



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE  
FACULTAD DE MATEMÁTICAS / DEPARTAMENTO DE ESTADÍSTICA

**Extensiones Asimétricas de Distribuciones  
Simétricas : Propiedades e Inferencia**

Por

**Héctor Wladimir Gómez Geraldo**

Tesis presentada a la Facultad de Matemáticas de la  
Pontificia Universidad Católica de Chile,  
como un requisito para optar al grado de  
Doctor en Estadística.

Orientador : **Reinaldo Boris Arellano Valle**

Marzo, 2004

Santiago, Chile

# Extensiones Asimétricas de Distribuciones Simétricas : Propiedades e Inferencia

Por

Héctor Wladimir Gómez Geraldo

Tesis presentada a la Facultad de Matemáticas de la  
Pontificia Universidad Católica de Chile,  
como un requisito para optar al grado de  
Doctor en Estadística.

Orientador : Reinaldo Boris Arellano Valle

Marzo, 2004

Santiago, Chile

Comisión Informante:

- Dra. Marisa Yadlin
- Dr. Heleno Bolfarine
- Dr. Manuel Galea
- Dr. Fernando Quintana
- Dr. Reinaldo Arellano

# Índice General

<b>1</b>	<b>Introducción</b>	<b>1</b>
1.1	Formulación del Problema . . . . .	1
1.2	Definición de los objetivos . . . . .	2
1.3	Resumen de los contenidos . . . . .	3
<b>2</b>	<b>Modelos Skew-Simétricos</b>	<b>5</b>
2.1	Ejemplo . . . . .	7
<b>3</b>	<b>Modelos Skew-Simétricos-Generalizados</b>	<b>10</b>
3.1	Introducción . . . . .	10
3.2	Definición y Propiedades Básicas . . . . .	11
3.3	Función de distribución . . . . .	16
3.4	Algunas Propiedades Importantes . . . . .	18
3.5	Momentos . . . . .	21

3.6	Localización y Escala . . . . .	24
3.7	Ejemplo . . . . .	26
<b>4</b>	<b>Familia General de Distribuciones Epsilon-Skew-Simétricas</b>	<b>30</b>
4.1	Introducción . . . . .	30
4.2	Familia General de Modelos Epsilon-Skew . . . . .	31
4.3	Algunos Casos Especiales . . . . .	36
<b>5</b>	<b>Inferencia Estadística en la Familia Epsilon-Skew</b>	<b>41</b>
5.1	Introducción . . . . .	41
5.2	Estimación por el Método de Momentos . . . . .	42
5.3	Estimación de Máxima Verosimilitud en el Modelo ESEP . . . . .	44
5.3.1	Ejemplos . . . . .	49
5.4	Estimación de Máxima Verosimilitud para el Modelo Epsilon-Skew-t . . . . .	53
5.4.1	Matriz de Covarianza Asintótica . . . . .	54
5.4.2	Ejemplos . . . . .	59
<b>6</b>	<b>Extensiones</b>	<b>68</b>

# Agradecimientos

Quiero dar las gracias a todos los que me dieron su apoyo para finalizar una de las tareas importantes de mi vida profesional. Muy en especial,

- Al Prof. Dr. Reinaldo Boris Arellano Valle por darme la confianza y la motivación para realizar este trabajo de investigación. También darle gracias por su paciencia y disposición que tuvo siempre durante todo el transcurso de este período.
- Al Prof. Dr. Fernando Quintana, cuyas sugerencias y contribuciones fueron fundamentales para la obtención de este trabajo, por su calidad científica y académica que tuve la oportunidad de aprender un poco más en estos años.
- Al Prof. Dr. Heleno Bolfarine, quien siempre estuvo dispuesto a la discusión de algunos resultados de esta tesis.
- Al Prof. Dr. Guido del Pino, quien en el comienzo de mis estudios en la PUC fué un apoyo fundamental para poder realizar mis estudios, como para tomar la decisión de seguir el doctorado.
- Al Profesor Nabor Castillo, quien me orientó y motivó en seguir el camino de la estadística.
- Al Departamento de Estadística de la Pontificia Universidad Católica de Chile, por la formación general que me proporciono y la oportunidad de interactuar con investigadores de otras áreas de la estadística como Pilar y Wilfredo.

- A los colegas del Departamento de Matemática de la Universidad de Atacama por su apoyo durante estos años.
- A mis amigos que encontré durante estos años en la PUC, a Mario, Junior, Carlos, Luis, Claudio, Ignacio, Oslando, Osvaldo, Rodrigo y Guillermo.
- Al Proyecto MECESUP 0103 de la Pontificia Universidad Católica de Chile por el financiamiento de esta tesis.

A mi amor Neveka y mis hijos Yoly, Josué y Paz Belén.

# Capítulo 1

## Introducción

### 1.1 Formulación del Problema

En los últimos años han aparecido modelos estadísticos paramétricos que permiten modelar datos provenientes de distribuciones asimétricos unimodales. Usualmente, este problema ha sido tratado mediante transformaciones de los datos, de modo que las observaciones transformadas pueden estudiarse mediante el modelo normal. Este procedimiento presenta al menos dos inconvenientes: el primero se relaciona con la determinación de la transformación requerida para obtener la normalidad, y el segundo con la interpretación de los resultados obtenidos a partir de los datos transformados. Estos aspectos muestran claramente la conveniencia de disponer de modelos que in-

corporen parámetros en su estructura paramétrica para modelar lo asimétrico de los datos. De esta forma Azzalini (1985) introduce el modelo skew-normal, en el cual estudia sus propiedades básicas, estimadores de máxima verosimilitud y matriz de información de Fisher, Azzalini (1986) estudia una clase general de distribuciones skew, que incluyen como casos particulares a las distribuciones skew-normal, skew-t, skew-logística, etc. Por otro lado Mudholkar y Hutson (2000) introducen otro modelo normal asimétrico que lo llaman epsilon-skew-normal. Los modelos skew-normal y epsilon-skew-normal aunque tiene distinta estructura poseen los mismos rangos de los coeficientes de asimetría y kurtosis.

El objetivo principal de esta tesis es presentar y estudiar nuevos modelos asimétricos más generales que los descritos anteriormente, es decir, construir modelos con asimetría más flexible que los conocidos.

## 1.2 Definición de los objetivos

El objetivo principal de esta tesis es desarrollar modelos estadísticos paramétricos más generales que los mencionados en la sección anterior, es decir, el trabajo se centra en la construcción de modelos skew con asimetría más flexible. De esta forma, los objetivos específicos de esta tesis pueden ser resumidos como sigue:

1. Proponer una familia de modelos más general que la familia estudiada por Azzalini(1986)
2. Ilustrar de forma unificada la inferencia estadística para los nuevos modelos skew con los de Azzalini(1986)
3. Extender el modelo epsilon-skew-normal estudiado por Mudholkar y Hutson (2000) a una clase general de distribuciones epsilon-skew
4. Realizar inferencia estadística en la nueva clase de distribuciones epsilon-skew
5. Proponer nuevas extensiones de los modelos descritos en los objetivos anteriores.

### 1.3 Resumen de los contenidos

Para cumplir los objetivos definidos anteriormente, este trabajo es organizado en cuatro capítulos, siendo los capítulos 3, 4 y 5 los más fundamentales en el cumplimiento de los objetivos propuestos. Así, el contenido de estos cuatro capítulos son resumidos como sigue.

En el capítulo 2 es definida la clase de distribuciones skew introducida por Azzalini (1985,1986) y que Gupta y otros (2001) estudia las propiedades básicas de algunos de estos modelos.

En el capítulo 3 extendemos la familia skew introducida por Azzalini (1985,1986). El motivo principal de esta extensión es flexibilizar el tipo de asimetría de los modelos de Azzalini (1985,1986). En esta nueva familia skew estudiamos sus propiedades básicas, momentos y estimadores de máxima verosimilitud. El capítulo finaliza con una aplicación a un conjunto de datos.

El capítulo 4 esta dedicado a la extensión del modelo epsilon-skew-normal introducido por Mudholkar y Hutson (2000). En este capítulo formulamos una clase general de distribuciones epsilon-skew, y estudiamos sus propiedades básicas, representación estocástica y momentos.

En el capítulo 5 esta dedicado al estudio de inferencia estadística, lo hacemos por el método de momentos en el caso general y el método de máxima verosimilitud lo aplicamos en los casos particulares de los modelos epsilon-skew-exponencial-potencia y epsilon-skew-t. También entregamos algunos resultados asintóticos de los estimadores y hacemos aplicaciones a conjuntos de datos reales.

El capítulo 6 propone extensiones de los modelos desarrollados en los capítulos anteriores. El objetivo de estas nuevas extensiones es flexibilizar aún más la asimetría de los modelos estudiados, ya que existen muchos conjuntos de datos asimétricos con asimetría distinta a los modelos paramétricos asimétricos que se conocen en la literatura actual.

## Capítulo 2

### Modelos Skew-Simétricos

La teoría de distribuciones elípticas ha sido desarrollada por Kelker (1970), Cambanis, Huang y Simons (1981), entre otros y recopilada en el texto de Fang, Kotz y Ng (1990). En el caso univariado, estas distribuciones son llamadas simétricas y se han utilizado en múltiples aplicaciones, en las cuales se debe cumplir el supuesto de simetría y para eso aplican algún tipo de transformación en los datos. Ahora para evitar este tipo de transformación en los datos, en los últimos años se han generado distintas distribuciones llamadas skew. La distribución skew-normal introducida por varios autores, por ejemplo Azzalini (1985,1986), O'Hagan y Leonard (1976), Aigner y otros (1977), Andel, Netuka y Zvára (1984). Esta distribución incluye la distribución normal y posee propiedades similares a la familia normal. Sin embargo, esta familia

es más flexible para modelar conjuntos de datos con cierta asimetría. La construcción de este modelo y de otros modelos como el skew-exponencial-potencia, skew-t, skew-logístico, etc. se basa en el siguiente lema ( ver Azzalini, 1985)

**Lema 2.1** *Sea  $f$  una función densidad simétrica alrededor de cero, y  $G$  función de distribución absolutamente continua tal que  $G'$  es simétrica alrededor de cero.*

*Entonces*

$$g(x|\lambda) = 2f(x)G(\lambda x), \quad \text{con } x \in \mathbb{R} \quad (2.1)$$

*es función densidad de la variable aleatoria  $X$  para cualquier  $\lambda$  real. La denotamos por  $X \sim Sf(\lambda)$*

**Proposición 2.1** *Usando la notación de arriba se cumplen las siguientes propiedades básicas.*

1.  $g(x|\lambda = 0) = f(x)$
2. *Cuando  $\lambda \rightarrow +\infty$ ,  $g(x|\lambda)$  tiende a la densidad  $2f(x)I\{x \geq 0\}$*
3. *Cuando  $\lambda \rightarrow -\infty$ ,  $g(x|\lambda)$  tiende a la densidad  $2f(x)I\{x < 0\}$*
4. *Si  $X \sim Sf(\lambda)$  entonces  $-X \sim Sf(-\lambda)$*
5.  $g(x|-\lambda) + g(x|\lambda) = 2f(x)$

## 2.1 Ejemplo

Consideremos a  $f = \phi$  y  $G = \Phi$ , entonces se obtiene el modelo skew-normal. Lo denotamos por  $X \sim SN(\lambda)$  y su densidad es:

$$g(x|\lambda) = 2\phi(x)\Phi(\lambda x) \quad \text{con } x \in \mathbb{R} \text{ y } \lambda \in \mathbb{R} \quad (2.2)$$

donde  $\phi$  y  $\Phi$  son la densidad  $N(0, 1)$  y su función de distribución respectivamente.

Algunas propiedades de esta distribución son:

$$E(X) = \sqrt{2/\pi} \frac{\lambda}{\sqrt{1+\lambda^2}} \quad \text{Var}(X) = 1 - \frac{2\lambda^2}{\pi(1+\lambda^2)} \quad (2.3)$$

$$\sqrt{\beta_1} = 0.5(4 - \pi) \left[ \frac{E^2(X)}{\text{Var}(X)} \right]^{3/2} \quad \beta_2 = 2(\pi - 3) \left[ \frac{E^2(X)}{\text{Var}(X)} \right]^2 \quad (2.4)$$

$$M_X(t) = 2 \exp(t^2/2) \Phi \left( \frac{t\lambda}{\sqrt{1+\lambda^2}} \right) \quad (2.5)$$

donde  $\sqrt{\beta_1}$ ,  $\beta_2$  son los coeficientes de asimetría y kurtosis respectivamente y  $M_X(t)$  la función generadora de momentos de la variable aleatoria  $X$ .

De (2.4) se tiene que :

$$-0.9953 \leq \sqrt{\beta_1} \leq 0.9953 \quad \text{y} \quad 3.0000 \leq \beta_2 \leq 3.8692 \quad (2.6)$$

En el caso univariado, por ejemplo, podemos decir que Henze (1986) entrega la representación estocástica de esta distribución, y utilizando esta representación calcula los momentos impares. Azzalini (1986) estudia modelos más generales que el modelo skew-normal. Arnold y otros (1993) utiliza el modelo skew-normal para hacer truncaciones en el modelo normal, Pewsey (2000) estudia los problemas de inferencia que tiene este modelo y Gupta y Chen (2001) entregan un test de bondad de ajuste, Liseo (1991) realiza inferencia Bayesiana, Liseo y Loperfido (2002) realizan análisis Bayesiano y Arellano-Valle, Gómez e Iglesias (2004) estudian mezclas de distribuciones skew.

El caso multivariado es considerado en Azzalini y Dalla-Valle (1996), propiedades del modelo skew-normal multivariado son estudiadas en Azzalini y Capitonio (1999), Arnold y Beaver (2000) estudian el modelo skew-cauchy, extensiones para distribuciones skew-elípticas multivariadas son derivadas por Azzalini y Capitonio (1999) y Branco y Dey (2001), Genton, He y Liu (2001) entrega los momentos de los vectores aleatorios del modelo skew-normal multivariado (ver también Loperfido (2001)), Liseo y Loperfido (2003) generalizan los resultados de O'Hagan y Leonard (1976) al caso multivariado, Arellano-Valle, del Pino y San Martín (2002) muestra que muchas de las propiedades de la distribuciones skew-normal univariada y multivariada se cumplen para una clase general de distribuciones, definidas en términos de independencia

condicional en el signo y el valor absoluto, una clase general llamada distribuciones skew fundamental es considerada en Arellano-Valle y Genton (2003)

# Capítulo 3

## Modelos

### Skew-Simétricos-Generalizados

#### 3.1 Introducción

En este capítulo introducimos una nueva familia de densidades que hace más flexible el lado que queda con poca área en la distribución skew-normal (cuando el parámetro de asimetría crece). Esta familia de densidades la llamamos skew-simétricos-generalizados y la denotamos por  $\{SfG(\lambda_1, \lambda_2); \lambda_1 \in \mathbb{R} \text{ y } \lambda_2 \geq 0\}$ .

Este capítulo se desarrolla de la siguiente manera. En la sección 3.2 entregamos la definición y propiedades básicas de esta nueva clase general de distribuciones

skew-generalizadas-simétricas y estudiamos sus propiedades básicas, en la sección 3.3 entregamos la función distribución y algunos resultados asintóticos, en la sección 3.4 entregamos algunos resultados importantes de esta clase general, en la sección 3.5 entregamos las expresiones de los momentos, en la sección 3.6 damos la extensión a una familia de localización y escala y en la sección 3.7 damos una ilustración del modelo skew-normal-generalizado y lo comparamos con el modelo skew-normal.

## 3.2 Definición y Propiedades Básicas

La construcción de esta nueva familia de distribuciones skew se basa en la extensión del Lema 2.1, el siguiente lema realiza este trabajo. Se entregan dos demostraciones, la primera es directa y la segunda utiliza un resultado del artículo de Azzalini y Capitonio (2003).

**Lema 3.1** *Sea  $f$  una función densidad simétrica alrededor de cero, y  $G$  función de distribución absolutamente continua tal que  $G'$  es simétrica alrededor de cero.*

*Entonces*

$$h(x|\lambda_1, \lambda_2) = 2f(x)G\left(\frac{\lambda_1 x}{\sqrt{1 + \lambda_2 x^2}}\right), \quad \text{con } x \in \mathbb{R} \quad (3.1)$$

*es función densidad de la variable aleatoria  $X$  para cualquier  $\lambda_1$  real y  $\lambda_2 \geq 0$ .*

*Denotamos esto por  $X \sim SfG(\lambda_1, \lambda_2)$ . Un caso especial de esta distribución es*

cuando  $\lambda_2 = \lambda_1^2$ , este caso lo llamamos skew-simétrica-curvada (SfC) y es denotado por  $X \sim SfC(\lambda_1)$ .

**Dem.1** Sea  $s(\lambda_1, \lambda_2) = \int_{-\infty}^{\infty} h(x|\lambda_1, \lambda_2)dx$ . Probemos que  $s(\lambda_1, \lambda_2) = 1 \forall \lambda_1, \lambda_2$ .

