mu(\theta_0) + av(\theta)]$ , puesto que  $r, s \in \mathcal{Q}$  que satisfacen (4.6.2) son arbitrarios.

En resumen, se ha demostrado que  $\beta_n^+(a) \rightarrow P_{\theta_0}[X_1 > \mu(\theta_0) + av(\theta)]$  a lo largo de una trayectoria arbitraria  $\{X_i\}$  para la cual  $\Omega_{**} \cap \Omega_*$  ocurre, y entonces por (4.6.1)

$$\lim_{n \rightarrow \infty} \beta_n^+(a) = P_{\theta_0}[X_1 > \mu(\theta_0) + av(\theta)] \quad P_{\theta_0}\text{-a.s.},$$

La afirmación para  $\{\beta_n^-(a)\}$  puede establecerse por medio de argumentos similares.

(ii) Sea  $\{\hat{\theta}_n\}$  una secuencia de estimadores de verosimilitud máxima y dado  $a > \sqrt{3}$ , defina los eventos

$$\begin{aligned} \Omega_\beta(a) &= [\lim_{n \rightarrow \infty} \beta_n^+(a) = P_{\theta_0}[X_1 > \mu(\theta_0) + av(\theta_0)]] \\ &\quad \cap [\lim_{n \rightarrow \infty} \beta_n^-(a) = P_{\theta_0}[X_1 < \mu(\theta_0) - av(\theta_0)]], \end{aligned}$$

y

$$\Omega_0 = \Omega_\beta(a) \cap \Omega_* \cap \left[ \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} [L_n(\hat{\theta}_n; X_1^n) \geq L_n(\theta; X_1^n) \text{ para toda } \theta \in \Theta] \right].$$

Note que

$$P_{\theta_0}[\Omega_0] = 1, \quad (4.6.3)$$

por la parte (i), (4.4.9), y la Definición 4.2.1. Ahora, sea  $X_1, X_2, \dots$ , una trayectoria arbitraria pero fija tal que  $\Omega_0$  ocurre, y observe que los hechos (a)–(c) enumerados abajo se cumplen a lo largo de esta trayectoria:

(a) Usando que  $\Omega_* \cap \Omega_\beta$  ocurre cuando  $X_1, X_2, \dots$  es observada, de (4.4.8) y la parte (i) se sigue que existe  $N_1 > 0$  tal que:

Para  $n > N_1$

$$0 < \frac{v(\theta_0)}{2} \leq S_n \leq 2v(\theta_0), \quad \mu(\theta_0) - 1 \leq \bar{X}_n \leq \mu(\theta_0) + 1,$$

$$\beta_n^+(a) > \frac{1}{\alpha^+(a, \theta_0)}, \quad \text{y} \quad \beta_n^-(a) > \frac{1}{\alpha^-(a, \theta_0)},$$



donde

$$\alpha^+(a, \theta_0) = \frac{2}{P_{\theta_0}} [X_1 > \mu(\theta_0) + av(\theta_0)]$$

y

$$\alpha^-(a, \theta_0) = \frac{2}{P_{\theta_0}} [X_1 < \mu(\theta_0) - av(\theta_0)].$$

(b) Dado que  $\bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} [L_n(\hat{\theta}_n; X_1^n) \geq L_n(\theta; X_1^n), \theta \in \Theta]$  ocurre a lo largo de la trayectoria  $X_1, X_2, \dots$ , existe un entero  $N_2$  tal que  $\hat{\theta}_n$  maximiza  $L_n(\cdot; X_1^n)$  cuando  $n > N_2$ .

