DISTRIBUCIÓN SKEW-NORMAL: IDENTIFICABILIDAD Y REDUCCIÓN, Y ENFOQUE BAYESIANO DE MEZCLAS

Luis Mauricio Castro Cepero

SOMETIDA COMO REQUISITO PARA
OPTAR AL GRADO DE
DOCTOR EN ESTADÍSTICA
EN LA
PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
SANTIAGO, CHILE
14 DE AGOSTO DE 2008

© Copyright Luis Mauricio Castro Cepero, 2008

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE DEPARTAMENTO DE ESTADÍSTICA

Los firmantes certifican que han leído y recomendado a la Facultad de Facultad de Matemáticas la aceptación de la Tesis titulada "Distribución Skew-Normal: Identificabilidad y Reducción, y Enfoque Bayesiano de Mezclas" escrita por Luis Mauricio Castro Cepero como requerimiento para optar al grado de Doctor en Estadística.

	Fecha: 14 de Agosto de 2008
Examinador Externo:	Rosangela Loschi
	Rosangera Dosem
Orientadores:	Division Co. M. W.
	Ernesto San Martín
	Reinaldo Arellano-Valle
Comité Examinador:	
	Alicia Carriquiry
	Héctor Gómez

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

Fecha: 14 de Agosto de 2008

Autor:

Luis Mauricio Castro Cepero

Título:

Distribución Skew-Normal: Identificabilidad y Reducción,

y Enfoque Bayesiano de Mezclas

Departmento: Estadística

Grado: Doctor

Convocatoria: Agosto

Año: 2008

Se le concede permiso a la Pontificia Universidad Católica de Chile para hacer circular y copiar, con propósitos no comerciales, este documento según lo requieran personas y/o instituciones.

Firma del Autor

EL AUTOR SE RESERVA LOS DERECHOS DE OTRAS PUBLICACIONES. LA TESIS NI EXTRACTOS EXTENSOS DE ELLA PUEDEN SER IMPRESOS O REPRODUCIDOS SIN EL PERMISO ESCRITO DEL AUTOR.

EL AUTOR CERTIFICA QUE EL PERMISO SE HA OBTENIDO PARA EL USO DE CUALQUIER MATERIAL REGISTRADO QUE APAREZCA EN ESTA TESIS (CON EXCEPCIÓN DE LOS EXTRACTOS QUE REQUIEREN SOLAMENTE EL RECONOCIMIENTO APROPIADO DE LOS ESTUDIANTES) Y QUE TODO USO ESTÉ RECONOCIDO CLARAMENTE.

Para Jenny, con mucho amor.

Gracias por todos estos años a mi lado!!!

"Todo lo puedo en Cristo que me fortalece"

Filipenses, 4:13

Tabla de Contenidos

T	abla c	de Contenidos	Ţ
Íı	ndice	de Tablas	VIII
Ĺ	ndice	de Figuras	IX
A	bstra	ct	X
A	grade	cimientos	XI
Ir	ıtrodu	acción	1
1.	Ider	ntificabilidad de la distribución skew-normal	6
	1.1.	Introducción	7
	1.2.	Distribución skew-normal	10
		Identificabilidad en los modelos estadísticos	
		Identificabilidad de la distribución skew-normal	
		1.4.1. Identificabilidad en el caso multivariado (\mathbb{R}^d , $d \geq 1$)	17
		1.4.2. Identificabilidad y construcción de la distribución skew-normal	
	1.5.	Aplicaciones de los resultados de identificabilidad a otras distribuciones	
		skew-normal	21
	1.6.	Resumen y conclusiones	
	1.7.	Agradecimientos	
	1.8.	Apéndice	
2.	Redi	ucción probabilística de la distribución skew-normal	28
	2.1.	Introducción	
	2.2.	Reducción probabilística	32
	2.3.	Distribución condicional skew-normal	36

	2.4	. It districts productional as in districts of the first terms of the	40
	2.5 2.6		46
	2.0		49
	2.1	Agradecimientos	50
3	. Inf	erencia Bayesiana en la distribución skew-normal	51
	3.1		52
	3.2.	•	55
		3.2.1. Distribución a priori general para el parámetro de asimetría	55
		1 1	59
	3.3.	Inferencia Bayesiana en la distribución skew-normal de localización-escala	65
		1 0 1	66
		3.3.2. Distribución a priori skew-normal para el parámetro de asimetría	67
	3.4.	Punto de cambio en el parámetro de asimetría	69
		<u> </u>	69
			71
	3.5.	Análisis de punto de cambio: datos simulados y de retornos de mercados	
			73
		3.5.1. Análisis de los datos simulados	74
			76
	3.6.	Resumen y conclusiones	79
	3.7.	Agradecimientos	30
	3.8.	Apéndice	31
4	Info	uonois Daviosiana on maralas da distribusianas abarr u correl	
1.	4.1.	•	36
	4.1.		37
	4.2.		1
		4.2.1. Definición y propiedades	
	4.3.		
		Mezclas de asimetría de distribuciones skew-normal multivariadas 10	
	4.4.	Inferencia Bayesiana en modelos de regresión skew-normal	
		4.4.1. Función de verosimilitud marginal de (β, ω^2) cuando $\lambda \sim \mathcal{N}(a, b^2)$ 100	
	1 5	4.4.2. Muestreo de Gibbs para el modelo de regresión skew-normal 108	3
	4.5.	Inferencia Bayesiana para mezclas de asimetría en modelos de regresión	
		lineal	
		4.5.1. Distribución skew-normal-generalizada	
		4.5.2. Distribución skew-normal-t	

	4.5.3. Distribución skew- t -normal	. 116
4.6.	Aplicación a los datos de atletas australianos	. 118
4.7.	Resumen y conclusiones	. 122
4.8.	Agradecimientos	. 123
Conclusiones generales		125
Bibliogr	rafía	128

Índice de cuadros

2.1.	Parámetros estimados vía máxima verosimilitud del modelo de regresión
	skew-normal
2.2.	Parámetros estimados vía máxima verosimilitud del modelo de regresión
	skew-normal, cuando la distribución de X_i es normal, $i=1\ldots,n$ 48
2.3.	Parámetros estimados vía máxima verosimilitud del modelo de regresión
	skew-normal, cuando la distribución de X_i es skew-normal, $i=1\ldots,n\ldots$ 48
3.1.	Estimaciones a posteriori de los parámetros del modelo de punto de cambio
	para datos simulados
3.2.	Estimaciones a posteriori de los parámetros del modelo de punto de cambio
	para los índices de los mercados Latino Americanos
4.1.	Estadísticas descriptivas para el conjunto de datos de atletas australianos:
	media muestral \overline{z} , desviación estándar muestral s , y coeficientes muestrales
	de asimetría y curtosis $\sqrt{b_1}$ y b_2 , respectivamente
4.2.	Estimaciones Bayesianas para el conjunto de datos de atletas australianos
	bajo el modelo de regresión lineal skew-normal
4.3.	Estimaciones Bayesianas para el conjunto de datos de atletas australianos
	bajo los modelos de regresión skew-normal-generalizado (\mathcal{SGN}), skew-
	normal- t (SNT), y skew- t -normal (STN)

Índice de figuras

3.1.	Datos simulados con distribución skew-normal	75
3.2.	Distribución a posteriori de λ_1 (línea sólida), λ_2 (línea punteada), K, ξ y ω .	76
4.1.	QQ-plot normal para los residuales estandarizados bajo el modelo de re-	
	gresión lineal normal.	120

Abstract

El presente trabajo busca responder, en primer lugar, en qué sentido las "extensiones" de la distribución skew-normal extienden la distribución skew-normal de Azzalini y cuál es el significado estadístico de los parámetros de la distribución skew-normal. Ambas preguntas son aclaradas a través del análisis de identificabilidad de la distribución skew-normal.

En segundo lugar, se estudia el enfoque de reducción probabilística, con el fin de generar el modelo de regresión lineal skew-normal. Para esto se analiza la descomposición marginal-condicional de la densidad conjunta skew-normal y se estudia la relación entre los espacios paramétricos asociados a los modelos marginal y condicional. Se proponen condiciones necesarias y suficientes sobre los parámetros del modelo conjunto con el fin de generar una variable exógena.

En tercer lugar, se considera la inferencia Bayesiana para los parámetros de localización, escala y asimetría de la distribución skew-normal, asumiendo diferentes distribuciones a priori. Representaciones estocásticas de las distribuciones a posteriori del parámetro de asimetría son obtenidas, bajo ciertas especificaciones de la función de verosimilitud. Los resultados obtenidos son aplicados en el contexto del problema de identificación de un punto de cambio en el parámetro de asimetría en una serie de datos determinada.

Finalmente, se desarrolla el análisis Bayesiano de los modelos de regresión skewnormal, skew-normal-generalizado, skew-normal-t y skew-t-normal bajo determinadas especificaciones de las distribuciones a priori de los parámetros de dichos modelos. Se propone una conveniente representación jerárquica de estos modelos, realizando el análisis de las distribuciones a posteriori e ilustrando los resultados con un conjunto de datos reales.

Agradecimientos

Cuando comencé este largo camino, allá por el año 2004, supe que llegar a la meta no sería fácil. Sin embargo, hubiera sido menos fácil sin personas como las que nombraré a continuación y que merecen mi más sincero agradecimiento.

Primero quiero agradecer a Dios Todopoderoso y a la Virgen Santísima por todas las bendiciones que colmaron mi hogar durante estos años. Todo lo que soy y lo que tengo es obra de su Divina Gracia y Eterna Misericordia.

Quiero agradecer muy especialmente al Profesor Dr. Ernesto San Martín, quien me enseñó a pensar como "estadístico" durante todos estos años y me motivó a profundizar en el estudio de los fundamentos de la estadística. Gracias Profesor San Martín por confiar en mi para el desarrollo de esta tesis. He sido muy afortunado al haberle tenido como maestro y poder aprender de sus enseñanzas dentro del campo de la estadística, matemática, filosofía y educación. Además quiero hacer extensivo este agradecimiento a su esposa e hijos, quienes me hicieron sentir como en casa en esta larga estadía en Santiago de Chile.

También deseo agradecer de manera también especial al Profesor y amigo Dr. Reinaldo Arellano, quien me animó a seguir los estudios de Doctorado y confió en mi para el desarrollo de los temas de la tesis. Gracias Reinaldo por enseñarme a ser independiente y afrontar las dificultades sin declinar. Gracias por la paciencia que tuviste cuando las cosas no salían. Y gracias por la especial amistad que hemos entablado durante todo este tiempo.

Al Profesor Dr. Guido del Pino, quien me dió ánimo para seguir luchando frente a mis limitaciones evidenciadas en el Doctorado. Don Guido me enseño a perseverar y sacar lo mejor de mi.

A la Dra. Pilar Iglesias, que me dió la oportunidad de trabajar a su lado, alentándome a seguir incluso cuando las fuerzas la abandonaban.

A la Dra. Rosangela Loschi de la Universidade Federal de Minas Gerais y al Dr. Marc Genton de University of Geneva por la oportunidad que me dieron de trabajar en conjunto.

A todos los profesores y personas que forman parte del Departamento de Estadística

de la Facultad de Matemática de la Pontificia Universidad Católica de Chile. Tengo en mi corazón un especial recuerdo de cada uno.

También deseo agradecer a todos los integrantes del Departamento de Estadística de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Concepción por el importante apoyo recibido en este último tiempo.

También agradezco a todos los integrantes del Centro de Estudios y Desarrollo de Talentos de la Pontificia Universidad Católica de Chile (PENTA UC).

Además quiero agradecer a la Comisión Nacional de Investigación Científica y Tecnológica de Chile (CONICYT) por el financiamiento recibido para realizar mis estudios de Doctorado.

Un agradecimiento especial se lo quiero hacer a mi familia en Perú. Primero a mis padres, Lucho y Ana. A ellos les agradezco por darme la fuerza que necesitaba para terminar el largo recorrido del Doctorado y por que me animaron a no rendirme ni bajar los brazos ante la adversidad. Sin ellos no sería nada de mi hoy. También quiero agradecer, con todo el corazón, a mi hermana Viviana y a su esposo Raúl. Ellos siempre estuvieron a mi lado en los momentos difíciles. Muchísimas gracias por todos los consejos que me dieron.

También quiero agradecer a mis cuatro abuelos que, gracias a Dios, pueden ser testigos de este logro. A Iris, Agustín, Lila y Eduardo. A todos mis tíos, tías y primos de las Familias Castro y Cepero. De todos ellos tengo recuerdos muy lindos en estos momentos. Quiero hacer una especial mención a mis padrinos Oscar y Mirtha. Ellos me ayudaron cuando más lo necesitaba. Definitivamente, sin su apoyo, en todo el sentido de la palabra, no hubiera podido terminar esta empresa.

Quiero agradecer a mis suegros, cuñados, cuñadas, sobrinos y en general a todos los miembros de las Familias Flores y Troncos, quienes viven en el norte del Perú.

No puedo dejar de mencionar a todos mis amigos aquí en Chile. Gracias a ellos pude sentirme como en casa. Gracias a la Señora María Roque que me adoptó como un hijo. Gracias Señora María por ayudarme desinteresadamente hasta el día de hoy. También quiero agradecer a todos y cada uno de los integrantes de la Familia Araya. Sin duda, el cariño que recibí de cada uno de ellos me sirvió para recargar el ánimo cuando este se me agotaba. También a los miembros de la Familia Mercado-Olivares, en especial a Don Osvaldo y a Doña Angélica. Siempre supieron darme una palabra de aliento en cada momento difícil.

A mis amigos del Postgrado de Estadística y Matemática de la Pontificia Universidad Católica de Chile, en especial para Gabriela, Michel, Natalia, Jorge, Ximena, Maria Paz y Aldo. Quiero hacer una mención especial para mi compañera de oficina Karla Muñoz. A ella le agradezco profundamente por toda la ayuda que me brindó en el desarrollo de esta

tesis, además de la paciencia que tuvo conmigo al compartir oficina por cuatro largos años.

También deseo agradecer a mis amigos Gabriel Gutierrez, Mansaku Maeda y familia, Héctor Gómez y familia, Ignacio Vidal y familia y Rolando de la Cruz y familia; por los consejos que me dieron y los buenos momentos que vivimos juntos. Quiero hacer un agradecimiento especial para mi amigo, compatriota y prácticamente hermano David Sirlopú. David, de ti aprendí a tener paciencia y a tomar las cosas con calma. Además, gracias a ti estos años se me hicieron mucho más llevaderos y entretenidos.

Finalmente quiero agradecer y dedicar este trabajo a mi esposa Jenny. Gracias Jenny por toda la paciencia que has tenido conmigo durante estos últimos años. Tus lecciones de vida fueron provechosas para mi. Gracias a ti aprendí mucho más de la vida, del amor y del respeto. Aprendí a reir y ser alegre. Aprendí a ser mucho más hombre. Y lo más importante, aprendí que más allá de todo, está el amor que nos tenemos.

Santiago, Chile 14 de agosto de 2008 Luis Mauricio Castro Cepero

Introducción

En los últimos años, el estudio de modelos estadísticos que introducen parámetros que representan asimetría y curtosis ha sido el foco de atención de la investigación estadística. Esta tendencia se justifica toda vez que algunos conjuntos de datos, en diferentes áreas de la ciencia como la medicina, biología, economía o finanzas, presentan evidencias de sesgo o colas pesadas en su función de distribución empírica.

Muchos de estos modelos se generan a partir de funciones de densidad de variables aleatorias con distribución simétrica que son multiplicadas por otras funciones, llamadas funciones de asimetría, las cuales dependen de ciertos parámetros. Otros tantos se generan a partir de la reducción por condicionamiento de una función de densidad conjunta simétrica. En ambos casos, el ejemplo más popular de este tipo de modelos encontrado en la literatura estadística, es la distribución skew-normal introducida por Azzalini [11] en 1985.

La distribución skew-normal de Azzalini [11] ha sido estudiada en detalle tanto desde la perspectiva clásica cuanto la Bayesiana. Algunas de estas propuestas se pueden encontrar en Azzalini y Capitanio [13], Genton [48], Azzalini [12], Arellano-Valle y Azzalini [2] y demás referencias dentro de estos artículos.

Paralelamente, muchos otros autores han centrado esfuerzos en estudiar distribuciones asimétricas que extienden la distribución skew-normal introducida por Azzalini [11], asi como sus propiedades. En algunos casos, el proceso de construcción de estas otras distribuciones es similar al mencionado anteriormente.

Sin embargo, estas extensiones de la distribución skew-normal de Azzalini [11] presentan algunos problemas de identificabilidad (ver por ejemplo Arellano-Valle y Azzalini [2] para una discusión sobre el tema). Aunque este hecho no ha sido estudiado formalmente, estos problemas se presentan debido a que la parametrización utilizada en las funciones de densidad de estas extensiones corresponde a la parametrización de la distribución simétrica original antes de realizar la reducción por condicionamiento. La falta de identificabilidad de un modelo estadístico origina que el uso del mismo se limite en lo que a la aplicación con datos reales se refiere. Por el contrario, un modelo identificado y correctamente especificado permite su uso en diversas aplicaciones y desde diferentes perspectivas tanto clásicamente cuanto Bayesianamente.

En este contexto, la presente investigación tiene dos objetivos principales: el primero busca estudiar el significado estadístico de los parámetros de la distribución skew-normal a través de un estudio formal de la identificación de los parámetros de la misma, tomando como punto de partida la generación de esta distribución vía la reducción por condicionamiento de un modelo conjunto normal. El segundo objetivo principal es realizar aplicaciones, por un lado clásicas de la distribución skew-normal desde el punto de vista de la

reducción de experimentos condicionales y por el otro, Bayesianas realizando inferencia en el parámetro de asimetría y en los parámetros de una distribución obtenida a partir de una mezcla de distribuciones skew-normal en el parámetro de asimetría.

Por tal razón, el Capítulo 1 de la investigación aborda el problema de identificación y la interpretación estadística de los parámetros de la distribución skew-normal. Se estudia con detalle la identificación de los parámetros que indexan la distribución de probabilidad skew-normal determinando en que casos las parametrizaciones son o no identificadas y explicitando el significado estadístico que tiene cada uno de los parámetros, en especial de aquél que modela la asimetría de los datos. Se determina el parámetro mínimo suficiente que indexa la distribución skew-normal y basado en él, se propone una parametrización identificada. Además, los resultados son aplicados a determinadas extensiones de la distribución skew-normal propuestas en la literatura estadística.

El Capítulo 2 busca estudiar la reducción probabilística de la distribución conjunta skew-normal basada en la distribución condicional. El objetivo de esta parte de la investigación es basar la inferencia del modelo skew-normal en aquél modelo que subyace a una reducción por condicionamiento, es decir, el modelo de regresión lineal. La hipótesis considerada en este capítulo es que las variables aleatorias siguen una distribución skew-normal y la reducción es estudiada en base a la descomposición marginal-condicional de la misma. El capítulo propone imponer condiciones necesarias y suficientes sobre los parámetros del modelo conjunto con el fin de generar un corte, y por lo tanto, una variable exógena. El

resultado principal es la caracterización de una densidad conjunta skew-normal que garantiza que los espacios paramétricos de los modelos marginal y condicional que genera no presentan relaciones funcionales entre ellos. Al finalizar, se propone un estudio de simulación con el objetivo de ilustrar las consecuencias de no tener una variable exógena en la especificación del modelo condicional skew-normal.

Finalmente, los Capítulos 3 y 4 estudian los modelos skew-normal a través de las mezclas de estas distribuciones en el parámetro de asimetría y bajo la perspectiva Bayesiana. Específicamente, el Capítulo 3 estudia el caso en el cual se considera una distribución general de mezclas, es decir, una distribución a priori. Se obtienen expresiones para la densidad predictiva y posteriori del parámetro de asimetría, además de los momentos a posteriori. El resultado principal es la obtención de una representación estocástica para la distribución a posteriori del parámetro de asimetría, la cual permite implementar eficientemente el algoritmo de muestreo de Gibbs para la generación de muestras aleatorias de la distribución antes mencionada. Una aplicación interesante mostrada en este capítulo es la relacionada con la detección de puntos de cambio en el parámetro de asimetría. El Capítulo 4 por su parte, muestra la inferencia Bayesiana para un modelo generado cuando la distribución de mezclas del parámetro de asimetría es normal. El modelo es llamado skew-generalizadonormal y ha sido estudiado en detalle por Arellano-Valle *et al.* [10]. El resultado principal en esta parte de la investigación es la especificación de una conveniente representación

jerárquica para este modelo, la cual permite la implementación del muestreo de Gibbs para obtener realizaciones de la distribución posterior. Además, se estudian distribuciones de mezcla del parámetro de asimetría sobre \mathbb{R}^+ , específicamente distribuciones gamma, obteniéndose otros modelos asimétricos que también son ajustados através de una representación jerárquica específica. La aplicación de los resultados obtenidos en este capítulo se realiza con un conjunto de datos reales.

Capítulo 1

Identificabilidad de la distribución skew-normal

"El proposito de este artículo es sugerir la reformulación del problema de especificación, apropiado a muchas aplicaciones de los métodos estadísticos, y poner énfasis en la aparición de un nuevo grupo de problemas, llamados problemas de identificación."

Tjalling Koopmans y Olav Reiersøl, 1950

El presente capítulo busca responder las siguientes dos preguntas estrechamente relacionadas: ¿en qué sentido, las "extensiones" de las distribuciones skew-normal extienden la distribución skew-normal de Azzalini? y, ¿cuál es el significado estadístico de los parámetros de la distribución skew-normal?. Estas preguntas son rigurosamente aclaradas a través del análisis de identificabilidad de la distribución skew-normal. El análisis se centra en la distribución skew-normal y no en la distribución normal de la cual la primera se obtiene como una reducción por condicionamiento.

1.1. Introducción

En los últimos años, el desarrollo de familias de distribuciones de probabilidad y el estudio de sus propiedades ha sido el centro de interés de la literatura estadística. En la práctica, muchas de estas distribuciones permiten modelar conjuntos de datos que presentan cierta evidencia de asimetría o colas pesadas a través de su distribución de probabilidad empírica, incluyendo parámetros que representan asimetría o curtosis respectivamente. Este hecho es posible gracias al significado estadístico que estos parámetros tienen en tales modelos.

Una de estas familias de distribuciones, la cual ha tenido un gran desarrollo en estos años, es la familia skew-normal. Introducida por Azzalini [11] en 1985, esta familia incluye la distribución normal como un caso particular y permite modelar la asimetría presente en un conjunto de datos. La correspondiente función de densidad de esta distribución de probabilidad está dada por

$$f_Z(z) = 2\phi(z)\Phi\left(\frac{\delta z}{\sqrt{1-\delta^2}}\right), \qquad z \in \mathbb{R},$$
 (1.1.1)

donde $\phi(\cdot)$ y $\Phi(\cdot)$ denotan la función de densidad y la función de distribución acumulada de la distribución $\mathcal{N}(0,1)$ respectivamente. El parámetro $\delta \in (-1,1)$ es considerado como el parámetro que regula la asimetría; en particular, cuando $\delta = 0$, (1.1.1) corresponde a la distribución normal estándar. En el caso de considerar una transformación lineal afin $Y = \xi + \omega Z$, se obtiene una distribución skew-normal indexada por los parámetros $(\xi, \omega^2, \delta) \in$

 $\mathbb{R} \times \mathbb{R}^+ \times (-1,1)$. Aquí, el significado estadístico de ξ y ω es asegurado por la propia definición de la transformación afín, es decir, ξ corresponde a un parámetro de localización, mientras que ω representa un parámetro de escala. La versión multivariada de esta densidad es estudiada en detalle por Azzalini y Dalla Valle [15], y Azzalini y Capitanio [13].

Recientemente, se han introducido numerosas extensiones de la distribución skewnormal de Azzalini [11] como por ejemplo, la distribución hierarchical skew-normal (Liseo y Loperfido [59]), la closed skew-normal (Gonzáles-Farías et al. [51]), la fundamental skew-normal (Arellano-Valle y Genton [8]). Más aún, Arellano-Valle y Azzalini [2] unifican estas distribuciones en la llamada, unified skew-normal. Sin embargo, y desde el punto de vista de la modelación estadística, la siguente pregunta parece tener un especial interés: ¿en qué sentido las "extensiones" de la distribución skew-normal extienden realmente la distribución skew-normal de Azzalini?. Para dar respuesta a esta pregunta es importante tener en cuenta que, dado que un modelo estadístico es una familia de distribuciones de muestreo indexadas por un parámetro (véase Fisher [40]; Basu [23]; Cox y Hinkley [32]; Lehman [58]; Bernardo y Smith [25]; McCullaugh [66]), una extensión de un modelo estadístico involucra parametros los cuales, por un lado, no están considerados en el modelo estadístico inicial y, por el otro, tienen una clara interpretación estadística. Este último aspecto no solo justifica la existencia de la extensión del modelo sino que se convierte en un aspecto de absoluta relevancia para las aplicaciones estadísticas en las que el modelo extendido pueda trabajar. Entonces, la pregunta planteada anteriormente está estrechamente

relacionada con el significado estadístico de los parámetros que indexan las "extensiones" de las distribución skew-normal.

De acuerdo con el principio de verosimilitud, el significado estadístico de una parametrización debe estar basado en la información de muestreo, la cual está representada por la familia de distribuciones de muestreo indexadas por tal parametrización (para más detalles véase Basu [23]). Consecuentemente, la relación entre información de muestreo y parametrización debe ser uno a uno. Claramente, esto último corresponde a la definición de parametrización identificada. De lo anterior se puede concluir que una parametrización no identificada no puede ser estadísticamente interpretable, y por lo tanto, es poco útil en aplicaciones.

Por ejemplo, considérese el modelo de análisis de varianza de una sola vía (ANOVA), el cual es especificado como sigue:

$$X_{ij} \sim \mathcal{N}(\xi_j, \omega^2), \qquad \xi_j = \alpha + \theta_j,$$
 (1.1.2)

con $i=1,\ldots,n_j$ y $j=1,\ldots,m$. El correspondiente modelo estadístico es indexado por $(\alpha,\theta,\omega^2)\in\mathbb{R}\times\mathbb{R}^m\times\mathbb{R}^+$, donde $\theta=(\theta_1,\ldots,\theta_m)$. Sin restricciones de identificación estos parámetros no serían estadísticamente interpretables, o en otras palabras, su interpretación estadística depende de las restricciones de identificación. Por lo tanto, si $\theta_1=0$, entonces $\alpha=\xi_1=E(X_1|\alpha,\theta,\omega^2)$, es decir, la media de la primera población, mientras que $\theta_j=\xi_j-\xi_1=E(X_j|\alpha,\theta,\sigma^2)-E(X_1|\alpha,\theta,\sigma^2)$, es decir, la diferencia entre la media de la población j-ésima y la media de la primera población. Sin embargo,

si $\sum_{j=1}^{m} \theta_{j} = 0$, entonces $\alpha = \bar{\xi} = \frac{1}{n} \sum_{j} E(X_{j} | \alpha, \theta, \omega^{2})$ representa la media general, y $\theta_{j} = \xi_{j} - \bar{\xi} = E(X_{j} | \alpha, \theta, \omega^{2}) - \bar{\xi}$, es decir, θ_{j} corresponde a la media de la j-ésima población corregida por la media general. Por lo tanto, la interpretación de los parámetros depende de la restricción de identificación; una vez que una restricción es adoptada, la información de muestreo proporciona el significado de la parametrización.

La pregunta de investigación planteada al inicio del capítulo es resuelta una vez que la identificación de los parámetros correspondientes es establecida. Al final del capítulo, se estudia la identificación de los parámetros de otras distribuciones skew-normal tales como la closed skew-normal, la hierarchical skew-normal y la unified skew-normal. En la Sección 1.2 se introduce el esquema básico bajo el cual se define la distribución skew-normal. En la Sección 1.3 se discute el concepto de identificación con la intención de mostrar la razón por la cual el análisis de identificación está basado en los cumulantes de la distribución de probabilidad. Cabe destacar que el resultado central del capítulo es introducido en las Secciones 1.4 y 1.5. Finalmente, en la Sección 1.6 se resumen los principales resultados y conclusiones de este capítulo.

1.2. Distribución skew-normal

El esquema común, bajo el cual las distribuciones closed skew-normal, hierarchical skew-normal y unified skew-normal son obtenidas está basado en la reducción por condicionamiento de una distribución normal multivariada. Sean U_0 y U_1 dos vectores aleatorios

de dimensión 1×1 y $d \times 1$, respectivamente, se asume que

$$\mathbf{U} = \begin{pmatrix} U_0 \\ \mathbf{U}_1 \end{pmatrix} \sim \mathcal{N}_{1+d} \left(\begin{pmatrix} 0 \\ \mathbf{0} \end{pmatrix}, \Omega^* = \begin{pmatrix} \tau^2 & \boldsymbol{\delta}^T \\ \boldsymbol{\delta} & \bar{\Omega} \end{pmatrix} \right), \tag{1.2.1}$$

donde $\tau^2 = Var(U_0|\boldsymbol{\delta}, \tau^2, \bar{\Omega}), \, \bar{\Omega} = Var(\mathbf{U}_1|\boldsymbol{\delta}, \tau^2, \bar{\Omega}) \, \mathbf{y} \, \boldsymbol{\delta} = Cov(\mathbf{U}_1, U_0|\boldsymbol{\delta}, \tau^2, \bar{\Omega}).$

Nota 1.2.1. Es importante destacar que, solamente en este capítulo, las esperanzas de muestreo (y, por lo tanto, las varianzas y covarianzas) son escritas como esperanzas condicionales dado ciertos parámetros específicos. Esta forma de escribir precisa con respecto a qué modelo estadístico se definen las esperanzas de muestreo.

El siguiente resultado permite realizar la construcción de una distribución asimétrica, particularmente la distribución skew-normal.

Lema 1.2.1. Sea $\mathbf{Z}_0 \in \mathbb{R}^d$ un vector aleatorio con función de densidad $f_{\mathbf{Z}_0}$. Sea además $\mathbf{W} \in \mathbb{R}^m$ un vector aleatorio. Se define $\mathbf{Z} \stackrel{d}{=} [\mathbf{Z}_0 | \mathbf{W} \geq \mathbf{0}]$ donde $\stackrel{d}{=}$ significa "igualdad en distribución". Entonces, la función de densidad de Z está dada por,

$$f_{\mathbf{Z}}(\mathbf{z}) = f_{\mathbf{Z}_0} \frac{P(\mathbf{W} \ge \mathbf{0} | \mathbf{Z}_0 = \mathbf{z}_0)}{P(\mathbf{W} \ge \mathbf{0})}$$

Dem. Para la demostración ver Arellano-Valle et al. [7].