Como  $G$  es una función acotada y  $f \in L^1$ , entonces utilizando el Teorema de la Convergencia Dominada y el hecho que  $\frac{2x}{\sqrt{1+\lambda_2x^2}}f(x)G'\left(\frac{\lambda_1x}{\sqrt{1+\lambda_2x^2}}\right)$  es una función impar se tiene

$$\frac{d}{d\lambda_1}s(\lambda_1, \lambda_2) = \int_{-\infty}^{\infty} \frac{2x}{\sqrt{1+\lambda_2x^2}}f(x)G'\left(\frac{\lambda_1x}{\sqrt{1+\lambda_2x^2}}\right)dx = 0, \text{ entonces } s(\lambda_1, \lambda_2) = cte.$$

con respecto a  $\lambda_1$ , y como  $s(0, \lambda_2) = 1$ , luego tenemos que  $s(\lambda_1, \lambda_2) = 1 \forall \lambda_1, \lambda_2$ .  $\square$

**Dem.2** Sea  $W(X) = \frac{\lambda_1x}{\sqrt{1+\lambda_2x^2}}$  entonces podemos observar que  $W$  es una función impar, luego aplicando la Proposición 1 del artículo de Azzalini y Capitanio (2003) se tiene que  $h(x|\lambda_1, \lambda_2)$  es una densidad.  $\square$

**Comentarios 3.1** En el caso en que  $f = \phi$  y  $G = \Phi$ , se tiene la distribución Skew-normal-Generalizada que fue introducida por Arellano, Gómez y Quintana (2004b), en la construcción de este modelo se utilizó el siguiente resultado.

Si  $Z \sim N(\mu, \sigma^2)$  entonces

$$E[\Phi(Z)] = \Phi\left(\frac{\mu}{\sqrt{1+\sigma^2}}\right) \quad (3.2)$$

este resultado es entregado por Ellison (1964). También se prueba en este artículo que la distribución skew-normal-generalizada es una mezcla entre una distribución skew-

normal y una normal en el parámetro de asimetría, es decir, si  $X | Y = y \sim SN(y)$  e  $Y \sim N(\lambda_1, \lambda_2)$ , entonces  $X \sim SNG(\lambda_1, \lambda_2)$ . Utilizando esta representación podemos generar muestras aleatorias de la distribución skew-normal-generalizada, para ello se hace uso del método de composición (ver Tanner) y la librería de S-PLUS (Sn) de Azzalini para la densidad skew-normal que se encuentra ubicado en su página web.

<http://azzalini.stat.unipd.it>

La figura 3.1 muestra las formas que tiene la densidad SNG para distintos valores de  $(\lambda_1, \lambda_2)$ .

La figura 3.2 compara los modelos  $SNC(\lambda_1)$  y  $SN(\lambda_1)$  para distintos valores de  $\lambda_1$ .

**Proposición 3.1** *Teniendo en cuenta la notación anterior se cumplen las siguientes propiedades básicas.*

1.  $h(x | \lambda_1 = 0, \lambda_2) = f(x)$  y  $h(x | \lambda_1, \lambda_2 = 0) = g(x | \lambda_1)$
2. Cuando  $\lambda_1 \rightarrow \pm\infty$ ,  $h(x | \lambda_1, \lambda_1^2)$  tiende a la densidad  $2f(x)G(\pm \text{sgn}(x))$
3. Si  $X \sim SfG(\lambda_1, \lambda_2)$  entonces  $-X \sim SfG(-\lambda_1, \lambda_2)$
4.  $h(x | -\lambda_1, \lambda_2) + h(x | \lambda_1, \lambda_2) = 2f(x)$

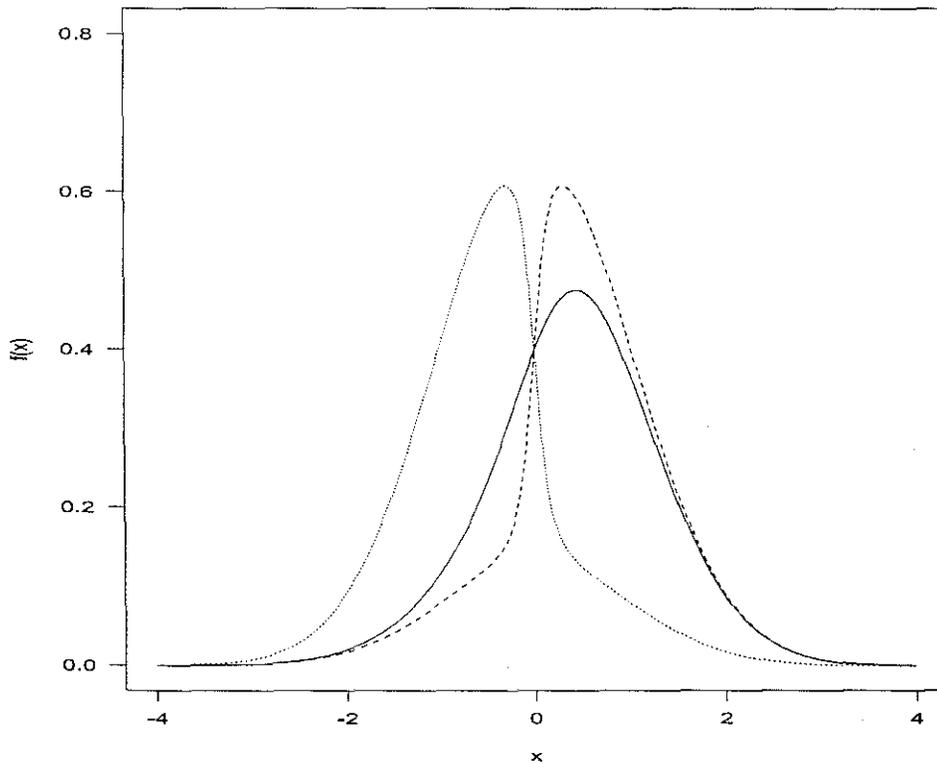


Figura 3.1: Ejemplos de la densidad  $SNG(\lambda_1, \lambda_2)$  para  $(\lambda_1, \lambda_2) = (-5, 20)$  (línea punteada),  $(\lambda_1, \lambda_2) = (1, 1)$  (línea sólida),  $(\lambda_1, \lambda_2) = (10, 40)$  (línea discontinua)

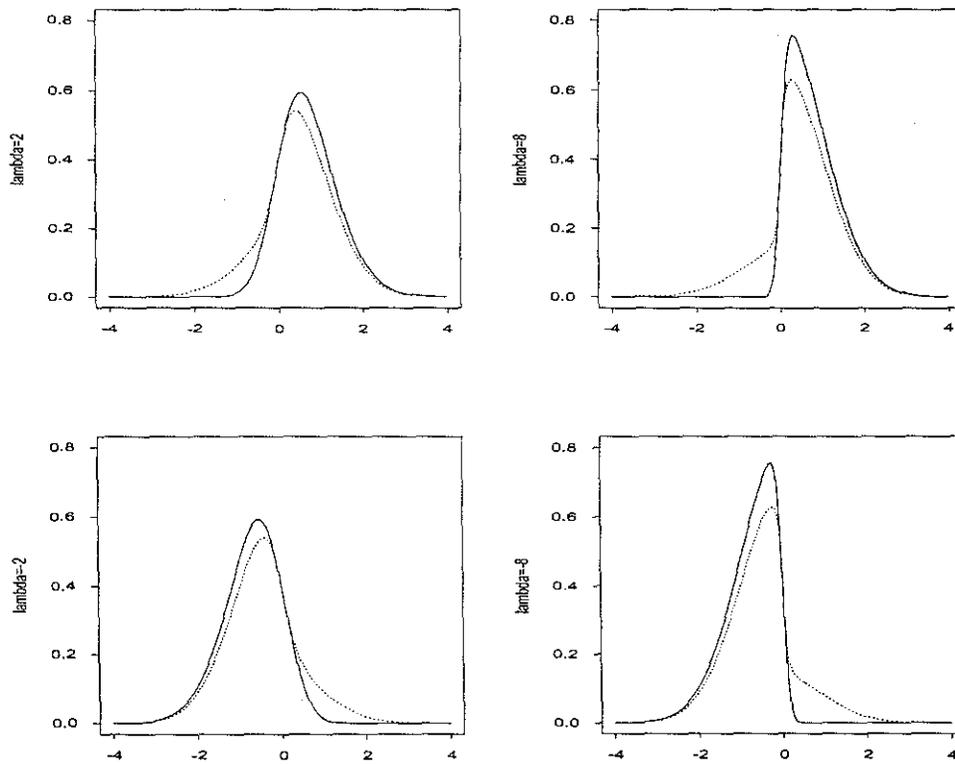


Figura 3.2: Los modelos  $SNC(\lambda_1)$  (línea punteada) y  $SN(\lambda_1)$  (línea sólida) para distintos valores de  $\lambda_1$ .

### 3.3 Función de distribución

La función de distribución de (3.1) la denotamos por

$$H(x|\lambda_1, \lambda_2) = \int_{-\infty}^x 2f(t)G\left(\frac{\lambda_1 t}{\sqrt{1 + \lambda_2 t^2}}\right) dt \quad (3.3)$$

**Proposición 3.2** *Utilizando (3.3) se cumplen las siguientes propiedades:*

1.  $H(x|\lambda_1 = 0, \lambda_2) = F(x)$ , donde  $F(x)$  es la fda de  $f(x)$ .

2.  $H(-x|\lambda_1, \lambda_2) = 1 - H(x; -\lambda_1, \lambda_2)$

3.  $\lim_{\lambda_1 \rightarrow +\infty} H(x|\lambda_1, \lambda_1^2) = \begin{cases} 2(1 - G(1))G(x) & \text{si } x < 0 \\ 1 - 2G(1)(1 - G(x)) & \text{si } x \geq 0 \end{cases}$

4.  $\lim_{\lambda_1 \rightarrow -\infty} H(x|\lambda_1, \lambda_1^2) = \begin{cases} 2G(1)G(x) & \text{si } x < 0 \\ 2[G(1) + (1 - G(1))G(x) - 1/2] & \text{si } x \geq 0 \end{cases}$

La tabla 3.1 muestra que la aproximación asintótica a  $H(x|\lambda_1, \lambda_1^2)$ , en el caso normal, se obtiene haciendo  $\lambda_1 \pm \infty$  para valores moderados de  $\lambda_1$ .

La tabla 3.2 muestra que la distribución  $SNC(\lambda_1, \lambda_1^2)$  acumula más área al lado izquierdo de cero que el modelo  $SN(\lambda_1)$ .

**Comentarios 3.2** *En la Proposición 3.2 se establece que la motivación básica de la construcción de la familia Sfc es por ejemplo, cuando el  $\lambda_1$  crece positivamente*

$\lambda_1$	$H(x \lambda_1, \lambda_1^2)$		
	$x = -2$	$x = -1$	$x = -0.5$
1	0.00815	0.06580	0.14974
2	0.00747	0.05502	0.11722
3	0.00733	0.05251	0.10748
5	0.00726	0.05114	0.10158
10	0.00723	0.05054	0.09850
$\lim_{\lambda_1 \rightarrow \infty} H(x \lambda_1, \lambda_1^2)$	0.007219	0.05034	0.0979

Tabla 3.1: Valores exactos de la fda  $H(x|\lambda_1, \lambda_1^2)$  para algunos valores de  $\lambda_1$ , junto con la aproximación asintótica dada por la Proposición 3.2

$\lambda_1$	$H_X(0)$	$H_Y(0)$
1	0.25	0.306
2	0.148	0.249
3	0.102	0.224
5	0.062	0.200
10	0.032	0.181
15	0.021	0.174

Tabla 3.2: Probabilidades acumuladas al lado izquierdo de 0 para variables aleatorias  $X \sim SN(\lambda_1)$   $Y \sim SNC(\lambda_1, \lambda_1^2)$  para valores diferentes de  $\lambda_1$ .

la densidad tiene la cola más pesada, en el lado izquierdo, que la familia  $Sf$  y la distribución límite es distinta a la half- $f$ , que es la distribución límite de la familia  $Sf$ .

Las familias  $\{Sf(\lambda_1); \lambda_1 \in \mathbb{R}\}$  y  $\{SfC(\lambda_1); \lambda_1 \in \mathbb{R}\}$  son identificables y ambas contienen a la familia simétrica univariada estándar  $f$  cuando  $\lambda_1 = 0$ .

### 3.4 Algunas Propiedades Importantes

Derivamos algunas propiedades importantes de la clase general  $SfG$ . Se muestra por ejemplo que, al igual que la familia definida en (2.1), la distribución del valor

absoluto de una variable  $SfG$  es half-simétrica. En particular, esto implica que todos los momentos de la familia son finitos, y los momentos pares coinciden con los de la distribución simétrica  $f$ .

**Proposición 3.3** Sean  $X \sim SfG(\lambda_1, \lambda_2)$  y  $w$  una función par, es decir  $w(-x) = w(x)$ , entonces la distribución de  $w(X)$  no depende de  $\lambda$ .

**Dem:** Sean  $Y = w(X)$  y  $M_Y(t)$  la función generadora de momentos de la variable  $Y$ , luego utilizando las propiedades 4 y 5 de la Proposición 3.1, y el hecho que  $w$  es par se tiene

$$\begin{aligned}
 M_Y(t) &= E(e^{tY}) = \int_{-\infty}^{\infty} e^{tw(x)} h(x | \lambda_1, \lambda_2) dx \\
 &= \int_{-\infty}^0 e^{tw(x)} h(x | \lambda_1, \lambda_2) dx + \int_0^{\infty} e^{tw(x)} h(x | \lambda_1, \lambda_2) dx \\
 &= \int_0^{\infty} e^{tw(x)} h(x | -\lambda_1, \lambda_2) dx + \int_0^{\infty} e^{tw(x)} h(x | \lambda_1, \lambda_2) dx \\
 &= \int_0^{\infty} 2e^{tw(x)} f(x) dx.
 \end{aligned}$$

observamos que la función generadora de momentos no depende de  $\lambda$ .  $\square$

**Proposición 3.4** Sean  $X \sim f$  e  $Y \sim SfG(\lambda_1, \lambda_2)$  y  $r$  una función par, entonces  $r(X)$  y  $r(Y)$  tienen la misma distribución.

**Dem:** Por la proposición 3.3 se tiene que

$$M_{r(Y)}(t) = \int_0^{\infty} 2e^{r(y)t} f(y) dy$$

y

$$\begin{aligned}M_{r(X)}(t) &= \int_{-\infty}^{\infty} e^{r(x)t} f(x) dx \\ &= \int_{-\infty}^0 e^{r(x)t} f(x) dx + \int_0^{\infty} e^{r(x)t} f(x) dx \\ &= \int_0^{\infty} 2e^{r(x)t} f(x) dx.\end{aligned}$$

ya que  $r$  y  $f$  son funciones pares.  $\square$

**Proposición 3.5** *Sea  $X \sim SfG(\lambda_1, \lambda_2)$  y  $Y \sim f$ . Entonces  $|X|$  y  $|Y|$  son idénticamente distribuidas, es decir  $|X| \stackrel{d}{=} |Y| \sim Hf(0, 1)$ , donde  $Hf(0, 1)$  denota la distribución half- $f$  estándar.*

**Dem:** Por la Proposición 3.4 se tiene que  $|X|$  y  $|Y|$  son idénticamente distribuidas, ya que  $r = |\cdot|$  es una función par. Ahora por otro lado es conocido que  $Z = |Y|$  tiene densidad  $2f(z)I\{z \geq 0\}$ , también sabemos por la Proposición 3.3 que  $|X|$  no depende de  $\lambda$ , luego encontramos su distribución, sea  $W = |X|$  entonces tenemos

$$\begin{aligned}f_W(w) &= h(w|\lambda_1, \lambda_2) + h(-w|\lambda_1, \lambda_2) \\ &= 2f(w)G\left(\frac{\lambda_1 w}{\sqrt{1 + \lambda_2 w^2}}\right) + 2f(-w)G\left(\frac{-\lambda_1 w}{\sqrt{1 + \lambda_2 w^2}}\right) \\ &= 2f(w) \quad \text{con } w > 0,\end{aligned}$$

que es la distribución half- $f$  estándar.  $\square$

### 3.5 Momentos

Tomando  $r(x) = x^k$  para  $k$  par en la Proposición 3.4 concluimos que los momentos pares de las familias  $SfG$  y  $Sf$  son los mismos. Esto también implica la existencia de los momentos impares, pero estos no tienen expresión explícita para una determinada densidad  $f$ . Podemos dar una expresión implícita, y para eso utilizamos la siguiente fórmula. Para  $k = 0, 1, 2, \dots$  se tiene

$$a_k(\lambda_1, \lambda_2) = \int_0^\infty u^k f(\sqrt{u}) G\left(\frac{\lambda_1 \sqrt{u}}{\sqrt{1 + \lambda_2 u}}\right) du. \quad (3.4)$$

Note que para todo  $\lambda_2 \geq 0$  se tiene  $a_k(0, \lambda_2) = a_k(0, 0)$ .

**Proposición 3.6** Si  $X \sim SfG(\lambda_1, \lambda_2)$  entonces para  $k = 0, 1, 2, \dots$  se tiene

$$E(X^{2k+1}) = 2\{a_k(\lambda_1, \lambda_2) - a_k(0, 0)\}. \quad (3.5)$$

**Dem:**

$$\begin{aligned} E(X^{2k+1}) &= 2 \int_0^\infty 2x^{2k+1} f(x) G\left(\frac{\lambda_1 x}{\sqrt{1 + \lambda_2 x^2}}\right) dx - 2 \int_0^\infty x^{2k+1} f(x) dx \\ &= 2a_k(\lambda_1, \lambda_2) - 2a_k(0, 0). \end{aligned}$$

factorizando por dos se tiene el resultado.  $\square$

El siguiente lema se utiliza para relacionar los momentos de las familias  $SfG$  y  $Sf$ .