(c) Recordando que  $a > \sqrt{3}$ , los hechos (a) y (b) anteriores y el Teorema 4.2.1(ii) conjuntamente implican que los componentes de  $\hat{\theta} = (\hat{\sigma}_n, \hat{\xi}_n, \hat{\lambda}_n)$  satisfacen las relaciones (4.2.6) y (4.2.7) cuando  $n > \max\{N_1, N_2\} =: N_3$ , y usando lo mostrado en el hecho (a) anterior, se sigue que para  $n > N_3$ ,

$$\frac{1}{2}v(\theta_0) \leq \hat{\sigma}_n \leq 4v(\theta_0), \quad \mu(\theta_0) - 1 - 2\sqrt{3}v(\theta_0) \leq \hat{\xi}_n \leq \mu(\theta_0) + 1 + 2\sqrt{3}v(\theta_0),$$

y

$$\frac{2}{a - \sqrt{3}} \Phi^{-1} \left( \left( \frac{1}{2} \right)^{\alpha^+(a, \theta_0)} \right) \leq \hat{\lambda}_n \leq \frac{2}{\sqrt{3} - a} \Phi^{-1} \left( \left( \frac{1}{2} \right)^{\alpha^-(a, \theta_0)} \right),$$

esto es,

$$\hat{\theta}_n = (\hat{\xi}_n, \hat{\sigma}_n, \hat{\lambda}_n) \in K_{\theta_0}, \quad n > N_3, \quad (4.6.4)$$

donde

$$K_{\theta_0} := \left[ \frac{v(\theta_0)}{2}, 4v(\theta_0) \right] \times [\mu(\theta_0) - 1 - 2\sqrt{3}, \mu(\theta_0) + 1 + 2\sqrt{3}] \\ \times \frac{2}{a - \sqrt{3}} \left[ \Phi^{-1} \left( \left( \frac{1}{2} \right)^{\alpha^+(a, \theta_0)} \right), -\Phi^{-1} \left( \left( \frac{1}{2} \right)^{\alpha^-(a, \theta_0)} \right) \right] \subset \Theta$$

es un conjunto compacto. Por (4.6.4), cuando la trayectoria  $X_1, X_2, \dots$  es observada, el evento  $\bigcap_{k=N_3+1}^{\infty} [\hat{\theta}_k \in K_{\theta_0}] \subset \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} [\hat{\theta}_k \in K_{\theta_0}]$  ocurre. En resumen, se ha demostrado que si una trayectoria muestral  $X_1, X_2, \dots$  es tal que  $\Omega_0$  ocurre, entonces  $\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} [\hat{\theta}_k \in K_{\theta_0}]$  también ocurre, así que

$$\Omega_0 \subset \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} [\hat{\theta}_k \in K_{\theta_0}],$$

y entonces  $P_{\theta_0} \left[ \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} [\hat{\theta}_k \in K_{\theta_0}] \right] = 1$ ; vea (4.6.3).  $\square$

La demostración del Teorema 4.6.1 se obtendrá combinando los resultados preliminares con las propiedades de continuidad y con la desigualdad de Kullback en el siguiente lema.

**Lema 4.6.2.**

(i) Dado un conjunto compacto  $K \subset \Theta$ , existe un polinomio de segundo orden  $p_K(\cdot)$  tal que

$$|\ell(\theta; x)| + \|(\ell_{\xi}(\theta; x), \ell_{\sigma}(\theta; x), \ell_{\lambda}(\theta; x))'\| \leq p_K(|x|), \quad \theta \in K, \quad x \in \mathbb{R},$$

donde  $\|\mathbf{a}\|$  denota la norma euclidiana de  $\mathbf{a} \in \mathbb{R}^3$ .

(ii) Dado  $\theta_0 \in \Theta$ , el mapeo  $\nu \mapsto E_{\theta_0}[\ell(\nu; X_1)]$  es finito y continuo en  $\nu \in \Theta$ , y

(iii) Para  $\theta_0, \nu \in \Theta$ ,  $E_{\theta_0}[\ell(\nu; X_1)] < E_{\theta_0}[\ell(\theta_0; X_1)]$  si  $\nu \neq \theta_0$ .