Usando el Lema 1.2.1, se tiene que la densidad del vector aleatorio $\mathbf{Z} \stackrel{d}{=} [\mathbf{U}_1 | U_0 > 0]$ está, por lo tanto, dada por

$$f_{\mathbf{Z}}(\mathbf{z}|\boldsymbol{\delta}, \tau^2, \bar{\Omega}) = 2\phi_d(\mathbf{z}; 0, \bar{\Omega})\Phi_1\left(\boldsymbol{\delta}^T \bar{\Omega}^{-1}\mathbf{z}; 0, \tau^2 - \boldsymbol{\delta}^T \bar{\Omega}^{-1}\boldsymbol{\delta}\right), \quad \mathbf{z} \in \mathbb{R}^d, \quad (1.2.2)$$

donde $\phi_d(\mathbf{z}; \boldsymbol{\mu}, \Sigma)$ y $\Phi_d(\mathbf{z}; \boldsymbol{\mu}, \Sigma)$ son la función de densidad y la función de distribución acumulada de la distribúción $\mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$, respectivamente, evaluada en z. La parametrización inducida para la densidad (1.2.2) es $(\boldsymbol{\delta}, \tau^2, \bar{\Omega}) \in \mathbb{R}^d \times \mathbb{R}^+ \times \mathcal{C}_{d \times d}$, donde $\mathcal{C}_{d \times d}$ es el cono

de las matrices definidas positivas de dimensión $d \times d$ (para detalles sobre los conos de matrices, véase Berman y Plemmons [24]). Dado que $Var(\mathbf{U}_1|U_0, \delta, \tau^2, \bar{\Omega}) = \tau^2 - \delta^T \bar{\Omega}^{-1} \delta$, se tiene que, en (1.2.2), $\tau^2 - \delta^T \bar{\Omega}^{-1} \delta > 0$.

La familia de distribuciones skew-normal indexada por $(\delta, \tau^2, \bar{\Omega})$ es denotada como $\mathcal{SN}_d(\delta, \tau^2, \bar{\Omega})$. Es importante resaltar que esta parametrización $(\delta, \tau^2, \bar{\Omega})$ no es necesariamente identificada, aunque los parámetros de la distribución normal multivariada de dimensión (1+d) son identificados. De hecho, cuando la distribución normal es reducida por condicionamiento, se pierde la identificación de la distribución resultante; para más ejemplos al respecto, véase Florens *et al.* [43], Capítulo 4. Entonces, la distribución skew-normal no puede ser distinguida considerando solamente la parametrización inducida $(\delta, \tau^2, \bar{\Omega})$. En este sentido, el significado estadístico de δ, τ^2 y $\bar{\Omega}$ es explícito con respecto a la distribución normal (1.2.1), pero no con respecto a la distribución skew-normal (1.2.2).

La construcción propuesta por Azzalini para la distribución skew-normal es realizada asumiendo que Ω^* en (1.2.1) es una matriz de correlación (véase por ejemplo, Azzalini [12], Arellano-Valle y Azzalini [2] y referencias dentro de estos artículos). Sin embargo, el significado de Ω^* con respecto a la distribución skew-normal tampoco es del todo claro. En la Sección 1.4.2 se mostrará que tal elección propuesta por Azzalini es una restricción de identificación con respecto al proceso de muestreo.

Una propiedad relevante de la distribución skew-normal es que es cerrada bajo transformación lineales afines de la forma $Y = \xi + \omega Z$, donde $\xi \in \mathbb{R}^d$ es un parámetro de

localización y $\omega \in \mathcal{C}_{d \times d}$ es un parámetro de escala. De hecho, de (1.2.2) se sigue que

$$f_{\mathbf{Y}}(\mathbf{y}|\boldsymbol{\xi},\omega,\boldsymbol{\delta},\tau^2,\bar{\Omega}) = 2\phi_d(\mathbf{y};\boldsymbol{\xi},\omega\bar{\Omega}\omega)\Phi_1\left(\boldsymbol{\delta}^T\bar{\Omega}^{-1}\omega^{-1}(\mathbf{y}-\boldsymbol{\xi});0,\tau^2-\boldsymbol{\delta}^T\bar{\Omega}^{-1}\boldsymbol{\delta}\right),$$

con $\mathbf{y} \in \mathbb{R}^d$. En este caso, la parametrización inducida es $(\boldsymbol{\xi}, \omega, \boldsymbol{\delta}, \tau^2, \bar{\Omega}) \in \mathbb{R}^d \times \mathcal{C}_{d \times d} \times \mathbb{R}^d \times \mathbb{R}^+ \times \mathcal{C}_{d \times d}$, la cual no es identificada toda vez que $(\boldsymbol{\delta}, \tau^2, \bar{\Omega})$ tampoco lo es.

Azzalini y Capitanio [13] sugieren la siguiente transformación para una variable con distribución skew-normal. Dado $Z_0 \sim \mathcal{SN}(\delta, \tau^2, \bar{\Omega})$, con $Z_0 \in \mathbb{R}$, se define $Z^0 = (Z - E(Z))/\sqrt{Var(Z)}$. Luego, se tiene que $E(Z^0) = 0$ y $Var(Z^0) = 1$; y si $Y = \xi + \omega Z^0$, entonces $E(Y) = \xi$ y $Var(Y) = \omega^2$. Esta transformación tiene la ventaja de que en la distribución de probabilidad, ξ y ω son estadísticamente interpretables. Sin embargo, los parámetros δ , τ^2 y $\bar{\Omega}$ no son ni identificados ni interpretables con respecto al modelo estadístico skew-normal inducido por Y.

Con estas consideraciones, se concluye que la parametrización inducida de la distribución skew-normal no es estadísticamente interpretable, dificultando la aplicabilidad de este modelo a conjuntos de datos reales. Por lo tanto, a este nivel, se justifica realizar un análisis de identificación sobre la parametrización inducida de la distribución skew-normal. Sin embargo, previamente es necesario revisar algunos conceptos de identificación con el objetivo de mostrar por qué el análisis de identificación debe estar basado en la información de muestreo, la cual está completamente caracterizada por las esperanzas de muestreo o en otras palabras, los cumulantes de la distribución de probabilidad.

1.3. Identificabilidad en los modelos estadísticos

Típicamente, la identificabilidad paramétrica es vista como una condición necesaria para la existencia de estimadores consistentes e insesgados (véase, por ejemplo, Bunke y Bunke [26], Gabrielsen [45] y San Martín y Quintana [77]). Sin embargo, y tal como lo propone Koopmans y Reiersøl [56] y Rothenberg [73], la identificabilidad se relaciona estrechamente con el problema de especificación (Fisher [40]) en el sentido de encontrar condiciones bajo las cuales los parámetros que indexan la distribución de muestreo tengan significado empírico. Esto explica, de cierto modo, por qué la inferencia está limitada por el análisis de identificación: cuando una parametrización es estadísticamente ambigua (es decir, no es identificada) el significado de los estimadores no es del todo claro.

Un modelo estadístico es definido como una familia de distribuciones o probabilidades de muestreo P^{θ} indexadas por un parámetro $\theta \in \Theta$. En este contexto, la identificabilidad paramétrica corresponde a la inyectividad en el mapeo $\theta \longmapsto P^{\theta}$. Entonces, dado que la información proporcionada por el proceso de muestreo es resumida en el parámetro o parámetros que la indexan, tales parámetros y cualquier transformación biyectiva de los mismos proporcionan la misma información. Esto permite introducir una estructura de medibilidad en el espacio paramétrico a través de un σ -álgebra $\mathcal A$ de subconjuntos de Θ (véase Raoult [72], Barankin *et al.* [18], y Florens *et al.* [42]). Consecuentemente, dicha σ -álgebra en el espacio parámetrico permite definir un parámetro suficiente tal como sigue (véase Barankin [17]):

Definición 1.3.1. Una función $g(\theta)$ de un parámetro θ es suficiente para X si

$$g(\theta_1) = g(\theta_2) \Longrightarrow P^{\theta_1}(\mathbf{X}) = P^{\theta_2}(\mathbf{X}) \quad \forall \ \theta_1, \theta_2 \in \Theta.$$

La importancia de la suficiencia paramétrica es que la identificabilidad puede ser definida en términos de la suficiencia mínima paramétrica (véase Kadane [55], Picci [70] y Oulhaj y Mouchart [68]). Más aún, es posible conectar este concepto clásico con el enfoque Bayesiano toda vez que la distribución a priori del parámetro sea regular; para más detalles, véase Florens et al. [43] (Capítulo 4). Esto significa que la suficiencia paramétrica puede ser escrita usando independencia condicional. Por ejemplo, $X \perp H \theta | g(\theta)$ define a $g(\theta)$ como un parámetro suficiente para la información observada X. Tal condición implica que la distribución de X queda completamente determinada por $g(\theta)$, o en otras palabras, θ es redundante una vez que $q(\theta)$ es conocido. Por simetría de la relación de independencia condicional, se puede concluir que $q(\theta)$ es un parámetro suficiente si la distribución condicional de la parte redundante θ dado el parámetro suficiente $q(\theta)$ no es actualizada por la información de la muestra, es decir, $p(\theta|X, g(\theta)) = p(\theta|g(\theta))$. Por lo tanto, la parametrización θ no es "identificada" por los datos X. Esta situación puede evitarse si θ es un parámetro mínimo suficiente, es decir, si θ es un parámetro suficiente y otro parámetro suficiente es función de θ . En este sentido, si la parametrización de un modelo estadístico está basada en un parámetro mínimo suficiente, entonces la parametrización no contiene información redundante. Estas consideraciones motivan la siguiente definición (véase Florens y Rolin [44]):

Definición 1.3.2. Un parámetro suficiente $\psi = g(\theta)$ se dice identificado por X, si ψ es un parámetro mínimo suficiente para X.

Se puede demostrar que si el mapeo $\theta \longmapsto P^{\theta}$ es inyectivo, entonces θ es un parámetro mínimo suficiente para X. Para detalles de la demostración, véase Florens *et al.* [42] y Florens *et al.* [43] (Capítulo 4). Más aún, el parámetro mínimo suficiente es siempre una función de un número contable de esperanzas de muestreo (para detalles de la demostración véase Florens *et al.* [43], Capítulo 4). Esto significa que la parametrización de interés $\psi = g(\theta)$ es identificada si existen funciones medibles f y h tales que

$$\psi = h\{E(f(\mathbf{X})|\boldsymbol{\theta})\}. \tag{1.3.1}$$

En el contexto de los modelos skew-normal, este concepto de identificación permite usar los cumulantes de la distribución para el análisis de identificación, toda vez que éstos son funciones medibles de esperanzas de muestreo. Por ejemplo, los cumulantes están dados por

$$\kappa_{1_i} = E(X_i|\boldsymbol{\theta}), \quad \kappa_{2_i} = Var(X_i|\boldsymbol{\theta}), \quad \kappa_{2_{ij}} = Cov(X_i, X_j|\boldsymbol{\theta}),$$

$$\kappa_{3_i} = E(X_i^3|\boldsymbol{\theta}) - 3E(X_i|\boldsymbol{\theta})E(X_i^2|\boldsymbol{\theta}) + 2E(X_i|\boldsymbol{\theta})^3, \dots$$

En particular, $\kappa_{3_i} = \gamma_{1_i} \kappa_{2_i}^{3/2}$, donde γ_1 corresponde al índice de asimetría (véase Fisher [41]).

1.4. Identificabilidad de la distribución skew-normal

En esta sección se analiza la identificabilidad de la distribución skew-normal en el caso multivariado (\mathbb{R}^d). Una vez que la parametrización identificada es establecida, su significado estadístico es discutido y, consecuentemente, una parametrización para la distribución skew-normal es propuesta. Seguidamente, se busca dar respuesta a la siguiente pregunta: ¿cómo el análisis de identificación de la distribución skew-normal se relaciona con la restricción de que Ω^* en (1.2.1) sea una matriz de correlación? La importancia de esta pregunta viene del hecho de que dicha restricción es típicamente utilizada para obtener una distribución skew-normal a través de la reducción por condicionamiento de un modelo normal conjunto (véase Azzalini [12]).

1.4.1. Identificabilidad en el caso multivariado ($\mathbb{R}^d,\,d\geq 1$)

Sea $\mathbf{Z} \in \mathbb{R}^d$ un vector aleatorio con distribución $\mathcal{SN}_d(\boldsymbol{\delta}, \tau^2, \bar{\Omega})$, y función de densidad dada por (1.2.2). Sea

$$\bar{\Omega} = \begin{pmatrix} \bar{\omega}_{11} & \cdots & \bar{\omega}_{1d} \\ \vdots & \ddots & \vdots \\ \bar{\omega}_{d1} & \cdots & \bar{\omega}_{dd} \end{pmatrix}, \quad \boldsymbol{\delta} = \begin{pmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_d \end{pmatrix}.$$

Considerando la transformación lineal afin $\mathbf{Y} = \boldsymbol{\xi} + \omega \mathbf{Z}$, donde $\boldsymbol{\xi} = (\xi_1, \dots, \xi_d)^T \in \mathbb{R}^d$ y ω es una matriz simétrica definida positiva de dimensión $d \times d$ y cuyos elementos son denotados por ω_{ij} , con $i, j = 1, \dots, d$; los cumulantes de Y_i , con $i = 1, \dots, d$, están dados

por

$$\kappa_{1_i} = \xi_i + \sqrt{\frac{2}{\pi}} \frac{\delta_i^*}{\tau}, \qquad \kappa_{2_i} = \omega_{ii}^* - \frac{2}{\pi} \frac{{\delta_i^*}^2}{\tau^2}, \qquad \kappa_{3_i} = \sqrt{\frac{2}{\pi^3}} (4 - \pi) \frac{{\delta_i^*}^3}{\tau^3}; \qquad (1.4.1)$$

y para todo $i \neq j$,

$$\kappa_{2ij} = \omega_{ij}^* - \frac{2 \, \delta_i^* \delta_j^*}{\pi \, \tau^2},\tag{1.4.2}$$

donde ω_{ij}^* denota el elemento de la fila i y columna j de la matriz $\omega \bar{\Omega} \omega$ con $i, j = 1, \ldots, d$, $\delta_i^* = \omega_{i\bullet}^T \delta$, y $\omega_{i\bullet}$ es la i-ésima columna de la matriz ω . Usando (1.4.1) y (1.4.2), se tiene que

$$\xi_i = \kappa_{1_i} - \left\{ \frac{2\kappa_{3_i}}{4-\pi} \right\}^{1/3}, \qquad \omega_{ii}^* = \kappa_{2_i} + \left\{ \frac{2\kappa_{3_i}}{4-\pi} \right\}^{2/3}, \qquad \frac{\delta_i^*}{\tau} = \sqrt{\pi} \left\{ \frac{\kappa_{3_i}}{\sqrt{2}(4-\pi)} \right\}^{1/3}$$

у

$$\omega_{ij}^* = \kappa_{2_{ij}} + \left\{ \frac{2}{4-\pi} \right\}^{2/3} \kappa_{3_i}^{1/3} \kappa_{3_j}^{1/3}.$$

Dado que los cumulantes κ_{r_i} y $\kappa_{r_{ij}}$ con $i,j=1,\ldots,d$ son funciones de $E[f(\mathbf{Y})|\boldsymbol{\xi},\omega,\boldsymbol{\delta},\tau^2,\bar{\Omega}],$ donde f es una función medible, entonces el parámetro suficiente mínimo es $(\boldsymbol{\xi},\Lambda,\boldsymbol{\alpha})\in\Theta\subset\mathbb{R}^d\times\mathcal{C}_{d\times d}\times\mathbb{R}^d,$ donde

(i)
$$\alpha = \tau^{-1} \omega \delta$$
, (ii) $\Lambda = \omega \bar{\Omega} \omega$, (1.4.3)

el cual es identificado por una observación. La función de densidad (1.2.2) puede ser escrita en función de (ξ, Λ, α) , a saber,

$$f_{\mathbf{Y}}(\mathbf{y}|\boldsymbol{\xi},\omega,\boldsymbol{\delta},\tau^{2},\bar{\Omega}) = 2\phi_{d}(\mathbf{y};\boldsymbol{\xi},\Lambda)\Phi_{1}\left(\boldsymbol{\alpha}^{T}\Lambda^{-1}(\mathbf{y}-\boldsymbol{\xi});0,1-\boldsymbol{\alpha}^{T}\Lambda^{-1}\boldsymbol{\alpha}\right) \doteq f_{\mathbf{Y}}(\mathbf{y}|\boldsymbol{\xi},\Lambda,\boldsymbol{\alpha}),$$

$$(1.4.4)$$

implicando que el espacio paramétrico de (ξ, α, Λ) está dado por

$$\Theta = \{ (\boldsymbol{\xi}, \Lambda, \boldsymbol{\alpha}) \in \mathbb{R}^d \times \mathcal{C}_{d \times d} \times \mathbb{R}^d : 1 - \boldsymbol{\alpha}^T \Lambda^{-1} \boldsymbol{\alpha} > 0 \}.$$
 (1.4.5)

Finalmente, los parámetros identificados pueden ser estadísticamente interpretables con respecto al modelo estadístico skew-normal: ξ corresponde al vector de medias de Y corregido por una función del vector κ_3 ; Λ es la matriz de varianza-covarianza de Y corregida por la matriz cuyos elementos son funciones de κ_{3i} , con $i=1,\ldots,d$; y, finalmente, α es una función del vector κ_3 . De lo anterior, se obtiene el siguiente teorema:

Teorema 1.4.1. Sea $\mathbf{Z} \in \mathbb{R}^d$ un vector aleatorio con distribución skew-normal $\mathcal{SN}_d(\delta, \tau^2, \bar{\Omega})$, con $d \geq 1$. Considérese la transformación lineal afín $\mathbf{Y} = \boldsymbol{\xi} + \omega \mathbf{Z}$. Entonces, el parámetro $(\boldsymbol{\xi}, \Lambda, \alpha) \in \Theta$ asociado a la distribución skew-normal inducida por $\mathbf{Y} \in \mathbb{R}^d$, donde α y Λ son definidos por (1.4.3) y Θ es dado por (1.4.5), es identificado por una observación.

Nota 1.4.1. Es importante mencionar en este punto que una biyección entre los parámetros del modelo estadístico y los cumulantes teóricos permite, de alguna manera, proponer estimadores para tales parámetros basados en los cumulantes, de manera similar al método de momentos.

1.4.2. Identificabilidad y construcción de la distribución skew-normal

Tal como se mencionó anteriormente, el análisis de identificación realizado en la sección previa está basado en una distribución skew-normal definida a través de una normal multivariada (1.2.1) en el cual Ω^* es una matriz de varianza-covarianza, y no una matriz

de correlación como se considera típicamente en la literatura (véase, por ejemplo, Azzalini [12] y Arellano-Valle y Azzalini [2]). En la densidad normal (1.2.1), Ω^* es una matriz de correlación después de imponer las siguientes restricciones,

$$\tau^2=1,\quad \bar{\Omega}=1, \qquad \text{para } d=1;$$

$$\tau^2=1,\quad \bar{\omega}_{ii}=1 \ \forall i=1,\ldots,d, \quad \omega=diag\{\omega_1,\ldots,\omega_d\}, \ \text{para } d>1.$$

$$(1.4.6)$$

Bajo estas restricciones, se sigue que $\delta \in (-1,1)^d$, para $d \geq 1$. Usando el Teorema 1.4.1, se deduce que tales restricciones son restricciones adicionales de identificación con respecto a la distribución skew-normal inducida por Y, bajo las cuales el parámetro identificado está dado por $(\xi, \omega, \delta) \in \mathbb{R}^d \times \mathcal{C}_{d \times d} \times (-1, 1)^d$ para $d \geq 1$. En consecuencia, la densidad skew-normal parametrizada por (ξ, ω, δ) corresponde a la distribución skew-normal definida por Azzalini [11] en el caso univariado y Azzalini and Dalla Valle [15] en el caso multivariado con (d > 1).

En el caso univariado, el significado estadístico de la parametrización identificada de la distribución skew-normal inducida por Y es como sigue: ξ corresponde a la media de Y corregida por una función de κ_3 , ω^2 es la varianza de Y corregida por una función de κ_3 y, finalmente δ es una función de κ_3 , a saber

$$\delta = \frac{\sqrt{\pi}\kappa_3^{1/3}}{\omega\{\sqrt{2}(4-\pi)\}^{1/3}} = \frac{\sqrt{\pi}\kappa_3^{1/3}}{\{2^{1/3}(4-\pi)\kappa_2 + 2\kappa_3^{2/3}\}^{1/2}}.$$
 (1.4.7)

Por lo tanto, δ es una función del índice de asimetría γ_1 definido por Fisher [41] y típicamente calculado en los programas computacionales más comunes, es decir,

$$\delta = \frac{\sqrt{\pi}\gamma_1^{1/3}}{\{2^{1/3}(4-\pi) + 2\gamma_1^{2/3}\}^{1/2}}.$$

El análisis en el caso multivariado es análogo.

De acuerdo a lo anterior, las restricciones de identificación (1.4.6) no pueden ser empíricamente interpretadas con respecto al proceso de muestreo generado por datos pertenecientes a una determinada distribución skew-normal. Por esta razón, cuando se utiliza una densidad skew-normal parametrizada por (ξ, ω, δ) , se debe asumir que los datos fueron generados por un proceso normal no observado con una estructura específica definida por (1.4.6) sin ninguna posibilidad de justificar empíricamente este supuesto.

1.5. Aplicaciones de los resultados de identificabilidad a otras distribuciones skew-normal

Arellano-Valle y Azzalini [2] sugieren, sin una demostración formal, que tanto la distribución closed skew-normal como la hierarchical skew-normal tienen problemas de identificación. En esta sección, la identificabilidad de estas distribuciones es establecida utilizando los resultados obtenidos en la Sección 1.4.

En el caso univariado, la distribución closed skew-normal introducida por Gonzáles-Farías *et al.* [51] es obtenida tal como se describe en la Sección 1.2 con

$$\Omega^* = \left(\begin{array}{cc} \Gamma + D^2 \bar{\Sigma} & \mathrm{D}\bar{\Sigma} \\ \mathrm{D}\bar{\Sigma} & \bar{\Sigma} \end{array} \right).$$

Considerando una transformación lineal afín $Y=\xi+\omega Z$, se obtiene la siguiente función

de densidad

$$f_Y(y|\xi,\omega^2,D,\Gamma,\bar{\Sigma}) = 2\phi_1(y;\xi,\omega^2\bar{\Sigma})\Phi_1(D\omega^{-1}(y-\xi);0,\Gamma), \quad y \in \mathbb{R},$$
 (1.5.1)

donde $(\xi, \omega^2, D, \Gamma, \bar{\Sigma}) \in \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}^+$. Aplicando el Teorema 1.4.1, el parámetro identificado es $(\xi, \Lambda, \alpha) \in \Theta \subset \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}$, donde $\Lambda = \omega^2 \bar{\Sigma}$, $\alpha = \omega D \bar{\Sigma}/(\Gamma + D^2 \bar{\Sigma})^{1/2}$ y el espacio paramétrico Θ está dado por (1.4.5). Cuando se reescribe la función de densidad (1.5.1) en términos del parámetro identificado, la distribución skew-normal original caracterizada por la función de densidad (1.4.4) es obtenida.

Los parámetros D y ω^2 pueden ser identificados bajo restricciones de identificación adicionales, a saber, $\Gamma + D^2 \bar{\Sigma} = 1$ y $\bar{\Sigma} = 1$. De esa manera, el parámetro identificado está dado por $(\xi, \omega^2, D) \in \mathbb{R} \times \mathbb{R}^+ \times (-1, 1)$ y la densidad (1.5.1) es reescrita como

$$f_Y(y|\xi,\omega^2,D) = 2\phi_1(y;\xi,\omega^2)\Phi_1(D\omega^{-1}(y-\xi);0,1-D^2), \quad y \in \mathbb{R},$$
 (1.5.2)

la cual corresponde a la distribución skew-normal de Azzalini [11]. Sin embargo, las restricciones de identificación $\Gamma + D^2 \bar{\Sigma} = 1$ y $\bar{\Sigma} = 1$ no son estadísticamente interpretables con respecto a la distribución closed skew-normal.

Otra familia de distribuciones asimétricas es la llamada closed-2 skew-normal. La versión univariada es obtenida considerando, en (1.2.1), una matriz de covarianzas de la forma

$$\Omega^* = \left(\begin{array}{cc} \tau^2 & \tau^2 D \\ \tau^2 D & \Psi + \tau^2 D^2 \end{array} \right).$$

Haciendo $Z=[U_1|U_0>0]$ en el modelo normal y considerando $Y=\xi+\omega Z$, se obtiene

la distribución closed-2 skew-normal con función de densidad dada por

$$f_Y(y|\xi,\omega^2,D,\tau^2,\Psi) = 2\phi_1(y;\xi,\omega^2(\Psi+\tau^2D^2))\Phi_1\left(\frac{\tau^2D(y-\xi)}{\omega(\Psi+\tau^2D^2)};0,\frac{\Psi\tau^2}{\Psi+\tau^2D^2}\right),$$
(1.5.3)

con $y \in \mathbb{R}$. Por lo tanto, usando el Teorema 1.4.1, se sigue que $(\xi, \Lambda, \alpha) \in \Theta \subset \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}$ es el parámetro identificado, donde $\Lambda = \omega^2(\Psi + \tau^2 D^2)$, $\alpha = \omega D \tau$ y el espacio paramétrico Θ está dado por (1.4.5). Como en el caso anterior, la función de densidad (1.5.3) se puede reescribir en términos del parámetro identificado.

Nuevamente, las restricciones de identificación adicionales $\Psi + \tau^2 D^2 = \tau^2 = 1$ introducidas, no son interpretables con respecto a la distribución closed-2 skew-normal.

La distribución hierarchical skew-normal univariada introducida por Liseo y Loperfido [59] es obtenida considerando en (1.2.1) la siguiente matriz de covarianzas:

$$\left(\begin{array}{cc} \Upsilon C^2 & -\Upsilon C \\ -\Upsilon C & \Upsilon + \Sigma \end{array}\right).$$

En este caso, el análisis de identificación es similar al de la distribución closed-2 skewnormal. Finalmente, la distribución unified skew-normal, propuesta por Arellano-Valle y Azzalini [2] es definida considerando en (1.2.1) una matriz de correlación de la forma:

$$\left(\begin{array}{cc} 1 & \delta \\ \delta & 1 \end{array}\right).$$

Aplicando el Teorema 1.4.1, se concluye que la versión univariada de esta distribución, bajo una transformación de localización-escala es indexada por el parámetro mínimo suficiente $(\xi, \omega^2, \delta) \in \mathbb{R} \times \mathbb{R}^+ \times (-1, 1)$.

1.6. Resumen y conclusiones

La distribución skew-normal es típicamente definida como la reducción por condicionamiento de una distribución normal multivariada del tipo (1.2.1). La distribución skew-normal de Azzalini se define asumiendo que Ω^* , en la distribución normal (1.2.1), es una matriz de correlación. Las extensiones de la distribución skew-normal se obtienen al imponer estructuras alternativas en Ω^* , tal como se ha discutido en la Sección 1.5. La parametrización usada para indexar la distribución skew-normal está basada en la parametrización de la distribución normal original. Sin embargo, tal parametrización no es identificada, aunque los parámetros de la distribución normal sean identificados y, consecuentemente, interpretables; es precisamente el proceso de reducción el que origina los problemas de identificación.

Desde que la parametrización inducida indexa el proceso de muestreo skew-normal, su significado debe estar basado en éste y no en un proceso normal no observable. El significado estadístico de la parametrización inducida puede ser establecido una vez que la identificabilidad de la misma, con respecto a la distribución skew-normal, sea establecida. Este análisis, realizado en las Secciones 1.4 y 1.5, origina las siguientes conclusiones:

1. Los parámetros δ , τ^2 y $\bar{\Omega}$ asociados a la matriz de varianzas-covarianzas de la distribución normal original, son redundantes y, por lo tanto, no interpretables con respecto a la distribución skew-normal.

- 2. Los parámetros identificados de la versión de localización-escala de la distribución skew-normal son tres, y la interpretación de los mismos es la siguiente: la media verdadera de los datos corregida por una función del cumulante κ₃; la varianza verdadera de los datos corregida por una función de κ₃; y un parámetro que es función de κ₃.
- 3. Cuando los parámetros de las distribuciones closed skew-normal, closed-2 skew-normal y hierarchical skew-normal, todas ellas en sus versiones univariadas, son identificados, tales distribuciones se reducen a la misma distribución skew-normal. Este hecho implica que todas esas familias fallan como extensiones de la distribución skew-normal de Azzalini.

Las distribuciones skew-normal propuestas por Azzalini [11] y Azzalini y Dalla Valle [15] pueden ser recuperadas introduciendo restricciones de identificación adicionales, a saber, que la matriz Ω^* en la distribución normal multivariada (1.2.1) sea una matriz de correlación y no una matriz de varianzas-covarianzas. Sin embargo, tales restricciones no son estadísticamente interpretables con respecto al proceso de muestreo que genera los datos, es decir, la distribución skew-normal. Aunque estas restricciones sean interpretables con respecto a la distribución normal original, desde el punto de vista de la modelación estadística, este hecho se vuelve irrelevante dado que la distribución de probabilidad de la cual provienen lo datos es la skew-normal.

1.7. Agradecimientos

Quiero expresar mi gratitud a Ernesto San Martín (Pontificia Universidad Católica de Chile) por sus comentarios, sugerencias y aportes realizados en este capítulo.

1.8. Apéndice

1. Apéndice A

De Arellano-Valle y Azzalini [2], se tiene que la función generadora de momentos del modelo $\mathcal{SN}_d(\boldsymbol{\xi}, \omega, \boldsymbol{\delta}, \tau^2, \bar{\Omega})$, con $\mathbf{Y} \in \mathbb{R}^d$ y función de densidad (1.2), está dada por:

$$M(\mathbf{t}) = 2 \exp \left\{ \boldsymbol{\xi}^T \mathbf{t} + \frac{1}{2} \mathbf{t}^T \omega \bar{\Omega} \omega \mathbf{t} \right\} \Phi(\boldsymbol{\delta}^T \omega \mathbf{t}; 0, \tau^2)$$
(1.8.1)

La función generadora de cumulantes está definida como $K(\mathbf{t}) = \log M(\mathbf{t})$, por lo que según (1.8.1), se tiene que,

$$K(\mathbf{t}) = \boldsymbol{\xi}^T \mathbf{t} + \frac{1}{2} \mathbf{t}^T \omega \bar{\Omega} \omega \mathbf{t} + \log \Phi(\boldsymbol{\delta}^T \omega \mathbf{t}; 0, \tau^2) + \log 2$$
 (1.8.2)

Utilizando (1.8.2), se tiene que,

$$\kappa_{1_i} = E(Y_i) = K'(0), \qquad \kappa_{2_i} = Var(Y_i) = K''(0), \qquad \kappa_{3_i} = K'''(0),$$

donde $K^r(0)$ indica la r-ésima derivada de K(t) con respecto a t_i evaluada en $t_i=0$, con $i=1,\ldots,d$.