**Lema 3.2** Para cualquier  $\lambda_1 > 0$ ,  $\lambda_2 > 0$  y  $k = 0, 1, 2, \dots$

$$a_k(0, 0) \leq a_k(\lambda_1, \lambda_2) \leq 2a_k(0, 0) G\left(\frac{\lambda_1}{\sqrt{\lambda_2}}\right) \quad (3.6)$$

y

$$a_k(0, 0) \leq a_k(\lambda_1, \lambda_2) \leq a_k(\lambda_1, 0) \leq 2a_k(0, 0). \quad (3.7)$$

**Dem:** Ya que  $G$  es una función creciente, tenemos que para cualquier  $\lambda_1 > 0$  y  $u > 0$  se tiene que

$$\frac{1}{2} \leq G\left(\frac{\lambda_1\sqrt{u}}{\sqrt{1+\lambda_2u}}\right) = G\left(\frac{\lambda_1}{\sqrt{\lambda_2}} \frac{\sqrt{\lambda_2u}}{\sqrt{1+\lambda_2u}}\right) \leq G\left(\frac{\lambda_1}{\sqrt{\lambda_2}}\right).$$

El argumento para mostrar (3.7) es similar, pero usamos

$$\frac{1}{2} \leq G\left(\frac{\lambda_1\sqrt{u}}{\sqrt{1+\lambda_2u}}\right) \leq G(\lambda_1\sqrt{u}).$$

Se puede ver fácilmente que las desigualdades (3.6) y (3.7) son inversas cuando  $\lambda_1 < 0$ . Por lo tanto, combinando (3.5) y (3.6) se sigue que  $E(X) > 0$  si y sólo si  $\lambda_1 > 0$ .  $\square$

Esta propiedad es generalizada en la siguiente proposición.

**Proposición 3.7** Si  $X \sim SfG(\lambda_1, \lambda_2)$  e  $Y \sim Sf(\lambda_1)$  entonces

$$|E(X^r)| \leq |E(Y^r)| \quad \forall \lambda_1 \in \mathbb{R}, \lambda_2 > 0 \text{ y } r = 1, 2, \dots$$

**Dem:** La igualdad es trivial para todo  $r$  cuando  $\lambda_1 = 0$ . Cuando  $r$  es par se puede ver fácilmente desde la Proposición 3.4 que  $E(X^r) = E(Y^r)$ . Cuando  $r$  es impar, consideramos primero el caso cuando  $\lambda_1 > 0$ . Por (3.5) se tiene  $E(X^r) =$

$2\{a_{2(r-1)}(\lambda_1, \lambda_2) - a_{2(r-1)}(0, 0)\}$  y  $E(Y^r) = 2\{a_{2(r-1)}(\lambda_1, 0) - a_{2(r-1)}(0, 0)\}$  y el resultado se sigue desde (3.7). Finalmente, el caso cuando  $\lambda_1 < 0$  es muy similar.

□

**Comentarios 3.3** *Por la Proposición 3.7 observamos que la varianza de la familia  $Sf$  es menor que la varianza de la familia  $SfG$ , y ésta menor que la varianza de la familia simétrica estándar  $f$ .*

*Por otro lado, por ejemplo en el caso skew-normal-curvado, los coeficientes de asimetría y kurtosis se encuentran en los siguientes intervalos:*

$$-0.375 \leq \sqrt{\beta_1} \leq 0.375 \quad \text{y} \quad 3.0000 \leq \beta_2 \leq 4.331 \quad (3.8)$$

*Podemos observar que el intervalo de asimetría de la distribución skew-normal es más ancho, no así el intervalo de kurtosis que es más angosto que el de la skew-normal-curvada. Así podemos utilizar estos modelos para conjuntos de datos en los cuales unos de lados tiene más área que el de la skew-normal, es decir, tiene más kurtosis. Además la asimetría de la familia  $SfG$  es menor o igual a la asimetría de la familia  $Sf$ , lo interesante de la familia  $SfG$  es que puede capturar distintos tipos de asimetría, además del tipo de asimetría que modela bien la familia  $Sf$ .*

### 3.6 Localización y Escala

En la siguiente definición hacemos una extensión de la familia  $SfG$  a una familia de localización y escala.

**Definición 3.1** Sea  $X \sim SfG(\lambda_1, \lambda_2)$ . La familia  $SfG$  de localización y escala, es definida como la distribución de  $Z = \mu + \sigma X$  para  $\mu \in \mathbb{R}$  y  $\sigma > 0$ . La correspondiente densidad es dada por:

$$h(z|\boldsymbol{\theta}) = \frac{2}{\sigma} f\left(\frac{z-\mu}{\sigma}\right) G\left[\frac{\lambda_1(z-\mu)}{\sqrt{\sigma^2 + \lambda_2(z-\mu)^2}}\right], \quad (3.9)$$

donde  $\boldsymbol{\theta} = (\mu, \sigma, \lambda_1, \lambda_2)$  y la denotamos por  $Z \sim SfG(\mu, \sigma, \lambda_1, \lambda_2)$ .

**Proposición 3.8** Sean  $Z|Y=y \sim SfC(\mu, \sigma, y)$  e  $Y \sim \frac{1}{\tau}f(y/\tau)$ . Entonces:

$$(a) Z \sim \frac{1}{\sigma}f\left(\frac{z-\mu}{\sigma}\right) \quad (b) Y|Z=z \sim SfC(0, \tau^2, \frac{\tau}{\sigma}(z-\mu)).$$

Dem:(a)

$$\begin{aligned} h_Z(z) &= \frac{2}{\sigma\tau} \int_{-\infty}^{\infty} f\left(\frac{z-\mu}{\sigma}\right) G\left(\frac{y(z-\mu)}{\sqrt{\sigma^2 + y^2(z-\mu)^2}}\right) f\left(\frac{y}{\tau}\right) dy \\ &= \frac{1}{\sigma} f\left(\frac{z-\mu}{\sigma}\right) \int_{-\infty}^{\infty} 2f(t) G\left(\frac{\tau(z-\mu)t}{\sqrt{\sigma^2 + \tau^2(z-\mu)^2 t^2}}\right) dt \\ &= \frac{1}{\sigma} f\left(\frac{z-\mu}{\sigma}\right). \end{aligned}$$

(b)

$$\begin{aligned} h_{Y|Z}(y) &= \frac{\frac{2}{\sigma\tau} f\left(\frac{z-\mu}{\sigma}\right) G\left(\frac{y(z-\mu)}{\sqrt{\sigma^2+y^2(z-\mu)^2}}\right) f\left(\frac{y}{\tau}\right)}{\frac{1}{\sigma} f\left(\frac{z-\mu}{\sigma}\right)} \\ &= \frac{2}{\tau} f\left(\frac{y}{\tau}\right) G\left(\frac{\frac{y}{\tau}[\tau(z-\mu)]}{\sqrt{\sigma^2+\frac{y^2}{\tau^2}[\tau(z-\mu)]^2}}\right) \end{aligned}$$

que es una distribución skew-simétrica-curvada.  $\square$

**Proposición 3.9** Sean  $Z|(Y_1 = y_1, Y_2 = y_2) \sim SfG(\mu, \sigma, y_1, y_2)$ , donde la distribución conjunta de  $(Y_1, Y_2)$  es definido como  $Y_1|Y_2 = y_2 \sim \frac{1}{\tau(y_2)}f(y_1/\tau(y_2))$  e  $Y_2 \sim f(y_2)$ . Entonces  $Z$  y  $Y_2$  son independientes,  $Z \sim \frac{1}{\sigma}f\left(\frac{z-\mu}{\sigma}\right)$ , y

$$h_{Y_1|Z}(y_1) = \frac{2}{\tau(y_2)} \int_0^\infty f(y_1/\tau(y_2)) G\left(\frac{\tau(y_2)(z-\mu)}{\sqrt{\sigma^2+y_2(z-\mu)^2}} \frac{y_1}{\tau(y_2)}\right) f(y_2) dy_2$$

**Dem:** Note primero que

$$\begin{aligned} h_{Z|Y_2}(z) &= \int_{-\infty}^\infty \frac{2}{\sigma} f\left(\frac{z-\mu}{\sigma}\right) G\left(\frac{y_1(z-\mu)}{\sqrt{\sigma^2+y_2(z-\mu)^2}}\right) \frac{1}{\tau(y_2)} f(y_1/\tau(y_2)) dy_1 \\ &= \frac{1}{\sigma} f\left(\frac{z-\mu}{\sigma}\right) \int_{-\infty}^\infty 2f(t) G\left(\frac{\tau(y_2)(z-\mu)t}{\sqrt{\sigma^2+y_2(z-\mu)^2}}\right) dt \\ &= \frac{1}{\sigma} f\left(\frac{z-\mu}{\sigma}\right) \end{aligned}$$

Como esta condicional no depende de  $y_2$ , tenemos que  $Z$  y  $Y_2$  son independientes y también que  $Z \sim \frac{1}{\sigma}f\left(\frac{z-\mu}{\sigma}\right)$ . Es fácil ver que

$$h_{Y_1|Z, Y_2}(y_1) = \frac{2}{\tau(y_2)} f(y_1/\tau(y_2)) G\left(\frac{\tau(y_2)(z-\mu)}{\sqrt{\sigma^2+y_2(z-\mu)^2}} \frac{y_1}{\tau(y_2)}\right),$$

es decir,

$$Y_1 | (Y_2 = y_2, Z = z) \sim SGf \left( 0, \tau^2(y_2); \frac{\tau(y_2)(z - \mu)}{\sqrt{\sigma^2 + y_2(z - \mu)^2}} \right)$$

y el resto de la demostración usa el hecho que  $Z$  y  $Y_2$  son independientes, esto implica que

$$h_{Y_1|Z}(y_1) = \int h_{Y_1|Z,Y_2}(y_1) f_{Y_2}(y_2) dy_2.$$

luego utilizando  $h_{Y_1|Z,Y_2}(y_1)$  en la última expresión se tiene el resultado.  $\square$

### 3.7 Ejemplo

Consideremos un conjunto de datos que corresponde a alturas (en centímetros) de 100 atletas Australianos. El conjunto de datos se encuentra en Cook y Weisberg (1994) y estan disponibles en <http://stat.umidp.it/SN/index.html>

En la tabla 3.3 presentamos un resumen de estadísticas descriptivas básicas.

Estimamos los parámetros numéricamente maximizando la función log-verosimilitud

$$\log L = -\frac{n}{2} \log \frac{\pi \sigma^2}{2} - \frac{1}{2\sigma^2} \sum_{i=1}^n (z_i - \mu)^2 + \sum_{i=1}^n \log \Phi \left[ \frac{\lambda_1(z_i - \mu)}{\sqrt{\sigma^2 + \lambda_2(z_i - \mu)^2}} \right]$$

$n$	$\bar{Z}$	$S^2$	$\gamma_1$	$\gamma_2$
100	174.594	67.934	-0.568	4.321

Tabla 3.3: *Estadísticas descriptivas para la alturas de 100 atletas Australiano, donde  $\gamma_1$  y  $\gamma_2$  representan los coeficientes de asimetría y kurtosis muestral respectivamente.*

con respecto a los componentes de  $\theta = (\mu, \sigma, \lambda_1, \lambda_2)$ . Los resultados son resumidos en la tabla 3.4, considerando el modelo  $SNG(\mu, \sigma, \lambda_1, \lambda_2)$  y dos casos especiales: normal y skew-normal.

Parámetros Estimados	N	SN	SNG
$\mu$	174.594000	174.583808	170.320485
$\sigma^2$	67.254564	67.254561	85.517551
$\lambda_1$	—	0.001559	4.380545
$\lambda_2$	—	—	24.183754
log-verosimilitud	-352.3181	-352.3181	-347.2392

Tabla 3.4: *Estimación paramétrica de máxima verosimilitud para las alturas de los atletas Australianos bajo el modelo SNG y dos sub-modelos particulares:  $Normal(\lambda_1 = \lambda_2 = 0)$ ,  $skew-normal(\lambda_2 = 0)$ .*

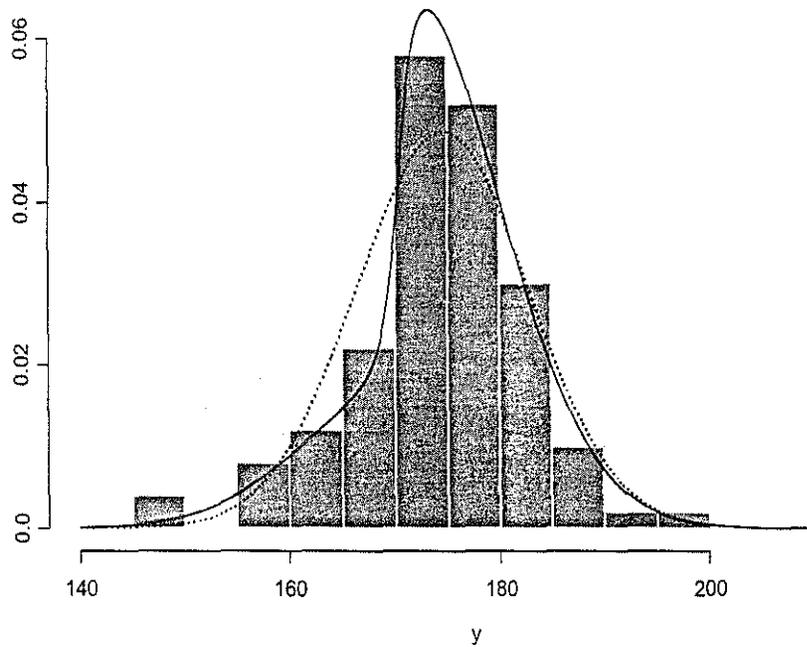


Figura 3.3: *Histograma de las 100 alturas de atletas Australianos. Las líneas representan ajustes de distribuciones usando la estimación de máxima verosimilitud:  $SNG(\hat{\mu}, \hat{\sigma}, \hat{\lambda}_1, \hat{\lambda}_2)$  (línea sólida) y  $SN(\hat{\mu}, \hat{\sigma}, \hat{\lambda}_1)$  (línea punteada)*

## Capítulo 4

### Familia General de Distribuciones

### Epsilon-Skew-Simétricas

#### 4.1 Introducción

El siguiente capítulo es organizado como sigue. En la Sección 4.2 entregamos la definición de la nueva familia, como también la extensión a una familia de localización y escala. Obtenemos la representación estocástica, la cual facilita la entrega de los momentos de toda la familia y los coeficientes de asimetría y kurtosis. En la Sección 4.3 aplicamos esta construcción a algunos casos especiales, tales como las distribuciones exponencial-potencia,  $t$  y logística.

## 4.2 Familia General de Modelos Epsilon-Skew

La idea básica es entregada en Fechner (1897) y es formar una distribución juntando en  $x = 0$  dos distribuciones half-normal con diferentes parámetros de escala. Esto es crucial para el desarrollo presentado en Mudholkar y Hutson (2000). Extendemos esta idea en la siguiente definición

**Definición 4.1** *Sea  $f$  una densidad simétrica. La clase de distribuciones epsilon-skew es indexada por el parámetro de asimetría  $|\varepsilon| < 1$  y tiene densidad definida por*

$$h(x|\varepsilon) = f\left(\frac{x}{1+\varepsilon}\right) I\{x < 0\} + f\left(\frac{x}{1-\varepsilon}\right) I\{x \geq 0\}. \quad (4.1)$$

*Si  $X$  tiene densidad (4.1) decimos que  $X$  es una variable aleatoria epsilon-skew y la denotamos por  $X \sim ES(f, \varepsilon)$ , o alternativamente, por  $X \sim ESf(\varepsilon)$ .*

**Comentarios 4.1** *La definición 4.1 puede ser reformulada como sigue. Sea  $f^*$  una densidad con soporte en  $\mathbb{R}^+$  y  $F^*$  su fda. Entonces (4.1) podemos escribir*

$$h^*(x|\varepsilon) = \frac{1}{2}f^*\left(-\frac{x}{1+\varepsilon}\right) I\{x < 0\} + \frac{1}{2}f^*\left(\frac{x}{1-\varepsilon}\right) I\{x \geq 0\}. \quad (4.2)$$

*Si  $f^* = 2f$ , con  $f(x) = f(-x)$  para  $x \in \mathbb{R}$  entonces tenemos (4.1) desde (4.2).*

*Recíprocamente, si*

$$f(x) = h^*(x|0) = \frac{1}{2}(f^*(-x)I\{x < 0\} + f^*(x)I\{x \geq 0\}),$$

se tiene (4.2) en terminos de (4.1).