**Demostración.** Sea  $H(\cdot)$  como en (4.2.4). Puesto que  $H'(z) = \varphi(z)/\Phi(z)$ , no es difícil ver que  $H'(z) \rightarrow 0$  cuando  $z \rightarrow \infty$  y  $-H'(z)/z \rightarrow 1$  cuando  $z \rightarrow -\infty$ . Se sigue que hay una constante positiva  $c$  que satisface

$$|H'(z)| \leq c|z|, \quad z \in \mathbb{R},$$

y el teorema del valor medio conduce a

$$|H(z)| \leq |H(0)| + c|z|^2, \quad z \in \mathbb{R}.$$

(i) Combinando lo mostrado arriba y (4.2.1), se sigue que

$$|\ell(\theta; x)| \leq \frac{(x - \xi)^2}{2\sigma^2} + |\log(\sigma)| + |H(0)| + c \frac{\lambda^2(x - \xi)^2}{\sigma^2}.$$

Por otra parte, cálculos directos usando (4.2.1) llevan a  $\ell_\sigma(\theta; x) = \frac{1}{\sigma} + \frac{(x-\xi)^2}{\sigma^3}$ ,  $\ell_\xi(\theta; x) = \frac{(x-\xi)}{\sigma^2}$ , y  $\ell_\lambda(\theta; x) = H' \left( \lambda \frac{(x-\xi)}{\sigma} \right) \frac{(x-\xi)}{\sigma}$ , así que  $|\ell_\lambda(\theta; x)| \leq c|\lambda| \frac{(x-\xi)^2}{\sigma^2}$ . Entonces,

$$\|(\ell_\xi(\theta; x), \ell_\sigma(\theta; x), \ell_\lambda(\theta; x))'\| \leq \frac{|x-\xi|}{\sigma^2} + \frac{1}{\sigma} + \frac{(x-\xi)^2}{\sigma^3} + c|\lambda| \frac{(x-\xi)^2}{\sigma^2}.$$

Ahora, la conclusión se sigue al observar que si  $\theta \in K \subset \Theta$  entonces los componentes de  $\theta = (\sigma, \xi, \lambda)$  son acotados y  $\sigma$  es acotado lejos del cero.

(ii) Poniendo  $K = \{\theta\}$  en la parte (i), se sigue que  $|\ell(\theta; x)|$  es acotado por arriba por un polinomio de segundo orden en  $|x|$ , y entonces  $E_{\theta_0}[|\ell(\theta; X_1)|] < \infty$ , dado que los momentos de una distribución normal sesgada son finitos (Dalla Valle, 2004). Para verificar la continuidad, sea  $\theta \in \Theta$  fija, y seleccione  $\varepsilon > 0$ , tal que la bola compacta  $B(\theta, \varepsilon)$  con centro  $\theta$  y radio  $\varepsilon$  está contenida en  $\Theta$ . En este caso, el teorema del valor medio y la parte (i) implican conjuntamente que  $|\ell(\nu; X_1) - \ell(\theta; X_1)| \leq p(|X_1|)\|\theta - \nu\|$ , donde  $p(\cdot)$  es el polinomio de segundo orden asociado a  $B(\theta, \varepsilon)$  en la parte (i), y entonces  $|E_{\theta_0}[\ell(\nu; X_1)] - E_{\theta_0}[\ell(\theta; X_1)]| \leq \|\theta - \nu\|E_{\theta_0}[p(|X_1|)] \rightarrow 0$  cuando  $\nu \rightarrow \theta$ .