Capítulo 2

Reducción probabilística de la distribución skew-normal

"Para establecer una formulación distinta de los problemas estadísticos, es necesario definir la tarea que el estadístico fija en sí mismo: brevemente, y en su forma más concreta, el objeto de los métodos estadísticos es la reducción de los datos"

Ronald A. Fisher, 1922

En el presente capítulo se estudia el enfoque de reducción probabilística, con el fin de generar el modelo de regresión lineal skew-normal. Para esto se considera la descomposición marginal-condicional de la densidad conjunta y se estudia la relación entre los espacios paramétricos asociados a los modelos marginal y condicional. En este contexto, condiciones necesarias y suficientes sobre los parámetros del modelo conjunto son analizadas con el fin de generar un corte, y por lo tanto, una variable exógena. Finalmente, un estudio de simulación es desarrollado con el objetivo de ilustrar los resultados obtenidos.

2.1. Introducción

En 1922, Fisher [40], menciona que el objetivo central de los métodos estadísticos es la reducción de los datos. En ese contexto, Fisher [40] identifica tres problemas concernientes a la reducción de los datos, a saber, el problema de especificación, relacionado con la elección de la forma matemática de la distribución de probabilidad; el problema de estimación, relacionado con la elección del método para encontrar un estadístico proveniente de la muestra y que permita estimar los valores de los parámetros; y el problema de distribución, relacionado con las distribuciones de probabilidad que siguen los estadísticos.

Para Fisher [40], el problema de especificación de un modelo estadístico tiene un particular interés en el contexto de la reducción de datos. Es decir, un modelo estadístico debe ser capaz de resumir toda la información relevante obtenida en el proceso de muestreo. La consecuencia inmediata de la mala especificación de un modelo estadístico es la generación de conclusiones equivocadas sobre los parámetros poblacionales.

Sin embargo, la tarea de especificar correctamente un modelo estadístico el cual permita resumir la información disponible de manera eficiente no es nada fácil y trae consigo una serie de dificultades. Es por eso que la siguiente pregunta parece tener especial relevancia además de mucha naturalidad: ¿qué tipo de modelo estadístico proponer tal que resuma la información probabilística relevante de los datos, sabiendo que la descripción más general de éstos se encuentra en la distribución de probabilidad conjunta de las variables aleatorias observadas?. En otras palabras, ¿que modelo estadístico puede obtener información

relevante de manera resumida?

El modelo de regresión es uno de los modelos estadísticos más comunes y utilizados en la literatura estadística. El objetivo de este modelo es resumir la información referida a la causalidad entre determinadas variables através de relaciones lineales entre las mismas. Uno de los representantes más conocido de este tipo de modelo es el modelo de regresión lineal normal homocedástico, a saber,

$$Y_i = \alpha + \mathbf{X}_i^T \beta + \epsilon_i \qquad \epsilon_i \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \omega^2),$$
 (2.1.1)

con $Y_i \in \mathbb{R}$ y $X_i \in \mathbb{R}^k$, para $i=1,\ldots,n$. La información probabilística del modelo (2.1.1) está referida principalmente al proceso de muestreo que genera la secuencia ϵ_i , $i=1,\ldots,n$, a saber, normalidad, independencia e idéntica distribución (la notación *i.i.d* hace referencia a la independencia e idéntica distribución y será utilizada en lo que sigue del presente documento).

El modelo (2.1.1) puede ser también interpretado como la reducción de un proceso de muestreo conjuntamente normal. De hecho, sea $Z_i = (Y_i, \mathbf{X}_i)^T, Y_i \in \mathbb{R}, \mathbf{X}_i \in \mathbb{R}^k,$ $i = 1, \ldots, n$, una secuencia de vectores aleatorios independientes. Entonces, se tiene que

$$f_{\mathbf{Z}}(z_1, ..., z_n | \boldsymbol{\theta}) = \prod_{i=1}^n f_Y(y_i | \mathbf{x}_i, \boldsymbol{\theta}_1) f_{\mathbf{X}}(\mathbf{x}_i | \boldsymbol{\theta}_2).$$
 (2.1.2)

donde θ , θ_1 y θ_2 son los parámetros del modelo conjunto, condicional y marginal respectivamente. Bajo el supuesto de que $Z_i = (Y_i, \mathbf{X}_i)^T \overset{i.i.d.}{\sim} \mathcal{N}_{1+k}(\boldsymbol{\xi}, \omega), i = 1, \ldots, n,$ con $(\boldsymbol{\xi}, \omega)$

particionado como

$$\boldsymbol{\xi} = \begin{pmatrix} \xi_1 \\ \boldsymbol{\xi}_2 \end{pmatrix}, \qquad \omega = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{pmatrix}, \tag{2.1.3}$$

se obtiene que $\theta_1=(\alpha,\beta,\omega^2)$ donde $\beta=\omega_{22}^{-1}\omega_{21}$, $\alpha=\xi_1-\beta^T\xi_2$, $\omega^2=\omega_{11}-\omega_{12}\omega_{22}^{-1}\omega_{21}$ y $\theta_2=(\xi_2,\omega_{22})$. Más aún $\mathbf{X}_i\stackrel{i.i.d.}{\sim}\mathcal{N}_k(\xi_2,\omega_{22})$ y $[Y_i|\mathbf{X}_i]\stackrel{ind.}{\sim}\mathcal{N}(\alpha+\mathbf{X}_i^T\beta,\omega^2)$, para todo $i=1,\ldots,n$. La reducción del modelo conjunto en el modelo condicional permite generar el modelo de regresión lineal propuesto en (2.1.1), en el cual se tiene que

$$E(Y_i|\mathbf{X}_i,\boldsymbol{\theta}_1) = \alpha + \mathbf{X}_i^T\boldsymbol{\beta},$$

$$V(Y_i|\mathbf{X}_i,\boldsymbol{\theta}_1) = \omega^2.$$

El supuesto de normalidad en el proceso de muestreo que genera \mathbb{Z} permite realizar una reducción del modelo conjunto en un modelo condicional $[Y|\mathbb{X}]$, el cual es conocido como modelo de regresión, desechando la información comprendida en el modelo marginal generado por \mathbb{X} . Esto último se debe a dos aspectos importantes (los cuales serán definidos más adelante), a saber, el corte (Barndorff-Nielsen [19, 20]) y la exogeneidad (Engle et al. [38]). Claramente, suponer otra distribución distinta a la normal implica perder estas últimas características (para más detalles véase Spanos [78]) y generar otros modelos estadísticos através de la reducción.

En este contexto, el objetivo de este capítulo es cambiar el supuesto de normalidad en la distribución del vector Z considerando una distribución skew-normal multivariada y proponiendo una reducción del modelo conjunto skew-normal basado en su descomposición

marginal-condicional. Incialmente, la reducción del modelo skew-normal genera un modelo de regresión no-lineal, no existiendo corte y ni exogeneidad. Luego, imponiendo algunas restricciones sobre el espacio paramétrico de los modelos condicional y marginal, se generan nuevos modelos de regresión, recuperándose el corte y por lo tanto la exogeneidad de la variable marginal.

Para tal fin, en la Sección 2.2, se estudia algunos conceptos preliminares relacionados con la reducción probabilística en términos de los modelos condicionales, el corte y la exogeneidad. En la Sección 2.3, el análisis se centra en la determinación de la descomposición marginal-condicional que caracteriza a la distribución skew-normal. En la Sección 2.4, se estudia la reducción probabilística de la distribución skew-normal y el modelo de regresión que subyace a esta reducción. Finalmente, en la Sección 2.5 se realiza una simulación con el fin de ilustrar los resultados obtenidos en las secciones previas.

2.2. Reducción probabilística

El concepto de reducción de datos tiene una gran importancia dentro del contexto de la inferencia estadística. De hecho, y ya desde 1922, Fisher [40] estableció que la reducción de los datos es el objetivo principal de los métodos estadísticos. Algunos años más tarde, Basu [23] formalizó las ideas de Fisher y propuso la reducción de un modelo estadístico en un par de experimentos complementarios (llamados marginal y condicional). En esa dirección, Basu [23] formuló el principio condicional el cual establece que, bajo la existencia de

un estadístico ancilar para un modelo, la información en el modelo conjunto es la misma que en el modelo condicional. Sin embargo, no siempre se puede garantizar la existencia de un estadístico ancilar y, por lo tanto, conseguir la reducción propuesta por Basu [23] implica algunas consideraciones sobre los espacios paramétricos asociados a los modelos marginal y condicional. Consecuentemente, el problema de reducción de un modelo conjunto se reduce a restringir tales espacios.

En ese contexto, diferentes soluciones han sido propuestas para el problema de la reducción de los modelos estadísticos; por ejemplo, Lancaster [57] en econometría y Andersen [1] en psicometría. De otro lado, al menos dos tipos de soluciones han sido propuestas en estadística; la primera, caracteriza una parametrización ortogonal para obtener una matriz de información diagonal por bloque (Cox and Reid [31]) y la segunda, utiliza el concepto de corte introducido por Barndorff-Nielsen [19, 20], el cual es definido de la siguiente manera:

Definición 2.2.1. Sea $\mathbf{Z} = (\mathbf{Y}^T, \mathbf{X}^T)^T$ un vector aleatorio con función de densidad $f_{\mathbf{Z}}(\mathbf{z}|\boldsymbol{\theta})$, donde el parámetro $\boldsymbol{\theta}$ pertenece al espacio paramétrico $\boldsymbol{\Theta}$. Considérese una partición $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ de $\boldsymbol{\theta}$ tal que

- (i) La distribución condicional de [Y|X] depende de θ solo através de θ_1 .
- (ii) La distribución marginal de X depende de θ solo através de θ_2 .
- (iii) (θ_1, θ_2) están en variación libre, es decir, $\Theta_1 \times \Theta_2 = \Theta$, donde Θ_1 y Θ_2 son los espacios paramétricos de θ_1 y θ_2 , respectivamente.

Bajo estas condiciones, se dice que X genera un corte. Más aún, X es un vector aleatorio exógeno con respecto a θ_1 (véase Engle *et al.* [38]; Florens *et al.* [43]; Spanos [78]).

Nótese que el punto (iii) de la Definición 2.2.1 implica que existe una biyección entre θ y (θ_1, θ_2) . Por ejemplo, en el caso normal, dado $\theta = (\xi, \omega)$, se puede determinar que $\theta_1 = (\alpha, \beta, \omega^2)$ y $\theta_2 = (\xi_2, \omega_{22})$. De otro lado, dado θ_1 y θ_2 se tiene que,

$$\boldsymbol{\xi} = \begin{pmatrix} \alpha + \boldsymbol{\beta}^T \boldsymbol{\xi}_2 \\ \boldsymbol{\xi}_2 \end{pmatrix}, \qquad \omega = \begin{pmatrix} \omega^2 + \boldsymbol{\beta}^T \omega_{22} \boldsymbol{\beta} & \boldsymbol{\beta}^T \omega_{22} \\ \omega_{22} \boldsymbol{\beta} & \omega_{22} \end{pmatrix},$$

y por lo tanto se recupera $\theta=(\xi,\omega)$ (para más detalles véase San Martín [76]).

El concepto de corte es típicamente utilizado en econometría para definir una variable exógena, o de manera informal, una variable no-aleatoria. Mas precisamente, si la distribución condicional de [Y|X] es invariante ante cambios en la distribución de X, entonces X es exógeno para θ_1 (véase Engle *et al.* [38], Maddala [65], Pearl [69]). Más aún, se puede reducir la información contenida en la distribución conjunta en la distribución condicional y por lo tanto, realizar inferencia sin considerar la información disponible en el modelo marginal (Christensen y Kiefer [29]). Es importante mencionar que esto último sólo se puede realizar cuando la distribución de Z es normal (Spanos [78]). En consecuencia, bajo el supuesto de normalidad de Z (véase Spanos [78]), el modelo estadístico subyacente a la reducción condicional es el modelo de regresión lineal normal homocedástico.

Desde el punto de vista del modelamiento, el enfoque de reducción probabilística propuesto por Spanos [79, 80] está centrado en especificar un modelo que capture lo esencial de los datos y proporcione un resumen confiable de la información probabilística, es decir, todo aquello relacionado al proceso de muestreo que generó las variables aleatorias observadas. Claramente, la reducción probabilística de un modelo estadístico está relacionada estrechamente a los conceptos estudiados por Fisher [40] y Basu [23] sobre la reducción de los datos y la reducción de los modelos, respectivamente.

Un ejemplo importante en el cual no se puede obtener corte, y por lo tanto exogeneidad, es el modelo t-Student. En este caso se tiene $Z_i = (Y_i, \mathbf{X}_i)^T \stackrel{i.i.d.}{\sim} t_{1+k}(\boldsymbol{\xi}, \omega, \varsigma, \nu),$ $i=1,\ldots,n,$ donde $t_m(\boldsymbol{\xi},\omega,\varsigma,\nu)$ denota la distribución t-Student generalizada con vector de localización ξ , matriz de dispersión ω , parámetro ζ y ν grados de libertad, considerada por Arellano-Valle y Bolfarine [3]. Considérese la misma partición (2.1.3) para ξ y ω . De los resultados mostrados en Arellano-Valle y Bolfarine [3] se tiene que $\theta_1=(\alpha,m{eta},\omega^2,\varsigma^*,m{\xi}_2,\omega_{22},
u^*),$ donde $\alpha,\,m{eta}$ y ω^2 son definidos al igual que en el modelo normal y $\varsigma^* = \varsigma + \boldsymbol{\xi}_2^T \omega_{22}^{-1} \boldsymbol{\xi}_2$ y $\nu^* = \nu + k$; y $\boldsymbol{\theta}_2 = (\boldsymbol{\xi}_2, \omega_{22}, \varsigma, \nu)$. En este caso, se puede verificar que θ no está en biyección con θ_1 y θ_2 , pues para un valor particular de ν en el modelo marginal, se tiene otro en el modelo condicional, a saber, $\nu = \nu^* - k$. De igual manera sucede con el parámetro ς , el cual se escribe en el modelo marginal como ς , mientras que en el modelo condicional se define como $\varsigma = \varsigma^* - \xi_2^T \omega_{22}^{-1} \xi_2$. Luego, se concluye que (θ_1, θ_2) no varían libremente. Consecuéntemente, no hay corte en el modelo conjunto t-Student y por lo tanto, X_i no es exógeno para θ_1 . Más aún, de Arellano-Valle y Bolfarine [3] se tiene $\operatorname{que} \mathbf{X}_{i} \overset{i.i.d.}{\sim} t_{k}(\boldsymbol{\xi}_{2}, \omega_{22}, \varsigma, \nu) \mathbf{y} \left[Y_{i} | \mathbf{X}_{i} \right] \overset{ind.}{\sim} t(\alpha + \mathbf{X}_{i}^{T} \boldsymbol{\beta}, \omega^{2}, \varsigma^{*} + \mathbf{X}_{i}^{T} \omega_{22}^{-1} \mathbf{X}_{i} - 2 \mathbf{X}_{i}^{T} \omega_{22}^{-1} \boldsymbol{\xi}_{2}, \nu^{*}),$

 $i=1,\ldots,n,$ y que

$$E(Y_{i}|\mathbf{X}_{i}, \boldsymbol{\theta}_{1}) = \alpha + \mathbf{X}_{i}^{T}\boldsymbol{\beta},$$

$$V(Y_{i}|\mathbf{X}_{i}, \boldsymbol{\theta}_{1}) = \omega^{2} \left(\frac{\varsigma^{*} + \mathbf{X}_{i}^{T}\omega_{22}^{-1}\mathbf{X}_{i} - 2\mathbf{X}_{i}^{T}\omega_{22}^{-1}\boldsymbol{\xi}_{2}}{\nu^{*} - 2} \right), \quad (\nu^{*} > 2).$$

implicando que la reducción de la distribución t-Student generalizada genera el modelo lineal t-Student heterocedástico (Spanos [78]). Nótese que en este caso, la estimación de los parámetros del modelo condicional depende de los parámetros del modelo marginal. Esta situación implica que el proceso de inferencia sobre los parámetros del modelo condicional no puede hacerse separadamente del modelo marginal.

2.3. Distribución condicional skew-normal

En esta sección se considera la descomposición marginal-condicional del modelo conjunto skew-normal. Para tal fin, se considera primero la construcción de la distribución skew-normal tal como se propuso en el Capítulo 1. Sean U_0 y U_1 dos vectores aleatorios de dimensión $m \times 1$ y $d \times 1$, respectivamente, y suponga que

$$\mathbf{U} = \begin{pmatrix} \mathbf{U_0} \\ \mathbf{U_1} \end{pmatrix} \sim \mathcal{N}_{m+d} \begin{pmatrix} \begin{pmatrix} 0 \\ \mathbf{0} \end{pmatrix}, \Omega^* = \begin{pmatrix} I_m & \Delta^T \\ \Delta & \bar{\Omega} \end{pmatrix} \end{pmatrix}, \tag{2.3.1}$$

donde Ω^* es una matriz de correlación. Sea \mathbf{Z}_0 un vector aleatorio tal que $\mathbf{Z}_0 \stackrel{d}{=} [\mathbf{U}_1 | \mathbf{U}_0 > 0]$. Entonces, utilizando el Lema 1.2.1 propuesto en el Capítulo 1, la función de densidad

de $\mathbf{Z} = \boldsymbol{\xi} + \omega \mathbf{Z}_0$ está dada por

$$f_{\mathbf{Z}}(\mathbf{z}|\boldsymbol{\xi},\Omega,\Delta) = 2^{m}\phi_{d}(\mathbf{z};\boldsymbol{\xi},\Omega)\Phi_{m}\left(\Delta^{T}\bar{\Omega}^{-1}\omega^{-1}(\mathbf{z}-\boldsymbol{\xi});I_{m}-\Delta^{T}\bar{\Omega}^{-1}\Delta\right), \qquad \mathbf{z} \in \mathbb{R}^{d}$$
(2.3.2)

donde $\boldsymbol{\xi}$ es un vector de localización, ω es una matriz diagonal formada por las desviaciones estándar de Ω y $\Omega = \omega \bar{\Omega} \omega$, donde $\bar{\Omega}$ es una matriz de correlación. El parámetro que caracteriza la densidad 2.3.2 es dado por $\boldsymbol{\theta} = (\boldsymbol{\xi}, \Omega, \Delta) \in \mathbb{R}^d \times \mathcal{C}_{d \times d} \times \mathcal{M}_{d \times m}$, donde $\mathcal{M}_{p \times q}$ representa el espacio de las matrices rectangulares de dimensión $p \times q$. La notación utilizada para referirse a la distribución de \mathbf{Z} es $\mathcal{SN}_{d,m}(\boldsymbol{\xi},\Omega,\Delta)$. En el caso de que m=1, se utilizará la notación $\mathcal{SN}_d(\boldsymbol{\xi},\Omega,\delta)$ considerada además en el resto del presente documento. Nota 2.3.1. Nótese que, por los resultados obtenidos en el Capítulo 1, la parametrización $\boldsymbol{\theta}$ de la densidad (2.3.2) es identificada, toda vez que para la construcción de la distribución del vector aleatorio \mathbf{Z}_0 , realizando una reducción por condicionamiento de una distribución normal, se considera una matriz de correlación y no una matriz de varianzas covarianzas.

El siguiente lema caracteriza la descomposición marginal-condicional del vector aleatorio Z.

Lema 2.3.1. Sea U_{11} y U_{12} una partición de U_1 con dimensiones $d_1 \times 1$ y $d_2 \times 1$ respectivamente, tal que $d_1 + d_2 = d$. Considérese la siguiente partición inducida:

$$\boldsymbol{\xi} = \begin{pmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \end{pmatrix}, \quad \Delta = \begin{pmatrix} \Delta_1 \\ \Delta_2 \end{pmatrix}, \quad \bar{\Omega} = \begin{pmatrix} \bar{\Omega}_{11} & \bar{\Omega}_{12} \\ \bar{\Omega}_{21} & \bar{\Omega}_{22} \end{pmatrix}, \quad \omega = \begin{pmatrix} \omega_1 & 0 \\ 0 & \omega_2 \end{pmatrix}.$$

Sean además, Y y X vectores aleatorios de dimensión $d_1 \times 1$ y $d_2 \times 1$, respectivamente, tal que $Z = (Y, X)^T \stackrel{d}{=} [U_{11}, U_{12} | U_0 > 0]$, con $Z \in \mathbb{R}^d$. Entonces, $X \stackrel{d}{=} [U_{12} | U_0 > 0]$ y

$$[Y|X] \stackrel{d}{=} [U_{11}|U_0 > 0, U_{12} = x], y \text{ por lo tanto}$$

$$\begin{split} f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x},\theta_1) &= \phi_{d_1}(\mathbf{y};\alpha+\mathbf{x}^T\boldsymbol{\beta},\Omega_{1,2}) \\ &\times \frac{\Phi_m(\Delta_2^T\bar{\Omega}_{22}^{-1}\omega_2^{-1}(\mathbf{x}-\boldsymbol{\xi}_2)+\Delta_{1,2}^T\bar{\Omega}_{1,2}^{-1}\omega_1^{-1}(\mathbf{y}-\alpha-\mathbf{x}^T\boldsymbol{\beta});I_m-\Delta_2^T\bar{\Omega}_{22}^{-1}\Delta_2-\Delta_{1,2}^T\bar{\Omega}_{1,2}^{-1}\Delta_{1,2})}{\Phi_m(\Delta_2^T\bar{\Omega}_{22}^{-1}\omega_2^{-1}(\mathbf{x}-\boldsymbol{\xi}_2);I_m-\Delta_2^T\bar{\Omega}_{22}^{-1}\Delta_2)} \\ f_{\mathbf{X}}(\mathbf{x}|\theta_2) &= 2^m\phi_{d_2}(\mathbf{x};\boldsymbol{\xi}_2,\Omega_{22})\Phi_m(\Delta_2^T\bar{\Omega}_{22}^{-1}\omega_2^{-1}(\mathbf{x}-\boldsymbol{\xi}_2);I_m-\Delta_2^T\bar{\Omega}_{22}^{-1}\Delta_2) \end{split}$$

donde $\alpha = \boldsymbol{\xi}_1 - \Omega_{12}\Omega_{22}^{-1}\boldsymbol{\xi}_2, \boldsymbol{\beta} = \Omega_{22}^{-1}\Omega_{21}, \bar{\Omega}_{1,2} = \bar{\Omega}_{11} - \bar{\Omega}_{12}\bar{\Omega}_{22}^{-1}\bar{\Omega}_{21}, \omega_1\bar{\Omega}_{1,2}\omega_1 = \Omega_{1,2} = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}, \omega_i\bar{\Omega}_{1j}\omega_j = \Omega_{ij}, para\ i, j = 1, 2\ y\ \Delta_{1,2} = \Delta_1 - \bar{\Omega}_{12}\bar{\Omega}_{22}^{-1}\Delta_2. \ \theta_1 \ es \ el$ $parámetro\ del\ modelo\ condicional\ y\ está\ dado\ por\ el\ vector\ (\alpha, \boldsymbol{\beta}, \Omega_{1,2}, \omega_1, \Delta_{1,2}, \boldsymbol{\xi}_2, \Omega_{22}, \Delta_2) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_1} \times \mathcal{C}_{d_1 \times d_1} \times \mathcal{C}_{d_1 \times d_1} \times \mathcal{M}_{d_1 \times m} \times \mathbb{R}^{d_2} \times \mathcal{C}_{d_2 \times d_2} \times \mathcal{M}_{d_2 \times m}; \ mientras\ que\ \theta_2 \ es\ el$ $parámetro\ del\ modelo\ marginal\ y\ está\ dado\ por\ (\boldsymbol{\xi}_2, \Omega_{22}, \Delta_2) \in \mathbb{R}^{d_2} \times \mathcal{C}_{d_2 \times d_2} \times \mathcal{M}_{d_2 \times m}.$

Dem. Para la demostración ver Arellano-Valle y Azzalini [2]. □

Nota 2.3.2. Tanto la densidad condicional de [Y|X], cuanto la densidad marginal de X pertenecen a la clase de distribuciones unified skew-normal introducida por Arellano-Valle y Azzalini [2].

Nota 2.3.3. Las parametrizaciones de los modelos marginal y condicional son claramente parametrizaciones inducidas. En lo que sigue del capítulo siempre se hará referencia a esta parametrización salvo que se indique lo contrario mencionando, la parametrización identificada y su respectiva restricción de identificación.

Ahora, sea $\mathbf{Z}_i = (Y_i, \mathbf{X}_i)^T$, un vector aleatorio tal que Y_i es una variable aleatoria univariada en \mathbb{R} y \mathbf{X}_i un vector aleatorio de dimensión $k \times 1$. Supóngase que para cada $i=1,\ldots,n$, los vectores \mathbf{Z}_i son independientes y con distribución idéntica, skew-normal con vector de localización $\boldsymbol{\xi}$, matriz de escala Ω y vector de asimetría $\boldsymbol{\delta}$, es decir, $\mathbf{Z}_i \overset{i.i.d.}{\sim} \mathcal{SN}_{1+k}(\boldsymbol{\xi},\Omega,\boldsymbol{\delta})$. La densidad condicional que genera este modelo conjunto skew-normal

sigue de la aplicación de los resultados en obtenidos en el Lema 2.3.1. Utilizando dicha densidad y tomando $\Omega_{1,2} = \omega^2$, $\Delta_{1,2} = \delta_{1,2}$, $\Delta_2 = \delta_2$, se obtiene la función generadora de cumulantes del modelo condicional skew-normal (véase Arellano-Valle y Azzalini [2]),

$$K(t) = (\alpha + \mathbf{X}_{i}^{T}\boldsymbol{\beta})t + \frac{\omega^{2}t^{2}}{2} + \log\Phi(\boldsymbol{\delta}_{2}^{T}\bar{\Omega}_{22}^{-1}\omega_{2}^{-1}(\mathbf{X}_{i} - \boldsymbol{\xi}_{2}) + \delta_{1,2}\omega_{1}t; 1 - \boldsymbol{\delta}_{2}^{T}\bar{\Omega}_{22}^{-1}\boldsymbol{\delta}_{2})$$
$$-\log\Phi(\boldsymbol{\delta}_{2}^{T}\bar{\Omega}_{22}^{-1}\omega_{2}^{-1}(\mathbf{X}_{i} - \boldsymbol{\xi}_{2}); 1 - \boldsymbol{\delta}_{2}^{T}\bar{\Omega}_{22}^{-1}\boldsymbol{\delta}_{2}). \tag{2.3.3}$$

Usando (2.3.3), se obtiene

$$E(Y_i|\mathbf{X}_i,\theta_1) = \alpha + \mathbf{X}_i^T \boldsymbol{\beta} + \frac{\delta_{1,2}\omega_1}{\sqrt{1 - \delta_2^T \bar{\Omega}_{22}^{-1} \delta_2}} \zeta_1 \left(\frac{\delta_2^T \bar{\Omega}_{22}^{-1} \omega_2^{-1} (\mathbf{X}_i - \boldsymbol{\mu}_2)}{\sqrt{1 - \delta_2^T \bar{\Omega}_{22}^{-1} \delta_2}} \right) (2.3.4)$$

$$V(Y_i|\mathbf{X}_i,\theta_1) = \omega^2 + \frac{\delta_{1,2}^2 \omega_1^2}{1 - \delta_2^T \bar{\Omega}_{22}^{-1} \delta_2} \zeta_2 \left(\frac{\delta_2^T \bar{\Omega}_{22}^{-1} \omega_2^{-1} (\mathbf{X}_i - \mu_2)}{\sqrt{1 - \delta_2^T \bar{\Omega}_{22}^{-1} \delta_2}} \right); \tag{2.3.5}$$

donde $\zeta_r(x)$ es la r-ésima derivada de $\zeta_0(x) = \log\{2\Phi(w)\}$, y

$$\zeta_1(0) = \sqrt{\frac{2}{\pi}} \quad \text{y} \quad \zeta_2(0) = \frac{2}{\pi},$$

$$y \zeta_2(w) = -\zeta_1(w)\{w + \zeta_1(w)\}.$$

La expresión (2.3.4) muestra que la reducción por condicionamiento del modelo skewnormal genera un modelo de regresión skew-normal, en el cual la relación entre las variables Y_i y \mathbf{X}_i , con $i=1,\ldots,n$ es no lineal, toda vez que la función ζ_1 es no lineal como función de \mathbf{X}_i , para $i=1,\ldots,n$. De otro lado, la expresión (2.3.5) muestra que la varianza de $[Y_i|\mathbf{X}_i]$ con $i=1,\ldots,n$ depende de \mathbf{X}_i y por lo tanto no es constante (heterocedasticidad). Además, $\theta_1=(\alpha,\beta,\omega^2,\omega_1,\delta_{1,2},\boldsymbol{\xi}_2,\Omega_{22},\delta_2)$ y $\theta_2=(\boldsymbol{\xi}_2,\Omega_{22},\delta_2)$ no varían libremente, pues θ_1 depende de $\boldsymbol{\xi}_2,\Omega_{22}$ y $\boldsymbol{\delta}_2$.

Nota 2.3.4. Es importante destacar que, aunque en el experimento conjunto skew-normal inducido por (Y, X) los parámetros son identificados, en el modelo condicional [Y|X] y marginal X la identificación no está garantizada y tampoco estudiada formalmente en el presente documento. Por lo tanto la parametrización utilizada en el resto del capítulo es la inducida (ver Capítulo 1). Sin embargo, en algunos casos se introducen restricciones de identificación en algunos modelos estadísticos de tal manera de evitar algunos problemas de identificabilidad. Estas restricciones son argumentadas por el análisis realizado en el capítulo anterior.

De la Definición 2.2.1, se concluye que X_i no es exógeno con respecto a θ_1 , y por lo tanto, en el modelo skew-normal (2.3.2) no es posible realizar inferencia considerando la información que proporciona el modelo condicional solamente. En virtud de lo anterior, en la siguiente sección se proponen condiciones necesarias y suficientes para lograr la variación libre entre los parámetros del modelo marginal y condicional skew-normal. En otras palabras, se caracteriza la exogeneidad de la variable sobre la cual se genera el modelo condicional, permitiéndose de esta manera separar la inferencia de los modelos marginal y condicional.

2.4. Reducción probabilística de la distribución skew-normal

Los resultados que se muestran a continuación caracterizan el corte y la exogeneidad en el modelo condicional skew-normal. Como subproducto, se obtienen algunos modelos de regresión skew-normal.