Desde (4.1), denotamos a  $F$  como la fda de  $f$  y obtenemos inmediatamente la fda de la familia de distribuciones epsilon-skew

$$H(x|\varepsilon) = \begin{cases} (1 + \varepsilon)F\left(\frac{x}{1+\varepsilon}\right) & \text{si } x < 0 \\ \varepsilon + (1 - \varepsilon)F\left(\frac{x}{1-\varepsilon}\right) & \text{si } x \geq 0 \end{cases} \quad (4.3)$$

Por un simple cálculo obtenemos

$$H^{-1}(x|\varepsilon) = \begin{cases} (1 + \varepsilon)F^{-1}\left(\frac{x}{1+\varepsilon}\right) & \text{si } 0 < x < (1 + \varepsilon)/2 \\ (1 - \varepsilon)F^{-1}\left(\frac{x-\varepsilon}{1-\varepsilon}\right) & \text{si } (1 + \varepsilon)/2 \leq x \leq 1 \end{cases} \quad (4.4)$$

Luego la mediana de la familia epsilon-skew es

$$H^{-1}(1/2|\varepsilon) = \begin{cases} (1 + \varepsilon)F^{-1}\left(\frac{1}{2(1+\varepsilon)}\right) & \text{si } \varepsilon > 0 \\ (1 - \varepsilon)F^{-1}\left(\frac{1-2\varepsilon}{2(1-\varepsilon)}\right) & \text{si } \varepsilon \leq 0 \end{cases} \quad (4.5)$$

Algunas propiedades básicas de la  $ES(f, \varepsilon)$  son entregadas en la siguiente proposición.

**Proposición 4.1** *Utilizando la notación entregada anteriormente, se cumplen las siguientes propiedades:*

1.  $h(x|\varepsilon = 0) = f(x)$  y  $H(0|\varepsilon) = \frac{1+\varepsilon}{2}$ .

2. Si  $X \sim ES(f, \varepsilon)$  entonces

$$\frac{P(X \geq 0|\varepsilon)}{P(X < 0|\varepsilon)} = \frac{1 - \varepsilon}{1 + \varepsilon}.$$

3. Si  $X \sim ES(f, \varepsilon)$  entonces  $-X \sim ES(f, -\varepsilon)$ .

4.  $\lim_{\varepsilon \rightarrow 1} h(x|\varepsilon) = f(x/2)I\{x < 0\}$  y  $\lim_{\varepsilon \rightarrow -1} h(x|\varepsilon) = f(x/2)I\{x \geq 0\}$ .

El primer resultado de la Proposición 4.1 establece que esta nueva familia contiene a la familia de distribuciones simétricas continuas univariadas cuando  $\varepsilon = 0$ .

En la siguiente Proposición mostramos que una variable aleatoria  $X$  con función densidad dada por (4.1) puede ser representada como el producto de dos variables aleatorias independientes.

**Proposición 4.2** Para cualquier  $|\varepsilon| < 1$  y densidad simétrica  $f$  tenemos que  $X \sim ESf(\varepsilon)$  si y sólo si existen dos variables aleatorias independientes  $V$  y  $U_\varepsilon$  con  $V \sim 2f(x)I\{x \geq 0\}$  y  $P(U_\varepsilon = 1 - \varepsilon) = (1 - \varepsilon)/2$ ,  $P(U_\varepsilon = -(1 + \varepsilon)) = (1 + \varepsilon)/2$  tal que  $X = U_\varepsilon V$ .

**Dem:** La condición de suficiencia puede obtenerse directamente calculando la función densidad de  $U_\varepsilon V$ . Para mostrar la condición necesaria, definimos

$$U_\varepsilon = \begin{cases} 1 - \varepsilon & \text{si } X \geq 0 \\ -(1 + \varepsilon) & \text{si } X < 0 \end{cases} \quad V = \begin{cases} \frac{X}{1 - \varepsilon} & \text{si } X \geq 0 \\ \frac{-X}{1 + \varepsilon} & \text{si } X < 0 \end{cases}$$

Entonces se tiene que  $U_\varepsilon$  y  $V$  son independientes y tienen sus distribuciones marginales apropiadas.  $\square$

**Comentarios 4.2** La representación estocástica entregada en la Proposición 4.2 y su demostración muestran una alternativa de construir variables aleatorias  $X \sim ES(f, \varepsilon)$ : generadas simplemente de  $Y \sim f$ , y  $U_\varepsilon$  con  $P(U_\varepsilon = 1 - \varepsilon) = (1 - \varepsilon)/2 = 1 - P(U_\varepsilon = -(1 + \varepsilon))$  independiente de  $Y$ , y el conjunto  $X = U_\varepsilon|Y|$ . También podemos reescribir  $U = U_\varepsilon/|U_\varepsilon| = X/|X|$ , donde  $P(U = 1) = \frac{1-\varepsilon}{2} = 1 - P(U = -1)$ , y haciendo  $R = |U_\varepsilon|V = |X|$  obtenemos  $X = RU$ , con  $R$  y  $U$  independientes.

En la siguiente definición extendemos esta clase a una familia de localización y escala.

**Definición 4.2** Sea  $X \sim ES(f, \varepsilon)$ . La familia de distribuciones epsilon-skew de localización y escala, es definida como la distribución de  $Z = \mu - \sigma X$  para  $x \in \mathbb{R}$  y  $\sigma > 0$ . La correspondiente densidad es dada por

$$h(z|\boldsymbol{\theta}) = \frac{1}{\sigma} f\left(\frac{z - \mu}{\sigma(1 + \varepsilon)}\right) I\{z < \mu\} + \frac{1}{\sigma} f\left(\frac{z - \mu}{\sigma(1 - \varepsilon)}\right) I\{z \geq \mu\}, \quad (4.6)$$

donde  $\boldsymbol{\theta} = (\mu, \sigma, \varepsilon)$  y denotamos esto por  $Z \sim ES(f, \mu, \sigma, \varepsilon)$ , o  $Z \sim ESf(\mu, \sigma, \varepsilon)$  o  $Z \sim ESf(\boldsymbol{\theta})$ .

Propiedades básicas de la familia de distribuciones epsilon-skew de localización y escala pueden ser obtenidas de la misma forma como antes. Por ejemplo, la mediana

es dada por

$$H^{-1}(1/2|\varepsilon) = \begin{cases} \mu + \sigma(1 + \varepsilon)F^{-1}\left(\frac{1}{2(1+\varepsilon)}\right) & \text{si } \varepsilon > 0 \\ \mu + \sigma(1 - \varepsilon)F^{-1}\left(\frac{1-2\varepsilon}{2(1-\varepsilon)}\right) & \text{si } \varepsilon \leq 0. \end{cases} \quad (4.7)$$

Utilizando la representación estocástica dada en la Proposición 4.2 es fácil calcular los momentos de las distribuciones epsilon-skew. Para  $r \in \{1, 2, \dots\}$  denotemos

$$a_r = \int_0^\infty x^r f(x) dx \quad \text{and} \quad b_r = \int_{-\infty}^\infty |x|^r f(x) dx = 2a_r,$$

es decir,  $b_r = E(V^r)$  en la notación de la Proposición 4.2. Es claro que  $b_r < \infty$  es una condición necesaria y suficiente para la existencia de todos los momentos de orden  $r$  de (4.1) o (4.6). Así asumiendo la existencia de estos momentos, entregamos una expresión para ellos.

**Proposición 4.3** Sean  $X \sim ESf(0, 1, \varepsilon)$  y  $Z \sim ESf(\mu, \sigma, \varepsilon)$ . Entonces

$$\mu_r = E(X^r) = \begin{cases} - \sum_{k=0}^{(r-1)/2} \binom{r+1}{2k+1} \varepsilon^{r-2k} b_r & \text{si } r \text{ es impar} \\ \sum_{k=0}^{r/2} \binom{r+1}{2k+1} \varepsilon^{r-2k} b_r & \text{si } r \text{ es par,} \end{cases}$$

y

$$E(Z^r) = \sum_{k=0}^r \binom{r}{k} \sigma^r \mu^{r-k} \mu_r, \quad \text{para } r = 1, 2, \dots$$

**Dem:** El resultado se sigue por la independencia de  $U_\varepsilon$  y  $V$  en la Proposición 4.2, y el Teorema del Binomio.  $\square$

Como consecuencia inmediata de esta Proposición 4.3, tenemos la esperanza, varianza, coeficientes de asimetría y kurtosis

$$E(Z) = \mu - 2\varepsilon\sigma b_1 \quad \text{y} \quad \text{Var}(Z) = \sigma^2 \{(1 + 3\varepsilon^2)b_2 - 4(\varepsilon b_1)^2\}. \quad (4.8)$$

$$\sqrt{\beta_1} = \frac{\mu_3 - 3\mu_1\mu_2 + 2\mu_1^3}{\{(1 + 3\varepsilon^2)b_2 - 4(\varepsilon b_1)^2\}^{3/2}} \quad \text{y} \quad \beta_2 = \frac{\mu_4 - 4\mu_1\mu_3 + 6\mu_1^2\mu_2 - 3\mu_1^4}{\{(1 + 3\varepsilon^2)b_2 - 4(\varepsilon b_1)^2\}^2}, \quad (4.9)$$

donde  $\mu_r, r = 1, 2, \dots$  son definidos en la Proposición 4.3.

### 4.3 Algunos Casos Especiales

La tabla 4.1 entrega resultados específicos para casos particulares de desidades  $f$ : (i)  $f(x) = c_1 \exp\{-\frac{1}{2}|x|^{2/(1+\beta)}\}$  con  $c_1^{-1} = 2^{(3+\beta)/2}\Gamma((3+\beta)/2)$  (exponencial potencia); (ii)  $f(x) = c_2(1+x^2/\nu)^{-(1+\nu)/2}$  con  $c_2 = (\nu\pi)^{-1/2}\Gamma(\nu/2)^{-1}\Gamma((\nu+1)/2)$  (t de Student); (iii)  $f(x) = e^{-x}/(1+e^{-x})^2$  (logística); y (iv)  $f(x) = (2\Gamma(\nu))^{-1}\beta^\nu|x|^{\nu-1}e^{-\beta|x|}$  (gamma simetrizada).

**Comentarios 4.3** *Notemos que la distribución exponencial potencia contiene a las distribuciones normal y doble exponencial cuando  $\beta = 0$  y  $\beta = 1$  respectivamente. También obtenemos mayor flexibilidad en los coeficientes de asimetría y kurtosis. Para  $\beta > 0$  se tienen colas más pesadas que la distribución normal, y ocurre lo contrario cuando  $\beta < 0$ .*

En la tabla 4.2 ilustra los rangos de asimetría y kurtosis calculado como (4.9) que pueden obtenerse para diferentes funciones de densidad  $f(x)$ . Observamos que los rangos para la distribución  $ES(\phi, \varepsilon)$  son los mismos que los de la distribución  $SN(\lambda)$  de Azzalini (1985). Desde (4.1) podemos ver que la moda de la distribución  $ES(\phi, \varepsilon)$  esta localizada en  $x = 0$ , mientras que para la distribución  $SN(\lambda)$  varía con  $\lambda$ . En suma, si  $X \sim SN(\lambda)$  entonces  $X^2 \sim \chi^2(1)$ , es la distribución chi-cuadrada con 1 grado de libertad. Por lo tanto, los momentos pares de  $X$  no dependen del parámetro de asimetría  $\lambda$ , en la cual no es el caso del modelo  $ES(f, \varepsilon)$ .

En la figura 4.1 se muestra los gráficos para las densidades  $ES(f, \varepsilon)$  para diferentes  $\varepsilon$  y  $f(x)$ . El gráfico que se encuentra en el lado superior izquierdo del cuadro corresponde a la distribución  $ES_{t(5)}$ , la cual tiene las colas más pesadas que la distribución  $ESN(\varepsilon)$  estudiada por Mudholkar y Hutson (2000). El gráfico que se encuentra en el medio del lado superior es el caso cuando la densidad es  $ESGS(2, 1)$ , en la cual tiene una forma bimodal. Notemos que estos modelos no son simétricos con respecto al origen, y la localización depende del valor de  $\varepsilon$ . Los demás gráficos representan varias formas que resultan cuando  $h(x)$  es la distribución skew-exponencial-potencia para diferentes valores de  $\beta$ .

$f(x)$	$b_r$	$E(Z)$	$\text{Var}(Z)$
EP	$\frac{2^{r(1+\beta)/2}\Gamma((1+\beta)(1+r)/2)}{\Gamma((1+\beta)/2)}$	$\mu - \frac{2^{(3+\beta)/2}\varepsilon\sigma\Gamma(1+\beta)}{\Gamma((1+\beta)/2)}$	$\frac{\sigma^2 2^{1+\beta}}{\Gamma((1+\beta)/2)} \left\{ (1 + 3\varepsilon^2)\Gamma\left(\frac{3(1+\beta)}{2}\right) - \frac{4\varepsilon^2\Gamma(1+\beta)^2}{\Gamma((1+\beta)/2)} \right\}$
t (*)	$2c_2\nu^{\frac{(r+1)}{2}} \sum_{j=0}^{\frac{(r-1)}{2}} \frac{(-1)^{r-1-2j}}{(\nu-2j-1)} \binom{(r-1)}{j},$ si $r$ es impar $\frac{\nu^{r/2} \cdot 1 \cdot 3 \cdots (r-1)}{(\nu-r)(\nu-r+2)\cdots(\nu-2)},$ si $r$ es par	$\mu - \frac{4c_2\nu\varepsilon\sigma}{\nu-1}$	$\sigma^2 \left\{ \frac{\nu(1+3\varepsilon^2)}{\nu-2} - \frac{16(c_2\varepsilon\nu)^2}{(\nu-1)^2} \right\}$
L(**)	$b_1 = 2\log(2)$ $b_r = \frac{2\Gamma(r+1)(2^r-2)\zeta(r)}{2^r},$ $r = 2, 3, \dots$	$\mu - 4\varepsilon\sigma\log(2)$	$\sigma^2 \left\{ \frac{(1+3\varepsilon^2)\pi^2}{3} - 16(\varepsilon\log(2))^2 \right\}$
GS	$\frac{\Gamma(r+\nu)}{\beta^r\Gamma(\nu)}$	$\mu - \frac{2\varepsilon\sigma\nu}{\beta}$	$\sigma^2 \left\{ \frac{(1+3\varepsilon^2)\nu(\nu+1)-4(\varepsilon\nu)^2}{\beta^2} \right\}$

Tabla 4.1: Resúmenes para algunos casos de interés. Si  $Z \sim ES(f, \mu, \sigma, \varepsilon)$  para  $f(x)$  indicado en la primera columna. (\*) En este caso,  $b_r < \infty$  si y sólo si  $\nu > r$ . (\*\*) Si  $\zeta(r) = \sum_{j=1}^{\infty} \frac{1}{j^r}$  es la función Zeta de Riemann (Johnson y otros (1995)),  $r = 2, 3, \dots$

$f(x)$	Asimetría	Kurtosis
Exponencial Potencia(0)	$-0.995272 < \sqrt{\beta_1} < 0.995272$	$3 < \beta_2 < 3.86918$
Exponencial Potencia(0.5)	$-1.503593 < \sqrt{\beta_1} < 1.503593$	$4.2 < \beta_2 < 6.0082$
Exponencial Potencia(-0.5)	$-0.442787 < \sqrt{\beta_1} < 0.442787$	$2.18 < \beta_2 < 2.44$
$t_{(5)}$	$-2.54964 < \sqrt{\beta_1} < 2.54964$	$9 < \beta_2 < 23.109$
Logística	$-1.540329 < \sqrt{\beta_1} < 1.540329$	$4.2 < \beta_2 < 6.583736$
Gamma Simetrizada(1,2)	$-2 < \sqrt{\beta_1} < 2$	$6 < \beta_2 < 9$
Gamma Simetrizada(2,1)	$-1.414214 < \sqrt{\beta_1} < 1.414214$	$\frac{10}{3} < \beta_2 < 6$

Tabla 4.2: Posibles valores para los coeficientes de asimetría y kurtosis para algunos casos particulares cuando el parámetro de asimetría  $\varepsilon$  se encuentra en los rangos de  $(-1, 1)$ .

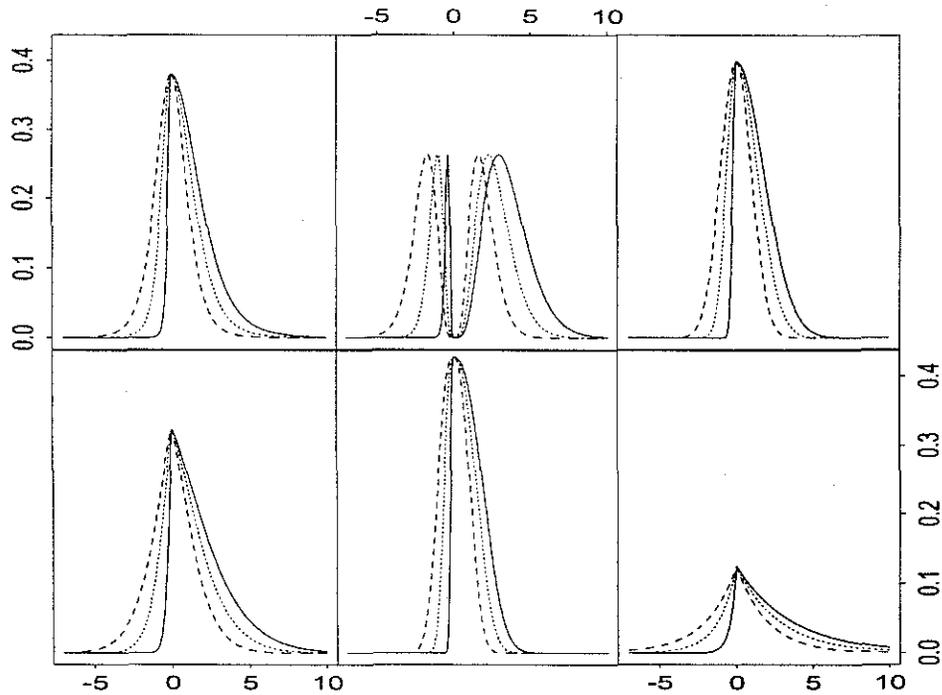


Figura 4.1: Ejemplos de la densidad  $ES(f, \varepsilon)$  para  $\varepsilon = 0$  (línea discontinua),  $\varepsilon = -0.5$  (línea punteada) y  $\varepsilon = -0.9$  (línea sólida). Desde la izquierda a derecha, el cuadro superior muestra los casos cuando  $f(x)$  es: (i) la densidad  $EST_{(5)}$ ; (ii) la densidad  $ESGS(2,1)$ ; y (iii) la densidad  $ESEP(0,1,\varepsilon,0)$ . Desde la izquierda a derecha, el cuadro inferior muestra los casos cuando  $f(x)$  es la densidad  $ESEP(0,1,\varepsilon,\beta)$  con (i)  $\beta = 0.5$ ; (ii)  $\beta = -0.5$ ; y (iii)  $\beta = 1$ .