(iii) Usando fórmulas para los momentos de una densidad normal sesgada (Dalla Valle, 2004), no es difícil ver que la parametrización  $\theta \mapsto \rho(\cdot; \theta)$  es identificable, esto es,  $\int_{\mathbb{R}} |\rho(x; \theta) - \rho(x; \theta_1)| dx \neq 0$  si  $\theta \neq \theta_1$ . Entonces, si  $\theta_0 \neq \nu$  son miembros arbitrarios de  $\Theta$ , la concavidad estricta de la función logaritmo lleva, via la desigualdad de Jensen, a

$$\begin{aligned} E_{\theta_0}[\ell(\nu; X_1) - \ell(\theta_0; X_1)] &= E_{\theta_0} \left[ \log \left( \frac{\rho(\nu; X_1)}{\rho(\theta_0; X_1)} \right) \right] \\ &< \log \left( E_{\theta_0} \left[ \frac{\rho(\nu; X_1)}{\rho(\theta_0; X_1)} \right] \right) \\ &= \log \left( \int_{\mathbb{R}} \rho(\nu; x) dx \right) = \log(1) = 0, \end{aligned}$$

y la conclusión se sigue usando que  $E_{\theta_0}[\ell(\theta_0; X_1)]$  es finita.  $\square$

**Demostración del Teorema 4.6.1.** Sea  $\{\hat{\theta}_n\}$  una secuencia de estimadores de verosimilitud máxima y, dado un parámetro fijo  $\theta_0 \in \Theta$ , sea  $K_{\theta_0}$  un subconjunto compacto de  $\Theta$  que satisface la conclusión del Lema 4.6.1(ii). Ahora, seleccione  $\delta > 0$  lo suficientemente pequeña como para que la  $\delta$ -vecindad de  $K_{\theta_0} \cup \{\theta_0\}$  esté contenida en  $\Theta$ , esto es,

$$K := \{\nu \in \mathbb{R}^n \mid \|\nu - \theta\| \leq \delta \text{ para alguna } \theta \in K_{\theta_0} \cup \{\theta_0\}\} \subset \Theta; \quad (4.6.5)$$

note que  $K$  es un subconjunto compacto de  $\Theta$ . Como antes, sea  $\mathcal{Q}$  el conjunto de los números racionales, y sea  $p_K$  un polinomio de segundo orden que satisfaga la conclusión del Lemma 4.6.2(i) para este conjunto compacto. Ahora, defina los eventos  $\Omega_i$ ,  $i = 1, 2$ , como sigue:

$$\begin{aligned} \Omega_1 := & \left[ \lim_{n \rightarrow \infty} L_n(\nu; X_1^n) = E_{\theta_0}[\ell(\nu; X_1)] \right], \nu \in \{\theta_0\} \cup (\Theta \cap \mathcal{Q}^3) \\ & \cap \left[ \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n p_K(|X_i|) = E_{\theta_0}[p_K(|X_1|)] \right]; \end{aligned} \quad (4.6.6)$$

$$\begin{aligned} \Omega_2 := & \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} \left[ L_k(\hat{\theta}_k; X_1^k) \geq L_k(\theta; X_1^k), \theta \in \Theta \right] \\ & \cap \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} [\hat{\theta}_k \in K_{\theta_0}]. \end{aligned}$$

Note que por la ley fuerte de los grandes números  $P_{\theta_0}[\Omega_1] = 1$ , mientras que, por la Definición 4.2.1 y por el Lema 4.6.1  $P_{\theta_0}[\Omega_2] = 1$ , así para establecer el teorema es suficiente demostrar que

$$\Omega_1 \cap \Omega_2 \subset \left[ \lim_{n \rightarrow \infty} \hat{\theta}_n = \theta_0 \right]. \quad (4.6.7)$$

Para establecer esta inclusión, sea  $X_1, X_2, X_3, \dots$  una trayectoria arbitraria pero fija tal que  $\Omega_1 \cap \Omega_2$  ocurre, y note que los siguientes hechos (a)–(d) se cumplen a lo largo de esta trayectoria:

(a) Dado que  $\Omega_2$  ocurre a lo largo de  $X_1, X_2, \dots$ , existe un entero  $N > 0$  tal que

$$L_n(\theta_0; X_1^n) \leq L_n(\hat{\theta}_n; X_1^n) \quad \text{y} \quad \hat{\theta}_n \in K_{\theta_0}, \quad n > N. \quad (4.6.8)$$