Proposición 2.4.1. Sea $\mathbf{Z}_i = (Y_i, \mathbf{X}_i)^T$, con $Y_i \in \mathbb{R}$, $\mathbf{X}_i \in \mathbb{R}^k$ para $i = 1, \dots, n$, independientes y con distribución idéntica skew-normal de parámetros $\boldsymbol{\xi}$, Ω y $\boldsymbol{\delta}$. Si marginalmente el vector aleatorio \mathbf{X}_i tiene distribución normal, entonces es exógeno para el parámetro del modelo condicional $\boldsymbol{\theta}_1 = (\alpha, \boldsymbol{\beta}, \omega^2, \omega_1, \delta_1)$. Más aún, la reducción por condicionamiento del modelo conjunto determinado por \mathbf{Z}_i , genera el modelo de regresión lineal skew-normal homocedástico, es decir,

$$Y_i = \alpha + \mathbf{X}_i^T \boldsymbol{\beta} + \epsilon_i; \qquad \epsilon_i \stackrel{i.i.d.}{\sim} \mathcal{SN}(\omega^2, \omega_1, \delta_1)$$
 (2.4.1)

y

$$E(Y_i|\mathbf{X}_i, \boldsymbol{\theta}_1) = \alpha + \mathbf{X}_i^T \boldsymbol{\beta} + \sqrt{\frac{2}{\pi}} \delta_1 \omega_1,$$

$$V(Y_i|\mathbf{X}_i, \boldsymbol{\theta}_1) = \omega^2 + \frac{2}{\pi} \delta_1^2 \omega_1^2;$$

 $para i = 1, \ldots, n.$

Dem. Usando el Lema 2.3.1 con $\delta_2=0$, se obtiene que $U_{12} \! \perp \! \! \! \perp \! \! \! \! \! \perp U_0$ y por lo tanto $X_i \stackrel{d}{=} U_{12} \sim \mathcal{N}_k(\xi_2,\Omega_{22})$, para $i=1,\ldots,n$. Además,

$$f_{Y|\mathbf{X}_i}(y|\mathbf{x}_i, \boldsymbol{\theta}_1) = 2\phi(y; \alpha + \mathbf{x}_i^T \boldsymbol{\beta}, \omega^2) \Phi\left(\frac{\delta_1 \omega_1 \omega^{-2} (y - \alpha - \mathbf{x}_i^T \boldsymbol{\beta})}{\sqrt{1 - \delta_1^2 \omega_1^2 \omega^{-2}}}\right).$$

Usando la Definición 2.2.1, se concluye que $\theta_1 = (\alpha, \beta, \omega^2, \omega_1, \delta_1)$ y $\theta_2 = (\xi_2, \Omega_{22})$ están en variación libre y por lo tanto X_i es exógeno para θ_1 . De (2.3.3) se tiene que, la función generadora de cumulantes del modelo condicional está dada por,

$$K(t) = (\alpha + \mathbf{x}_i^T \boldsymbol{\beta})t + \frac{1}{2}\omega^2 t^2 + \log\{2\Phi(\delta_1 \omega_1 t)\}.$$

Finalmente, considerando las dos primeras derivadas de esta función con respecto a t y haciendo t=0 se obtienen expresiones para la esperanza y varianza condicional.

Típicamente, y de acuerdo a los resultados de identificación obtenidos en el Capítulo 1, la restricción de identificación que se utiliza en la especificación del modelo skew-normal condicional es la siguiente:

$$\omega = \omega_1;$$

de esta manera, el parámetro identificado está dado por el vector $(\alpha, \beta, \omega^2, \delta_1)$.

Proposición 2.4.2. Sea $\mathbf{Z}_i = (Y_i, \mathbf{X}_i)^T$, con $Y_i \in \mathbb{R}$, $\mathbf{X}_i \in \mathbb{R}^k$ para $i = 1, \ldots, n$, independientes y con distribución idéntica skew-normal de parámetros $\boldsymbol{\xi}$, Ω y $\boldsymbol{\delta}$. Si condicionalmente en \mathbf{X}_i , Y_i tiene distribución normal; entonces \mathbf{X}_i es exógeno para $\boldsymbol{\theta}_1 = (\alpha, \boldsymbol{\beta}, \omega^2)$. Más aún, la reducción por condicionamiento del modelo conjunto determinado por \mathbf{Z}_i , genera el modelo de regresión normal lineal homocedástico, es decir,

$$Y_i = \alpha + \mathbf{X}_i^T \boldsymbol{\beta} + \epsilon_i; \qquad \epsilon_i \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \omega^2)$$
 (2.4.2)

y

$$E(Y_i|\mathbf{X}_i,\boldsymbol{\theta}_1) = \alpha + \mathbf{X}_i^T\boldsymbol{\beta},$$

 $V(Y_i|\mathbf{X}_i,\boldsymbol{\theta}_1) = \omega^2;$

 $para i = 1, \ldots, n.$

Dem. Considérese el Lema 2.3.1 con $\delta_1 = \bar{\Omega}_{12}\bar{\Omega}_{22}^{-1}\boldsymbol{\delta}_2$. En la densidad condicional de $[Y_i|\mathbf{X}_i]$ se tiene que $\delta_{1,2}=0$ y por lo tanto

$$f_{\mathbf{Y}_{i}|\mathbf{X}_{i}}(y_{i}|\mathbf{x}_{i},\boldsymbol{\theta}_{1}) = \phi(y_{i};\boldsymbol{\alpha} + \mathbf{x}_{i}^{T}\boldsymbol{\beta},\boldsymbol{\omega}^{2}), \quad \mathbf{y}$$

$$f_{\mathbf{X}_{i}}(\mathbf{x}_{i}|\boldsymbol{\theta}_{2}) = 2\phi_{k}(\mathbf{x}_{i};\boldsymbol{\xi}_{2},\Omega_{22})\Phi\left(\frac{\boldsymbol{\delta}_{2}^{T}\bar{\Omega}_{22}^{-1}\boldsymbol{\omega}_{2}^{-1}(\mathbf{x}_{i} - \boldsymbol{\xi}_{2})}{1 - \boldsymbol{\delta}_{2}^{T}\bar{\Omega}_{22}^{-1}\boldsymbol{\delta}_{2}^{T}}\right).$$

De la Definición 2.2.1, se concluye que $\theta_1 = (\alpha, \beta, \omega^2)$ y $\theta_2 = (\xi_2, \Omega_{22}, \delta_2)$ están en variación libre y que X_i es exógeno para θ_1 . Finalmente, el cálculo de los momentos condicionales se realiza utilizando el modelo normal.

La siguiente proposición caracteriza la función de densidad conjunta del modelo skewnormal de manera tal que los parámetros del modelo marginal y condicional varíen libremente.

Proposición 2.4.3. El vector $\mathbf{Z}_i = (Y_i, \mathbf{X}_i)^T$, con $Y_i \in \mathbb{R}$ y $\mathbf{X}_i \in \mathbb{R}^k$, para $i = 1, \dots, n$, tiene descomposición marginal-condicional de la forma,

$$f_{Y_{i}|\mathbf{X}_{i}}(y_{i}|\mathbf{x}_{i},\boldsymbol{\theta}_{1}) = \phi(y_{i};\alpha + \mathbf{x}_{i}^{T}\boldsymbol{\beta},\omega^{2}) \frac{\Phi\left(\frac{\boldsymbol{\delta}_{2}^{T}\bar{\Omega}_{22}^{-1}\omega_{2}^{-1}(\mathbf{x}_{i}-\boldsymbol{\xi}_{2})+\delta_{1,2}\omega^{-1}(y_{i}-\alpha-\mathbf{x}_{i}^{T}\boldsymbol{\beta})}{\sqrt{1-\boldsymbol{\delta}_{2}^{T}\bar{\Omega}_{22}^{-1}\boldsymbol{\delta}_{2}-\delta_{1,2}^{2}}}\right)} \Phi\left(\frac{\boldsymbol{\delta}_{2}^{T}\bar{\Omega}_{22}^{-1}\boldsymbol{\delta}_{2}-\delta_{1,2}^{2}}{\sqrt{1-\boldsymbol{\delta}_{2}^{T}\bar{\Omega}_{22}^{-1}\boldsymbol{\delta}_{2}}}\right) (2.4.3)$$

$$f_{\mathbf{X}_{i}}(\mathbf{x}_{i}|\boldsymbol{\theta}_{2}) = 2\phi_{k}(\mathbf{x}_{i};\boldsymbol{\xi}_{2}^{*},\bar{\Omega}_{22}^{*})\Phi\left(\frac{\boldsymbol{\delta}_{2}^{*T}\bar{\Omega}_{22}^{*-1}\omega_{2}^{*-1}(\mathbf{x}_{i}-\boldsymbol{\xi}_{2}^{*})}{\sqrt{1-\boldsymbol{\delta}_{2}^{*T}\bar{\Omega}_{22}^{*-1}\boldsymbol{\delta}_{2}^{*}}}\right), \tag{2.4.4}$$

con $\theta_1 = (\alpha, \beta, \omega^2, \delta_{1,2}, \xi_2, \Omega_{22}, \delta_2)$ y $\theta_2 = (\xi_2^*, \Omega_{22}^*, \delta_2^*)$ en variación libre si solo si, la distribución de \mathbf{Z}_i es de la forma

$$f_{\mathbf{Z}_i}(\mathbf{z}_i|\boldsymbol{\theta}) = \phi_{1+k}(\mathbf{z}_i; \boldsymbol{\mu}, \Omega) \frac{\Phi_2(\boldsymbol{\beta}^* \mathbf{z}_i - \boldsymbol{\alpha}^*; \mathbf{0}, I_2)}{\Phi_2(\boldsymbol{\beta}^{**} \mathbf{z}_i - \boldsymbol{\alpha}^{**}; \mathbf{0}, I_2)};$$
(2.4.5)

donde

$$\mu = \begin{pmatrix} \alpha + \beta^T \xi_2^* \\ \xi_2^* \end{pmatrix}, \Omega = \begin{pmatrix} \omega^2 + \beta^T \Omega_{22}^* \beta & \beta^T \Omega_{22}^* \\ \Omega_{22}^* \beta & \Omega_{22}^* \end{pmatrix}, \tag{2.4.6}$$

$$\alpha^* = \begin{pmatrix} \delta_2^T \bar{\Omega}_{22}^{-1} \omega_2^{-1} \xi_2 + \delta_{1,2} \omega^{-1} \alpha \\ {\delta_2^*}^T \bar{\Omega}_{22}^{*-1} \omega_2^{*-1} \xi_2^* \end{pmatrix}, \beta^* = \begin{pmatrix} H_1^{-1} \delta_{1,2} \omega^{-1} & H_1^{-1} (\delta_2^T \bar{\Omega}_{22}^{-1} \omega_2^{-1} - \delta_{1,2} \omega^{-1} \beta^T) \\ 0 & H_2^{-1} \delta_2^* \bar{\Omega}_{22}^{*-1} \omega_2^{*-1} \end{pmatrix}, (2.4.7)$$

$$\alpha^{**} = \begin{pmatrix} 0 \\ \frac{\delta_2^T \bar{\Omega}_{22}^{-1} \omega_2^{-1} \xi_2}{\sqrt{1 - \delta_2^T \bar{\Omega}_{22}^{-1} \delta_2}} \end{pmatrix}, \beta^{**} = \begin{pmatrix} 0 & 0 \\ 0 & \frac{\delta_2^T \bar{\Omega}_{22}^{-1} \omega_2^{-1}}{\sqrt{1 - \delta_2^T \bar{\Omega}_{22}^{-1} \delta_2}} \end{pmatrix};$$
(2.4.8)

 $H_1 = \sqrt{1 - \delta_2^T \bar{\Omega}_{22}^{-1} \delta_2 - \delta_{1,2}^2} \ y \ H_2 = \sqrt{1 - \delta_2^{*T} \bar{\Omega}_{22}^{*-1} \delta_2^*}.$ Más aún, \mathbf{X}_i es exógeno para θ_1 y la reducción por condicionamiento del modelo conjunto determinado por \mathbf{Z}_i , genera el modelo de regresión no-lineal skew-normal heterocedástico, es decir,

$$Y_i = \alpha + \mathbf{X}_i^T \boldsymbol{\beta} + \epsilon_i \tag{2.4.9}$$

 $con \ \epsilon_i, i = 1, ..., n \ independientes \ con \ densidad \ (2.4.3). \ Además,$

$$E(Y_{i}|\mathbf{X}_{i}) = \alpha + \mathbf{X}_{i}^{T} \boldsymbol{\beta} + \frac{\delta_{1,2} \omega}{\sqrt{1 - \delta_{2}^{T} \bar{\Omega}_{22}^{-1} \delta_{2}}} \zeta_{1} \left(\frac{\delta_{2}^{T} \bar{\Omega}_{22}^{-1} \omega_{2}^{-1} (\mathbf{X}_{i} - \boldsymbol{\mu}_{2})}{\sqrt{1 - \delta_{2}^{T} \bar{\Omega}_{22}^{-1} \delta_{2}}} \right),$$

$$V(Y_{i}|\mathbf{X}_{i}) = \omega^{2} \left(1 + \frac{\delta_{1,2}^{2}}{1 - \delta_{2}^{T} \bar{\Omega}_{22}^{-1} \delta_{2}} \zeta_{2} \left(\frac{\delta_{2}^{T} \bar{\Omega}_{22}^{-1} \omega_{2}^{-1} (\mathbf{X}_{i} - \boldsymbol{\mu}_{2})}{\sqrt{1 - \delta_{2}^{T} \bar{\Omega}_{22}^{-1} \delta_{2}}} \right) \right).$$

Dem. La demostración sigue de la descomposición marginal condicional de la distribución conjunta, es decir,

$$\begin{split} f_{\mathbf{Z}_{i}}(\mathbf{z}_{i}|\boldsymbol{\theta}) &= f_{Y_{i}|\mathbf{X}_{i}}(y_{i}|\mathbf{x}_{i},\boldsymbol{\theta}_{1})f_{\mathbf{X}_{i}}(\mathbf{x}_{i}|\boldsymbol{\theta}_{2}) \\ &= \phi_{1}(y_{i};\alpha + \mathbf{x}_{i}^{T}\boldsymbol{\beta},\omega^{2})\phi_{k}(\mathbf{x}_{i};\boldsymbol{\xi}_{2}^{*},\Omega_{22}^{*}) \frac{\Phi\left(\frac{\delta_{2}^{T}\bar{\Omega}_{22}^{-1}\omega_{2}^{-1}(\mathbf{x}_{i}-\boldsymbol{\xi}_{2}^{*})+\delta_{1,2}\omega^{-1}(y_{i}-\alpha-\mathbf{x}_{i}^{T}\boldsymbol{\beta})}{\sqrt{1-\delta_{2}^{T}\bar{\Omega}_{22}^{-1}\delta_{2}-\delta_{1,2}^{2}}}\right)}{\Phi\left(\frac{\delta_{2}^{T}\bar{\Omega}_{22}^{-1}\omega_{2}^{-1}(\mathbf{x}_{i}-\boldsymbol{\xi}_{2}^{*})}{\sqrt{1-\delta_{2}^{T}\bar{\Omega}_{22}^{-1}\delta_{2}}}\right)}{\nabla\left(\frac{\delta_{2}^{T}\bar{\Omega}_{22}^{-1}\omega_{2}^{-1}(\mathbf{x}_{i}-\boldsymbol{\xi}_{2}^{*})}{\sqrt{1-\delta_{2}^{T}\bar{\Omega}_{22}^{-1}\delta_{2}}}\right)}\right)}. \end{split}$$

Usando algunas propiedades de la distribución normal multivariada, se tiene que

$$\phi_{1+k}(\mathbf{z}_i; \boldsymbol{\xi}, \Omega) = \phi_1(y_i; \alpha + \mathbf{x}_i^T \boldsymbol{\beta}, \omega^2) \phi_k(\mathbf{x}_i; \boldsymbol{\xi}_2^*, \Omega_{22}^*)$$

donde,

$$\boldsymbol{\xi} = \begin{pmatrix} \alpha + \beta^T \boldsymbol{\xi}_2^* \\ \boldsymbol{\xi}_2^* \end{pmatrix}, \Omega = \begin{pmatrix} \omega^2 + \boldsymbol{\beta}^T \Omega_{22}^* \boldsymbol{\beta} & \boldsymbol{\beta}^T \Omega_{22}^* \\ \Omega_{22}^* \boldsymbol{\beta} & \Omega_{22}^* \end{pmatrix}.$$

Ahora, si se considera las variables aleatorias W_1 y W_2 , independientes y con distribución normal estándar, entonces,

$$P(\mathbf{T} \le \boldsymbol{\beta}^* \mathbf{z}_i) = P(W_1 \le H_1^{-1} \delta_{1,2} \omega^{-1} y_i + H_1^{-1} (\boldsymbol{\delta}_2^T \bar{\Omega}_{22}^{-1} \omega_2^{-1} - \delta_{1,2} \omega^{-1} \boldsymbol{\beta}^T) \mathbf{x}_i) \times P(W_2 \le H_2^{-1} \boldsymbol{\delta}_2^{\star T} \bar{\Omega}_{22}^{\star T} \omega_2^{\star -1} \mathbf{x}_i),$$

donde $\mathbf{T} = \mathbf{W} + \boldsymbol{\alpha}^*$, $\mathbf{W} = (W_1, W_2)^T$ y $\boldsymbol{\beta}^*$ y $\boldsymbol{\alpha}^*$ están dados por (2.4.7). Por lo tanto, dado que $\mathbf{T} \sim \mathcal{N}_2(\boldsymbol{\alpha}^*, I_2)$, se tiene,

$$P(\mathbf{T} \leq \boldsymbol{\beta}^* \mathbf{z}_i) = \Phi_2(\boldsymbol{\beta}^* \mathbf{z}_i - \boldsymbol{\alpha}^*; \mathbf{0}, I_2).$$

De manera similar,

$$P(\mathbf{T}^* \leq \boldsymbol{\beta}^{**} \mathbf{z}_i) = P(W_1 \leq 0) P\left(W_2 \leq \frac{\boldsymbol{\delta}_2^T \bar{\Omega}_{22}^{-1} \omega_2^{-1}}{\sqrt{1 - \boldsymbol{\delta}_2^T \bar{\Omega}_{22}^{-1} \boldsymbol{\delta}_2}} \mathbf{x}_i\right),$$

donde $\mathbf{T}^* = \mathbf{W} + \boldsymbol{\alpha}^{**}$, $\mathbf{W} = (W_1, W_2)^T$ y $\boldsymbol{\beta}^{**}$ y $\boldsymbol{\alpha}^{**}$ están dados por (2.4.8). Luego, dado que $\mathbf{T}^* \sim \mathcal{N}_2(\boldsymbol{\alpha}^{**}, I_2)$, se tiene,

$$P(\mathbf{T} \leq \boldsymbol{\beta}^{**}\mathbf{z}_i) = \Phi_2(\boldsymbol{\beta}^{**}\mathbf{z}_i - \boldsymbol{\alpha}^{**}; \mathbf{0}, I_2).$$

Finalmente, la esperanza y la varianza condicional están dadas por (2.3.4-2.3.5).

El siguiente corolario muestra la distribución conjunta del vector $\mathbf{Z}_i = (Y_i, \mathbf{X}_i)^T$ de manera tal que la reducción al modelo condicional genere el modelo de regresión lineal skew-normal heterocedástico, sin imponer condición alguna sobre la distribución marginal del vector aleatorio \mathbf{X}_i .

Corolario 2.4.4. Si el vector aleatorio $\mathbf{Z}_i = (Y_i, \mathbf{X}_i)^T$, con $Y_i \in \mathbb{R}$ y $\mathbf{X}_i \in \mathbb{R}^k$, para $i = 1, \dots, n$, tiene distribución conjunta de la forma

$$f_{\mathbf{Z}_i}(\mathbf{z}_i|\boldsymbol{\theta}) = 2^2 \phi_{1+k}(\mathbf{z}_i; \boldsymbol{\mu}, \Omega) \Phi_2(\boldsymbol{\beta}^* \mathbf{z}_i - \boldsymbol{\alpha}^*; \mathbf{0}, I_2);$$
(2.4.10)

donde

$$\mu = \begin{pmatrix} \alpha + \beta^T \xi_2^* \\ \xi_2^* \end{pmatrix}, \Omega = \begin{pmatrix} \omega^2 + \beta^T \Omega_{22}^* \beta & \beta^T \Omega_{22}^* \\ \Omega_{22}^* \beta & \Omega_{22}^* \end{pmatrix}, \tag{2.4.11}$$

$$\alpha^* = \begin{pmatrix} \delta_{1,2}\omega^{-1}\alpha \\ \delta_2^{*T}\bar{\Omega}_{22}^{*-1}\omega_2^{*-1}\xi_2^* \end{pmatrix}, \beta^* = \begin{pmatrix} \frac{\delta_{1,2}\omega^{-1}}{\sqrt{1-\delta_{1,2}^2}} & -\frac{\delta_{1,2}\omega^{-1}}{\sqrt{1-\delta_{1,2}^2}} \\ 0 & \frac{\delta_2^*\bar{\Omega}_{22}^{*-1}\omega_2^{*-1}}{\sqrt{1-\delta_2^*T}\bar{\Omega}_{22}^{*-1}\delta_2^*} \end{pmatrix}; \quad (2.4.12)$$

entonces, la reducción por condicionamiento genera el modelo de regresión lineal skewnormal homocedástico (2.4.1), con la restricción de identificación $\omega = \omega_1$.

Dem. La demostración es inmediata, haciendo
$$\delta_2 = 0$$
 en la Proposición 2.4.3.

2.5. Simulación

El objetivo principal de esta sección es mostrar el efecto que tiene el supuesto de exogeneidad de la variable explicativa X_i , $i=1,\ldots,n$ en la estimación de los parámetros de un modelo de regresión skew-normal. Para tal fin, se considera dos escenarios de simulación, los cuales se muestran a continuación. Es importante mencionar que se utilizó la parametrización identificada de los modelos condicionales através de la restricción de identificación propuesta en la Proposición 2.4.1

1. Escenario I:

■ Simulación de un modelo no-exógeno: Dado X_i para $i=1,\ldots,250$ con distribución skew-normal de parámetros $\theta_2=(\xi_2=5,\Omega_{22}=2,\delta_2=0,7)$, se genera $[Y_i|X_i]$, para $i=1\ldots,250$; con distribución skew-normal de parámetros $\theta_1=(\alpha=1,5,\beta=0,5,\omega^2=1,25,\delta_{1,2}=0,55,\xi_2=5,\Omega_{22}=2,\delta_2=0,7)$.

- Simulación de un modelo exógeno: Dado X_i para $i=1,\ldots,250$ con distribución skew-normal de parámetros $\boldsymbol{\theta}_2=(\xi_2=5,\Omega_{22}=2,\delta_2=0)$, se genera $[Y_i|X_i],\ i=1\ldots,250$; con distribución skew-normal de parámetros $\boldsymbol{\theta}_1=(\alpha=1,5,\beta=0,5,\omega^2=1,25,\delta_{1,2}=0,55)$.
- Modelo estimado: Para los datos simulados bajo el modelo exógeno y no exógeno, se ajusta el modelo de regresión lineal skew-normal homocedástico,

$$Y_i = \alpha + \beta X_i + \epsilon_i; \quad i = 1, \dots, 250,$$

asumiendo $\epsilon_i \stackrel{i.i.d.}{\sim} \mathcal{SN}(0, \omega^2, \delta_{1,2}), i = 1, \dots, 250.$

- Aspectos de la estimación: Para la estimación de los parámetros se utilizó el método de máxima verosimilitud, utilizando diferentes valores iniciales para el proceso de maximización.
- Resultados: En el Cuadro 2.1 se reporta la media del estimador de Monte Carlo y el error cuadrático medio del mismo (entre paréntesis) para 10,000 simulaciones.

Parametros	Modelo no-exógeno	Modelo exógeno
α	2,307(0,806)	1,576(0,163)
β	0,357(0,022)	0,499(0,001)
ω^2	1,094(0,100)	1,261(0,099)
$\delta_{1,2}$	0,407(0,148)	0,403(0,174)

Cuadro 2.1: Parámetros estimados vía máxima verosimilitud del modelo de regresión skewnormal.

2. Escenario II

• Simulación de un modelo skew-normal: Para diferentes distribuciones de la variable X_i , con i = 1, ..., 200, (normal y skew-normal) se genera la variable

condicional $[Y_i|X_i]$, $i=1,\ldots,200$; con distribución skew-normal de parámetros $\theta_1=(\alpha=1,5,\beta=0,3,\omega^2=1,\delta_{1,2}=1)$.

 Modelo estimado: Para los datos simulados, se ajusta el modelo de regresión lineal skew-normal homocedástico,

$$Y_i = \alpha + \beta X_i + \epsilon_i; \quad i = 1, \dots, 200,$$

asumiendo $\epsilon_i \overset{i.i.d.}{\sim} \mathcal{SN}(0, \omega^2, \delta_{1,2}), i = 1, \dots, 200.$

- Aspectos de la estimación: Para la estimación de los parámetros se utilizó el método de máxima verosimilitud, utilizando diferentes valores iniciales para el proceso de maximización.
- Resultados: En los Cuadros 2.2 y 2.3 se reporta la media del estimador de Monte Carlo y el error cuadrático medio del mismo (entre paréntesis) para 10,000 simulaciones.

Parámetros	$X_i \stackrel{i.i.d.}{\sim} \mathcal{N}(0,1)$	$X_i \stackrel{i.i.d.}{\sim} \mathcal{N}(-1,2)$	$X_i \stackrel{i.i.d.}{\sim} \mathcal{N}(3,0,5)$
α	1,5648(0,077)	1,5611(0,075)	1,5753(0,245)
β	0,3013(0,004)	0,2994(0,000001)	0,2982(0,017)
σ^2	1,1998(0,062)	1,2009(0,062)	1,2014(0,062)
$\delta_{1,2}$	1,1249(0,223)	1,1329(0,226)	$1,\!1152(0,\!241)$

Cuadro 2.2: Parámetros estimados vía máxima verosimilitud del modelo de regresión skewnormal, cuando la distribución de X_i es normal, $i=1\ldots,n$.

Parámetros	$X_i \stackrel{i.i.d.}{\sim} \mathcal{SN}(0, 1, 10)$	$X_i \stackrel{i.i.d.}{\sim} \mathcal{SN}(-1,2,-5)$	$X_i \stackrel{i.i.d.}{\sim} \mathcal{SN}(3,0,5,0,5)$
α	0,7182(0,656)	0,6317(0,834)	1,4316(0,230)
eta	0,7263(0,215)	$0,\!1280(0,\!037)$	0,3150(0,021)
σ^2	1,5019(0,260)	1,5213(0,280)	1,2347(0,067)
$\delta_{1,2}$	3,6523(7,315)	3,5818(6,948)	1,2737(0,153)

Cuadro 2.3: Parámetros estimados vía máxima verosimilitud del modelo de regresión skewnormal, cuando la distribución de X_i es skew-normal, $i=1\ldots,n$.

El escenario I muestra cómo las estimaciones de máxima verosimilitud de los parámetros del modelo de regresión skew-normal se ven afectadas cuando la variable independiente no es exógena. En este caso, los coeficientes de regresión tienden a ser sobre y sub-estimados, mientras que el parámetro de escala es sub-estimado. La estimación del parámetro de asimetría es muy similiar cuando la variable independiente es exógena y cuando no lo es.

Por otro lado, el Escenario II muestra el efecto de la exogeneidad en la construcción de la variable condicional [Y|X]. Cuando X tiene distribución normal, las estimaciones por el método de máxima verosimilitud son bastante similares a pesar de que los parámetros de escala y localización de dicha distribución cambian. Esto refleja la característica más importante de la exogeneidad a saber, que la distribución condicional de [Y|X] se mantiene invariante ante cambios en la distribución de X. Sin embargo, y como se puede apreciar en el Cuadro 2.3, si la distribución de la variable X no es normal, se pierde la exogeneidad y por lo tanto la invarianza de [Y|X]. Esto significa, que la estimación de los parámetros del modelo condicional se verá afectada ante cambios en la distribución de la variable marginal.

2.6. Resumen y conclusiones

El enfoque de reducción probabilística propone la especificación de un modelo estadístico que capture información relevante de los datos de manera confiable. En este sentido, el modelo estadístico propuesto debe funcionar bien en dos aspectos, a saber, la reducción de la información y la recuperación de información relevante de los datos.

Bajo este enfoque, y tomando como reducción probabilística el modelo condicional, el modelo de regresión que subyace a esta reducción claramente dependerá de la distribución conjunta sobre la cual tal reducción se realice.

En esta dirección, un claro ejemplo de lo anterior es el caso del modelo de regresión lineal normal homocedástico, el cual es originado asumiendo una distribución conjunta normal. Por otro lado, si se reemplaza la distribución normal por la *t*-Student, entonces se genera el modelo de regresión lineal *t*-Student heterocedástico. Cabe destacar que en este caso, no existe variación libre entre los parámetro de los modelos marginal y condicional y por lo tanto, hacer inferencia solamente en el modelo condicional implica cierta pérdida de información.

Más aún, al cambiar el supuesto de normalidad en la distribución conjunta por una distribución skew-normal multivariada, la reducción probabilística genera un modelo de regresión no-lineal, no existiendo corte y ni exogeneidad.

Sin embargo, al imponer ciertas restricciones sobre el espacio paramétrico de los modelos marginal y condiciona, se recupera el corte y la exogeneidad. En este caso se obtiene, bajo normalidad en la variable marginal, el modelo de regresión lineal skew-normal homocedástico. De otro lado, si el modelo condicional se especifica con distribución normal, se obtiene corte en el modelo y por lo tanto la exogeneidad de la variable marginal. Claramente, el modelo estadístico que subyace a esta reducción es el modelo de regresión lineal normal homocedástico.

Además, bajo una distribución conjunta específica skew-normal, la reducción probabilística del modelo conjunto genera el modelo de regresión lineal skew-normal homocedástico.

2.7. Agradecimientos

Quiero expresar mi gratitud a Ernesto San Martín (Pontificia Universidad Católica de Chile) por sus comentarios, sugerencias y aportes realizados en este capítulo.

Capítulo 3

Inferencia Bayesiana en la distribución skew-normal

"El objeto de esta memoria es diferente. Me propongo determinar la probabilidad de las causas de los eventos, tema que no ha sido puesto en consideración antes, pero que merece ser estudiado pues la ciencia del azar, bajo este punto de vista, puede ser útil en la vida civil."

Pierre S. Laplace, 1774

En el presente capítulo se considera la inferencia Bayesiana para los parámetros de localización, escala y asimetría de la distribución skew-normal de Azzalin [11] asumiendo diferentes distribuciones a priori. Se obtiene además, representaciones estocásticas para las distribuciones a posteriori y para el cálculo de los momentos a posteriori bajo ciertas especificaciones a priori. Los resultados obtenidos son aplicados con el objetivo de identificar un punto de cambio en el parámetro de asimetría en una serie de datos determinada. La ilustración de tales resultados es realizada con datos simulados y datos de mercados financieros emergentes Latino Americanos.

3.1. Introducción

El supuesto de normalidad puede ser un supuesto bastante restrictivo para datos provenientes de diferentes áreas de aplicación como por ejemplo finanzas, medio ambiente, medicina u otras más. Por ejemplo, es bien sabido que las distribuciones de probabilidad empíricas de los retornos de algunos mercados financieros emergentes son asimétricas y, en general, con colas más pesadas que la distribución normal.

En ese sentido, uno de los desafios relacionados al modelamiento de datos es encontrar clases de distribuciones flexibles capaces de representar este tipo de comportamiento en los datos, a saber, asimetría, colas pesadas o livianas, curtosis, etc. De acuerdo a esto, y tal como se estudió en el Capítulo 1, una de estas clases de distribuciones es la formalizada por Azzalini [11].