## Capítulo 5

# Inferencia Estadística en la Familia

## Epsilon-Skew

### 5.1 Introducción

Este capítulo está dedicado a la Inferencia Estadística para esta nueva familia epsilon-skew de localización y escala definida en (4.6). Estudiamos estimadores de momentos para la familia general, estimadores de máxima verosimilitud para los casos particulares de las distribuciones epsilon-skew-exponencial-potencia (ESEP) y epsilon-skew- $t$  (ESt). Aplicamos esta metodología a tres ejemplos para el modelo ESEP y tres ejemplos para el modelo ESt.

## 5.2 Estimación por el Método de Momentos

Sean  $Z_1, \dots, Z_n$  una muestra desde la densidad  $ESf(\mu, \sigma, \varepsilon)$  definida en (4.6). Entonces

$$\mu = E(Z) + 2\varepsilon\sigma b_1 \quad \text{y} \quad \sigma^2 = \frac{Var(Z)}{(1 + 3\varepsilon^2)b_2 - 4\varepsilon^2 b_1^2}. \quad (5.1)$$

Cuando  $f$  no depende de parámetros adicionales, los coeficientes  $b_i$  son completamente especificados, y usamos el coeficiente de asimetría muestral  $\gamma_1$  y resolvemos

$$\gamma_1 = \frac{-4\hat{\varepsilon}(1 + \hat{\varepsilon}^2)b_3 + 6\hat{\varepsilon}(1 + 3\hat{\varepsilon}^2)b_1 b_2 - 16\hat{\varepsilon}^3 b_1^3}{\{(1 + 3\hat{\varepsilon}^2)b_2 - 4\hat{\varepsilon}^2 b_1^2\}^{3/2}}, \quad (5.2)$$

para  $\hat{\varepsilon}$ , utilizando un método numérico apropiado. Sea  $\hat{\varepsilon}_M$  la solución. Luego, sustituimos en (5.1)  $\varepsilon$  por  $\hat{\varepsilon}_M$ ,  $E(Z)$  por la media muestral y  $Var(Z)$  por la varianza muestral, desde la cual los estimadores de momentos  $(\hat{\mu}_M, \hat{\sigma}_M^2, \hat{\varepsilon}_M)$  son obtenidos.

**Comentarios 5.1** *Un procedimiento similar se aplica cuando los coeficientes  $b_i$  son desconocidos, es decir,  $f$  depende de un parámetro adicional  $\beta$ . En este caso, utilizamos los coeficientes de asimetría ( $\gamma_1$ ) y el de kurtosis muestral ( $\gamma_2$ ), adicionando una ecuación (5.2)*

$$\gamma_2 = \frac{(1 + 10\hat{\varepsilon}^2 + 5\hat{\varepsilon}^4)\hat{b}_4 - 32\hat{\varepsilon}^2(1 + \hat{\varepsilon}^2)\hat{b}_1 \hat{b}_3 + 24\hat{\varepsilon}^2(1 + 3\hat{\varepsilon}^2)\hat{b}_1^2 \hat{b}_2 - 48\hat{\varepsilon}^4 \hat{b}_1^4}{\{(1 + 3\hat{\varepsilon}^2)b_2 - 4\hat{\varepsilon}^2 b_1^2\}^2}, \quad (5.3)$$

donde  $\hat{b}_r = b_r(\hat{\beta})$ . Resolviendo (5.2) y (5.3) para  $\varepsilon$  y  $\beta$ , usando algún método numérico, obtenemos  $\hat{\varepsilon}_M$  y  $\hat{\beta}_M$ , en la cual es usado en (5.1) y se encuentran  $\hat{\mu}_M$  y  $\hat{\sigma}_M^2$ . Notar en los casos  $b_r$  en (5.2) necesitamos reemplazar por  $\hat{b}_r$ ,  $r = 1, 2, 3$ .

En la siguiente Proposición establecemos normalidad asintótica de estos estimadores.

**Proposición 5.1** Sean  $Z_1, \dots, Z_n$  una muestra aleatoria desde  $ESf(\mu, \sigma, \varepsilon)$  con distribución  $f$  conocida. Sea  $\boldsymbol{\theta} = (\mu, \sigma^2, \varepsilon)$  y para  $k = 1, 2, 3$ , escribimos  $\mu'_k = \mu'_k(\boldsymbol{\theta}) = E(Z^k)$ . Si  $\mu'_6(\boldsymbol{\theta}) < \infty$  y  $\hat{\boldsymbol{\theta}}_M(n)$  son los correspondientes momentos estimados, tenemos

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_M(n) - \boldsymbol{\theta}) \xrightarrow{D} \mathcal{N}_3(0, \Gamma(\boldsymbol{\theta})),$$

donde  $\Gamma(\boldsymbol{\theta}) = H^{-1}(\boldsymbol{\theta})\Sigma[H^{-1}(\boldsymbol{\theta})]^T$ ,  $\Sigma = \{(\mu'_{i+j} - \mu'_i\mu'_j)_{i,j}\}$  y  $H(\boldsymbol{\theta})$  es dado por

$$\begin{pmatrix} 1 & \frac{-\varepsilon b_1}{\sigma} & -2\sigma b_1 \\ 2\mu - 4\sigma\varepsilon b_1 & \tau_3 b_2 - \frac{2\mu\varepsilon b_1}{\sigma} & -4\mu\sigma b_1 + 6\sigma^2\varepsilon b_2 \\ 3\mu^2 - 12\mu\sigma\varepsilon b_1 + 3\sigma^2\tau_3 b_2 & 3\mu\tau_3 b_2 - \frac{3\mu^2\varepsilon b_1}{\sigma} - 6\sigma\varepsilon\tau_1 b_3 & 18\mu\sigma^2\varepsilon b_2 - 6\mu^2\sigma b_1 - 4\sigma^3\tau_3 b_3 \end{pmatrix},$$

con  $\tau_k = (1 + k\varepsilon^2)$ .

**Dem:** Definiendo  $A(\boldsymbol{\theta}) = (\mu'_1(\boldsymbol{\theta}), \mu'_2(\boldsymbol{\theta}), \mu'_3(\boldsymbol{\theta}))$  y  $H(\boldsymbol{\theta}) = \partial A(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ . Se tiene que  $H(\boldsymbol{\theta})$  es una función continua de  $\boldsymbol{\theta}$  y tiene rango completo. Entonces el resultado es una consecuencia de una expansión de Taylor estándar de  $A(\boldsymbol{\theta})$ :

$$\sqrt{n}(M_n - A(\boldsymbol{\theta})) = \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})H(\boldsymbol{\theta})^T + \mathcal{O}_p(n),$$

donde  $M_n$  es el vector de los primeros tres momentos.  $\square$

### 5.3 Estimación de Máxima Verosimilitud en el Modelo Epsilon-Skew-Exponencial-Potencia

La Estimación de Máxima Verosimilitud (EMV) para la familia  $ESf(\mu, \sigma, \varepsilon)$  es, conceptualmente, directa. En la práctica, sin embargo, calcularlo representa alguna dificultad. Ilustramos estos cálculos, en el caso de la distribución epsilon-skew-exponencial-potencia (ESEP) con  $\beta$  conocido. Extendemos la metodología presentada en Mudholkar y Hutson (2000), la cual considera como caso particular la distribución epsilon-skew-normal (es decir,  $\beta = 0$ ).

Sean  $z_{(1)} \leq z_{(2)} \leq \dots \leq z_{(n)}$  los estadísticos de orden asociados a una muestra aleatoria simple  $Z_1, \dots, Z_n$  desde la distribución  $ESEP(\mu, \sigma, \varepsilon, \beta)$ . Denotemos  $z_{(0)} = -\infty$  y  $z_{(n+1)} = \infty$ . Sea  $k \equiv k(z_{(1)}, \dots, z_{(n)}, \mu)$  un entero tal que  $z_{(k)} < \mu < z_{(k+1)}$ . Por la continuidad de la distribución ESEP,  $k$  es definido con probabilidad 1, y el rango es  $\{0, 1, \dots, n\}$ . La función log-verosimilitud puede ser expresada como

$$l(\mu, \sigma^2, \varepsilon) = n \log(c_1) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^{\frac{2}{(1+\beta)}}} g_k(\mu, \varepsilon), \quad (5.4)$$

donde

$$g_k(\mu, \varepsilon) = \left\{ (1 + \varepsilon)^{-\frac{2}{(1+\beta)}} \sum_{i=1}^k |z_{(i)} - \mu|^{\frac{2}{(1+\beta)}} + (1 - \varepsilon)^{-\frac{2}{(1+\beta)}} \sum_{i=k+1}^n |z_{(i)} - \mu|^{\frac{2}{(1+\beta)}} \right\}. \quad (5.5)$$

El cálculo del EMV de  $\theta$  necesitamos maximizar (5.4), en la cual podemos fijar primero  $k$ , y entonces encontramos los correspondientes valores óptimos  $\hat{\theta}_k$ . El EMV, es  $\hat{\theta}_k^* = \arg \max l(\hat{\theta}_k)$  en el conjunto  $k = 0, \dots, n$ .

Los casos  $k = 0$  y  $k = n$  pueden resolverse fácilmente. Cuando  $k = 0$  entonces (5.4) se reduce a

$$-\frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^{\frac{2}{1+\beta}}} \left\{ \sum_{i=1}^n \frac{|z_{(i)} - \mu|^{\frac{2}{1+\beta}}}{(1 - \varepsilon)^{\frac{2}{1+\beta}}} \right\},$$

en la cual, en terminos de  $\varepsilon$ , se maximiza para  $\hat{\varepsilon}_0 = -1$ . En consecuencia,  $\mu < z_{(1)}$  implica que  $\hat{\mu}_0 = z_{(1)}$ , y desde esto se tiene que

$$\hat{\sigma}_0^2 = \left[ \frac{1}{2^{\frac{2}{1+\beta}} n(1+\beta)} \sum_{i=2}^n |z_{(i)} - z_{(1)}|^{\frac{2}{1+\beta}} \right]^{1+\beta}.$$

Similarmente, cuando  $k = n$  obtenemos  $\hat{\varepsilon}_n = 1$ ,  $\hat{\mu}_n = z_{(n)}$  y

$$\hat{\sigma}_n^2 = \left[ \frac{1}{2^{\frac{2}{1+\beta}} n(1+\beta)} \sum_{i=1}^{n-1} |z_{(i)} - z_{(n)}|^{\frac{2}{1+\beta}} \right]^{1+\beta}.$$

Cuando  $1 \leq k < n$ , necesitamos resolver las siguientes ecuaciones de verosimilitud

$$\frac{\partial}{\partial \mu} g_k(\mu, \varepsilon) = 0, \quad \frac{\partial}{\partial \varepsilon} g_k(\mu, \varepsilon) = 0, \quad \frac{\partial}{\partial \sigma^2} l(\mu, \sigma^2, \varepsilon) = 0.$$

Resolviendo, podemos encontrar fácilmente que

$$\hat{\varepsilon}_k = \frac{\left( \sum_{i=1}^k |z_{(i)} - \hat{\mu}_k|^{\frac{2}{1+\beta}} \right)^{\frac{1+\beta}{3+\beta}} - \left( \sum_{i=k+1}^n |z_{(i)} - \hat{\mu}_k|^{\frac{2}{1+\beta}} \right)^{\frac{1+\beta}{3+\beta}}}{\left( \sum_{i=1}^k |z_{(i)} - \hat{\mu}_k|^{\frac{2}{1+\beta}} \right)^{\frac{1+\beta}{3+\beta}} + \left( \sum_{i=k+1}^n |z_{(i)} - \hat{\mu}_k|^{\frac{2}{1+\beta}} \right)^{\frac{1+\beta}{3+\beta}}}$$

$$\hat{\sigma}_k^2 = \frac{1}{n(1+\beta)} \left\{ \sum_{i=1}^k \frac{|z_{(i)} - \hat{\mu}_k|^{\frac{2}{1+\beta}}}{(1 + \hat{\varepsilon}_k)^{\frac{2}{1+\beta}}} + \sum_{i=k+1}^n \frac{|z_{(i)} - \hat{\mu}_k|^{\frac{2}{1+\beta}}}{(1 - \hat{\varepsilon}_k)^{\frac{2}{1+\beta}}} \right\}$$

Adicionalmente, se tiene que  $g_k(\cdot)$  es diferenciable con respecto a  $\mu$  siempre que  $-1 < \beta < 1$ . Sin embargo, esto no implica que  $\frac{\partial}{\partial \mu} g_k(\mu, \varepsilon) = 0$  puede resolverse en el intervalo  $(z_{(k)}, z_{(k+1)})$ . Si este es el caso entonces  $\hat{\mu}_k$  se encuentra en  $z_{(k)}$  o  $z_{(k+1)}$  según sea  $\frac{\partial}{\partial \mu} g_k(\mu, \varepsilon)$  negativa o positiva  $(z_{(k)}, z_{(k+1)})$ . En otro lado, si  $g_k(\mu, \varepsilon)$  tiene un mínimo local en un punto interior del intervalo  $(z_{(k)}, z_{(k+1)})$  entonces este punto es solución de

$$\frac{\sum_{i=1}^k |z_{(i)} - \mu|^{\frac{1-\beta}{1+\beta}}}{\sum_{i=1}^k |z_{(i)} - \mu|^{\frac{1}{3+\beta}}} = \frac{\sum_{i=k+1}^n |z_{(i)} - \mu|^{\frac{1-\beta}{1+\beta}}}{\sum_{i=k+1}^n |z_{(i)} - \mu|^{\frac{1}{3+\beta}}}.$$

Juntando los resultados anteriores, se tiene mostrado lo siguiente:

**Proposición 5.2** *El EMV de  $\theta$  es  $(\hat{\mu}_k, \hat{\sigma}_k^2, \hat{\varepsilon}_k)$ , donde  $k$  es tal que  $l(\hat{\mu}_k, \hat{\sigma}_k^2, \hat{\varepsilon}_k) \geq l(\hat{\mu}_j, \hat{\sigma}_j^2, \hat{\varepsilon}_j)$  para todo  $j \in \{0, \dots, n\}$ .*

En la siguiente Proposición se tiene un resultado asintótico del EMV.

**Proposición 5.3** *Sea  $Z \sim ESf(\mu, \sigma, \varepsilon)$ . Asumimos que  $f$  tiene segundas derivadas continuas y que las siguientes cantidades existen para  $i = 0, 1, 2$  y  $j = 0, 1$ :*

$$\begin{aligned} \kappa_i \equiv \kappa_i(f) &= \int_0^\infty \left( \frac{(f'(t))^2}{f(t)} - f''(t) \right) t^i dt \\ s_j \equiv s_j(f) &= \int_0^\infty t^j f'(t) dt. \end{aligned}$$

Entonces el EMV  $\hat{\theta} = (\hat{\mu}, \hat{\sigma}^2, \hat{\varepsilon})$  es consistente y asintóticamente normal:  $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}_3(0, V)$  cuando  $n \rightarrow \infty$ , con

$$V = \begin{pmatrix} \frac{\sigma^2(1-\varepsilon^2)(\kappa_2-2s_1)}{2[\kappa_0(\kappa_2-2s_1)-(s_0-\kappa_1)^2]} & 0 & \frac{\sigma(1-\varepsilon^2)(s_0-\kappa_1)}{2[(s_0-\kappa_1)^2-\kappa_0(\kappa_2-2s_1)]} \\ 0 & \frac{2\sigma^4}{\kappa_2-3s_1-1} & 0 \\ \frac{\sigma(1-\varepsilon^2)(s_0-\kappa_1)}{2[(s_0-\kappa_1)^2-\kappa_0(\kappa_2-2s_1)]} & 0 & \frac{\kappa_0(1-\varepsilon^2)}{2[\kappa_0(\kappa_2-2s_1)-(s_0-\kappa_1)^2]} \end{pmatrix}.$$

**Dem:** El resultado se obtiene utilizando propiedades conocidas de los estimadores de máxima verosimilitud en teoría asintótica donde las condiciones de regularidad se cumplen. Ver Lehmann (1999).□

En la table 5.1 se muestran valores de  $\kappa_i$  y  $s_j$  para los ejemplos considerados en la Sección 4.3.