Entonces, por la compacidad de  $K_{\theta_0}$ , cada subsecuencia de  $\{\hat{\theta}_n\}$  tiene un punto límite en  $K_{\theta_0}$ . Sea  $\nu^*$  un punto límite arbitrario de  $\{\hat{\theta}_n\}$ , y seleccione

una subsecuencia  $\{\hat{\theta}_{n_k}\}$  tal que

$$n_k > N \text{ para toda } k, \quad n_k \rightarrow \infty, \quad \text{y} \quad \lim_{k \rightarrow \infty} \hat{\theta}_{n_k} = \nu^* \in K_{\theta_0}. \quad (4.6.9)$$

(b) Ahora, sea  $\varepsilon \in (0, \delta)$  arbitrario, y sea  $B(\nu^*, \varepsilon)$  una bola cerrada con centro  $\nu^*$  y radio  $\varepsilon$ . En este caso, existe un entero positivo  $r$  tal que

$$\hat{\theta}_{n_k} \in B(\nu^*, \varepsilon), \quad k > r;$$

más aún,  $B(\nu^*, \varepsilon) \subset K \subset \Theta$ , puesto que  $\varepsilon < \delta$  (ver (4.6.5)). Seleccionando

$$\nu(\varepsilon) \in B(\nu^*, \varepsilon) \cap \mathcal{Q}^3 \subset \Theta \cap \mathcal{Q}^3, \quad (4.6.10)$$

se sigue que  $\|\hat{\theta}_{n_k} - \nu(\varepsilon)\| \leq 2\varepsilon$  y que el segmento que une a  $\hat{\theta}_{n_k}$  con  $\nu(\varepsilon)$  está enteramente contenido en  $B(\nu^*, \varepsilon) \subset K$  cuando  $k > r$ ; en este caso, via el teorema del valor medio, el Lema 4.6.2(i) lleva a  $|\ell(\hat{\theta}_{n_k}; x) - \ell(\nu(\varepsilon); x)| \leq 2\varepsilon p_K(|x|)$  para toda  $x \in \mathbb{R}$ . Entonces por (4.2.2),

$$|L_{n_k}(\hat{\theta}_{n_k}; X_1^{n_k}) - L_{n_k}(\nu(\varepsilon); X_1^{n_k})| \leq 2\varepsilon \frac{1}{n_k} \sum_{i=1}^{n_k} p_K(|X_i|), \quad k > r.$$

Recordando que  $n_k > N$  para toda  $k$ , esta relación y (4.6.8) llevan a

$$L_{n_k}(\theta_0; X_1^{n_k}) \leq L_{n_k}(\nu(\varepsilon); X_1^{n_k}) + 2\varepsilon \frac{1}{n_k} \sum_{i=1}^{n_k} p_K(|X_i|), \quad k > r. \quad (4.6.11)$$

(c) Puesto que el evento  $\Omega_1$  ocurre cuando se observa la trayectoria  $X_1, X_2, \dots$ , se sigue que, cuando  $n \rightarrow \infty$ , se cumple la siguiente convergencia:

$$(i) \quad n^{-1} \sum_{i=1}^n p_K(|X_i|) \rightarrow E_{\theta_0} [p_K(|X_1|)],$$

$$(ii) \quad L_n(\theta_0; X_1^n) \rightarrow E_{\theta_0} [\ell(\theta_0; X_1)], \text{ y}$$

$$(iii) \quad L_n(\nu; X_1^n) \rightarrow E_{\theta_0} [\ell(\nu; X_1)] \text{ para cada } \nu \in \Theta \cap \mathcal{Q}^3.$$

Entonces, usando el que  $\nu(\varepsilon) \in \Theta \cap \mathcal{Q}^3$ , después de tomar el límite cuando  $k$  tiende a  $\infty$ , (4.6.11) implica

$$E_{\theta_0}[\ell(\theta_0, X_1)] \leq E_{\theta_0}[\ell(\nu(\varepsilon); X_1)] + 2\varepsilon E_{\theta_0}[p_K(|X_1|)]. \quad (4.6.12)$$