En el presente capítulo, se considera que la función de verosimilitud de los datos observados es un miembro de la clase de distribuciones skew-normal introducida por Azzalini [11], es decir, se asume que dada la parametrización identificada, con parámetros de localización, escala y asimetría, es decir, $\xi \in \mathbb{R}$, $\omega \in \mathbb{R}^+$ y $\lambda = \frac{\delta}{\sqrt{1-\delta^2}} \in \mathbb{R}$, la variable aleatoria X tiene distribución skew-normal $\mathcal{SN}(\xi,\omega^2,\lambda)$ con densidad dada por:

$$f(x|\xi,\omega^2,\lambda) = \frac{2}{\omega}\phi\left(\frac{x-\xi}{\omega}\right)\Phi\left(\lambda\left(\frac{x-\xi}{\omega}\right)\right), \quad x \in \mathbb{R}.$$
 (3.1.1)

Nótese que si $\xi=0$ y $\omega^2=1$ se obtiene la distribución skew-normal estándar, denotada por $\mathcal{SN}(\lambda)$.

La distribución en (3.1.1) es un caso particular de la familia de distribuciones univariada skew-normal-generalizada (SGN), introducida por Arellano-Valle *et al.* [10] y estudiada con detalle en el Capítulo 4. Particularmente, en dicho capítulo se establece que tal distribución puede ser representada como una mezcla de distribuciones skew-normal en el parámetro de asimetría (llamadas también mezclas de asimetría). Cabe destacar que el estudio de las mezclas de asimetría de la distribución skew-normal multivariada ha sido realizado por Arellano-Valle *et al.* [9] asumiendo una distribución de mezclas skew-normal para el parámetro de asimetría. En ese contexto, y desde el punto de vista Bayesiano, la distribución de mezcla es la distribución a priori del parámetro de asimetría λ .

El objetivo del presente capítulo es considerar mezclas de asimetría de distribuciones skew-normal (3.1.1) asumiendo como distribución de mezclas, o en otras palabras, como distribución a priori una distribución general. Producto de esto, se obtienen distribuciones predictivas y a posteriori de los parámetros identificados de la distribución skew-normal, generando clases extendidas de distribuciones asimétricas (skew). Además, bajo ciertas especificaciones de la función de verosimilitud y de las distribuciones a priori, se introducen representaciones estocásticas de las distribuciones a posteriori del parámetro de asimetría. Tales representaciones juegan un rol importante tanto desde el punto de vista teórico cuanto práctico. Por ejemplo, utilizando la representación estocástica de la distribución a posteriori del parámetro de asimetría se obtienen la media y varianza a posteriori de λ . Además,

estas representaciones estocásticas facilitan la implementación de algoritmos computacionales tales como el muestreo de Gibbs, el cual es utilizado en la solución del problema de detección de un punto de cambio en el parámetro de asimetría.

Finalmente, los resultados obtenidos en este capítulo son aplicados con el fin de identificar un punto de cambio en el parámetro de asimetría para una secuencia determinada de datos. Para ilustrar estos resultados se analiza un conjunto de dato simulados y de retornos de mercados financieros emergentes Latino Americanos. Es importante mencionar aquí, que un punto de cambio puede ser entendido como aquel instante en el que un cambio estructural es observado en una serie de datos o simplemente, cuando un dato atípico tiene cabida. La identificación de puntos de cambio en analísis de series de tiempo financiera es importante, toda vez que las tasas de ocurrencia de retornos atípicos son tomadas en cuenta en la evaluación del riesgo financiero. En general, en este tipo de serie temporal, es bastante común observar una alta tasa de ocurrencia de retornos atípicos. Este fenómeno es observado con frecuencia en mercados financieros emergentes dado que éstos son más susceptibles a los shocks que puedan occurrir en el mercado.

La identificación de puntos de cambio es un tópico de investigación bastante popular en muchas y diferentes áreas, como por ejemplo finanzas, genética, confiabilidad, medicina, entre otras. De hecho, muchos artículos han sido publicados en esta tema, entre los cuales destacan, desde el punto de vista clásico Dueker [36], Hawkins [52], Horváth y Kokoszka [53], y Jandhyala *et al.* [54]. Desde el punto de vista Bayesiano, los problemas de

puntos de cambio han sido estudiados por Barry y Hartigan [21, 22], Crowley [33], Quintana e Iglesias [71], Loschi y Cruz [62], Ruggeri y Sivaganesan [74], Fearnhead y Liu [39], Loschi *et al.* [63] entre muchos otros.

Antes de realizar la aplicación para detectar un puntos de cambio en el parámetro de asimetría en una serie de datos con distribución skew-normal, en la Sección 3.2 se introduce el análisis Bayesiano para el parámetro de asimétria asumiendo que la muestra aleatoria tiene distribución skew-normal estándar y considerando una distribución a priori general y una skew-normal para el parámetro de asimetría. En la Sección 3.3 se hace una extensión considerando una verosimilitud skew-normal de localización-escala. Luego, en la Sección 3.4 se considera la identificación de un punto de cambio en el parámetro de asimetría y en la la Sección 3.5 se aplican los resultados a un conjunto de datos simulados y de datos reales. Finalmente, la Sección 3.6 propone un resumen de los principales resultados y plantea algunas conclusiones como cierre del capítulo.

3.2. Inferencia Bayesiana en la distribución skew-normal estándar

3.2.1. Distribución a priori general para el parámetro de asimetría

En esta sección se considera mezclas de asimetría de distribuciones skew-normal independientes y se generalizan algunos resultados propuestos por Arellano-Valle *et al.* [9], considerando una distribución de mezclas general para el parámetro de asimetría λ . Como subproducto, se obtiene la inferencia a posteriori para λ .

Antes de iniciar el desarrollo de los cálculos, vale la pena introducir alguna notación útil que será usada a lo largo del presente capítulo. Se denota por $\mathcal{L}T\mathcal{N}_n(\mathbf{c}, \mu, \Sigma)$ la distribución de un vector aleatorio $\mathcal{N}_n(\mu, \Sigma)$ cuyos componentes son truncados por encima de $\mathbf{c} \in \mathbb{R}^n$, es decir, la distribución de $\mathbf{U}^{\mathbf{c}} = [\mathbf{U}|\mathbf{U}>\mathbf{c}]$, donde $\mathbf{U} \sim \mathcal{N}_n(\mu, \Sigma)$. Se denota por $t_n(\cdot; \mu, \Sigma, \nu)$ y $T_n(\cdot; \mu, \Sigma, \nu)$ a la función de densidad y función de distribución acumuladada respectivamente de la distribución t-Student con vector de localización μ , matriz de escala Σ y grados de libertad ν . Además, $E_{\lambda}(\cdot)$ denota la esperanza con respecto a la distribución de λ . Finalmente, $V_{\mathbf{x}}$ denota el vector aleatorio condicional $[V|\mathbf{X}=\mathbf{x}]$. En particular, $\lambda_{\mathbf{x}}$ denota $\lambda|\mathbf{x}$.

Se asume que, condicionalmente en λ , las variables aleatorias X_1, \ldots, X_n son independientes e idénticamente distribuídas con distribución skew-normal estándar. Es decir, para cada muestra observada $\mathbf{x} = (x_1, \ldots, x_n)^T$ del vector aleatorio $\mathbf{X} = (X_1, \ldots, X_n)^T$, la función de verosimilitud está dada por:

$$L(\mathbf{x}|\lambda) = 2^n \phi_n(\mathbf{x}) \Phi_n(\lambda \mathbf{x}), \quad \lambda \in \mathbb{R}, \tag{3.2.1}$$

donde $\phi_n(\mathbf{x}) = \prod_{i=1}^n \phi(x_i)$ y $\Phi_n(\lambda \mathbf{x}) = \prod_{i=1}^n \Phi(\lambda x_i)$.

Con el fin de proponer un método alternativo para obtener la media y varianza a posteriori de λ , para cada valor fijo x de X, se asume un vector aleatorio no observable \mathbf{Y}_x tal que

$$\mathbf{Y}_x = \lambda \mathbf{x} + \epsilon_*,\tag{3.2.2}$$

donde $\epsilon_* \sim \mathcal{N}_n(0, I_n)$ y es independiente de λ .

Proposición 3.2.1. Supóngase que, dado $\lambda \in \mathbb{R}$, las variables aleatorias X_1, \ldots, X_n son independientes con distribución idéntica skew-normal estándar y parámetro de asimetría λ . Se asume \mathbf{Y}_x definido en (3.2.2). Si λ tiene distribución a priori propia $\pi(\cdot)$, entonces se sigue que:

I) la función de densidad predictiva de X está dada por:

$$\pi(\mathbf{x}) = 2^n \phi_n(\mathbf{x}) E_{\lambda}[\Phi_n(\lambda \mathbf{x})]; \tag{3.2.3}$$

II) la función de densidad a posteriori de λ está dada por:

$$\pi(\lambda|\mathbf{x}) = \pi(\lambda) \frac{\Phi_n(\lambda \mathbf{x})}{E_{\lambda}[\Phi_n(\lambda \mathbf{x})]};$$
(3.2.4)

III) la esperanza a posteriori de $g(\lambda)$ (si existe) está dada por:

$$E[g(\lambda)|\mathbf{x}] = E\left\{E[g(\lambda)|\mathbf{Y}_x]|\mathbf{Y}_x \ge 0\right\},\tag{3.2.5}$$

donde q es una función medible e integrable.

Dem: La demostración de (i) y (ii) es directa de algunos resulados bien conocidos de cálculo de probabilidades y por lo tanto son omitidas. Para demostrar el resultado en (3.2.5), primero nótese de (3.2.2) que, para cada valor fijo \mathbf{x} , $\Phi_n(\lambda \mathbf{x}) = P(\mathbf{Y}_x \geq 0 | \lambda) = \int_{\{\mathbf{y} \geq \mathbf{0}\}} \pi(\mathbf{w} | \lambda) d\mathbf{y} \ \mathbf{y} \ P(\mathbf{Y}_x \geq \mathbf{0}) = E_{\lambda}(\Phi_n(\lambda \mathbf{x}))$, donde $\pi(\mathbf{y} | \lambda) = \phi_n(\mathbf{y}; \lambda \mathbf{x}, I_n)$ denota la

densidad condicional de \mathbf{Y}_x dado λ . Consecuentemente, para cualquier función g integrable y medible se sigue de (3.2.4), que

$$\begin{split} E(g(\lambda)|\mathbf{x}) &= (P(\mathbf{Y}_x \geq \mathbf{0}))^{-1} \int_{-\infty}^{\infty} g(\lambda) \pi(\lambda) \Phi_n(\lambda \mathbf{x}) d\lambda \\ &= (P(\mathbf{Y}_x \geq \mathbf{0}))^{-1} \int_{-\infty}^{\infty} g(\lambda) \pi(\lambda) \int_{\{\mathbf{y} \geq \mathbf{0}\}} \pi(\mathbf{y}|\lambda) d\mathbf{y} d\lambda \\ &= (P(\mathbf{Y}_x \geq \mathbf{0}))^{-1} \int_{\{\mathbf{y} \geq \mathbf{0}\}} \pi(\mathbf{y}) \int_{-\infty}^{\infty} g(\lambda) \pi(\lambda|\mathbf{y}) d\lambda d\mathbf{y} \\ &= \int_{\{\mathbf{y} \geq \mathbf{0}\}} \pi_0(\mathbf{y}) E(g(\lambda)|\mathbf{y}) d\mathbf{y}. \end{split}$$

La demostración concluye notando que $\pi_0(\mathbf{y}) = \pi(\mathbf{y})[P(\mathbf{Y}_x \geq \mathbf{0})]^{-1}$ es la distribución predictiva del vector \mathbf{Y}_x cuyas componentes son truncadas sobre $\mathbf{0}$.

De la Proposición 3.2.1, se puede notar que la distribución predictiva en (3.2.3) es un miembro de la clase de distribuciones fundamental skew-normal introducida por Arellano-Valle y Genton [8]. Además, la distribución posterior de λ es un tipo de distribución skew llamada skew- π , con función de asimetría $\Phi_n(\lambda \mathbf{x})$ y constante de normalización $E_\lambda(\Phi_n(\lambda \mathbf{x}))$. De hecho, la distribución a priori y a posteriori de λ están en la misma familia de distribuciones, a saber, $\mathcal{P} = \{\pi(\lambda|\psi) = K\pi(\lambda)\Phi(\lambda\psi) : \psi \in \mathcal{A}\}$ en la cual $K^{-1} = \int_{-\infty}^{\infty} \pi(\lambda)\Phi(\lambda\psi)d\lambda$ y \mathcal{A} denota el conjunto de índices de la distribución. Por lo tanto, \mathcal{P} y la familia en (3.1.1) son conjugadas (véase Arellano *et al.* [9]).

En la siguiente subsección se considera una priori skew-normal para el parámetro de asimetría. El resultado principal aquí es la representación estocástica de la distribución posterior de λ , además de la media y varianza a posteriori de la misma.

3.2.2. Distribución a priori skew-normal para el parámetro de asimetría

En esta sección, se asume que el parámetro de asimetría sigue una distribución skewnormal de localización-escala denotada por $\lambda \sim \mathcal{SN}(m,v,a)$, donde los parámetros de localización, escala y asimetría son, respectivamente, $m \in \mathbb{R}$, $v \in \mathbb{R}^+$ y $a \in \mathbb{R}$.

Tal como se mencionó en el Capítulo 1, esta familia de distribuciones permite representar diversas opiniones o creencias a priori, útiles en muchas circunstancias. Por ejemplo, incluye la priori $\mathcal{N}(m,v)$ en el caso de definir a=0, y además, la priori $\mathcal{SN}(a)$ cuando se declara m=0 y v=1.

El siguiente resultado muestra la densidad predictiva y a posteriori del parámetro de asimetría cuando éste tiene una distribución a priori skew-normal. En lo que va del capítulo, considérese $\mathbf{x}^0 = (\mathbf{x}^T \ 0)^T, \mathbf{x}^a = \left(\mathbf{x}^T \ \frac{a}{\sqrt{v}}\right)^T \mathbf{y} \ \mathbf{m}^* = m \left(\mathbf{0}^T \ \frac{a}{\sqrt{v}}\right)^T.$

Proposición 3.2.2. Supóngase que, dado $\lambda \in \mathbb{R}$, las variables aleatorias X_1, \ldots, X_n son independientes y con distribución idéntica $SN(\lambda)$. Si λ tiene distribución a priori skewnormal de localización-escala, entonces se tiene que:

I) la función de densidad predictiva de X es

$$\pi(\mathbf{x}) = 2^{n+1} \phi_n(\mathbf{x}) \Phi_{n+1}(m\mathbf{x}^0; \mathbf{0}, I_{n+1} + v\mathbf{x}^a \mathbf{x}^{aT});$$
(3.2.6)

II) la función de densidad a posteriori de λ es:

$$\pi(\lambda|\mathbf{x}) = \phi(\lambda; m, v) \frac{\Phi_{n+1}(\lambda \mathbf{x}^a; \mathbf{m}^*, I_{n+1})}{\Phi_{n+1}(m\mathbf{x}^0; \mathbf{0}, I_{n+1} + v\mathbf{x}^a\mathbf{x}^{aT})}.$$
 (3.2.7)

Los resultados de la Proposición 3.2.2 se obtienen directamente de la Proposición 3.2.1. Nótese que si $\lambda \sim SN(m,v,a)$, de (3.2.2) se tiene que $\mathbf{Y}_{\mathbf{x}} \sim SN_n(m(\mathbf{x}),v(\mathbf{x}),a(\mathbf{x}))$, donde $m(\mathbf{x})=m\mathbf{x}$, $v(\mathbf{x})=I_n+v\mathbf{x}\mathbf{x}^T$ y $a(\mathbf{x})=\frac{av^{-1/2}}{[(1+a^2v\mathbf{x}^T\mathbf{x})(1+v\mathbf{x}^T\mathbf{x})]^{1/2}}\mathbf{x}$.

Un resultado interesante es propuesto en el siguiente corolario. En él se considera una distribución normal como distribución a priori para λ .

Corolario 3.2.3. Supóngase que, dado $\lambda \in \mathbb{R}$, las variables aleatorias X_1, \ldots, X_n son independientes con distribución idéntica $SN(\lambda)$. Si λ tiene distribución a priori normal de localización-escala, es decir, $\lambda \sim SN(m, v, 0)$ en la cual $m \in \mathbb{R}$ y $v \in \mathbb{R}^+$; entonces se tiene que:

1) la función de densidad predictiva de X está dada por:

$$f(\mathbf{x}) = 2^n \phi_n(\mathbf{x}) \Phi_n(m\mathbf{x}; \mathbf{0}, I_n + v\mathbf{x}\mathbf{x}^T); \tag{3.2.8}$$

II) la función de densidad a posteriori de λ es:

$$\pi(\lambda|\mathbf{x}) = \phi(\lambda; m, v) \frac{\Phi_n(\lambda \mathbf{x})}{\Phi_n(m\mathbf{x}; \mathbf{0}, I_n + v\mathbf{x}\mathbf{x}^T)};$$
(3.2.9)

Los resultados propuestos en el Corolario 3.2.3 se obtienen de forma directa de la Proposición 3.2.1 tomando a=0. Además, es importante mencionar que, de acuerdo a lo establecido en la Proposición 3.2.2 y el Corolario 3.2.3, tanto la distribución posterior de λ cuanto la distribución predicitva son miembros de la familia unified skew-normal (SUN) introducida por Arellano-Valle y Azzalini [2].

Con el objetivo de obtener la media y varianza a posteriori de λ , se proponen los siguientes resultados. El primero, propuesto en el Lema 3.2.4, proporciona un resultado general para calcular el vector de medias y la matriz de varianza-covarianza de un vector aleatorio normal truncado $\mathbf{U}^{\mathbf{c}} \stackrel{d}{=} [\mathbf{U}|\mathbf{U} \geq \mathbf{c}]$, donde $\mathbf{U} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$. La demostración de este lema se encuentra en el Apéndice A.

Lema 3.2.4. Considérese un vector aleatorio $\mathbf{U}^{\mathbf{c}} \stackrel{d}{=} [\mathbf{U}|\mathbf{U} \geq \mathbf{c}]$, donde $\mathbf{U} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, es decir, se asume que $\mathbf{U}^{\mathbf{c}} \sim \mathcal{L}T\mathcal{N}_n(\mathbf{c}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$. Entonces, el vector de esperanzas y la matriz de varianza-covarianza de $\mathbf{U}^{\mathbf{c}}$ están dados, respectivamente, por:

$$E(\mathbf{U^{c}}) = \boldsymbol{\mu} + \boldsymbol{\Sigma} \frac{\Phi'_{n}(\boldsymbol{\mu}; \mathbf{c}, \boldsymbol{\Sigma})}{\Phi_{n}(\boldsymbol{\mu}; \mathbf{c}, \boldsymbol{\Sigma})},$$

$$V(\mathbf{U^{c}}) = \boldsymbol{\Sigma} + \boldsymbol{\Sigma} \left\{ \frac{\Phi''_{n}(\boldsymbol{\mu}; \mathbf{c}, \boldsymbol{\Sigma})}{\Phi_{n}(\boldsymbol{\mu}; \mathbf{c}, \boldsymbol{\Sigma})} - \left(\frac{\Phi'_{n}(\boldsymbol{\mu}; \mathbf{c}, \boldsymbol{\Sigma})}{\Phi_{n}(\boldsymbol{\mu}; \mathbf{c}, \boldsymbol{\Sigma})} \right) \left(\frac{\Phi'_{n}(\boldsymbol{\mu}; \mathbf{c}, \boldsymbol{\Sigma})}{\Phi_{n}(\boldsymbol{\mu}; \mathbf{c}, \boldsymbol{\Sigma})} \right)^{T} \right\} \boldsymbol{\Sigma},$$

$$donde \, \Phi'_{n}(\mathbf{s}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{\partial}{\partial \mathbf{s}} \Phi_{n}(\mathbf{s}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \, \boldsymbol{y} \, \Phi''_{n}(\mathbf{s}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{\partial^{2}}{\partial \mathbf{s} \partial \mathbf{s}^{T}} \Phi_{n}(\mathbf{s}; \boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

El segundo resultado, propuesto en la Proposión 3.2.5, establece la representación estocástica de la distribución a posteriori de λ , la cual es útil para el calculo de los momentos a posteriori y para la ejecución del algoritmo que permite la implementación computacional introducida en la Sección 3.4.2.

Proposición 3.2.5. Sea $W \in \mathbb{R}$ y $\mathbf{U}^0 \in \mathbb{R}^{n+1}$ vectores aleatorios que, condicionalmente en \mathbf{X} , son independientes. Se asume que $W_{\mathbf{x}} \sim \mathcal{N}(m, v(1 + v\mathbf{x}_a^T\mathbf{x}_a))$ y $\mathbf{U}_{\mathbf{x}}^0 \sim \mathcal{LTN}_{n+1}(\mathbf{0}, m\mathbf{x}^0, I_{n+1} + v\mathbf{x}^a\mathbf{x}^{aT})$. Entonces, si $\lambda \sim \mathcal{SN}(m, v, a)$, se sigue que:

$$\lambda_{\mathbf{x}} \stackrel{d}{=} \frac{W_{\mathbf{x}} + v \mathbf{x}^{aT} \mathbf{U}_{\mathbf{x}}^{\mathbf{b0}}}{1 + v \mathbf{x}^{aT} \mathbf{x}^{a}}.$$
 (3.2.10)

Dem: Considerando el enfoque através de la función generadora de momentos (f.g.m.), se necesita mostrar que $M_{\lambda_{\mathbf{x}}}(t) = M_{W_{\mathbf{x}}}\left(\frac{t}{1+v\mathbf{x}^{aT_{\mathbf{x}}}}\right)M_{\mathbf{U}_{\mathbf{x}}^{0}}\left(\frac{v\mathbf{x}t}{1+v\mathbf{x}^{aT_{\mathbf{x}}}}\right)$, para todo $t \in \mathbb{R}$. En efecto, la f.g.m. de $\lambda_{\mathbf{x}}$ está dada por:

$$M_{\lambda_{\mathbf{x}}}(t) = \exp\left\{mt + \frac{1}{2}vt^2\right\} \frac{\Phi_n(m\mathbf{x}^0 + vt\mathbf{x}^a; \mathbf{0}, I_n + v\mathbf{x}^a\mathbf{x}^{aT})}{\Phi_n(m\mathbf{x}^0; \mathbf{0}, I_n + v\mathbf{x}^a\mathbf{x}^{aT})},$$

(para ver la demostración de este hecho, véase el Apéndice B). De otro lado, de la ecuación (3.8.1) en el Apéndice A, se tiene que

$$M_{\mathbf{U}_{\mathbf{x}}^{\mathbf{0}}}\left(\frac{v\mathbf{x}^{a}t}{1+v\mathbf{x}^{aT}\mathbf{x}}\right) = \exp\left\{tm\mathbf{x}^{\mathbf{0}} + \frac{1}{2}\frac{v^{2}\mathbf{x}^{aT}\mathbf{x}^{a}t^{2}}{1+v\mathbf{x}^{aT}\mathbf{x}^{a}}\right\} \frac{\Phi_{n}(m\mathbf{x}^{\mathbf{0}} + v\mathbf{x}^{a}t; \mathbf{0}, I_{n} + v\mathbf{x}^{a}\mathbf{x}^{aT})}{\Phi_{n}(m\mathbf{x}^{\mathbf{0}}; \mathbf{0}, I_{n} + v\mathbf{x}^{a}\mathbf{x}^{aT})}.$$

Luego, la demostración sigue de forma directa después de algunos cálculos algebráicos.

Corolario 3.2.6. Sea $W \in \mathbb{R}$ y $\mathbf{U} \in \mathbb{R}^n$ dos vectores aleatorios que, condicionalmente en \mathbf{X} , son independientes. Se asume que $W_{\mathbf{x}} \sim \mathcal{N}(m, v(1 + v\mathbf{x}^T\mathbf{x}))$ y que $\mathbf{U}_{\mathbf{x}}^{\mathbf{0}} \sim \mathcal{L}T\mathcal{N}_n(\mathbf{0}, m\mathbf{x}, I_n + v\mathbf{x}\mathbf{x}^T)$. Entonces, si $\lambda \sim \mathcal{S}\mathcal{N}(m, v, 0)$, se sigue que:

$$\lambda_{\mathbf{x}} \stackrel{d}{=} \frac{W_{\mathbf{x}} + v\mathbf{x}^{T}\mathbf{U}_{\mathbf{x}}^{0}}{1 + v\mathbf{x}^{T}\mathbf{x}}.$$
 (3.2.11)

Dem: La demostración es directa de la Proposición 3.2.5, considerando a=0.

Alternativamente, estas representaciones pueden ser demostradas considerando los resultados introducidos por Arellano-Valle y Azzalini [2].

La siguiente proposición entrega expresiones explícitas para la media y varianza a posteriori del parámetro de asimetría λ .

Proposición 3.2.7. Supóngase que, dado $\lambda \in \mathbb{R}$, X_1, \ldots, X_n son variables aleatorias independientes con distribución idéntica $SN(\lambda)$. Si λ tiene distribución a priori skewnormal, entonces se tiene que:

I) la media a posteriori de λ está dada por:

$$E(\lambda|\mathbf{x}) = m + v\mathbf{x}^{aT}\boldsymbol{\xi},\tag{3.2.12}$$

donde el i-ésimo componente del vector ξ es

$$\phi(mx_{i}^{0}; 0, 1 + vx_{i}^{a^{2}}) \frac{\Phi_{n}\left(m\mathbf{x}_{(i)}^{0}; \frac{mvx_{i}^{0}x_{i}^{a}\mathbf{x}_{(i)}^{a}}{1 + vx_{i}^{a^{2}}}, I_{n} + \frac{v\mathbf{x}_{(i)}^{a}\mathbf{x}_{(i)}^{a}}{1 + vx_{i}^{a^{2}}}\right)}{\Phi_{n+1}(m\mathbf{x}^{0}; 0, I_{n+1} + v\mathbf{x}^{a}\mathbf{x}^{aT})},$$

 $y \mathbf{x}_{(i)}^{0} y \mathbf{x}_{(i)}^{a}$ son, respectivamente, subvectores de $\mathbf{x}^{0} y \mathbf{x}^{a}$ sin la i-ésima componente;

II) la varianza a posteriori de λ está dada por:

$$Var(\lambda|\mathbf{x}) = v + v^2 \mathbf{x}^{aT} (\Gamma - \xi \xi^T) \mathbf{x}^a, \tag{3.2.13}$$

donde el elemento γ_{ij} de la matriz Γ es

$$\gamma_{ij} = \begin{cases} \frac{\phi(mx_{i}^{0};0,1+vx_{a}^{a2})}{\Phi_{n+1}(m\mathbf{x}^{0};0,I_{n+1}+v\mathbf{x}^{a}\mathbf{x}^{aT})} \times \\ \left[\left(\frac{-mx_{i}^{0}}{1+vx_{i}^{a2}} \right) \Phi_{n} \left(m\mathbf{x}_{(i)}^{0}; \frac{mvx_{i}^{0}x_{i}^{a}\mathbf{x}_{(i)}^{a}}{1+vx_{i}^{a2}}, I_{n} + \frac{v\mathbf{x}_{(i)}^{a}\mathbf{x}_{(i)}^{a}}{1+vx_{i}^{a2}} \right) \\ + \frac{\partial}{\partial mx_{i}^{0}} \Phi_{n} \left(m\mathbf{x}_{(i)}^{0}; \frac{mvx_{i}^{0}x_{i}^{a}\mathbf{x}_{(i)}^{a}}{1+vx_{i}^{a2}}, I_{n} + \frac{v\mathbf{x}_{(i)}^{a}\mathbf{x}_{(i)}^{a}}{1+vx_{i}^{a2}} \right) \right], \qquad i = j \\ \frac{\phi(mx_{i}^{0};0,1+vx_{i}^{a2})}{\Phi_{n+1}(m\mathbf{x}^{0};0,I_{n+1}+v\mathbf{x}^{a}\mathbf{x}^{aT})} \phi \left(mx_{j}^{0}; \frac{mvx_{i}^{a}x_{j}^{a}x_{i}^{0}}{1+vx_{i}^{a2}}, \frac{1+v(x_{i}^{a2}+x_{j}^{a2})}{1+vx_{i}^{a2}} \right) \\ \times \Phi_{n-2} \left(m\mathbf{x}_{(ij)}^{0}; \frac{mv\mathbf{x}_{(ij)}^{a}(x_{i}^{a}x_{i}^{0}+x_{j}^{a}x_{j}^{0})}{1+v(x_{i}^{a2}+x_{j}^{a2})}, I_{n-1} + \frac{v\mathbf{x}_{(ij)}^{a}\mathbf{x}_{(ij)}^{a}\mathbf{x}_{(ij)}^{a}\mathbf{x}_{(ij)}^{T}}{1+v(x_{i}^{a2}+x_{j}^{a2})} \right), \quad i \neq j, \end{cases}$$

 $y \mathbf{x}_{(ij)}^0 y \mathbf{x}_{(ij)}^a$ son, respectivamente subvectores de $\mathbf{x}^0 y \mathbf{x}^a$ sin las i-ésima y j-ésima componente.

Dem: (i) Utilizando la expresión (3.2.10) y asumiendo independencia, se tiene que:

$$E(\lambda|\mathbf{x}) = \frac{E(W_{\mathbf{x}}) + v\mathbf{x}^{aT}E(\mathbf{U_{\mathbf{x}}^{0}})}{1 + v\mathbf{x}^{aT}\mathbf{x}^{a}}.$$

Como consecuencia del Lema 3.2.4, la esperanza de $\mathbf{U}_{\mathbf{x}}^{\mathbf{0}}$ está dada por:

$$E(\mathbf{U_x^0}) = (I_n + v\mathbf{x}^a\mathbf{x}^{aT}) \frac{\Phi'_n(m\mathbf{x}^0; \mathbf{0}, I_n + v\mathbf{x}^a\mathbf{x}^{aT})}{\Phi(m\mathbf{x}^0; \mathbf{0}, I_n + v\mathbf{x}\mathbf{x}^T)}.$$

La demostración concluye aplicando el Teorema 3.8.3 dado en el Apéndice C.

(ii) Similarmente, se tiene que la varianza a posteriori de λ es:

$$Var(\lambda|\mathbf{x}) = \frac{Var(W_{\mathbf{x}}) + v^2 \mathbf{x}^{aT} V(\mathbf{U}_{\mathbf{x}}^0) \mathbf{x}^a}{(1 + v \mathbf{x}^{aT} \mathbf{x}^a)^2}.$$

La demostración es consecuencia del Teorema 3.8.3, Apéndice C y del Lema 3.2.4 el cual proporciona $Var(\mathbf{U_x^0})$.