$f(x)$	$s_0$	$s_1$	$\kappa_0$	$\kappa_1$	$\kappa_2$
Exponencial potencia	$-c_1$	$-\frac{1}{2}$	$\frac{c_1(1-\beta)\Gamma((1-\beta)/2)}{2^{(1+\beta)/2}(1+\beta)}$ , $\beta < 1$	$\frac{c_1(1-\beta)}{1+\beta}$	$\frac{1-\beta}{2(1+\beta)}$
$t_{(\nu)}$	$-c_2$	$-\frac{1}{2}$	$\frac{\nu+1}{2(\nu+3)}$	$\frac{c_2(\nu-1)}{3+\nu}$	$\frac{\nu-3}{2(\nu+3)}$
Logística	$-\frac{1}{4}$	$-\frac{1}{2}$	$\frac{1}{6}$	$-\frac{1}{12} + \frac{\log(2)}{2}$	$-\frac{1}{3} + \frac{\pi^2}{18}$
Gamma Simetrizada	$-\frac{1}{2}$	0	$\frac{\beta^2}{2(\nu-2)}$ , $\nu > 2$	$\frac{\beta}{2}$ , $\nu > 1$	$\frac{\nu-1}{2}$

Tabla 5.1: Valores de  $\kappa_i$  y  $s_j$  para los ejemplos considerados en la Sección 4.3.

**Corolario 5.1** Si  $Z \sim ESEP(\mu, \sigma, \varepsilon, \beta)$  con  $\beta$  conocido entonces la matriz de covarianza asintótica  $V$  entregada en la Proposición 5.3 es dada por

$$V = \begin{pmatrix} \frac{2^{1+\beta}\sigma^2(1+\beta)(3+\beta)(1-\varepsilon^2)\Gamma(\frac{3+\beta}{2})^2}{(1-\beta)(3+\beta)\Gamma(\frac{1-\beta}{2})\Gamma(\frac{3+\beta}{2})-4} & 0 & \frac{2^{\frac{3+\beta}{2}}\sigma(1+\beta)(1-\varepsilon^2)\Gamma(\frac{3+\beta}{2})}{(1-\beta)(3+\beta)\Gamma(\frac{1-\beta}{2})\Gamma(\frac{3+\beta}{2})-4} \\ 0 & 2(1+\beta)\sigma^4 & 0 \\ \frac{2^{\frac{3+\beta}{2}}\sigma(1+\beta)(1-\varepsilon^2)\Gamma(\frac{3+\beta}{2})}{(1-\beta)(3+\beta)\Gamma(\frac{1-\beta}{2})\Gamma(\frac{3+\beta}{2})-4} & 0 & \frac{(1-\beta^2)(1-\varepsilon^2)\Gamma(\frac{1-\beta}{2})\Gamma(\frac{3+\beta}{2})}{(1-\beta)(3+\beta)\Gamma(\frac{1-\beta}{2})\Gamma(\frac{3+\beta}{2})-4} \end{pmatrix}.$$

**Comentarios 5.2** Podemos también extender el Teorema 4.11 de Mudholkar y Hutson (2000). Necesitamos, tomar la transformación

$$g(\hat{\mu}, \hat{\sigma}^2, \hat{\varepsilon}) = \left( \hat{\mu}, \sqrt{\frac{\kappa_2 - 3s_1 - 1}{2}} \log(\hat{\sigma}^2), \sqrt{\frac{2[\kappa_0(\kappa_2 - 2s_1) - (s_0 - \kappa_1)^2]}{\kappa_0}} \arcsin(\hat{\varepsilon}) \right),$$

entonces cuando  $n \rightarrow \infty$ , se tiene

$$\sqrt{n}(g(\hat{\theta}) - g(\theta)) \xrightarrow{\mathcal{D}} \mathcal{N}_3 \left[ 0, \begin{pmatrix} \frac{\sigma^2(1-\varepsilon^2)(\kappa_2 - 2s_1)}{2[\kappa_0(\kappa_2 - 2s_1) - (s_0 - \kappa_1)^2]} & 0 & \frac{\sigma(\kappa_1 - s_0)\sqrt{1-\varepsilon^2}}{\sqrt{2[\kappa_0(\kappa_2 - 2s_1) - (s_0 - \kappa_1)^2]}} \\ 0 & 1 & 0 \\ \frac{\sigma(\kappa_1 - s_0)\sqrt{1-\varepsilon^2}}{\sqrt{2[\kappa_0(\kappa_2 - 2s_1) - (s_0 - \kappa_1)^2]}} & 0 & 1 \end{pmatrix} \right].$$

Podemos utilizar este resultado en la construcción de intervalos de confianza aproximados para parámetros individuales. Por ejemplo,

$$\left( \sin \left\{ \arcsin(\hat{\varepsilon}) - \Phi^{-1} \left( 1 - \frac{\gamma}{2} \right) \sqrt{\frac{\kappa_0}{2n[\kappa_0(\kappa_2 - 2s_1) - (s_0 - \kappa_1)^2]}} \right\}, \right. \\ \left. \sin \left\{ \arcsin(\hat{\varepsilon}) + \Phi^{-1} \left( 1 - \frac{\gamma}{2} \right) \sqrt{\frac{\kappa_0}{2n[\kappa_0(\kappa_2 - 2s_1) - (s_0 - \kappa_1)^2]}} \right\} \right)$$

es un intervalo de confianza aproximado  $100(1 - \gamma)\%$  para  $\varepsilon$ .

Finalmente, el caso donde  $\beta$  es desconocido puede ser conceptualmente manejable con una dificultad extra. Así se requiere resolver un conjunto de ecuaciones de verosimilitud que incluya la variable extra y la ecuación es dada por  $\frac{\partial}{\partial \beta} l(\mu, \sigma^2, \varepsilon, \beta) = 0$ .

### 5.3.1 Ejemplos

Ahora aplicamos los métodos discutidos anteriormente a tres ejemplos. El primer conjunto de datos son 219 alturas de volcanes usadas en Mudholkar y Hutson (2000). El segundo ejemplo es un conjunto de datos que fueron tomadas a 202 atletas Australianos y son entregadas en el libro de Cook y Weisberg (1994). Concretamente, estas medidas son concentración de plasma ferritin. El tercer ejemplo considera un conjunto de datos que es proporcionado por el Instituto Nacional Chileno de Estadísticas (INE), y consiste en 250 edades de las madres cuando tienen a sus hijos.

La Tabla 5.2 presentan algún resumen estadístico para los tres conjuntos de datos. Los resúmenes sugieren que los dos primeros ejemplos corresponden a densidades leptokurticas, mientras que el tercero corresponde a platicurtica. Podemos ver por otro lado que los tres ejemplos presentan asimetría positiva. Utilizamos el modelo epsilon-skew-exponencial-potencia (ESEP) para los tres ejemplos, usando los métodos de momentos (MM) y máxima verosimilitud descritos en las Secciones anteriores. Los resultados son resumidos en la Tabla 5.3, también incluyen intervalos de confianza

Ejemplo	$n$	$\bar{X}$	$S^2$	$\gamma_1$	$\gamma_2$
1	219	70.246	1850.56	0.8403	3.4821
2	202	76.876	2256.37	1.2902	4.4863
3	250	26.408	49.27	0.2828	2.3976

Tabla 5.2: *Resumen de estadísticas descriptivas para los tres ejemplos.*

asintóticos del 95% de coeficiente de confianza para todas las estimaciones.

Mostramos los gráficos de los datos y las densidades estimadas pueden encontrarse en Figura 5.1. Por motivo de comparar también incluimos el EMV de la distribución epsilon-skew-normal obtenida cuando  $\beta = 0$  en la distribución general ESEP. Para el primer ejemplo, las tres densidades son practicamente la misma. El correspondiente intervalo de confianza (asintótico) para  $\beta$  sugiere que el modelo epsilon-skew-normal es el más adecuado. En contraste, para los datos del plasma ferritin, el EMV de la distribución ESEP presenta resultados más satisfactorios, con  $\beta$  significativamente distinto de cero. En el tercer ejemplo observamos un intervalo de confianza para  $\beta$  que no contiene al cero y se concentra en el lado negativo, esto nos indica que el modelo ESEP platikurtico es adecuado para este conjunto de datos.

Ejemplos	Parámetros	Método de Estimación	
		MM	EMV
1 ( $n = 219$ )	$\mu$	27.7734 (-8.9291,64.4760)	26.7744 (16.2510,37.2978)
	$\sigma^2$	1588.71 (133.74,2983.42)	1768.43 (872.38,2664.48)
	$\varepsilon$	-0.6789 (-1.0000,-0.0844)	-0.6841 (-0.8346,-0.5336)
	$\beta$	-0.0093 (-0.6050,0.5864)	-0.0762 (-0.3559,0.2035)
2 ( $n = 202$ )	$\mu$	(*)	35.0000 (28.8139,41.1861)
	$\sigma^2$	(*)	615.98 (122.11,1109.85)
	$\varepsilon$	(*)	-0.6952 (-0.8063,-0.5841)
	$\beta$	(*)	0.5028 (0.1073,0.8985)
3 ( $n = 250$ )	$\mu$	23.3330 (20.4451,26.2216)	21.5229 (18.2990,24.7469)
	$\sigma^2$	70.63 (39.23,102.03)	89.62 (58.96,120.28)
	$\varepsilon$	-0.2705 (-0.4954,-0.0456)	-0.4260 (-0.6730,-0.1790)
	$\beta$	-0.2365 (-0.4988,0.0258)	-0.4217 (-0.6274,-0.2159)

Tabla 5.3: *Estimación de los parámetros e intervalos de confianza asintóticos para cada uno de los tres ejemplos descritos en la Sección 5.3.1, usando el modelo epsilon-skew-exponencial-potencia. La estimación se realiza vía el método de momentos (MM), y de máxima verosimilitud (EMV). (\*) No existe solución en el sistema de ecuaciones que definen los estimadores de momentos.*

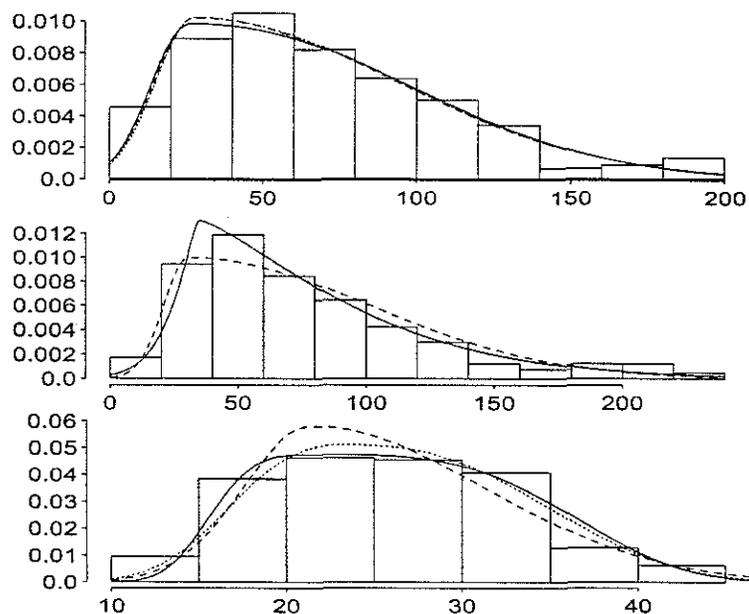


Figura 5.1: La densidad ESEP es considerada para los tres ejemplos. Las estimaciones son: máxima verosimilitud (línea sólida), método de momentos (línea punteada) y de máxima verosimilitud restringida a  $\beta = 0$  (línea discontinua). El gráfico que está en la parte superior representa el conjunto de medidas de alturas de los volcanes, el gráfico del medio corresponde a la concentración de plasma ferritina y el gráfico que se encuentra en la parte inferior corresponde a la edad de las madres que tienen hijos en el año 1996.

## 5.4 Estimación de Máxima Verosimilitud para el Modelo Epsilon-Skew-t

En esta sección discutimos los estimadores de máxima verosimilitud para la distribución epsilon-skew-t. Entregamos además algunos resultados asintóticos de estos estimadores. La distribución epsilon-skew-normal es un caso particular de la distribución epsilon-skew-t (es decir,  $\nu = \infty$ ).

Sean  $z_{(1)} \leq z_{(2)} \leq \dots \leq z_{(n)}$  estadísticos de orden asociados a una muestra aleatoria  $Z_1, Z_2, \dots, Z_n$  desde la distribución  $Est(\mu, \sigma, \varepsilon, \nu)$ . Denotamos  $z_{(0)} = -\infty$  y  $z_{(n+1)} = \infty$ . Sea  $k = k(z_{(1)}, z_{(2)}, \dots, z_{(n)}, \mu)$  un entero tal que  $z_{(k)} < \mu < z_{(k+1)}$ . Por la continuidad de la distribución  $Est$ ,  $k$  es definido en el recorrido  $\{0, 1, 2, \dots, n\}$ . La función log-verosimilitud puede ser expresada como

$$l(\mu, \sigma^2, \varepsilon, \nu) = -\frac{n}{2} \log(\pi) - \frac{n}{2} \log(\nu) + n \log \left( \Gamma \left( \frac{\nu+1}{2} \right) \right) - n \log \left( \Gamma \left( \frac{\nu}{2} \right) \right) - \frac{n}{2} \log(\sigma^2) - \left( \frac{1+\nu}{2} \right) g(\mu, \sigma^2, \varepsilon, \nu), \quad (5.6)$$

donde

$$g(\mu, \sigma^2, \varepsilon, \nu) = \sum_{i=1}^k \log \left( 1 + \frac{1}{\nu} \left( \frac{z_{(i)} - \mu}{\sigma(1+\varepsilon)} \right)^2 \right) + \sum_{i=k+1}^n \log \left( 1 + \frac{1}{\nu} \left( \frac{z_{(i)} - \mu}{\sigma(1-\varepsilon)} \right)^2 \right).$$

Para calcular el EMV de  $\theta = (\mu, \sigma^2, \varepsilon, \nu)$  necesitamos maximizar (5.6), en el cual fijamos primero  $k$  y encontramos los valores óptimos correspondientes  $\hat{\theta}_k$ . El EMV, es  $\hat{\theta}_k = \arg \max l(\hat{\theta}_k)$  para el conjunto  $k = 0, 1, \dots, n$ .

Cuando  $1 \leq k \leq n$ , necesitamos resolver las ecuaciones de verosimilitud  $\frac{\partial l}{\partial \mu} = 0$ ;  $\frac{\partial l}{\partial \sigma^2} = 0$ ;  $\frac{\partial l}{\partial \varepsilon} = 0$ ;  $\frac{\partial l}{\partial \nu} = 0$ , se tiene el siguiente sistema de ecuaciones

$$\begin{aligned}
\sum_{i=1}^k \frac{z_{(i)} - \mu}{\nu \sigma^2 (1 + \varepsilon)^2 + (z_{(i)} - \mu)^2} + \sum_{i=k+1}^n \frac{z_{(i)} - \mu}{\nu \sigma^2 (1 - \varepsilon)^2 + (z_{(i)} - \mu)^2} &= 0 \\
\sum_{i=1}^k \frac{(z_{(i)} - \mu)^2}{\nu \sigma^2 (1 + \varepsilon)^2 + (z_{(i)} - \mu)^2} + \sum_{i=k+1}^n \frac{(z_{(i)} - \mu)^2}{\nu \sigma^2 (1 - \varepsilon)^2 + (z_{(i)} - \mu)^2} &= \frac{n}{\nu + 1} \\
\sum_{i=1}^k \frac{(z_{(i)} - \mu)^2}{\nu \sigma^2 (1 + \varepsilon)^2 + (z_{(i)} - \mu)^2} - \sum_{i=k+1}^n \frac{(z_{(i)} - \mu)^2}{\nu \sigma^2 (1 - \varepsilon)^2 + (z_{(i)} - \mu)^2} &= \frac{n\varepsilon}{\nu + 1} \\
n\psi\left(\frac{\nu}{2}\right) - n\psi\left(\frac{\nu+1}{2}\right) - g(\mu, \sigma^2, \varepsilon, \nu) &= 0,
\end{aligned} \tag{5.7}$$

donde  $\psi(x)$  es la función digamma, el cual se resuelve por algún método numérico apropiado.

**Proposición 5.4** *El EMV de  $\theta$  es  $(\hat{\mu}_k, \hat{\sigma}_k^2, \hat{\varepsilon}_k, \hat{\nu}_k)$ , donde  $k$  es tal que  $l(\hat{\mu}_k, \hat{\sigma}_k^2, \hat{\varepsilon}_k, \hat{\nu}_k) \geq l(\hat{\mu}_j, \hat{\sigma}_j^2, \hat{\varepsilon}_j, \hat{\nu}_j)$  para todo  $j = \{0, 1, \dots, n\}$ .*

### 5.4.1 Matriz de Covarianza Asintótica

En lo que sigue, entregamos la matriz de información de Fisher para la distribución epsilon-skew- $t$  y hacemos notar el caso particular de la distribución epsilon-skew-cauchy (ESC). Luego entregamos un resultado asintótico de los EMV.

**Caso I.** Cuando los grados de libertad son conocidos ( $\nu = \nu_0$ )

**Proposición 5.5** *Sea  $Z \sim ESt(\mu, \sigma, \varepsilon, \nu_0)$ . Entonces la matriz de información de*

Fisher  $I(\boldsymbol{\theta})$  es dada por

$$I(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\nu_0+1}{\sigma^2(1-\varepsilon^2)(\nu_0+3)} & 0 & -\frac{4c(\nu_0)(\nu_0+1)}{\sigma(1-\varepsilon^2)(\nu_0+3)} \\ & \frac{\nu_0}{2\sigma^4(\nu_0+3)} & 0 \\ & & \frac{3(\nu_0+1)}{(1-\varepsilon^2)(\nu_0+3)} \end{pmatrix}, \quad (5.8)$$

donde  $c(\nu_0) = (\nu_0\pi)^{-1/2}\Gamma(\nu_0/2)^{-1}\Gamma((\nu_0+1)/2)$ .