(d) Debido a que  $E_{\theta_0}[p_K(|X_1|)]$  es finita, se obtiene que

$$2\varepsilon E_{\theta_0}[p_K(|X_1|)] \rightarrow 0 \quad \text{cuando } \varepsilon \searrow 0.$$

Por otra parte, (4.6.10),  $\nu(\varepsilon) \rightarrow \nu^*$  cuando  $\varepsilon \searrow 0$ , y entonces el resultado de continuidad en el Lema 4.6.2(ii) conduce a que  $\lim_{\varepsilon \searrow 0} E_{\theta_0}[\ell(\nu(\varepsilon); X_1)] = E_{\theta_0}[\ell(\nu^*; X_1)]$ . Por lo tanto, tomando el límite cuando  $\varepsilon$  decrece a 0 en (4.6.12), se sigue que  $E_{\theta_0}[\ell(\theta_0; X_1)] \leq E_{\theta_0}[\ell(\nu^*; X_1)]$ , y entonces por el Lema 4.6.2(iii),  $\nu^* = \theta_0$ . Puesto que en este argumento  $\nu^*$  es un punto límite arbitrario de  $\{\hat{\theta}_n\}$ , se sigue que

$$\lim_{n \rightarrow \infty} \hat{\theta}_n = \theta_0 \quad (4.6.13)$$

En resumen, se ha demostrado que si la trayectoria  $X_1, X_2, \dots$  es tal que si  $\Omega_1 \cap \Omega_2$  ocurre, entonces la convergencia (4.6.13) tiene lugar. Esto establece la inclusión en (4.6.7) y, como ya se mencionó, completa la demostración del Teorema 4.6.1.  $\square$

# LITERATURA CITADA

- [1] Azzalini, A. (1985). A Class of Distributions Which Includes the Normal Ones, *Scandinavian Journal of Statistics*, **12**, 171–178.
- [2] Azzalini, A., (1986). Further Results on a Class of Distributions Which Includes the Normal Ones, *Statistica*, **46**, 199–208.
- [3] Azzalini, A., Capitanio, A., (1999). Statistical Applications of the Multivariate Skew Normal Distribution, *Journal of the Royal Statistical Society, Ser. B*, **61**, 579–602.
- [4] Azzalini, A., Dalla Valle, A., (1996). The Multivariate Skew- Normal Distribution, *Biometrika*, **83**, 715–726.
- [5] Billingsley, P., (1995). *Probability and Measure*, Third Edition, Wiley, New York.
- [6] Chiogna, M., (2005). A note on the asymptotic distribution of the maximum likelihood estimator for the scalar skew-normal distribution, *Stat. Meth. & Appl.* **14**, 331–341.
- [7] Dalla Valle A., (2004). The Skew-Normal Distribution, in Genton M. G. (Ed.) *Skew Elliptical Distributions and Their Applications*, Chapman & Hall, London.
- [8] Genton M. G., (2004). *Skew Elliptical Distributions and Their Applications*, Chapman & Hall, London.
- [9] Lehmann E. L., and Casella G. (1998). *Theory of Point Estimation*, 2nd Ed., Springer, New York.
- [10] Newey W., McFadden, D., (1993). Estimation in large samples. In D. McFadden and R. Engler (eds), *Handbook of Econometrics*, Vol. 4., North-Holland, Amsterdam.

- [11] Rotnitzky, A., Cox, D. R., Bottai, M., and Robins, J., (2000), Likelihood-Based Inference With Singular Information Matrix, *Bernoulli*, **6**, 243–284.
- [12] Sartori, N., (2006). Bias prevention of maximum likelihood estimates for scalar skew normal and skew t distributions, *Journal of Statistical Planning and Inference*, **136**, 4259–4275.
- [13] Shao, J., (1999). *Mathematical Statistics*, Springer, New York.