Corolario 3.2.8. Supónagse que, dado $\lambda \in \mathbb{R}$, las variables aleatorias X_1, \ldots, X_n son independientes y con distribución idéntica $SN(\lambda)$. Si $\lambda \sim SN(m, v, 0)$ donde $m \in \mathbb{R}$ y $v \in \mathbb{R}^+$, entonces se tiene que:

I) la media a posteriori de λ está dada por:

$$E(\lambda|\mathbf{x}) = m + v\mathbf{x}^T\boldsymbol{\xi},\tag{3.2.14}$$

donde la i-ésima componente de $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n)^T$ es:

$$\xi_i = \phi(mx_i; 0, 1 + vx_i^2) \frac{\Phi_{n-1}\left(\frac{m\mathbf{x}_{(i)}}{1 + vx_i^2}; 0, I_{n-1} + \frac{v\mathbf{x}_{(i)}\mathbf{x}_{(i)}^T}{1 + vx_i^2}\right)}{\Phi_n(m\mathbf{x}; 0, I_n + v\mathbf{x}\mathbf{x}^T)};$$

II) la varianza a posteriori de λ está dada por:

$$Var(\lambda|\mathbf{x}) = v + v^2 \mathbf{x}^T (\Gamma - \xi \xi^T) \mathbf{x}, \tag{3.2.15}$$

donde $\Gamma = (\gamma_{ij})$ es la matriz de $n \times n$ cuyos elementos son:

$$\gamma_{ij} = \begin{cases} \frac{\phi(mx_i; 0, 1 + vx_i^2)}{\Phi_n(mx; 0, I_n + vxx^T)} \left[\left(\frac{-mx_i}{1 + vx_i^2} \right) \Phi_{n-1} \left(\frac{m\mathbf{x}_{(i)}}{1 + vx_i^2}; \mathbf{0}, I_{n-1} + \frac{v\mathbf{x}_{(i)}\mathbf{x}_{(i)}^T}{1 + vx_i^2} \right) \\ + \frac{\partial}{\partial mx_i} \Phi_{n-1} \left(\frac{m\mathbf{x}_{(i)}}{1 + vx_i^2}; \mathbf{0}, I_{n-1} + \frac{v\mathbf{x}_{(i)}\mathbf{x}_{(i)}^T}{1 + vx_i^2} \right) \right], & i = j \\ \frac{\phi(mx_i; 0, 1 + vx_i^2)}{\Phi_n(m\mathbf{x}; 0, I_n + v\mathbf{x}\mathbf{x}^T)} \phi \left(\frac{mx_j}{1 + vx_i^2}; \mathbf{0}, \frac{1 + v(x_i^2 + x_j^2)}{1 + vx_i^2} \right) \\ \times \Phi_{n-2} \left(\frac{m\mathbf{x}_{(ij)}}{1 + v(x_i^2 + x_j^2)}; \mathbf{0}, I_{n-2} + \frac{v\mathbf{x}_{(ij)}\mathbf{x}_{(ij)}^T}{1 + v(x_i^2 + x_j^2)} \right), & i \neq j. \end{cases}$$

Nota 3.2.1. Nótese que aplicando (3.2.5), la Proposición 3.2.1, y los resultados del Lema 3.2.4, se obtiene otra manera de calcular la media y varianza a posteriori de λ . De (3.2.2) se obtiene que $\mathbf{Y}_x|\lambda \sim \mathcal{N}_n(\lambda \mathbf{x}, I_n)$. Consecuentemente, asumiendo que $\lambda \sim \mathcal{N}(m, v)$, se tiene que

$$\lambda | \mathbf{y} \sim \mathcal{N}\left(\frac{m + v\mathbf{x}^T\mathbf{y}}{1 + v\mathbf{x}^T\mathbf{x}}, \frac{v}{1 + v\mathbf{x}^T\mathbf{x}}\right).$$

Entonces, de (3.2.5), la media a posteriori de λ está dada por:

$$E(\lambda|\mathbf{x}) = E\{E(\lambda|\mathbf{Y}_x)|\mathbf{Y}_x \ge \mathbf{0}\} = \frac{m + v\mathbf{x}^T E(\mathbf{Y}_x^{\mathbf{0}})}{1 + v\mathbf{x}^T \mathbf{x}},$$

donde la esperanza truncada $E(\mathbf{Y}_x^0) = E(\mathbf{Y}_x | \mathbf{Y}_x \geq \mathbf{0})$ es proporcionada por el 3.2.4. De manera similar, se puede calcular $E(\lambda^2 | \mathbf{x})$ y, por lo tanto, la varianza a posteriori de λ puede ser obtenida.

3.3. Inferencia Bayesiana en la distribución skew-normal de localización-escala

En esta sección, se considera mezclas de asimetría de distribuciones independientes skew-normal de localización-escala, es decir, se asume que, condicionalmente en ξ , ω^2

y λ , las variables aleatorias X_1, \ldots, X_n son independientes y con distribución idéntica $\mathcal{SN}(\xi, \omega^2, \lambda)$. Por lo tanto, para cada muestra observada $\mathbf{x} = (x_1, \ldots, x_n)^T$ del vector aleatorio $\mathbf{X} = (X_1, \ldots, X_n)^T$, la función de verosimilitud bajo consideración está dada por:

$$L(\mathbf{x}|\xi,\omega^2,\lambda) = \left(\frac{2}{\omega}\right)^n \phi_n \left(\frac{\mathbf{x} - \xi \mathbf{1}_n}{\omega}\right) \Phi_n \left(\lambda \left(\frac{\mathbf{x} - \xi \mathbf{1}_n}{\omega}\right)\right). \tag{3.3.1}$$

Un resultado importante de esta sección es el kernel de la densidad posterior para los parámetros de localización, escala y asimetría asumiendo independencia a priori entre (ξ, ω^2) y λ . Además, se considera una priori impropia para (ξ, ω^2) y diferentes prioris (propias) para λ .

Es importante mencionar que bajo la anterior especificación Bayesiana del modelo, la distribución a posteriori conjunta $\pi(\xi, \omega^2, \lambda)$ es propia (ver Arellano-Valle *et al.* [5] para la demostración de este hecho).

En lo que resta del capítulo, denótese la norma Euclidiana del vector a por ||a||.

3.3.1. Distribución a priori general para el parámetro de asimetría

Proposición 3.3.1. Supóngase que, dado $\xi \in \mathbb{R}$, $\omega^2 \in \mathbb{R}^+$ y $\lambda \in \mathbb{R}$, las variables aleatorias X_1, \ldots, X_n son independientes y con distribución idéntica $SN(\xi, \omega^2, \lambda)$. Si ξ y ω^2 tienen distribución a priori conjunta impropia $\pi(\xi, \omega^2) = \omega^{-2}$ y λ tiene distribución a priori propia $\pi(\lambda)$, entonces

I) la función de densidad a posteriori de (ξ,ω^2) es proporcional a

$$\pi(\xi, \omega^2 | \mathbf{x}) \propto \frac{1}{\omega^{n+2}} \phi_n \left(\frac{\mathbf{x} - \xi \mathbf{1}_n}{\omega} \right) E_{\lambda} \left(\Phi_n \left(\lambda \left(\frac{\mathbf{x} - \xi \mathbf{1}_n}{\omega} \right) \right) \right)$$

II) la función de densidad a posteriori de λ es proporcional a

$$\pi(\lambda|\mathbf{x}) \propto \pi(\lambda) T_n \left(\lambda(\mathbf{x} - \bar{x}\mathbf{1}_n); \mathbf{0}, \left(\frac{||\mathbf{x}||^2 + n\bar{x}^2}{n-1}\right) [I_n + n^{-1}\lambda^2 \mathbf{1}_n \mathbf{1}_n^T], n-1\right)$$

Dem: (i) La demostración es directa de (3.3.1), teniendo en cuenta que la distribución conjunta de $(\mathbf{x}, \xi, \omega^2, \lambda)$ es

$$\pi(\xi, \omega^2, \lambda, \mathbf{x}) = \frac{2^n}{\omega^{n+2}} \phi_n \left(\frac{\mathbf{x} - \xi \mathbf{1}_n}{\omega} \right) \Phi_n \left(\lambda \left(\frac{\mathbf{x} - \xi \mathbf{1}_n}{\omega} \right) \right) \pi(\lambda). \tag{3.3.2}$$

(ii) Se sigue de (3.3.2), el Lema 3.8.1 y algunos resultados del cálculo de probabilidad que la densidad posterior de λ es proporcional a,

$$\begin{split} \pi(\lambda|\mathbf{x}) &\propto \pi(\lambda) \int_{\omega^{2} \in \mathbb{R}^{+}} \int_{\xi \in \mathbb{R}} \frac{1}{\omega^{n+2}} \exp\left\{\frac{||\mathbf{x} - \xi \mathbf{1}_{n}||^{2}}{2\omega^{2}}\right\} \Phi_{n}\left(\lambda\left(\frac{\mathbf{x} - \xi \mathbf{1}_{n}}{\omega}\right)\right) d\xi d\omega^{2} \\ &= \pi(\lambda) \int_{\omega^{2} \in \mathbb{R}^{+}} \frac{1}{\omega^{n+1}} \exp\left\{\frac{||\mathbf{x}||^{2} - n\bar{x}}{2\omega^{2}}\right\} \int_{\xi \in \mathbb{R}} \frac{\sqrt{n}}{\omega} \phi\left(\frac{\xi - \bar{x}}{\omega/\sqrt{n}}\right) \Phi_{n}\left(\lambda\left(\frac{\mathbf{x} - \xi \mathbf{1}_{n}}{\omega}\right)\right) d\xi d\omega^{2} \\ &= \pi(\lambda) \int_{\omega^{2} \in \mathbb{R}^{+}} \frac{1}{\omega^{n+1}} \exp\left\{\frac{||\mathbf{x}||^{2} - n\bar{x}}{2\omega^{2}}\right\} \Phi_{n}\left(\lambda(\mathbf{x} - \bar{x}\mathbf{1}_{n}); 0, \omega^{2}[I_{n} + n^{-1}\lambda^{2}\mathbf{1}_{n}\mathbf{1}_{n}^{T}]\right) d\omega^{2} \\ &= \pi(\lambda) \int_{\mathbf{t} \leq \lambda(\mathbf{x} - \bar{x}\mathbf{1}_{n})} \int_{\omega^{2} \in \mathbb{R}^{+}} \frac{1}{\omega^{n+1}} \exp\left\{\frac{||\mathbf{x}||^{2} - n\bar{x}}{2\omega^{2}}\right\} \phi_{n}\left(\mathbf{t}; 0, \omega^{2}[I_{n} + n^{-1}\lambda^{2}\mathbf{1}_{n}\mathbf{1}_{n}^{T}]\right) d\mathbf{t} d\omega^{2} \\ &= \pi(\lambda) \int_{\mathbf{t} \leq \lambda(\mathbf{x} - \bar{x}\mathbf{1}_{n})} t_{n}\left(\mathbf{t}; 0, \left(\frac{||\mathbf{x}||^{2} + n\bar{x}^{2}}{n - 1}\right)[I_{n} + n^{-1}\lambda^{2}\mathbf{1}_{n}\mathbf{1}_{n}^{T}], n - 1\right) d\mathbf{t}, \end{split}$$

concluyendo la demostración.

3.3.2. Distribución a priori skew-normal para el parámetro de asimetría

En esta subsección se asume que λ tienen distribución a priori skew-normal de localización-escala, es decir, $\lambda \sim \mathcal{SN}(m, v, a)$.

Proposición 3.3.2. Supóngase que, dado $\xi \in \mathbb{R}$, $\omega^2 \in \mathbb{R}^+$ y $\lambda \in \mathbb{R}$, las variables aleatorias X_1, \ldots, X_n son independientes con distribución idéntica $SN(\xi, \omega^2, \lambda)$. Si ξ y ω^2 tienen distribución a priori impropia $\pi(\xi, \omega^2) = \omega^{-2}$ y $\lambda \sim SN(m, v, a)$, donde $m \in \mathbb{R}$, $v \in \mathbb{R}^+$ y $a \in \mathbb{R}$, entonces

I) la función de densidad a posteriori conjunta de (ξ, ω^2) es proporcional a

$$\pi(\xi, \omega^{2}|\mathbf{x}) \propto \frac{1}{\omega^{n+2}} \phi_{n} \left(\frac{\mathbf{x} - \xi \mathbf{1}_{n}}{\omega} \right) \Phi_{n+1} \left(m \mathbf{x}_{\xi, \omega}^{a}; \mathbf{m}^{*}, I_{n+1} + v \mathbf{x}_{\xi, \omega}^{a} \mathbf{x}_{\xi, \omega}^{a}^{T} \right)$$

$$donde \ \mathbf{x}_{\xi, \omega}^{a} = \left(\left(\frac{\mathbf{x} - \xi \mathbf{1}_{n}}{\omega} \right)^{T} \frac{a}{\sqrt{v}} \right)^{T}.$$

II) la función de densidad a posteriori de λ es proporcional a

$$\pi(\lambda|\mathbf{x}) \propto \phi\left(\frac{\lambda-m}{v}\right) \Phi_n\left(\frac{a(\lambda-m)}{v}\right)$$
$$T_n\left(\lambda(\mathbf{x}-\bar{x}\mathbf{1}_n); \mathbf{0}, \left(\frac{||\mathbf{x}||^2+n\bar{x}}{n-1}\right) [I_n+n^{-1}\lambda^2\mathbf{1}_n\mathbf{1}_n^T], n-1\right)$$

Dem: La demostración de (i) y (ii) es directa de la Proposición 3.3.1.

Corolario 3.3.3. Supóngase que, dado $\xi \in \mathbb{R}$, $\omega^2 \in \mathbb{R}^+$ y $\lambda \in \mathbb{R}$, las variables aleatorias X_1, \ldots, X_n son independientes con distribución idéntica $SN(\xi, \omega, \lambda)$. Si ξ y ω^2 tienen distribución a priori impropia $\pi(\xi, \omega^2) = \omega^{-2}$ y $\lambda \sim \mathcal{N}(m, v)$, entonces

I) la función de densidad a posteriori conjunta de (ξ,ω^2) es proporcional a

$$\pi(\xi, \omega^{2} | \mathbf{x}) \propto \frac{1}{\omega^{n+2}} \phi_{n} \left(\frac{\mathbf{x} - \xi \mathbf{1}_{n}}{\omega} \right)$$

$$\Phi_{n+1} \left(\frac{m(\mathbf{x} - \xi \mathbf{1}_{n})}{\omega}; I_{n} + v \left(\frac{\mathbf{x} - \xi \mathbf{1}_{n}}{\omega} \right) \left(\frac{\mathbf{x} - \xi \mathbf{1}_{n}}{\omega} \right)^{T} \right).$$

II) la función de densidad a posteriori de λ es proporcional a

$$\pi(\lambda|\mathbf{x}) \propto \phi\left(\frac{\lambda - m}{v}\right)$$

$$T_n\left(\lambda(\mathbf{x} - \bar{x}\mathbf{1}_n); \mathbf{0}, \left(\frac{||\mathbf{x}||^2 + n\bar{x}}{n-1}\right) [I_n + n^{-1}\lambda^2 \mathbf{1}_n \mathbf{1}_n^T], n-1\right)$$

Dem: La demostración de (i) y (ii) es directa de la Proposición 3.3.2 tomando a = 0.

3.4. Punto de cambio en el parámetro de asimetría

Los resultados introducidos en las secciones previas pueden ser extendidos en el contexto de la identificación de un punto de cambio en el parámetro de asimetría, cuando un conjunto de datos es observado de manera secuencial.

3.4.1. Especificación del modelo

Denótese por K el instante aleatorio en el cual un cambio sucede y por λ_1 y λ_2 , los parámetros de asimetría antes y después del punto de cambio, respectivamente. Se asume que, dado ξ , ω^2 , λ_1 , λ_2 y K=k, los vectores aleatorios $\mathbf{X}_{(k)}=(X_1,\ldots,X_k)$ y $\mathbf{X}_{(n-k)}=(X_{k+1},\ldots,X_n)$ son independientes y que, para $k\in\{1,\ldots,n\}$, se sigue que:

$$\mathbf{X}_{(k)}|\lambda_1, \lambda_2, k \stackrel{ind.}{\sim} \mathcal{SN}(\xi, \omega^2, \lambda_1);$$

 $\mathbf{X}_{(n-k)}|\lambda_1, \lambda_2, k \stackrel{ind.}{\sim} \mathcal{SN}(\xi, \omega^2, \lambda_2).$

Consecuentemente, la función de verosimilitud en consideración es:

$$L(\mathbf{x}|\xi,\omega^2,\lambda_1,\lambda_2,K=k) = \left(\frac{2}{\omega}\right)^n \phi_n(\mathbf{z}_{(n)}) \Phi_{(k)}(\lambda_1 \mathbf{z}_{(k)}) \Phi_{(n-k)}(\lambda_2 \mathbf{z}_{(n-k)}),$$
 (3.4.1) donde $\mathbf{z}_{(l)} = (\mathbf{x}_{(l)} - \xi \mathbf{1}_l)/\omega$.

Con el objetivo de construir la distribución a priori conjunta de $(\xi, \omega^2, \lambda_1, \lambda_2, K)$, se asume que K es independiente de ξ , ω^2 , λ_1 y λ_2 . Además, dado K=k, asúmase que el vector de parámetros (ξ, ω^2) es independiente de los parámetros de asimetría λ_1 y λ_2 y que λ_1 es independiente de λ_2 . Considérese que, a priori, $\pi(\xi, \omega^2) = \omega^{-2}$ y denótese por $\pi(k)$, $\pi(\lambda_1)$ y $\pi(\lambda_2)$, las especificaciones a priori para K, λ_1 y λ_2 , respectivamente.

Como consecuencia de tales especificaciones, se sigue que la distribución predictiva de x es dada por:

$$\pi(\mathbf{x}) = 2^n \sum_{k=1}^n \pi(k) \int_{-\infty}^{\infty} \int_0^{\infty} \omega^{-(n+2)} \phi_n(\mathbf{z}_{(n)}) H_{\lambda_1, \lambda_2}(\mathbf{z}) d\omega^2 d\xi,$$

donde $H_{\lambda_1,\lambda_2}(\mathbf{z}) = E_{\lambda_1}\left(\Phi_k(\lambda_1\mathbf{z}_{(k)})\right)E_{\lambda_2}\left(\Phi_{n-k}(\lambda_2\mathbf{z}_{(n-k)})\right)$ y la esperanza $E_{\lambda_i}(\cdot)$ es calculada bajo la distribución a priori de λ_i , i=1,2. La aplicación directa del Teorema de Bayes proporciona las distribuciones a posteriori de K, λ_1 y (ξ,ω^2) respectivamente:

$$\pi(K = k | \mathbf{x}) = \frac{\pi(k) \int_{-\infty}^{\infty} \int_{0}^{\infty} \omega^{-(n+2)} \phi_{n}(\mathbf{z}_{(n)}) H_{\lambda_{1},\lambda_{2}}(\mathbf{z}) d\omega^{2} d\xi}{\sum_{k=1}^{n} \pi(k) \int_{-\infty}^{\infty} \int_{0}^{\infty} \omega^{-(n+2)} \phi_{n}(\mathbf{z}_{(n)}) H_{\lambda_{1},\lambda_{2}}(\mathbf{z}) d\omega^{2} d\xi}, \qquad (3.4.2)$$

$$\pi(\lambda_{1} | \mathbf{x}) = \frac{\pi(\lambda_{1}) \sum_{k=1}^{n} \pi(k) \int_{-\infty}^{\infty} \int_{0}^{\infty} \omega^{-(n+2)} \phi_{n}(\mathbf{z}_{(n)}) G_{\lambda_{2}}(\mathbf{z}, \lambda_{1}) d\omega^{2} d\xi}{\sum_{k=1}^{n} \pi(k) \int_{-\infty}^{\infty} \int_{0}^{\infty} \omega^{-(n+2)} \phi_{n}(\mathbf{z}_{(n)}) H_{\lambda_{1},\lambda_{2}}(\mathbf{z}) d\omega^{2} d\xi}, \qquad (3.4.3)$$

$$\pi(\xi, \omega^{2} | \mathbf{x}) = \frac{\omega^{-(n+2)} \phi_{n}(\mathbf{z}_{(n)}) \sum_{k=1}^{n} [\pi(k) H_{\lambda_{1},\lambda_{2}}(\mathbf{z})]}{\sum_{k=1}^{n} \pi(k) \int_{-\infty}^{\infty} \int_{0}^{\infty} \omega^{-(n+2)} \phi_{n}(\mathbf{z}_{(n)}) H_{\lambda_{1},\lambda_{2}}(\mathbf{z}) d\omega^{2} d\xi}, \qquad (3.4.4)$$

donde $G_{\lambda_2}(\mathbf{z}, \lambda_1) = \Phi_k(\lambda_1 \mathbf{z}_{(k)}) E_{\lambda_2} \left(\Phi_{n-k}(\lambda_2 \mathbf{z}_{(n-k)}) \right)$. Más aún, sea $\mathbf{Y}_{\mathbf{z}_{(k)}} = \lambda_1 \mathbf{z}_{(k)} + \epsilon_*$ y $\epsilon_* \sim \mathcal{N}_k(\mathbf{0}, I_k)$. Entonces, de la Proposición 3.2.1 se sigue que:

$$E_{\lambda_1}\left(\lambda_1 \frac{\Phi_k(\lambda_1 \mathbf{z}_{(k)})}{E_{\lambda_1}\left(\Phi_k(\lambda_1 \mathbf{z}_{(k)})\right)}\right) = E\left\{E(\lambda_1|\mathbf{Y}_{\mathbf{z}_{(k)}})|\mathbf{Y}_{\mathbf{z}_{(k)}} \ge 0\right\} = E(\lambda_1|\mathbf{z}_{(k)}).$$

Consecuentemente, la esperanza a posteriori del parámetro de asimetría λ_1 es dada por:

$$E(\lambda_1|\mathbf{x}) = \frac{\sum_{k=1}^n \pi(k) \int_{-\infty}^{\infty} \int_0^{\infty} \omega^{-(n+2)} \phi_n(\mathbf{z}_n) E(\lambda_1|\mathbf{z}_{(k)}) H_{\lambda_1,\lambda_2}(\mathbf{z}) d\omega^2 d\xi}{\sum_{k=1}^n \int_{-\infty}^{\infty} \int_0^{\infty} \omega^{-(n+2)} \phi_n(\mathbf{z}_{(n)}) \pi(k) H_{\lambda_1,\lambda_2}(\mathbf{z}) d\omega^2 d\xi}.$$
 (3.4.5)

La distribución posterior de λ_2 y su esperanza puede ser obtenida de manera similar.

En la aplicación siguiente se asume, como caso particular, que $\lambda_i \sim \mathcal{N}(m_i, v_i)$ para i=1,2. En este caso, la distribución a posteriori y la esperanza a posteriori de λ_1 se

obtienen de (3.4.2) to (3.4.5) notando que:

$$E_{\lambda_1} \left[\Phi_k(\lambda_1 \mathbf{z}_{(k)}) \right] = \Phi_k(m_1 \mathbf{z}_{(k)}; \mathbf{0}, I_k + v \mathbf{z}_{(k)} \mathbf{z}_{(k)}^T),$$

$$E_{\lambda_2} \left[\Phi_{n-k}(\lambda_2 \mathbf{z}_{(n-k)}) \right] = \Phi_{n-k}(m_2 \mathbf{z}_{(n-k)}; \mathbf{0}, I_{n-k} + v \mathbf{z}_{(n-k)} \mathbf{z}_{(n-k)}^T),$$

$$E(\lambda_1 | \mathbf{z}_k) = \frac{m_1 + v_1 \mathbf{z}_{(k)}^T E(\mathbf{Y}_x^0)}{1 + v \mathbf{z}_{(k)}^T \mathbf{z}_{(k)}},$$

donde $E(\mathbf{Y}_{z_k}^0) = E(\mathbf{Y}_{z_k}|\mathbf{Y}_{z_k} \geq \mathbf{0})$ es propuesto en el Lema 3.2.4.

3.4.2. Muestreo de Gibbs

Nótese de (3.4.2), (3.4.3) y (3.4.4) que es computacionalmente difícil obtener las posterioris exactas. Por tal razón, se propone la implementación del algoritmo de Metropolis Hastings dentro del Muestreo de Gibbs para obtener las estimaciones de tales distribuciones utilizando las distribuciones condicionales completas ya que estas son analíticamente manejables y además tienen formas cerradas:

$$\pi(K = k | \xi, \omega^{2}, \lambda_{1}, \lambda_{2}, \mathbf{x}) \propto \pi(k) \Phi_{k}(\lambda_{1} \mathbf{z}_{(k)}) \Phi_{n-k}(\lambda_{2} \mathbf{z}_{(n-k)})$$

$$\pi(\lambda_{1} | \xi, \omega^{2}, \lambda_{2}, K = k, \mathbf{x}) \propto \pi(\lambda_{1}) \Phi_{k}(\lambda_{1} \mathbf{z}_{(k)}),$$

$$\pi(\lambda_{2} | \xi, \omega^{2}, \lambda_{1}, K = k, \mathbf{x}) \propto \pi(\lambda_{2}) \Phi_{n-k}(\lambda_{2} \mathbf{z}_{(n-k)}),$$

$$\pi(\xi | \omega^{2}, \lambda_{1}, \lambda_{2}, K = k, \mathbf{x}) \propto \phi_{n}(\mathbf{z}_{(n)}) \Phi_{k}(\lambda_{1} \mathbf{z}_{(k)}) \Phi_{n-k}(\lambda_{2} \mathbf{z}_{(n-k)}),$$

$$\pi(\omega^{2} | \xi, \lambda_{1}, \lambda_{2}, K = k, \mathbf{x}) \propto \omega^{-(n+2)} \phi_{n}(\mathbf{z}_{(n)}) \Phi_{k}(\lambda_{1} \mathbf{z}_{(k)}) \Phi_{n-k}(\lambda_{2} \mathbf{z}_{(n-k)}).$$

Particulamente, para distribuciones a priori skew-normal para los parámetros de asimetría, es decir, considerando $\lambda_i \sim \mathcal{N}(m_i, v_i, a_i)$, i = 1, 2, las distribuciones condicionales completas son:

$$\pi(\lambda_{1}|\xi,\omega^{2},\lambda_{2},K=k,\mathbf{x}) \propto \phi(\lambda_{1},m_{1},v_{1})\Phi_{k+1}(\lambda_{1}\mathbf{z}_{(k)}^{*}),$$

$$\pi(\lambda_{2}|\xi,\omega^{2},\lambda_{1},K=k,\mathbf{x}) \propto \phi(\lambda_{2},m_{2},v_{2})\Phi_{n-k+1}(\lambda_{2}\mathbf{z}_{(n-k)}^{*}),$$

$$\pi(\xi|\omega^{2},\lambda_{1},\lambda_{2},K=k,\mathbf{x}) \propto \phi\left(\frac{\xi-\bar{x}}{\omega/\sqrt{n}}\right)\Phi_{n}(\xi\lambda^{*}+\mathbf{x}^{*}),$$

$$\pi(\omega^{2}|\xi,\lambda_{1},\lambda_{2},K=k,\mathbf{x}) \propto \mathcal{IG}\left(\frac{n}{2},\frac{||\mathbf{x}-\xi\mathbf{1}_{n}||^{2}}{2}\right)\Phi_{k}(\lambda_{1}\mathbf{z}_{(k)})\Phi_{n-k}(\lambda_{2}\mathbf{z}_{(n-k)}),$$

donde IG denota la distribución gamma-inversa.

Además,
$$\mathbf{z}_{(k)}^* = (z_1, \dots, z_k, \frac{a_1}{\sqrt{v_1}}), \mathbf{z}_{(n-k)}^* = (z_{k+1}, \dots, z_n, \frac{a_2}{\sqrt{v_2}}), \lambda^* = \left(-\frac{\lambda_1 \mathbf{1}_k^T}{\omega} - \frac{\lambda_2 \mathbf{1}_{n-k}^T}{\omega}\right)^T$$
 y $\mathbf{x}^* = \left(-\frac{\lambda_1}{\omega} \mathbf{x}_{(k)}^T - \frac{\lambda_2}{\omega} \mathbf{x}_{(n-k)}^T\right)^T$. Las distribuciones condicionales completas para prioris normales se obtienen asumiendo $a_1 = a_2 = 0$.

Nótese que, para todas las especificaciones a priori de los parámetros de asimetría, las distribuciones condicionales completas de λ_1 , λ_2 y ξ son distribuciones skew-normal. Aunque, no es fácil generar muestras aleatorias de tales distribuciones, la representación estocástica mostrada en la Section 3.2.2 juega un rol fundamental en el desempeño de esta tarea.

La generación de muestras de la distribución condicional completa de ω^2 es realizada considerando el algoritmo Metropolis-Hastings, a través de un candidato con distribución Gamma, mientras que la generación de muestras de la distribución condicional de K es realizada a través del enfoque sugerido por Carlin *et al.* [27].

3.5. Análisis de punto de cambio: datos simulados y de retornos de mercados financieros

El objetivo de esta sección es aplicar los resultados obtenidos en el capítulo en la identificación de un punto de cambio en los retornos de índices de la bolsa de cuatro importantes mercados financieros Latino Americanos. Estas series han sido analizadas previamente por Loschi *et al.* [61], quienes consideraron modelos de partición producto (Barry y Hartigan [21]) para indentificar múltiples puntos de cambio en la media de los retornos y en las volatilidades de estos mercados. Loschi *et al.* [61] asumen que la serie de retornos tiene distribución normal y concluyen que todos los mercados experimentan cambios en la media de los retornos y la volatilidad en periodos similares.

Tal como se mencionara en la Introducción, el supuesto de normalidad puede ser muy restrictivo en el caso de los mercados financieros emergentes, toda vez que las distribuciones empíricas de éstos exhiben asimetrías y colas más pesadas o livianas que las de la distribución normal. Más aún, es bien sabido que la media y la varianza pueden ser influenciadas por la asimetría de la distribución. Consecuentemente, cambios en el parámetro de asimetría pueden originar tales cambios.

En lo que sigue de esta sección, se asume una función de verosimilitud skew-normal de localización-escala para describir el comportamiento de los datos. Por lo tanto, los resultados introducidos en las secciones previas son utilizados. El foco de atención es la identificación de un punto de cambio en el parámetro de asimetría de tal distribución.

Con el fin de evaluar la eficiencia del enfoque de punto de cambio, previamente se analiza un conjunto de datos simulados.

En ambos casos, se asume una distribución a priori uniforme para K y una distribución

a priori conjunta impropia para (ξ, ω^2) (ver Sección 3.3).

Para la realización del muestreo de Gibbs, se consideró 20,000 iteraciones. Luego que la convergencia se alcanzara, se descartaron las 5,000 iteraciones iniciales, consideradas como periodo de quema. Para controlar la correlación entre iteraciones. se consideró un salto de tamaño 10.