**Dem.** El resultado se sigue calculando la esperanza de

$$-\frac{\partial^2 \log h(z|\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}.$$

□

**Comentarios 5.3** *Note que la matriz de información de Fisher es no singular para  $-1 < \varepsilon < 1$  así el modelo epsilon-skew-t no presenta los problemas reportados en Azzalini y Capitanio (1999) con respecto a la distribución skew-normal, en la cual la matriz de información de Fisher es singular para  $\lambda = 0$ . Problemas similares también se tienen con la distribución skew-t (Azzalini y Capitanio, 2003). Como se muestra en Rotnovitsky y otros (2002) una consecuencia de este hecho es que la distribución muestral asintótica de la razón de verosimilitud no es chicuadrado.*

**Corolario 5.2** *Sea  $Z \sim ESC(\mu, \sigma, \varepsilon)$ . Entonces la matriz de información de Fisher es dada por*

$$I_0(\theta) = \begin{pmatrix} \frac{1}{2\sigma^2(1-\varepsilon^2)} & 0 & -\frac{2}{\sigma\pi(1-\varepsilon^2)} \\ & \frac{1}{8\sigma^4} & 0 \\ & & \frac{3}{2(1-\varepsilon^2)} \end{pmatrix}. \quad (5.9)$$

**Proposición 5.6** Sea  $Z \sim ESt(\mu, \sigma, \varepsilon, \nu_0)$ . Entonces el EMV  $\hat{\theta} = (\hat{\mu}, \hat{\sigma}^2, \hat{\varepsilon})$  es consistente y asintóticamente normal.

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\mathcal{D}} N_3(0, V) \text{ , cuando } n \rightarrow \infty,$$

donde  $V = I^{-1}$  y es dada por

$$V = \begin{pmatrix} \frac{3\sigma^2(1-\varepsilon^2)(\nu_0+3)}{(3-16c^2)(\nu_0+1)} & 0 & \frac{4c\sigma(1-\varepsilon^2)(\nu_0+3)}{(3-16c^2)(\nu_0+1)} \\ & \frac{2\sigma^4(\nu_0+3)}{\nu_0} & 0 \\ & & \frac{(1-\varepsilon^2)(\nu_0+3)}{(3-16c^2)(\nu_0+1)} \end{pmatrix}. \quad (5.10)$$

**Dem.** La demostración se sigue desde propiedades conocidas de estimadores de máxima verosimilitud bajo condiciones de regularidad estándar (Lehmann, 1999).  $\square$

**Corolario 5.3** Sea  $Z \sim ESC(\mu, \sigma, \varepsilon)$ . Entonces el EMV  $\hat{\theta} = (\hat{\mu}, \hat{\sigma}^2, \hat{\varepsilon})$  es consistente y asintóticamente normal.

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{\mathcal{D}} N_3(0, I_0^{-1}) \text{ , cuando } n \rightarrow \infty,$$

donde  $V_0 = I_0^{-1}$  y es dada por

$$V_0 = \begin{pmatrix} \frac{6\sigma^2(1-\varepsilon^2)}{3-16c^2} & 0 & \frac{8c\sigma(1-\varepsilon^2)}{3-16c^2} \\ & 8\sigma^4 & 0 \\ & & \frac{2(1-\varepsilon^2)}{3-16c^2} \end{pmatrix}. \quad (5.11)$$

**Caso II.** Cuando los grados de libertad son desconocidos. Este caso es estudiado por Gómez, Torres y Bolfarine (2003), y los resultados más importantes de este trabajo son los siguientes.

**Proposición 5.7** Sea  $Z \sim ESt(\mu, \sigma, \varepsilon, \nu)$ . Entonces la matriz de información de Fisher  $I(\boldsymbol{\theta})$  es dada por

$$I(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\nu+1}{\sigma^2(1-\varepsilon^2)(\nu+3)} & 0 & -\frac{4c(\nu)(\nu+1)}{\sigma(1-\varepsilon^2)(\nu+3)} & 0 \\ & \frac{\nu}{2\sigma^4(\nu+3)} & 0 & -\frac{1}{\sigma^2(\nu+1)(\nu+3)} \\ & & \frac{3(\nu+1)}{(1-\varepsilon^2)(\nu+3)} & 0 \\ & & & -h(\nu) + \frac{1}{2(\nu+3)} \end{pmatrix}, \quad (5.12)$$

con  $h(\nu) = \frac{1}{4} \left( \frac{2}{\nu^2} + \psi' \left( \frac{1+\nu}{2} \right) - \psi' \left( \frac{\nu}{2} \right) \right)$  donde  $\psi$  es la función digamma.

Dem. Se utilizan propiedades conocidas de estimadores de máxima verosimilitud asintóticos. Ver Lehmann(1999).  $\square$

**Proposición 5.8** Sea  $Z \sim ESt(\mu, \sigma, \varepsilon, \nu)$ . Entonces el EMV  $\hat{\theta} = (\hat{\mu}, \hat{\sigma}^2, \hat{\varepsilon}, \hat{\nu})$  es consistente y asintóticamente normal.

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} N_4(0, V) \text{ , cuando } n \rightarrow \infty,$$

donde  $V = I(\theta)^{-1}$  y es dada por

$$V = \begin{pmatrix} V_{\mu\mu} & V_{\mu\sigma^2} & V_{\mu\varepsilon} & V_{\mu\nu} \\ & V_{\sigma^2\sigma^2} & V_{\sigma^2\varepsilon} & V_{\sigma^2\nu} \\ & & V_{\varepsilon\varepsilon} & V_{\varepsilon\nu} \\ & & & V_{\nu\nu} \end{pmatrix}, \quad (5.13)$$

donde

$$V_{\mu\mu} = \frac{1}{D} \frac{3\nu(\nu+1)(1-2h(\nu+3))}{4\sigma^4(\nu+3)(1-\varepsilon^2)}$$

$$V_{\sigma^2\sigma^2} = \frac{1}{D} \frac{(3-16c^2)(\nu+1)^2(1-2h(\nu+3))}{2\sigma^2(\nu+3)^3(1-\varepsilon^2)^2}$$

$$V_{\varepsilon\varepsilon} = \frac{1}{D} \frac{\nu(\nu+1)^2(1-2h(\nu+3))-4(\nu+3)(1-\varepsilon^2)}{4\sigma^6(\nu+1)(\nu+3)^4(1-\varepsilon^2)^2}$$

$$V_{\nu\nu} = \frac{1}{D} \frac{\nu(\nu+1)^2(3-16c^2)}{2\sigma^6(1-\varepsilon^2)(\nu+3)}$$

$$V_{\mu\varepsilon} = \frac{1}{D} \frac{c\nu(\nu+1)^2(1-2h(\nu+3))-4c(\nu+3)(1-\varepsilon^2)}{\sigma^5(\nu+1)(\nu+3)^3(1-\varepsilon^2)}$$

$$V_{\sigma^2\nu} = \frac{1}{D} \frac{(3-16c^2)(\nu+1)}{\sigma^4(\nu+3)^3(1-\varepsilon^2)^2}$$

$$V_{\mu\sigma^2} = V_{\mu\nu} = V_{\sigma^2\varepsilon} = V_{\varepsilon\nu} = 0$$

$$\text{con } D = \frac{[\nu(\nu+1)^2(1-2h(\nu+3))-4j[3-16c^2]]}{4\sigma^6(\nu+3)^3(1-\varepsilon^2)^2}.$$

**Dem.** Se utilizan propiedades conocidas de estimadores de máxima verosimilitud asintóticos. Ver Lehmann(1999).  $\square$

### 5.4.2 Ejemplos

Ahora aplicamos esta metodología a tres ejemplos. El primer conjunto de datos consiste en 219 alturas de volcanes usadas en artículo de Modholkar y Hutson (2000). El segundo ejemplo es tomado del libro Cook and Weisberg (1994) que analiza las mediciones de 202 atletas Australianos. Concretamente analizamos 100 medidas de concentración del plasma ferritin a atletas mujeres. El tercer ejemplo considera el conjunto de datos proveniente de la Facultad de Ingeniería, Departamento de Minas, Universidad de Atacama. Este conjunto de datos consiste en 86 muestras de rocas, en la región de Atacama-CHILE, donde se mide a cada roca la cantidad de Niquel que contiene cada una de ellas.

La tabla 5.4 presenta algunas estadísticas básicas de estos tres ejemplos. En los tres ejemplos se tiene asimetría positiva, y en los ejemplos dos y tres se tiene mayor asimetría y kurtosis que en el modelo *ESN*. Observamos en el primer ejemplo que la asimetría y kurtosis muestral se encuentran dentro de los rangos del modelo *ESN*.

Ejemplo	$n$	$\bar{X}$	$S^2$	$\gamma_1$	$\gamma_2$
1	219	70.246	1850.56	0.840	3.482
2	100	56.960	958.261	1.366	5.767
3	86	21.337	276.861	2.440	12.043

Tabla 5.4: Resumen de estadísticas descriptivas para cada uno de los conjuntos de datos.

Utilizamos el modelo estudiado epsilon-skew-t para modelar estos tres ejemplos, hacemos la estimación de los parámetros dando conocidos y desconocidos los grados de libertad.

Así utilizando la sección anterior para  $\nu$  conocido, se tiene en la tabla 5.5 que el primer ejemplo corresponde a un modelo epsilon-skew-normal, ya que sólo en el infinito se maximiza la log-verosimilitud. En la tabla 5.6 se tiene los resultados del ejemplo 2, y observamos que con 10 grados de libertad la log-verosimilitud es maximizada, en la tabla 5.7 se tiene el ejemplo 3 que con tres grados de libertad se maximiza la log-verosimilitud

En la tabla 5.8 entregamos un resumen de las estimaciones de máxima verosimilitud y resultados asintóticos para los cuatro parámetros del modelo en los tres ejemplos. Observamos que en los ejemplos dos y tres el modelo epsilon-skew-t es

$\nu_0$	$\mu$	$\sigma$	$\varepsilon$	log-verosimilitud
1	31.389	24.526	-0.609	-1157.081
5	30.075	33.952	-0.631	-1120.16
10	29.205	36.321	-0.645	-1117.178
30	28.233	38.282	-0.662	-1115.827
50	28.020	38.723	-0.665	-1115.636
100	27.803	39.065	-0.669	-1115.513
500	27.629	39.347	-0.672	-1115.426
$\infty$	<b>27.586</b>	<b>39.419</b>	<b>-0.673</b>	<b>-1115.406</b>

Tabla 5.5: *Estimaciones para el ejemplo 1, cuando  $\nu$  es conocido.*

$\nu_0$	$\mu$	$\sigma$	$\varepsilon$	log-verosimilitud
1	30.568	15.475	-0.621	- 485.428
5	28.099	22.274	-0.684	- 470.894
9	27.014	23.879	-0.711	-470.375
<b>10</b>	<b>26.840</b>	<b>24.116</b>	<b>-0.715</b>	<b>-470.363</b>
11	26.706	24.318	-0.718	-470.364
20	26.191	25.329	-0.731	-470.511
30	25.965	25.803	-0.736	-470.651
50	25.394	26.217	-0.749	- 470.802

Tabla 5.6: *Estimaciones para el ejemplo 2, cuando  $\nu$  es conocido.*

$\nu_0$	$\mu$	$\sigma$	$\varepsilon$	log-verosimilitud
1	14.520	5.438	-0.288	- 343.464
2	13.243	7.542	-0.375	-339.883
<b>3</b>	<b>11.841</b>	<b>8.741</b>	<b>-0.461</b>	<b>-339.704</b>
4	10.592	9.506	-0.531	- 340.000
5	8.431	10.075	-0.642	-340.287
10	7.166	11.280	-0.709	-341.621
20	4.237	12.149	-0.842	- 342.877
30	4.167	12.492	-0.846	-343.491

Tabla 5.7: *Estimaciones para el ejemplo 3, cuando  $\nu$  es conocido.*

Parámetros estimados	Ejemplo 1	Ejemplo 2	Ejemplo 3
$\mu$	27.713 (0.806)	26.782 (4.190)	12.265 (1.868)
$\sigma^2$	1537.382 (3.780)	585.672 (94.154)	70.433 (16.430)
$\varepsilon$	-0.670 (0.074)	-0.716 (0.100)	-0.435 (0.129)
$\nu$	$\geq 170$ (1.257)	10.401 (0.515)	2.651 (0.337)
log-verosimilitud	-1115.467	-470.362	-339.668

Tabla 5.8: *Estimaciones de máxima verosimilitud (desviación estándar en parentesis) para los tres ejemplos, cuando  $\nu$  es desconocido.*

significativo, ya que las estimaciones de  $\nu$  son pequeñas y sus desviaciones estándar también lo son, mientras que en el ejemplo 1 la estimación de  $\nu$  es tan grande que es mejor utilizar el modelo epsilon-skew-normal, usado en Mudholkar y Hutson (2000).

En las siguientes figuras se muestra el ajuste a los tres conjuntos de datos por los estimadores de máxima verosimilitud del modelo epsilon-skew-t y el modelo epsilon-skew-normal. Observamos que el modelo *ES*t es más apropiado para los datos presentados en el histograma en los ejemplos 2 y 3 que el modelo *ES*N .

Por otro lado, estudios recientes que utilizan algunos de estos modelos de la familia epsilon-skew son entregados en los trabajos de Gómez y Arellano (2004), Arellano, Bolfarine y Gómez (2004) y Arellano, Gómez y Quintana (2004c, 2004d).

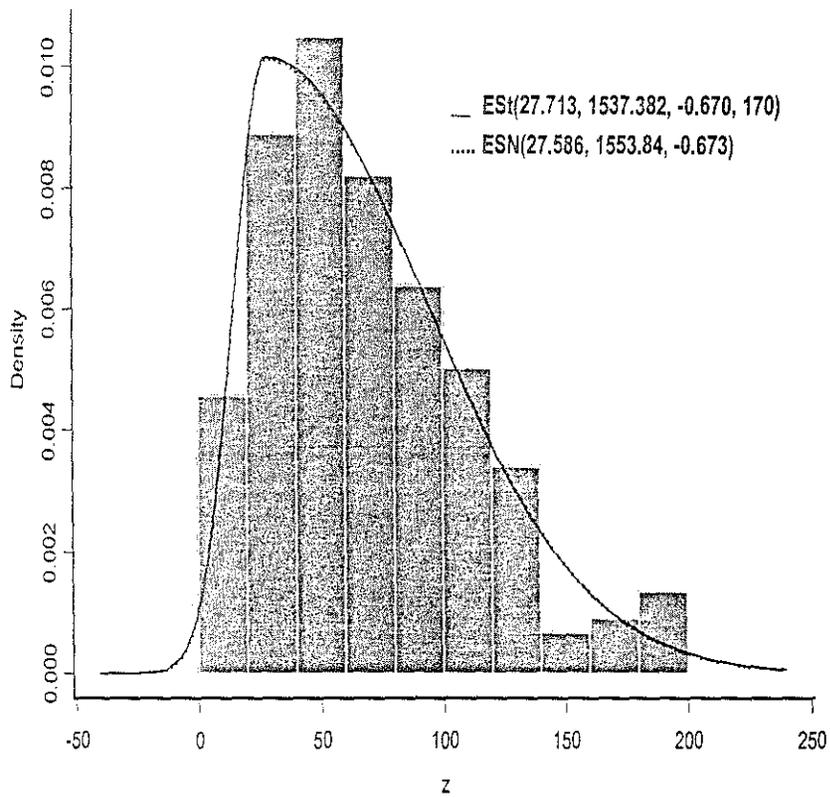


Figura 5.2: Distribución Epsilon-Skew-t versus Epsilon-Skew-Normal, para el ejemplo

1

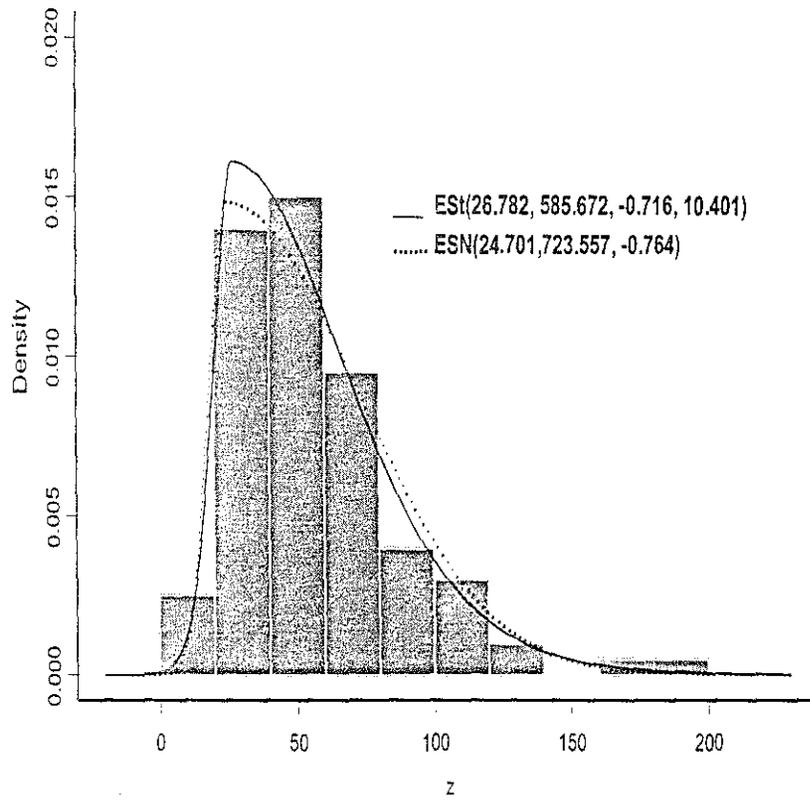


Figura 5.3: Distribución Epsilon-Skew-t versus Epsilon-Skew-Normal, para el ejemplo

2

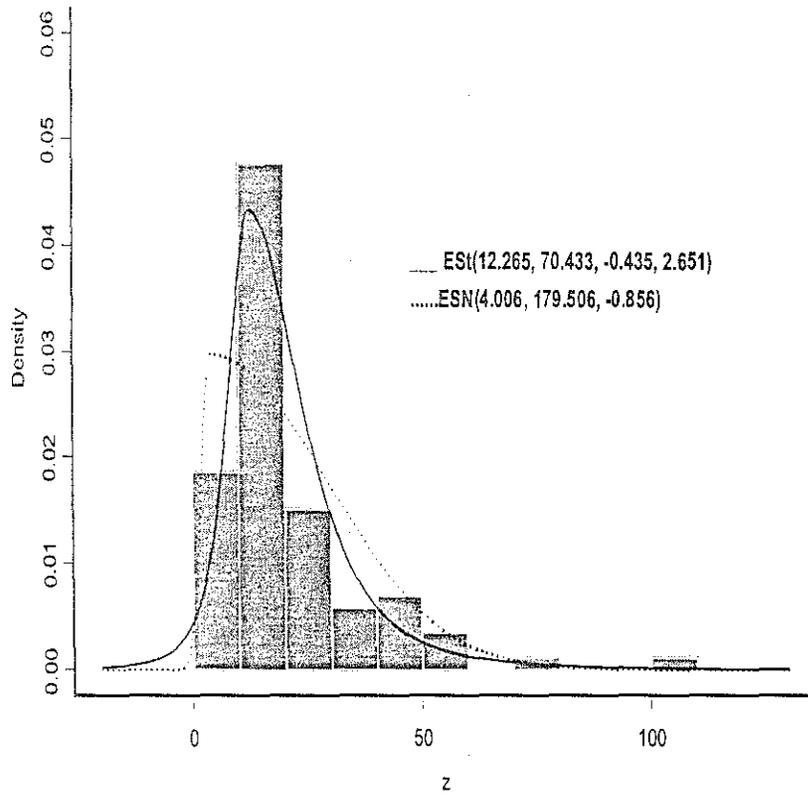


Figura 5.4: Distribución Epsilon-Skew-t versus Epsilon-Skew-Normal, para el ejemplo

3

# Capítulo 6

## Extensiones

En este capítulo se discuten nuevas extensiones de los modelos desarrollados en esta tesis. Estas nuevas extensiones, son inicios de trabajos recientes que aún no han sido sometido.

- En el artículo de Arellano, Gómez e Iglesias (2004) introducen una extensión multivariada del modelo *SNG* introducido por Arellano, Gómez y Quintana (2004b) y comentado en el tercer capítulo de esta tesis. La versión de la distribución skew-normal-generalizada multivariada es definida por:

$$f(\mathbf{x} | \lambda_1, \lambda_2) = 2^k \phi_n(\mathbf{x}) \Phi_k(\lambda_1 \mathbf{x} | I_n + \lambda_2 \mathbf{x} \mathbf{x}^T), \quad \mathbf{x} \in \mathbb{R}^k \quad (6.1)$$

en la cual contiene la fdp  $N_k(0, I_k)$  cuando  $\lambda_1 = \lambda_2 = 0$  y la fdp skew-normal-multivariada dada por  $2^k \phi_n(\mathbf{x}) \Phi_k(\lambda_1 \mathbf{x}, \cdot)$  cuando  $\lambda_2 = 0$ .

• La clase de distribuciones (4.1) puede ser extendidas como sigue. Considere  $U_\varepsilon$  como definidas en la Proposición 4.2, independiente de

$$V_\beta \sim \frac{f(x + \beta)}{1 - F(\beta)} I\{x \geq 0\},$$

y sea  $X = U_\varepsilon V_\beta$ . Así introducimos un nuevo tipo de densidades asimétricas, la cual denotamos por  $X \sim ESf2(\varepsilon, \beta)$ , distribución con dos parámetros  $|\varepsilon| < 1$  y  $\beta \in \mathbb{R}$ . Con esta notación tenemos inmediatamente que  $ESf(\varepsilon) \equiv ESf2(\varepsilon, 0)$ . Se puede mostrar que cuando  $\beta > 0$ , la distribución  $ESf2(\varepsilon, \beta)$  tiene más flexibilidad asimétrica que (4.1). Cuando  $\beta < 0$  las formas de las densidades  $ESf2(\varepsilon, \beta)$  son muy similares a la obtenidas usando (4.2).

• En el artículo de Arellano, Gómez y Quintana (2004a) introducen una extensión de la clase de distribuciones (4.1) considerando la variable discreta  $U_\alpha$ . La distribución de  $U_\alpha$  es completamente determinada por el parámetro de asimetría  $\alpha$ . La versión general de tal distribución es de la forma  $P(U_\alpha = a(\alpha)) = 1 - P(U_\alpha = -b(\alpha)) = \frac{a(\alpha)}{a(\alpha) + b(\alpha)}$ , donde  $a(\alpha) > 0$  y  $b(\alpha) > 0$ . La distribución skew puede ser generada usando las variables aleatorias de la forma  $X = U_\alpha \cdot V$ . Observamos que cuando  $a(\alpha) = 1 - \alpha$  y  $b(\alpha) = 1 + \alpha$ , con  $|\alpha| < 1$  se tiene la familia estudiada en el capítulo 4 de esta tesis y cuando  $a(\alpha) = \alpha$  y  $b(\alpha) = 1/\alpha$ , con  $\alpha > 0$  se tiene la familia estudiada por Fernández y Steel (1998).

• Otra extensión relevante es concerniente al caso multivariado. La idea básica

consiste en aplicar la construcción (4.1) a cada coordenada de la densidad  $f$  con la propiedad de ser simétrica en cada coordenada cuando una de ellas queda fija. Note primero que (4.1) puede reescribirse como  $h(x|\varepsilon) = f(x/(1 - \text{sgn}(x)\varepsilon))$ , donde  $\text{sgn}(x) = x/|x|$  si  $x \neq 0$  y  $\text{sgn}(0) = 0$ . Sea ahora  $f$  la densidad  $k$ -variada con la propiedad que  $f(x_i, \mathbf{x}_{-i}) = f(-x_i, \mathbf{x}_{-i})$  para todo  $i = 1, \dots, k$ ,  $x_i \in \mathbb{R}$ , y  $\mathbf{x}_{-i} \in \mathbb{R}^{k-1}$ , donde  $\mathbf{x}_{-i}$  denota el vector  $k$ -dimensional  $\mathbf{x}$  con la  $i$ -ésima coordenada removida. La clase de distribuciones skew  $k$ -variada con parámetro de asimetría  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_k) \in (-1, 1)^k$  es definida como

$$h(\mathbf{x}|\varepsilon) = f\left(\frac{x_1}{1 - \text{sgn}(x_1)\varepsilon_1}, \dots, \frac{x_k}{1 - \text{sgn}(x_k)\varepsilon_k}\right), \quad \mathbf{x} \in \mathbb{R}^k. \quad (6.2)$$

Denotemos al vector  $\mathbf{X} = (X_1, \dots, X_k)$  desde (6.2) como  $\mathbf{X} \sim ESM(f, k, \varepsilon)$ . En el caso bivariado (6.2) se tiene

$$\begin{aligned} h(x_1, x_2|\varepsilon) &= f\left(\frac{x_1}{1 + \varepsilon_1}, \frac{x_2}{1 + \varepsilon_2}\right) I\{x_1 < 0, x_2 < 0\} \\ &+ f\left(\frac{x_1}{1 + \varepsilon_1}, \frac{x_2}{1 - \varepsilon_2}\right) I\{x_1 < 0, x_2 \geq 0\} + f\left(\frac{x_1}{1 - \varepsilon_1}, \frac{x_2}{1 + \varepsilon_2}\right) I\{x_1 \geq 0, x_2 < 0\} \\ &+ f\left(\frac{x_1}{1 - \varepsilon_1}, \frac{x_2}{1 - \varepsilon_2}\right) I\{x_1 \geq 0, x_2 \geq 0\} \end{aligned}$$

La representación estocástica dada en la Proposición 4.2 puede también ser extendida naturalmente a la familia multivariada  $ESM(f, k, \varepsilon)$ . Usando esencialmente la misma demostración, se puede mostrar que un vector  $\mathbf{X} = (X_1, \dots, X_k) \sim ESM(f, k, \varepsilon)$

puede ser expresado como  $(U_1 \cdot V_1, \dots, U_k \cdot V_k)$  donde  $\mathbf{U} = (U_1, \dots, U_k)$  es independiente de  $\mathbf{V} = (V_1, \dots, V_k)$ . Por lo tanto, las componentes de  $\mathbf{U}$  son independientes con  $P(U_i = 1 - \varepsilon_i) = 1 - P(U_i = -1 - \varepsilon_i) = (1 - \varepsilon_i)/2$  para  $i = 1, \dots, k$  y  $\mathbf{V} \stackrel{d}{=} (|Y_1|, \dots, |Y_k|)$ , donde  $\mathbf{Y} = (Y_1, \dots, Y_k) \sim f$ . Basado en el trabajo de Arellano-Valle, del Pino y San Martín (2002), podemos decir que  $\mathbf{X}$  es un vector- $\mathcal{C}_\varepsilon$ , en la cual extiende la clase simétrica de vectores aleatorios- $\mathcal{C}$  e incluye asimetría. Muchas propiedades de la clase  $\mathcal{C}$  pueden ser fácilmente extendidas a la clase  $\mathcal{C}_\varepsilon$ . Por ejemplo, la distribución marginal de un vector aleatorio  $\mathcal{C}_\varepsilon$  pertenecen a la clase  $\mathcal{C}_\varepsilon$ .

# Bibliografía

- [1] Aigner, D. J., Lovell, C. A. K. and Schmidt, P. (1977), Formulation and estimation of stochastic frontier production function model. *J. Econometric* **12**, 21-37.
  
- [2] Andél, J., Netuka, I. and Zvára, K. (1984), On theshold autoregressive processes. *Kybernetika* **20**, 89-106.
  
- [3] Arellano-Valle, R. B. and del Pino, G. (2003), From Symmetric Distributions : A Unified Approach.
  
- [4] Arellano-Valle, R. B., del Pino, G. and San Martín, E. (2002), Definition and probabilistic properties of skew-distributions. *Statist. Prob. Letters* **58**, 111-121.
  
- [5] Arellano-Valle, R.B. and Genton, M. G. (2003), On Fundamental skew distributions. Technical Report, North Carolina State University, Institute of Statistics.

- [6] Arellano-Valle, R.B., Bolfarine, H. and Gómez, H.W. (2004). Epsilon Skew Normal Measurement Error Models.(Preprint)
- [7] Arellano-Valle, R.B., Gómez, H.W. and Iglesias, P. (2004). Shape mixtures of skewed-distributions: A Bayesian interpretation .(Preprint)
- [8] Arellano-Valle, R.B., Gómez, H.W. and Quintana, F.A. (2004a). Statistical Inference for a General Class of Asymmetric Distributions.To appear in *Journal of Statistical Planning and Inference*.
- [9] Arellano-Valle, R.B., Gómez, H.W. and Quintana, F.A. (2004b). A New Class of Skew-Normal Distributions. To appear in *Communications in Statistics : Theory and Methods*.
- [10] Arellano-Valle, R.B., Gómez, H.W. and Quintana, F.A. (2004c). A New Class of Skew-Generalized-Symmetrical Distributions. (Preprint)
- [11] Arellano-Valle, R.B., Gómez, H.W. and Quintana, F.A. (2004d). Regression Models with error Asymmetries. (Preprint)
- [12] Arnold, B.C. and Beaver, R.J. (2000b). The skew Cauchy distribution. *Statistics and Probability Letters*, **49**, 285-290.

- [13] Arnold, B.C. and Beaver, R.J. (2002). Skewed multivariate models related to hidden truncation and/or selective reporting. *Test*, **11**, 7-54.
- [14] Arnold, B.C., Beaver, R.J., Groeneveld, R. A., Meeker, W. Q. (1993). The non-truncated marginal of a truncated bivariate normal distribution. *Psychometrika*, **58**, 471-478.
- [15] Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist.* **12**, 171-8.
- [16] Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. *Statistica* **46**: 199-208.
- [17] Azzalini, A. & Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society*, **B**, **61**, 579-602.
- [18] Azzalini, A. & Capitanio, A. (2003). Distributions generate by perturbation of symmetry with emphasis on a multivariate skew-t distribution. *Journal of the Royal Statistical Society*, **B**, **65**, 367-389.
- [19] Azzalini, A. & Dalla Valle, A. (1996). The multivariate skew normal distribution. *Biometrika* **83**, 715-26.

- [20] Branco, M. and Dey, D. (2001). A general class of multivariate elliptical distributions. *Journal of Multivariate Analysis* **79**: 99-113.
- [21] Cambanis, S., Huang, S. and Simons, G. (1981). On the theory of elliptical contoured distributions. *J. Multivariate Anal.*, 11, 368-385.
- [22] Cook, R.D. and Weisberg, S. (1994). An introduction to regression graphics. Wiley.
- [23] Ellison, B. (1964), Two theorems for inferences about the normal distribution with applications in acceptance sampling. *Journal of the American Statistical Association* **59** , 89-95.
- [24] Fang, K. T., Kotz, S. and Ng, K. W. (1990). Symmetric Multivariate and Related Distributions. Chapman and Hall, London.
- [25] Fernández, C., Osiewalski, J. and Steel, M.F.J. (1995). Modelig and Inference With  $v$ -Spherical Distributions. *Journal of the American Statistical Association*, **90**, 1331-1340.
- [26] Fernández, C. and Steel, M.F.J. (1998). On Bayesian modeling of tails and skewness. *Journal of the American Stastical Association* **93**, 359-371.

- [27] Genton, M. G., Ho, L. and Liu, X. (2001), Moments of skew-normal random vectors and their quadratic forms. *Statist. Prob. Letters* **51**, 319-325.
- [28] Genton, M. G. and Loperfido, N. M. R. (2002), Generalized skew-elliptical distributions and their quadratic forms. Technical Report.
- [29] Gilks, W. R. and Wild, P. (1992), Adaptive rejection sampling for Gibbs sampling. *Journal of Applied Statistics*, **41**, 337-348.
- [30] Gómez, H. W. and Arellano, R. B. (2004), A extension of the skew-normal and epsilon-skew-normal models.(Preprint)
- [31] Gómez, H. W., Torres, F. J. and Bolfarine, H. (2003), Large-Sample Inference for the Epsilon-Skew-t. Relatório Técnico, Universidade de Sao Paulo, Instituto de Matemática e Estatística.
- [32] Gupta, A. K., Chang, F. C. and Huang, W. J., (2002). Some skew-symmetric models. *Random Operators Stochastic Equations.*, **10**, 133-140.
- [33] Gupta, R. C. and Chen, T. (2001). Goodness-of-fit tests for the skew-normal distribution. *Commun. Statist-Simula.*, **30**(4), 907-930.
- [34] Henze, N. (1986). A probabilistic representation of the skew-normal distribution. *Scand. J. Statist.* **13**, 271-5.

- [35] Johnson, N., Kotz, S. L. and Balakrishnan, N. (1995). Continuous univariate distributions. Second Edition, New York : Wiley & Sons.
- [36] Jones, M. C. and Faddy, M. J. (2003). A skew extension of the t distribution, with applications. *J. R. Statist. Soc. B*, **65**, 159-174.
- [37] Lehmann, E.L. (1999). Elements of Large-Sample Theory, New York: Springer.
- [38] Liseo, B. (1990). La classe delle densita normali sghembe: aspetti inferenziali da un punto di vista bayesiano. *Statistica.*, **L**, 59-70.
- [39] Liseo, B. and Loperfido, N. (2003). A Bayesian interpretation of the multivariate skew-normal distribution. *Statist. Prob. Letters*, **61**, 395-401.
- [40] Mudholkar, G.S. and Hutson, A.D. (2000). The epsilon-skew-normal distribution for analyzing near-normal data. *Journal of Statistical Planning and Inference*. **83**, 291-309.
- [41] O'Hagan, A. & Leonard, T. (1976). Bayes estimation subject to uncertainty about parameter constraints. *Biometrika*, **63**, 201-3.
- [42] Pewsey, A.(2000). Problems of inference for Azzalini's skew-normal distribution. *Journal of Applied Statistics*, Vol. 27, N° 7, 859-870.

- [43] Rotnovisky, A., Cox, D.R., Bottai, M. and Robbins, J. (2000). Likelihood-based inference with singular information matrix. *Bernoulli*, 6, 2, 243-284.
- [44] Tanner, M. A. (1996), Tools for Statistical Inference. Methods for the Exploration of Posterior Distributions and Likelihood Functions. Third Edition. New York: Springer-Verlag.