3.5.1. Análisis de los datos simulados

El conjunto de datos simulados consiste en una secuencia de tamaño n=120 de observaciones skew-normal, la cual experimenta un cambio en la posición 40 (véase Figura 3.1). Esta secuencia es generada de las siguientes distribuciones:

$$X_1, \dots, X_{40} | \xi, \omega^2, \lambda_1, \lambda_2, K \sim \mathcal{SN}(5, 2, 5^2, -2)$$

 $X_{41}, \dots, X_{120} | \xi, \omega^2, \lambda_1, \lambda_2, K \sim \mathcal{SN}(5, 2, 5^2, 2)$

Como especificaciones a priori, se asume que λ_1 y λ_2 son idénticamente distribuidas con distribución normal de media igual a 0 y varianza igual a 10. Los resultados a posteriori son resumidos en la Figura 3.2 y en el Cuadro 3.1

Cuadro 3.1: Estimaciones a posteriori de los parámetros del modelo de punto de cambio para datos simulados.

Param.	Media	SD	$\mathbf{P}_{2,5}$	$P_{97,5}$
ξ	5,04	0,19	4,67	5,43
ω	2,75	0,18	$2,\!41$	3,12
λ_1	-2,51	0,64	-3,89	-1,38
λ_2	2,03	0,42	1,26	2,93
K	39,68	1,43	37,0	43,0

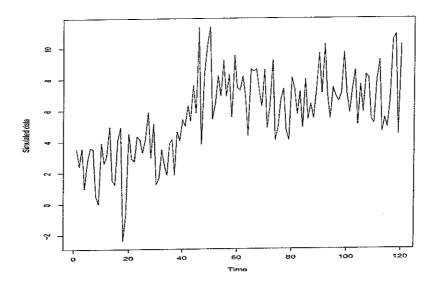


Figura 3.1: Datos simulados con distribución skew-normal.

Se aprecia de la Figura 3.2 que las posterioris de λ_1 y λ_2 son asimétricas. Si se compara con la posteriori de λ_1 , la posteriori de λ_2 tiende a poner más masa en valores altos. La posteriori de K indica que la probabilidad de que el cambio sea en el instante 40 es alta (cercana al 20,0 %) pero más probable es que el cambio tenga lugar antes. Se puede observar además que la media a posteriori de K (39,68) es muy cercana al verdadero valor del punto de cambio. Para el resto de parámetros, se puede notar que, aunque las medias a posteriori son muy cercanas a los parámetros verdaderos, se tiende a sobre estimarlos. Una posible explicación al respecto es la asimetría de las posterioris. Nótese además que todas las distribuciones a posteriori presentan varianzas más pequeñas que las distribuciones a priori consideradas.

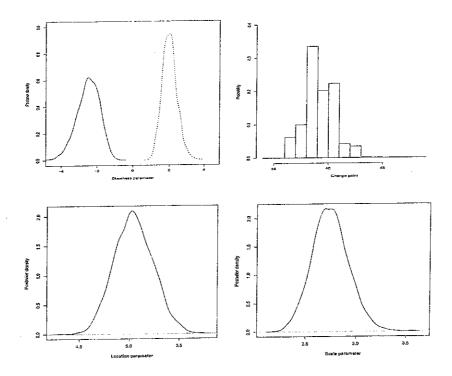


Figura 3.2: Distribución a posteriori de λ_1 (línea sólida), λ_2 (línea punteada), K, ξ y ω .

3.5.2. Análisis de datos de mercados emergentes Latino Americanos

Es bien sabido que los mercados financieros emergentes son más susceptibles a cambios en el escenario político que los mercados en desarrollo. Por tal razón, los índices en estos mercados tienden a presentar observaciones atípicas y cambios estructurales, usualmente en la volatilidad y en la media de los retornos.

En esta sección, se compara el comportamiento de 4 de los más importantes mercados financieros Latino Americanos. Para tal fin, se consideró la serie de retornos de los siguientes índices: MERVAL (Índice de Mercado de Valores de Buenos Aires) de Argentina, IBOVESPA (Índice da Bolsa de Valores do Estado de São Paulo) de Brazil, IPSA (Índice de Precios Selectivos de Acciones) de Chile y IPyC (Índice de Precios y Cotizaciones)

de Mexico. Los retornos fueron registrados de manera semanal, desde el 31 de octubre de 1995 hasta el 31 de octubre del año 2000.

La serie de retornos financieros está definida usando la transformación $R_t = (P_t - P_{t-1})/P_{t-1}$ donde P_t es el último precio observado en la semana t. Además, se asume que la serie de datos tiene distribución skew-normal con parámetros de localización, escala y asimetría desconocidos.

Se consideraron distribuciones a priori normal para los parámetros de asimetría λ_i , i=1,2. Para ambos parámetros, se asumió una gran escala, es decir, $v_i=10$, para todo i. El hiperparámetro de localización fue elegido utilizando el estimador de cumulantes del parámetro de asimetría de cada una de las series de retornos. Específicamente, se consideró $m_i=-1,51,\,-2,71,\,-1,05$ y -1,41 para MERVAL, IBOVESPA, IPSA y IPyC, respectivamente.

De acuerdo a la opinión de expertos, cambios en el comportamiento de los retornos financieros de determinados países son consecuencia de crisis o eventos que han ocurrido en otros países. Las grandes crisis financieras que han involucrado a los mercados emergentes, especialmente a los Latino Americanos, han ocurrido en enero de 1995 (crisis mexicana), agosto de 1997 (crisis asiática) y julio de 1998 (crisis rusa). Además, en enero de 1999, el gobernador estatal de Minas Gerais (Brazil) dejó de pagar la deuda del estado a otros países. Finalmente, la crisis argentina se desató alrededor de diciembre de 1999. Estos importantes eventos de países específicos, producen efectos que se disgregan en otros países con economías similares produciendo cambios en el comportamiento de las mismas (véase Lopes y Migon [60]).

Nótese que, del Cuadro 3.2 la media a posteriori para K indica que el cambio en el

¹Véase el Capítulo 1 de este documento.

índice IBOVESPA tuvo lugar alrededor de la observación 128 (4ta semana, julio de 1998). Además, las estimaciones a posteriori de los parámetros de asimetría muestran que ambos parámetros son muy diferentes y negativos. El cambio en el índice IPSA tuvo lugar alrededor de la observación 89 (4ta semana, julio de 1997). En este caso, las estimaciones a posteriori de los parámetros de asimetría muestran que ambos son bastante diferentes y positivos. Ambos cambios tienen relación con periodos cercanos a crisis internacionales.

Los otros índices no muestran evidencia de asimetría y quizás, los cambios en tales series sean originados por cambios en los parámetros de localización y/o escala.

Cuadro 3.2: Estimaciones a posteriori de los parámetros del modelo de punto de cambio para los índices de los mercados Latino Americanos.

Desc. Stat.	ξ	ω	λ_1	λ_2	K				
	MERVAL								
Media	0,01	0,06	-0.36	-0,41	106,76				
SD	0,03	0,01	1,65	0,85	$47,\!44$				
$P_{2,5}$	-0.03	0,05	-2,71	-1,62	4				
$P_{97,5}$	0,05	0,07	2,30	0,89	228				
IBOVESPA									
Media	0,05	0,07	-3,17	-0,95	128,16				
SD	0,01	0,01	0,51	0,21	12,75				
$P_{2,5}$	0,04	0,06	-4,19	-1,41	80				
$P_{97,5}$	0,06	0,08	-2,24	-0,61	137				
IPSA									
Media	-0,02	0,04	2,74	0,73	89,81				
SD	0,01	0,01	0,54	0,17	7,19				
$P_{2,5}$	-0.03	0,03	1,72	0,41	74				
$\mathbf{P_{97,5}}$	-0,01	0,05	3,81	1,08	103				
IPyC									
Media	0,01	0,06	0,17	-0,19	102,07				
SD	0,04	0,01	1,66	0,96	57,44				
$P_{2,5}$	-0.04	0,05	-2,11	-1,74	5				
P _{97,5}	0,05	0,08	2,75	1,11	237				

3.6. Resumen y conclusiones

En este capítulo se abordó el problema de la inferencia Bayesiana para el parámetro de asimetría de la distribución skew-normal standard y de localización-escala. Se consideró una priori conjunta impropia para los parámetros de localización y escala y, para el parámetro de asimetría, una priori general, skew-normal y normal. Bajo la familia de distribuciones skew-normal standard, se obtuvo las distribuciones a posteriori exactas además de representaciones estocásticas para las mismas. Utilizando tales representaciones, se encontró expresiones exactas para las medias y varianzas a posteriori del parámetro de asimetría. Estos resultados fueron aplicados en la identificación de un punto de cambio en el parámetro de asimetría en una serie de retornos de mercados financieros emergentes.

Es importante mencionar, que tanto las densidades predictivas cuanto las densidades a posteriori del parámetro de asimetría pertenecen a diferentes clases de distribuciones asimétricas, dependiendo de la elección a priori que se haga para este parámetro. De hecho, bajo distribuciones a priori skew-normal, se obtiene un modelo Bayesiano conjugado.

En el caso de la familia de distribuciones skew-normal de localización-escala, se obtuvo el kernel de la distribución a posteriori conjunta para los parámetros de localización y escala y para el parámetro de asimetría. En este caso, del parámetro de asimetría, este kernel pertenece a la familia de distribuciones skew-normal-t (véase Nadarajah y Kotz [67]).

Las representaciones estocásticas propuestas juegan un rol fundamental en la implementación de los algoritmos computacionales necesarios para la estimación de los parámetros desde el punto de vista Bayesiano. Estas representaciones permiten la puesta en práctica de un algoritmo híbrido, es decir, Metropolis-Hastings dentro del muestro de Gibbs para la estimación de los parámetros del modelo skew-normal en el contexto del problema de identificación de un punto de cambio en el parámetro de asimetría. Con respecto a la aplicación de los resultados, se percibió cierta evidencia a favor de un punto de cambio en el parámetro de asimetría para el mercado financiero de Brazil y el de Chile. En los mercados argentino y mexicano, la evidencia en favor de la asimetría es bastante débil.

3.7. Agradecimientos

Quiero expresar mi gratitud a Pilar Iglesias (Pontificia Universidad Católica de Chile) por sus comentarios y sugerencias en las versiones iniciales de este capítulo. También
a Rosangela Loschi (Universidade Federal de Minas Gerais) y a Reinaldo Arellano-Valle
(Pontificia Universidad Católica de Chile) por los comentarios, sugerencias y aportes realizados en este capítulo.

3.8. Apéndice

1. Apéndice A

En este apéndice, se demuestra el Lema 3.2.4. Se asume que $U^c \sim \mathcal{LTN}_n(c, \mu, \Sigma)$. De los resultados presentados en Domínguez-Molina *et al.* [34, 35] (véase además Tallis [82]) se tiene que la *fgm* de U^c está dada por:

$$M_{\mathbf{U}^{\mathbf{c}}}(\mathbf{t}) = e^{\mathbf{t}^{T}\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}^{T}\boldsymbol{\Sigma}\mathbf{t}} \frac{\Phi_{n}(\boldsymbol{\mu} + \boldsymbol{\Sigma}\mathbf{t}; \mathbf{c}, \boldsymbol{\Sigma})}{\Phi_{n}(\boldsymbol{\mu}; \mathbf{c}, \boldsymbol{\Sigma})}.$$
 (3.8.1)

Considérese la función de cumulantes $K_{\mathbf{U}^c}(\mathbf{t}) = \log M_{\mathbf{U}^c}(\mathbf{t})$ y sea $K'_{\mathbf{U}^c}(\mathbf{t}) = \frac{\partial}{\partial \mathbf{t}} K_{\mathbf{U}^c}(\mathbf{t})$ y $K''_{\mathbf{U}^c}(\mathbf{t}) = \frac{\partial}{\partial \mathbf{t}^T} K'_{\mathbf{U}^c}(\mathbf{t})$. De (3.8.1), se tiene que

$$K'_{\mathbf{U}^{\mathbf{c}}}(\mathbf{t}) = \mu + \Sigma \mathbf{t} + \Sigma \frac{\Phi'_{n}(\mu + \Sigma \mathbf{t}; \mathbf{c}, \Sigma)}{\Phi_{n}(\mu + \Sigma \mathbf{t}; \mathbf{c}, \Sigma)}$$
(3.8.2)

У

$$K_{\mathbf{U}^{\mathbf{c}}}''(\mathbf{t}) = \Sigma + \Sigma \frac{\partial}{\partial \mathbf{t}^{T}} \frac{\Phi_{n}'(\mu + \Sigma \mathbf{t}; \mathbf{c}, \Sigma)}{\Phi_{n}(\mu + \Sigma \mathbf{t}; \mathbf{c}, \Sigma)}$$

$$= \Sigma + \Sigma \left\{ \frac{\Phi_{n}''(\mu + \Sigma \mathbf{t}; \mathbf{c}, \Sigma)}{\Phi_{n}(\mu + \Sigma \mathbf{t}; \mathbf{c}, \Sigma)} \Sigma - \frac{\Phi_{n}'(\mu + \Sigma \mathbf{t}; \mathbf{c}, \Sigma)\Phi_{n}'(\mu + \Sigma \mathbf{t}; \mathbf{c}, \Sigma)^{T}}{[\Phi_{n}(\mu + \Sigma \mathbf{t}; \mathbf{c}, \Sigma)]^{2}} \Sigma \right\}.$$
(3.8.3)

Dado que $E(\mathbf{U^c}) = K'_{\mathbf{U^c}}(\mathbf{0})$ y $Var(\mathbf{U^c}) = K''_{\mathbf{U^c}}(\mathbf{0})$, la demostración concluye haciendo $\mathbf{t} = \mathbf{0}$ en las expresiones (3.8.2) y (3.8.3).

2. Apéndice B

En este apéndice se obtiene la fgm del parámetro de asimetría a posteriori $\lambda_{\mathbf{x}} = \lambda | \mathbf{x}$. Se considera el siguiente Lema el cual es usado para mostrar otros resultados del presente capítulo.

Lema 3.8.1. Si $U \sim \mathcal{N}_n(\mathbf{c}, \mathbf{C})$ es un vector aleatorio normal, no singular, entonces, para cada vector fijo \mathbf{a} de $m \times 1$ y para cada matriz fija \mathbf{A} de $m \times n$, se tiene que

$$E(\Phi_m(\mathbf{A}\mathbf{U} + \mathbf{a}; \mathbf{b}, B)) = \Phi_m(\mathbf{A}\mathbf{c} + \mathbf{a}; \mathbf{b}, B + \mathbf{A}\mathbf{C}\mathbf{A}^T).$$

La fgm de λ_x asumiendo una distribución a priori skew-normal, $\lambda \sim \mathcal{SN}(m,v,a)$, está dada por:

$$\begin{split} M_{\lambda_{\mathbf{x}}}(t) &= \int_{-\infty}^{\infty} e^{\lambda t} \phi(\lambda; m, v) \frac{\Phi_{n+1}(\lambda \mathbf{x}^{a}; \mathbf{m}^{*}, I_{n+1})}{\Phi_{n+1}(m \mathbf{x}^{0}; \mathbf{0}, I_{n+1} + v \mathbf{x}^{a} \mathbf{x}^{aT})} d\lambda \\ &= e^{tm + \frac{v^{2}v}{2}} \int_{-\infty}^{\infty} \phi(\lambda; m + tv, v) \frac{\Phi_{n+1}(\lambda \mathbf{x}^{a}; \mathbf{m}^{*}, I_{n+1})}{\Phi_{n+1}(m \mathbf{x}^{0}; \mathbf{0}, I_{n+1} + v \mathbf{x}^{a} \mathbf{x}^{aT})} d\lambda \\ &= e^{tm + \frac{v^{2}v}{2}} \frac{E[\Phi_{n+1}(\lambda \mathbf{x}^{a}; \mathbf{m}^{*}, I_{n+1})]}{\Phi_{n+1}(m \mathbf{x}^{0}; \mathbf{0}, I_{n+1} + v \mathbf{x}^{a} \mathbf{x}^{aT})}. \end{split}$$

Aplicando nuevamente el Lema 3.8.1 se tiene que:

$$M_{\lambda_{\mathbf{x}}}(t) = e^{tm + \frac{t^2v}{2}} \frac{\Phi_{n+1}(m\mathbf{x}^0 + tv\mathbf{x}^a; \mathbf{0}, I_{n+1} + v\mathbf{x}^a\mathbf{x}^{aT})}{\Phi_{n+1}(m\mathbf{x}^0; \mathbf{0}, I_{n+1} + v\mathbf{x}^a\mathbf{x}^{aT})}.$$

3. Apéndice C

En este apéndice, se presenta un teorema el cual es una herramienta importante para calcular el vector de medias y la matriz de varianza-covarianza de una distribución normal multivariada truncada. El siguiente lema preliminar es considerado en la demostración.

Denótese por $\mathbf{s}_{(i)}$ y $\mathbf{s}_{(ij)}$ los subvectores de $\mathbf{s} = (s_1, \dots, s_n)^T$ sin la componente s_i y (s_i, s_j) , respectivamente. Sea f y F la función de densidad y la función de distribución acumulada de un vector aleatorio $\mathbf{U} \in \mathbb{R}^n$ respectivamente. Por simplicidad, se denota por $f(s_i)$ $(F(s_i))$, la densidad (función de distribución acumulada) marginal

de la *i*-ésima componente y por $f(s_j|s_i)$ ($F(s_j|s_i)$), la densidad (función de distribución acumulada) condicional de la *j*-ésima componente dada la *i*-ésima componente de $U \in \mathbb{R}^n$. Las funciones $f'(s_i)$ y $F'(\cdot|s_i)$ son las derivadas parciales de f y F, respectivamente, con respecto a s_i .

Lema 3.8.2. Sea $U \in \mathbb{R}^n$ un vector aleatorio absolutamente continuo y se denota, respectivamente, por f y F a su función de densidad y su función de distribución acumulada. Suponga que f y F son funciones diferenciables. Entonces, para todo $\mathbf{s} = (s_1, \dots, s_n)^T \in \mathbb{R}^n$, se tiene que:

I)
$$\frac{\partial}{\partial s_i} F(\mathbf{s}) = f(s_i) F(\mathbf{s}_{(i)} | s_i);$$
II)
$$\frac{\partial^2}{\partial s_i^2} F(\mathbf{s}) = f'(s_i) F(\mathbf{s}_{(i)} | s_i) + f(s_i) F'(\mathbf{s}_{(i)} | s_i); y$$
III)
$$\frac{\partial^2}{\partial s_i \partial s_j} F(\mathbf{s}) = f(s_i) f(s_j | s_i) F(\mathbf{s}_{(j)} | s_i, s_j).$$

Dem: (i) Dado que $F(s) = P(U_i \le s_i | \mathbf{U}_{(i)} \le \mathbf{s}_{(i)}) P(\mathbf{U}_{(i)} \le \mathbf{s}_{(i)})$, se tiene

$$\frac{\partial}{\partial s_i} F(\mathbf{s}) = \frac{\partial}{\partial s_i} P(U_i \le s_i | \mathbf{U}_{(i)} \le \mathbf{s}_{(i)}) P(\mathbf{U}_{(i)} \le \mathbf{s}_{(i)})$$

$$= f(s_i | \mathbf{U}_{(i)} \le \mathbf{s}_{(i)}) P(\mathbf{U}_{(i)} \le \mathbf{s}_{(i)}). \tag{3.8.4}$$

Más aún, para el vector aleatorio $(V, W) \in \mathbb{R}^{p+q}$, tal que V tiene densidad f_V , es bien sabido que:

$$f(v|\mathbf{W} \in A) = f_{\mathbf{V}}(\mathbf{v}) \frac{P(\mathbf{W} \in A|\mathbf{V} = \mathbf{v})}{P(\mathbf{W} \in A)},$$
(3.8.5)

para cualquier subconjunto A de \mathbb{R}^q (véase Arellano-Valle *et al.* [7]). Entonces, se sigue de (3.8.4) y (3.8.5) que

$$\frac{\partial}{\partial s_i} F(\mathbf{s}) = f(s_i) F(\mathbf{s}_{(i)} | s_i),$$

(ii) Es directo de (i) que

$$\frac{\partial^2}{\partial s_i^2} F(\mathbf{s}) = \left\{ \frac{\partial}{\partial s_i} f(s_i) \right\} F(\mathbf{s}_{(i)}|s_i) + f(s_i) \frac{\partial}{\partial s_i} F(\mathbf{s}_{(i)}|s_i)$$
$$= f'(s_i) F(\mathbf{s}_{(i)}|s_i) + f(s_i) F'(\mathbf{s}_{(i)}|s_i).$$

(iii) Similarmente, se tiene que en (i), $F(\mathbf{s}_{(i)}|s_i) = P(U_j \leq s_j|U_i = s_i, \mathbf{U}_{(j)} \leq \mathbf{s}_{(j)})P(\mathbf{U}_{(j)} \leq \mathbf{s}_{(j)}|U_i = s_i)$. Entonces se sigue de (i) y (3.8.5) que:

$$\frac{\partial^2}{\partial s_i \partial s_j} F(\mathbf{s}) = \frac{\partial}{\partial s_j} f(s_i) F(\mathbf{s}_{(i)} | s_i)$$

$$= f(s_i) \frac{\partial}{\partial s_j} P(U_j \le s_j | U_i = s_i, \mathbf{U}_{(j)} \le \mathbf{s}_{(j)}) P(\mathbf{U}_{(j)} \le \mathbf{s}_{(j)} | U_i = s_i)$$

$$= f(s_i) f(s_j | U_i = s_i, \mathbf{U}_{(j)} \le \mathbf{s}_{(j)}) P(\mathbf{U}_{(j)} \le \mathbf{s}_{(j)} | U_i = s_i, \mathbf{U}_{(j)} \le \mathbf{s}_{(j)}) P(\mathbf{U}_{(j)} \le \mathbf{s}_{(j)} | U_i = s_i, \mathbf{U}_{(j)} \le \mathbf{s}_{(j)}) P(\mathbf{U}_{(j)} \le \mathbf{s}_{(j)} | U_i = s_i, \mathbf{U}_{(j)} \le \mathbf{s}_{(j)}) P(\mathbf{U}_{(j)} \le \mathbf{s}_{(j)} | U_i = s_i, \mathbf{U}_{(j)} \le \mathbf{s}_{(j)}) P(\mathbf{U}_{(j)} \le \mathbf{s}_{(j)} | U_i = s_i, \mathbf{U}_{(j)} \le \mathbf{s}_{(j)}) P(\mathbf{U}_{(j)} \le \mathbf{s}_{(j)} | U_i = s_i, \mathbf{U}_{(j)} \le \mathbf{s}_{(j)}) P(\mathbf{U}_{(j)} \le \mathbf{s}_{(j)} | U_i = s_i, \mathbf{U}_{(j)} \le \mathbf{s}_{(j)}) P(\mathbf{U}_{(j)} \le \mathbf{s}_{(j)} | U_i = s_i, \mathbf{U}_{(j)} \le \mathbf{s}_{(j)}) P(\mathbf{U}_{(j)} \le \mathbf{s}_{(j)} | U_i = s_i, \mathbf{U}_{(j)} \le \mathbf{s}_{(j)}) P(\mathbf{U}_{(j)} \le \mathbf{s}_{(j)} | U_i = s_i, \mathbf{U}_{(j)} \le \mathbf{s}_{(j)}) P(\mathbf{U}_{(j)} \le \mathbf{s}_{(j)} | U_i = s_i, \mathbf{U}_{(j)} \le \mathbf{s}_{(j)}) P(\mathbf{U}_{(j)} \le \mathbf{s}_{(j)} | U_i = s_i, \mathbf{U}_{(j)} \le \mathbf{s}_{(j)}) P(\mathbf{U}_{(j)} \le \mathbf{s}_{(j)} | U_i = s_i, \mathbf{U}_{(j)} \le \mathbf{s}_{(j)}) P(\mathbf{U}_{(j)} \le \mathbf{s}_{(j)} | U_i = s_i, \mathbf{U}_{(j)} \le \mathbf{s}_{(j)}) P(\mathbf{U}_{(j)} \le \mathbf{s}_{(j)} | U_i = s_i, \mathbf{U}_{(j)} \le \mathbf{s}_{(j)}) P(\mathbf{U}_{(j)} \le \mathbf{s}_{(j)} | U_i = s_i, \mathbf{U}_{(j)} \le \mathbf{s}_{(j)}) P(\mathbf{U}_{(j)} \le \mathbf{s}_{(j)} | U_i = s_i, \mathbf{U}_{(j)} \le \mathbf{s}_{(j)}) P(\mathbf{U}_{(j)} \le \mathbf{s}_{(j)} | U_i = s_i, \mathbf{U}_{(j)} \le \mathbf{s}_{(j)}) P(\mathbf{U}_{(j)} \le \mathbf{s}_{(j)} | U_i = s_i, \mathbf{U}_{(j)} \le \mathbf{s}_{(j)}) P(\mathbf{U}_{(j)} \le \mathbf{s}_{(j)} | U_i = s_i, \mathbf{U}_{(j)} \le \mathbf{s}_{(j)}) P(\mathbf{U}_{(j)} \le \mathbf{s}_{(j)} | U_i = s_i, \mathbf{U}_{(j)} \le \mathbf{s}_{(j)}) P(\mathbf{U}_{(j)} \le \mathbf{s}_{(j)} | U_i = s_i, \mathbf{U}_{(j)} \le \mathbf{s}_{(j)}) P(\mathbf{U}_{(j)} \le \mathbf{s}_{(j)} | U_i = s_i, \mathbf{U}_{(j)} \le \mathbf{s}_{(j)}) P(\mathbf{U}_{(j)} \le \mathbf{s}_{(j)} | U_i = s_i, \mathbf{U}_{(j)} \le \mathbf{s}_{(j)}) P(\mathbf{U}_{(j)} \le \mathbf{s}_{(j)} | U_i = s_i, \mathbf{U}_{(j)} \le \mathbf{s}_{(j)}) P(\mathbf{U}_{(j)} \le \mathbf{s}_{(j)} | U_i = s_i, \mathbf{U}_{(j)} \le \mathbf{s}_{(j)}) P(\mathbf{U}_{(j)} \le \mathbf{s}_{(j)} | U_i = s_i, \mathbf{U}_{(j)} \le \mathbf{s}_{(j)}) P(\mathbf{U}_{(j)} \le \mathbf{s}_{(j)} | U_i = s_i, \mathbf{U}_{(j)} \le \mathbf{s}_{(j)}) P(\mathbf{U}_{(j)$$

Algunos resultados útiles referidos a la distribución normal son propuestos en el siguiente teorema (véase además Domínguez-Molina *et al.* [34, 35]).

Teorema 3.8.3. Asúmase que $U \sim \mathcal{N}_n(\mu, \Sigma)$. Sea $U_{(i)}$ y $U_{(ij)}$ subvectores de U sin las componentes U_i y (U_i, U_j) , respectivamente. Considérese $\mu_{(i),i} = E(U_{(i)}|U_i = s_i)$, $\Sigma_{(i),i} = Var(U_{(i)}|U_i = s_i)$, $\mu_{(ij),ij} = E(U_{(ij)}|U_i = s_i, U_j = s_j)$ y $\Sigma_{(ij),ij} = Var(U_{(ij)}|U_i = s_i, U_j = s_j)$. Entonces, se tiene que:

I)
$$\frac{\partial}{\partial s_i} \Phi_n(\mathbf{s}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \phi(s_i; \mu_i, \sigma_{ii}) \Phi_{n-1}(\mathbf{s}_{(i)}; \boldsymbol{\mu}_{(i).i}, \boldsymbol{\Sigma}_{(i).i});$$

II)
$$\begin{split} \frac{\partial^2}{\partial s_i^2} \Phi_n(\mathbf{s}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \phi(s_i; \boldsymbol{\mu}_i, \sigma_{ii}) \bigg[\bigg(\frac{\mu_i - s_i}{\sigma_{ii}} \bigg) \Phi_{n-1}(\mathbf{s}_{(i)}; \boldsymbol{\mu}_{(i).i}, \boldsymbol{\Sigma}_{(i).i}) \\ &+ \frac{\partial}{\partial s_i} \Phi_{n-1}(\mathbf{s}_{(i)}; \boldsymbol{\mu}_{(i).i}, \boldsymbol{\Sigma}_{(i).i}) \bigg]; \end{split}$$

III)
$$\frac{\partial^2}{\partial s_i \partial s_j} \Phi_n(\mathbf{s}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \phi(s_i; \mu_i, \sigma_{ii}) \phi(s_j; \mu_{j,i}, \sigma_{j,i}) \Phi_{n-2}(\mathbf{s}_{(ij)}; \boldsymbol{\mu}_{(ij),ij}, \boldsymbol{\Sigma}_{(ij),ij})$$

donde para cada subconjunto I de $\{1,\ldots,n\}$, el vector de medias y la matriz de varianzas-covarianzas condicionales de $[\mathbf{U}_{(I)}|\mathbf{U}_I=\mathbf{s}_I]$ están dados, respectivamente, por: $\boldsymbol{\mu}_{(I).I}=\boldsymbol{\mu}_{(I)}+\boldsymbol{\Sigma}_{(I)I}\boldsymbol{\Sigma}_{II}^{-1}(\mathbf{s}_I-\boldsymbol{\mu}_I)$ and $\boldsymbol{\Sigma}_{(I).I}=\boldsymbol{\Sigma}_{(I)(I)}-\boldsymbol{\Sigma}_{(I)I}\boldsymbol{\Sigma}_{II}^{-1}\boldsymbol{\Sigma}_{I(I)}$.

Dem: La demostración sigue del Lema 3.8.2 y de la bien conocida descomposición marginal-condicional de la distribución normal multivariada. De hecho, si

$$\mathbf{U} = \left(egin{array}{c} \mathbf{U}_1 \ \mathbf{U}_2 \end{array}
ight) \sim \mathcal{N}_{l+(n-l)} \left(oldsymbol{\mu} = \left(egin{array}{c} oldsymbol{\mu}_1 \ oldsymbol{\mu}_2 \end{array}
ight) oldsymbol{\Sigma} = \left(egin{array}{cc} oldsymbol{\Sigma}_{11} & oldsymbol{\Sigma}_{12} \ oldsymbol{\Sigma}_{21} & oldsymbol{\Sigma}_{22} \end{array}
ight)
ight),$$

entonces $\phi_n(\mathbf{u}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \phi_l(\mathbf{u}_1; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}) \phi_{n-l}(\mathbf{u}_2; \boldsymbol{\mu}_{2,1}, \boldsymbol{\Sigma}_{2,1})$, donde $\boldsymbol{\mu}_{2,1} = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{u}_1 - \boldsymbol{\mu}_1)$ y $\boldsymbol{\Sigma}_{2,1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$ es el vector de medias y la matriz de varianzas-covarianzas de $[\mathbf{U}_2|\mathbf{U}_1 = \mathbf{u}_1]$, respectivamente.

Capítulo 4

Inferencia Bayesiana en mezclas de distribuciones skew-normal

"Otro importante caso de la disección de una curva de frecuencia pueden surgir, cuando la curva de frecuencia, sin ser asimétrica, se compone de la suma o diferencia de dos componentes... Este caso es aún más interesante e importante, ya que no es poco probable que se produzca en las investigaciones estadísticas."

Karl Pearson, 1894

En el presente capítulo se introduce una clase de distribuciones llamada mezclas de asimetría de distribuciones skew-normal y se estudia algunas de sus propiedades. En particular, se discute una interpretación Bayesiana de algunos resultados de invarianza. Se desarrolla el análisis Bayesiano de los modelos de regresión skew-normal, skew-normal-generalizado, skew-normal-t y skew-t-normal bajo determinadas especificaciones de las distribuciones a priori de los parámetros de dichos modelos y proponiendo una conveniente representación jerárquica, se realiza el análisis de las distribuciones a posteriori. Finalmente, se ilustra en el enfoque Bayesiano con un conjunto de datos reales.

4.1. Introducción

Tal como se mencionó en los Capítulos 1 y 3, la construcción de distribuciones multivariadas más allá de la normal ha tenido un creciente interés en los ultimos años. Algunas de estas propuestas pueden encontrarse en el libro editado por Genton [48], en Azzalini [12], en Arellano-Valle y Azzalini [2], y desde el punto de vista unificado, en Arellano-Valle *et al.* [4].

Una clase de estas distribuciones es la introducida por Wang et al. [83] y consiste en las llamadas distribuciones skew-simétricas (SS), cuya función de densidad es de la forma

$$2f(\mathbf{z})Q(\mathbf{z}), \qquad \mathbf{z} \in \mathbb{R}^n,$$
 (4.1.1)

donde $f:\mathbb{R}^n \to \mathbb{R}^+$ es una densidad continua, simétrica alrededor de cero, es decir, $f(-\mathbf{z}) = f(\mathbf{z})$ para todo $\mathbf{z} \in \mathbb{R}^n$, y $Q:\mathbb{R}^n \to [0,1]$ es una función de asimetría que satisface $Q(-\mathbf{z}) + Q(\mathbf{z}) = 1$ para todo $\mathbf{z} \in \mathbb{R}^n$. Un vector aleatorio \mathbf{Z} con densidad (4.1.1) es denotado por $\mathbf{Z} \sim \mathcal{SS}_{n,f,Q}$. Cuando f es una densidad que pertenece a la familia de distribuciones elípticas, la familia (4.1.1) define la distribución skew-elíptica-generalizada estudiada por Genton y Loperfido [49]. Ma y Genton [64] y Azzalini y Capitanio [14], mencionan que cualquier función de asimetría Q puede ser escrita como $Q(\mathbf{z}) = G(w(\mathbf{z}))$, donde $G:\mathbb{R} \to [0,1]$ es la función de distribución acumulada de una variable aleatoria continua simétrica alrededor de cero, y $w:\mathbb{R}^n \to \mathbb{R}$ es una función continua par, es decir,

 $w(-\mathbf{z}) = -w(\mathbf{z})$ para todo $\mathbf{z} \in \mathbb{R}^n$. Generalmente, se elige

$$Q(\mathbf{z}) = G(\boldsymbol{\lambda}^T \mathbf{z}), \qquad \mathbf{z} \in \mathbb{R}^n,$$
 (4.1.2)

donde el vector $\lambda \in \mathbb{R}^n$ controla la asimetría. Cuando $\lambda = 0$, (4.1.1) es una densidad simétrica. Un vector aleatorio Z con densidad (4.1.1) y función de asimetría (4.1.2) es denotado por $\mathbf{Z} \sim \mathcal{SS}_{n,f,G}(\lambda)$. En particular, cuando $f(\mathbf{z}) = \phi_n(\mathbf{z};0,I_n)$, y $G = \Phi(\lambda^T\mathbf{z})$, la densidad resultante (4.1.1) es la distribución skew-normal multivariada estándar $\mathcal{SN}_n(0,I_n,\lambda) = \mathcal{SN}_n(\lambda)$ definida por Azzalini y Dalla Valle [15]. Para n=1, ésta se reduce a la distibución skew-normal univariada estándar de Azzalini [11]. El uso de polinomios de orden par para la función w, propuesto por Ma y Genton [64] permite flexibilizar las distribuciones skew-simétricas, añadiendo multimodalidad.

En este capítulo, se considera elecciones alternativas de polinomios lineales o de orden par como funciones de asimetría Q. Para efectos de ilustración, considérese el caso univariado de (4.1.1). Una generalización interesante de (4.1.2) resulta de la elección de

$$w(z) = \frac{\alpha_1 z}{\sqrt{1 + \alpha_2^2 z^2}}, \qquad z \in \mathbb{R}, \tag{4.1.3}$$

donde $\alpha_1 \in \mathbb{R}$ y $\alpha_2 \geq 0$. El caso particular de $f = \phi(\cdot)$, y $G = \Phi(\cdot)$ ha sido estudiado por Arellano-Valle *et al.* [10], y es llamada distribución skew-normal-generalizada, denotada por $\mathcal{SGN}(\alpha_1, \alpha_2)$. Ahí, se muestra que esta distribución puede ser representada como una mezcla en el parámetro de asimetría de distribuciones skew-normal, donde la distribución de mezclas para el parámetro de asimetría es normal. Específicamente, si Z tiene

distribución skew-normal-generalizada de parámetros α_1 y α_2 , entonces existe una variable aleatoria S tal que

$$[Z|S=s] \sim \mathcal{SN}(s)$$
 y $S \sim \mathcal{N}(\alpha_1, \alpha_2)$. (4.1.4)

En otras palabras, esta representación permite identificar la función w(z) dentro de la clase definida por (4.1.1) cuando $f = \phi(\cdot)$ y $G = \Phi(\cdot)$. Otro caso importante se origina cuando se considera mezclas de escala y asimetría de la distribución skew-normal. Por ejemplo, la clase definida por

$$[Z|S_1 = s_1, S_2 = s_2] \sim \mathcal{SN}(0, s_2^{-1}, s_1^{-1/2}\alpha_1),$$
 (4.1.5)

donde S_1 y S_2 son variables independientes no negativas. Cualquier modelo de la forma $2f(z)G(\alpha_1z)$, para lo cual f y G son mezclas de escala de la distribución normal, pertenece a la clase definida por (4.1.5). Particularmente, la distribución skew-t en la forma introducida por Azzalini y Capitanio [14] con densidad $2t(z;\nu)T(\sqrt{\nu+1}\,z/\sqrt{\nu+z^2};\nu+1)$, donde $t(z;\nu)$ y $T(z;\nu)$ son la función de densidad y de distribución acumulada de la distribución t-Student, respectivamente, es obtenida cuando S_2 tiene distribución gamma, $\mathcal{G}(\nu/2,\nu/2)$ y $S_1\equiv 1$ (veáse Azzalini y Genton [16] para propiedades adicionales). Otros ejemplos obtenidos a partir de (4.1.5) son las llamadas distribuciones skew-normal-t y skew-t-normal, con densidades de la forma $2\phi(z)T(\alpha_1z)$ y $2t(z;\nu)\Phi(\alpha_1z)$, y denotadas por $\mathcal{SNT}(\alpha_1,\nu)$ y $\mathcal{STN}(\alpha_1,\nu)$, respectivamente. Estas distribuciones han sido estudiadas por Nadarajah y Kotz [67] y Gómez et al. [50], y pueden ser obtenidas de (4.1.5) considerando $S_1 \sim \mathcal{G}(\nu/2,\nu/2)$ y $S_2 \equiv 1$, y $S_2 = S_1 \sim \mathcal{G}(\nu/2,\nu/2)$, respectivamente.

Las representaciones en (4.1.4) y (4.1.5) son particularmente importantes desde el punto de vista Bayesiano, toda vez que proporcionan formulaciones jerárquicas de las versiones de localización y escala respectivas de (4.1.1). Más aún, ellas pueden ser interpretadas como especificaciones Bayesianas del modelo skew-normal con ciertas distribuciones a priori sobre los parámetros de asimetría/escala indicados arriba. Por ejemplo, (4.1.4) puede ser aplicado en la especificación Bayesiana del modelo skew-normal para una muestra aleatoria Y_1, \ldots, Y_n de $Y = \xi + \omega Z$, donde la distribución condicional de Z dado S es skew-normal con una priori normal para el parámetro de asimetría S. Entonces, sólo se necesita elicitar una priori para los parámetros de localización y escala (ξ, ω^2) para completar la especificación del modelo.

Este capítulo está referido a la subclase de distribuciones asimétricas que pueden ser representadas como mezclas de asimetría de la familia definida por (4.1.1) y (4.1.2). En particular, la representación (4.1.4) es extendida para todas las clases de distribuciones con densidad dada por (4.1.1) con (4.1.3), y extensiones multivariadas. Por simplicidad en la exposición, en la Sección 4.2 se considera el caso univariado, donde se discute las consecuencias de usar una distribución de mezcla simétrica de locación-escala. En la Sección 4.3, se considera la idea de mezclas de asimetría con el fin de identificar funciones de asimetría para la clase multivariada skew-simétrica (4.1.1). El procedimiento es ilustrado con la distribución skew-normal multivariada. En las Secciones 4.4 y 4.5, se desarrolla el análisis

Bayesiano para los modelos de regresión skew-normal, skew-normal-generalizado, skew-normal-t y skew-t-normal bajo ciertas especificaciones de las distribuciones a priori de los parámetros. En la Sección 4.6 se presenta una aplicación utilizando un conjunto de datos de atletas australianos. Finalmente, el resumen del capítulo y las principales conclusiones del mismo se presentan en la Sección 4.7.

4.2. Mezclas de asimetría de distribuciones skew univariadas

4.2.1. Definición y propiedades

Las mezclas de asimetría de distribuciones asimétricas univariadas forman una importante sub-clase de la familia de distribuciones skew-simétricas definida por (4.1.1) con n=1. Esto se debe a que en esta clase se puede especificar la función de asimetría Q partiendo de una clase más simple basada en (4.1.2). Esta subclase es obtenida como una mezcla de distribuciones asimétricas definidas por (4.1.1) y (4.1.2) en el parámetro de asimetría λ . Consecuentemente, la subclase resultante contiene distribuciones que son más flexibles que las originales.

Definición 4.2.1. La distribución de una variable aleatoria Y es una mezcla de asimetría de una distribución skew-simétrica (MASS) si existe una variable aleatoria S tal condicionalmente en S, Y tenga distribución skew-simétrica, es decir, $[Y|S=s] \sim \mathcal{SS}_{f,G}(\xi,\omega^2,s)$ para alguna densidad simétrica f (la cual puede depender de s) y donde G es una función

de distribución (la cual puede depender de $z=(y-\xi)/\omega$ y/o de s).

Cuando $f = \phi(\cdot)$, la distribución de la variable aleatoria Y es una mezcla de asimetría de distribuciones skew-normal (MASN). De la Definición 4.2.1 se sigue que la densidad condicional de Y dado S es de la forma (4.1.1) con (4.1.2). El siguiente resultado proporciona la distribución marginal de Y.

Proposición 4.2.1. Sea la distribución condicional de Y dado S skew-simétrica, es decir, $[Y|S=s] \sim \mathcal{SS}_{f,G}(\xi,\omega^2,s), s \in \mathbb{R}$, donde la densidad f no depende de s y S tiene función de distribución acumulada $H(s|\theta)$, indexada por un parámetro θ . Entonces, $Y \sim \mathcal{SS}_{f,Q}(\xi,\omega^2,\theta)$, es decir, la densidad es de la forma skew-simétrica (4.1.1), donde la función de asimetria Q está dada por

$$Q(z,\theta) = E[G(zS)] = \int_{-\infty}^{\infty} G(zs)dH(s|\theta). \tag{4.2.1}$$

Más aún, si H es absolutamente continua con densidad h=H', entonces la distribución condicional de S dado Y=y depende de (y,ξ,ω^2,θ) y está dada por

$$f(s|z,\theta) = \frac{h(s|\theta)G(zs)}{\int_{-\infty}^{\infty} h(s|\theta)G(zs)ds}.$$
 (4.2.2)

donde $z = (y - \xi)/\omega$.

Dem. La demostración es inmediata de la Definición 4.2.1 y el Teorema de Bayes.

Nótese que cualquier variable aleatoria Y con distribución skew-simétrica $\mathcal{SS}_{f,Q}(\xi,\omega^2,\theta)$ puede ser representada como $Y \stackrel{d}{=} \xi + \omega Z$, donde Z tiene distribución skew-simétrica $\mathcal{SS}_{f,Q}(0,1,\theta)$, y por lo tanto sus propiedades pueden ser analizadas utilizando la versión estandarizada. La función de distribución acumulada de mezclas $H(\cdot|\theta)$ puede ser elegida

arbitrariamente. Por lo tanto, se obtiene una subfamilia de particular importancia con una distribución discreta H. Esto permite definir mezclas finitas con funciones de asimetría de la forma

$$Q(z) = \sum_{k=1}^{K} a_k G(\lambda_k z), \tag{4.2.3}$$

donde $a_k \geq 0$, para todo k = 1, ..., K, con $\sum_{k=1}^K a_k = 1$, y $\lambda_k \in \mathbb{R}$. Se pueden utilizar funciones de asimetría de la forma (4.2.3) para obtener aproximaciones de la forma (4.2.1) con H continua, las cuales no pueden ser calculadas de forma explícita.

4.2.2. Distribución de mezcla simétrica de localización-escala

Otra interesante subfamilia de mezclas de asimetría de distribuciones asimétricas se obtiene cuando la función de asimetría (4.2.1) es calculada con una función de distribución simétrica de localización-escala *H*.

Para caracterizar esta subfamilia, considérese primero la siguiente situación general. Sea (X_0, Z_0, S) un vector aleatorio tal que, condicionalmente en S, la variable aleatoria $W = X_0 - SZ_0$ tiene distribución simétrica, implicando que para cada valor s de S

$$P(W \le 0|S = s) = \int_{-\infty}^{\infty} P(X_0 \le sz|Z_0 = z, S = s) f_{Z_0|S = s}(z) dz$$
$$= \int_{-\infty}^{\infty} F_{X_0|Z_0 = z, S = s}(sz) f_{Z_0|S = s}(z) dz$$
$$= 1/2,$$

y por lo tanto la función

$$f(z|s) = 2F_{X_0|Z_0=z,S=s}(sz)f_{Z_0|S=s}(z), \quad z \in \mathbb{R}, \tag{4.2.4}$$

define una función de densidad para cualquier valor de s. Nótese además que, bajo esta especificación, la función de distribución acumulada $F_{X_0|Z_0=z,S=s}=G_{(z,s)}$, y la densidad condicional $f_{Z_0|S=s}=f_{(s)}$, no son necesariamente simétricas y pueden depender de (z,s) y s, respectivamente. Sin embargo, si ellas son simétricas, es decir, $f_{(s)}(-y)=f_{(s)}(y)$ y $G_{(z,s)}(-y)=1-G_{(z,s)}(y)$, para todo y y cada valor de (z,s), y $G_{(z,s)}(-z,-s)=G_{(z,s)}(-z,s)=G_{(z,s)}(z,s)$ para todo (z,s), entonces (4.2.4) se vuelve la densidad de una distribución skew-simétrica $\mathcal{SS}_{f_{(s)},G_{(z,s)}}(s)$. Además,

$$G_{(z,s)} = F_{X_0|Z_0=z,S=s}(sz) = F_{W|Z_0=z,S=s}(0).$$

Se considera ahora la variable aleatoria definida por $Z \stackrel{d}{=} [Z_0|W \le 0]$. Nótese que f(z|s) en (4.2.4) es simplemente la densidad condicional de Z dado S, es decir, la densidad de $[Z_0|W \le 0, S=s]$. Entonces, la densidad de $Z \stackrel{d}{=} [Z_0|W \le 0]$ está dada por

$$f_{Z}(z) = \int_{-\infty}^{\infty} f(z|s) dH_{S}(s|\theta) = \int_{-\infty}^{\infty} 2F_{X_{0}|Z_{0}=z,S=s}(sz) f_{Z_{0}|S=s}(z) dH_{S}(s|\theta),$$

por lo que Z dado S tiene distribución skew-simétrica $SS_{f(s),G(z,s)}(s)$ cuando la densidad f(s) de Z_0 dado S y la función de distribución G(z,s) de X_0 dado Z_0 y S son simétricas. Si, adicionalmente, se asume que Z_0 y S son independientes, entonces $f(s) = f_{Z_0|S=s} = f_{Z_0} = f_{Z_0}$, no depende de S. Por lo tanto, la función de densidad de S se reduce a

$$f_Z(z) = 2f(z) \int_{-\infty}^{\infty} G_{(z,s)}(sz) dH_S(s|\theta) = 2f(z) F_{W|Z_0=z}(0).$$

Si Z_0 y S se asumen independientes, entonces el supuesto de simetría en f y $G_{(z,s)}$ para todo (z,s), implica que Z_0 condicionalmente en S tiene distribución skew-simétrica $\mathcal{SS}_{f,G_{(z,s)}}(s)$ y además, marginalmente Z tiene distribución skew-simétrica $\mathcal{SS}_{f,Q}(\theta)$, con

$$Q(z,\boldsymbol{\theta}) = F_{W|Z_0=z}(0) = \int_{-\infty}^{\infty} G_{(z,s)}(sz) dH(s|\boldsymbol{\theta}).$$

Adicionalmente, este supuesto de simetría implica que condicionalmente en S, la variable aleatoria $W = X_0 - SZ_0$ es además simétrica cualquiera sea la distribución de S.

Proposición 4.2.2. Sea (X_0, Z_0, S) un vector aleatorio tal que Z_0 tiene función de densidad f, S tiene función de distribución H, indexada por un parámetro θ , con Z_0 independiente de S y f es una función de densidad simétrica alrededor de cero. Sea $G_{(z,s)} = F_{X_0|Z_0=z,S=s}$ la función de distribución acumulada condicional de X_0 dado Z_0 y S. Suponga que $G_{(z,s)}(-y) = 1 - G_{(z,s)}(y)$ y $G_{(-z,-s)}(y) = G_{(-z,s)}(y) = G_{(z,-s)}(y) = G_{(z,s)}(y)$ para todo y y para cada valor de (z,s). Sea además Z una variable aleatoria tal que $Z \stackrel{d}{=} [Z_0|W \leq 0]$, donde $W = X_0 - SZ_0$. Entonces, condicionalmente en S, Z tiene distribución skew-simétrica $SS_{f,G_{(z,s)}}(s)$ y por lo tanto, marginalmente, Z tiene distribución skew-simétrica $SS_{f,Q}(\theta)$, con $Q(z,\theta) = F_{W|Z_0=z}(0) = \int_{-\infty}^{\infty} G_{(z,s)}(sz) dH(s|\theta)$.

Corolario 4.2.3. Bajo las condiciones de la Proposición 4.2.2, se tiene los siguientes resultados:

- I) Si X_0 es independiente de Z_0 , entonces dado S, Z tiene distribución skew-simétrica $\mathcal{SS}_{f,G(s)}(s)$ y por lo tanto la distribución marginal de Z es skew-simétrica $\mathcal{SS}_{f,Q}(\theta)$, con $G_{(s)} = F_{X_0|S=s}$ y $Q(z,\theta) = F_{W|Z_0=z}(0) = \int_{-\infty}^{\infty} G_{(s)}(sz)dH(s|\theta)$.
- II) Si X_0 con función de distribución G, Z_0 con función de densidad f y S con función

de distribución H indexada por un parámetro θ , son variables aleatorias independientes, con f(-x) = f(x) y G(-x) = 1 - G(x) para todo x, entonces Z dado S tiene distribución skew-simétrica $SS_{f,G}(s)$ y marginalmente, Z tiene distribución skew-simétrica $SS_{f,Q}(\theta)$, con $Q(z,\theta) = F_{W|Z_0=z}(0) = \int_{-\infty}^{\infty} G(sz)dH(s|\theta)$.

Bajo los supuestos planteados en la Proposición 4.2.2, se tiene que X_0 y Z_0 son variables aleatorias simétricas alrededor de cero, y que Z_0 y S son independientes. Esto implica que la variable aleatoria $W=X_0-SZ_0$ es también simétrica cualquiera sea la distribución de S. En particular, si S es asumida simétrica alrededor de cero, entonces condicionalmente en Z_0 , la variable aleatoria $W=X_0-SZ_0$ será también simétrica, es decir, $\int_{-\infty}^{\infty} G_{(z,s)}(sz)dH(s|\theta) = F_{W|Z_0=z}(0) = 1/2 \text{ para todo } z, \text{ y por lo tanto } Q(z,\theta) = 1/2,$ para todo z y θ . El siguiente corolario sigue de este último hecho.

Corolario 4.2.4. Sea $Z \stackrel{d}{=} [Z_0|X_0 - SZ_0 \le 0]$, donde X_0 , Z_0 y S son variables aleatorias simétricas alrededor de cero, con Z_0 y S independientes. Entonces, Z tiene la misma distribución simétrica que Z_0 , es decir, $[Z_0|X_0 - SZ_0 \le 0] \stackrel{d}{=} Z_0$ con función de densidad f.

4.2.3. MASS basadas en distribución de mezcla de localización-escala

En esta sección, se caracteriza la subfamilia MASS obtenida cuando la función de asimetría (4.2.1) considera una función de distribución acumulada de localización-escala H, es decir,

$$H(s|\theta) = H_0((s-\eta)/\tau), \quad \text{con} \quad \theta = (\eta, \tau) \in \mathbb{R} \times \mathbb{R}^+,$$

donde H_0 es una función de distribución acumulada estandarizada y simétrica alrededor de cero, es decir, la función de distribución acumulada de $S_0 = (S - \eta)/\tau$, donde S tiene función de distribución H indexada por los parámetros η y τ . En este caso, la función de asimetría (4.2.1) puede ser reescrita como

$$Q(z, \eta, \tau) = \mathbb{E}[G(\tau z S_0 + z \eta)] = \int_{-\infty}^{\infty} G(z\{\tau s_0 + \eta\}) dH_0(s_0). \tag{4.2.5}$$

Dos propiedades importantes de esta función de asimetría son:

- a) Si $\tau=0$ y G no depende de s_0 , entonces $Q(z,\eta,\tau)=G(\eta z)$ y (4.1.1) con (4.1.2) y $\lambda=\eta$.
- b) Si $\eta=0$, entonces $Q(z,\eta,\tau)=\mathrm{E}[G(z\tau S_0)]=1/2$, para cualquier valor de z y τ , lo cual es consecuencia de la simetría (alrededor de cero) de G y H_0 , véase el Corolario 4.2.4.

De otro lado, los resultados de la Proposición 4.2.2 proporcionan una expresión más general y conveniente para obtener una función de asimetría en (4.2.5), tal como se indica a continuación. De hecho, sea

$$W_0 = \frac{W + \eta Z_0}{\sqrt{1 + \tau^2 Z_0^2}} = \frac{X_0 - \tau S_0 Z_0}{\sqrt{1 + \tau^2 Z_0^2}},$$

la cual es la vesión simétrica estandarizada de $W=X_0-SZ_0$ donde X_0 es una variable aleatoria estandarizada independiente de S_0 y $S=\eta+\tau S_0$ es una variable aleatoria simétrica de localización-escala. Nótese que $W=\sqrt{1+\tau^2Z_0^2}~W_0-\eta Z_0$, y por lo tanto $Z\stackrel{d}{=}[Z_0|W_0\leq \eta Z_0/\sqrt{1+\tau^2Z_0^2}]$, cuya densidad está dada por (véase, por ejemplo,

Arellano-Valle et al. [7])

$$f_Z(z) = f(z) \frac{P(W_0 \le \frac{\eta Z_0}{\sqrt{1+\tau^2 Z_0^2}} | Z_0 = z)}{P(W_0 \le \frac{\eta Z_0}{\sqrt{1+\tau^2 Z_0^2}})},$$

donde $P(W_0 \le \frac{\eta Z_0}{\sqrt{1+\tau^2 Z_0^2}}) = P(W \le 0) = 1/2$. Este resultado es resumido en la siguiente proposición.

Proposición 4.2.5. Sea (X_0, Z_0, S) un vector aleatorio que satisface las mismas condiciones de la Proposición 4.2.2, con X_0 una variable aleatoria estandarizada independiente de S. Supóngase que $S=\eta+\tau S_0$, donde $\eta\in\mathbb{R}$ y $\tau>0$ son parámetros de localización y escala, respectivamente, y S_0 es una variable aleatoria simétrica estandarizada. Entonces, la variable aleatoria $Z \stackrel{d}{=} [Z_0|X_0-SZ_0\leq 0]$ tiene distribución skew-simétrica $SS_{f,Q}(\eta,\tau)$ con $Q(z,\eta,\tau)=F_{W_0|Z_0=z}\left(\eta z/\sqrt{1+\tau^2 z^2}\right)$, es decir, con densidad dada por

$$f_Z(z) = 2f(z)F_{W_0|Z_0=z}\left(\frac{\eta z}{\sqrt{1+\tau^2 z^2}}\right),$$
 (4.2.6)

donde $W_0 = \frac{X_0 - \tau S_0 Z_0}{\sqrt{1 + \tau^2 Z_0^2}}$ es una variable aleatoria simétrica estandarizada.

Nótese que si W_0 es independiente de Z_0 , entonces $F_{W_0|Z_0=z}=F_{W_0}=G_0$, y por lo tanto (4.2.6) se reduce a (4.1.1) con (4.1.3).

Dos consecuencias interesantes de asumir que la función de distribución acumulada de mezclas H es simétrica alrededor de cero, para todo s, son que, en primer lugar, la distribución marginal de Z_0 no es afectada por la elección de una función de distribución acumulada simétrica para la variable aleatoria de mezcla S (este hecho se relaciona con la propiedad b) del Corolario 4.4.2). En segundo lugar, la distribución condicional de la variable aleatoria de mezclas S dado Z_0 pertenece a la clase de distribuciones skew-simétricas

cuando H es absolutamente continua, es decir, cuando H tiene densidad h = H'. Ambos resultados son relevantes desde el punto de vista Bayesiano y son resumidos en la siguiente proposición, cuya demostración es directa de (4.2.6) y el Teorema de Bayes.

Proposición 4.2.6. Sea Z dado S con distribución skew-simétrica $SS_{f,G}(s)$, donde f no depende de s, y S tiene función de distribución H con parámetro θ . Si H es la función de distribución acumulada de una distribución simétrica alrededor de cero, entonces la distribución marginal de Z es simétrica alrededor de cero y tiene función de densidad f. Más aún, si H admite una función de densidad h = H', entonces S dado S tiene distribución skew-simétrica $SS_{h,G}(z,\theta)$.

4.2.4. Mezclas de asimetría de distribuciones skew-normal univaria-

Un ejemplo interesante de los resultados previos es la distribución skew-normal-generalizada definida por (4.1.1) y (4.1.3). Tal como se indicara en (4.1.4), este modelo puede ser especificado como una mezcla de asimetría de la distribución skew-normal, tomando una distribución de mezcla en el parámetro de asimetría (véase Arellano-Valle et~al.~[10]). De la Proposición 4.2.1, esto es equivalente a considerar $Y=\xi+\omega Z$, y suponer que condicionalmente en S,Z tiene distribución skew-simétrica $\mathcal{SS}_{f,G}(s)$ y S tiene función de distribución $H,~\mathrm{con}~f=\phi(\cdot),~G=\Phi(\cdot)$ y $H=\Phi(\frac{t-\eta}{\tau}),~\mathrm{es}~\mathrm{decir},~Z~\mathrm{dado}~S~\mathrm{tiene}~\mathrm{distribución}~\mathrm{skew-normal}~\mathrm{con}~\mathrm{parámetro}~\mathrm{de}~\mathrm{asimetría}~\mathrm{s}~\mathrm{y}~\mathrm{S}~\mathrm{distribución}~\mathrm{normal}~\mathrm{con}~\mathrm{parámetro}~\mathrm{s}~\mathrm{y}~\mathrm{y}^{-2}.$ Dado que $S_0=\frac{S-\eta}{\tau}~\mathrm{distribuye}~\mathrm{normal}~\mathrm{estándar},~\mathrm{entonces}~\mathrm{por}~\mathrm{(4.2.5)}~\mathrm{se}~\mathrm{sigue}~\mathrm{que}~\mathrm{(véase~además)}$

Ellison [37])

$$Q(z, \eta, \tau) = \mathrm{E}[\Phi(z\tau S_0 + z\eta)] = \Phi\left(\frac{\eta z}{\sqrt{1 + \tau^2 z^2}}\right),\,$$

resultando la siguiente densidad marginal para la variable aleatoria Z:

$$f_Z(z) = 2\phi(z)\Phi\left(\frac{\eta z}{\sqrt{1+\tau^2 z^2}}\right), \quad z \in \mathbb{R}.$$
 (4.2.7)

Esta es la distribución skew-normal-generalizada mencionada en la introducción, y denotada por $\mathcal{SGN}(\eta,\tau)$. Nótese que $\mathcal{SGN}(0,\tau)=\mathcal{N}(0,1)$, para cualquier τ , $\mathcal{SGN}(\eta,\infty)=\mathcal{N}(0,1)$, para cualquier η , y $\mathcal{SGN}(\eta,0)=\mathcal{SN}(\eta)$. Si Z tiene distribución skew-normal-generalizada con parámetros η y τ , entonces -Z tiene distribución skew-normal-generalizada con parámetros $-\eta$ y τ . Además,

$$[Z|S=s] \stackrel{d}{=} \frac{s}{\sqrt{1+s^2}}|U| + \frac{1}{\sqrt{1+s^2}}V,$$
 (4.2.8)

donde S tiene distribución normal con parámetros η y τ^2 y es independiente de U y V, las cuales son variables aleatorias independientes y con distribución idéntica normal estándar. La siguiente representación es útil desde el punto de vista computacional, pues significa que si Z tiene distribución skew-normal-generalizada de parámetros η y τ , entonces existen variables aleatorias S y T mutualmente independientes tal que:

i)
$$[Z|S=s, T=t] \sim \mathcal{N}\left(\frac{st}{\sqrt{1+s^2}}, \frac{1}{1+s^2}\right)$$
,

$$ii)$$
 $S \sim \mathcal{N}(\eta, \tau^2),$

$$iii)$$
 $T \sim \mathcal{HN}(0,1),$

donde $\mathcal{HN}(0,1)$ es la distribución half-normal. Esta especificación jerárquica puede ser utilizada para implementar métodos MCMC (desde el punto de vista Bayesiano) o el algoritmo EM (desde el punto de vista clásico), para realizar inferencia sobre (η, τ) .

Tal como se mencionara en la introducción, otros ejemplos que pertenecen a la clase $2f(z)G(\lambda z)$ se obtienen tomando $S=\sqrt{W}\lambda$, para alguna variable W no negativa. Más aún, la subclase de mezclas de escala y asimetría de distribuciones skew-normal puede ser introducida considerando (4.1.5). Por lo tanto, esta es la representación de la subclase $2f(z)G(\lambda z)$ donde f y G son la función de densidad y función de distribución acumulada las cuales son mezclas de escala de la distribución normal.

4.3. Mezclas de asimetría de distribuciones skew-normal multivariadas

A continuación se proponen algunas extensiones multivariadas de la distribución skewnormal-generalizada (4.2.7). En todos los casos, la idea de mezcla de asimetría discutida
en las secciones anteriores es adaptada en el caso de distribuciones skew-normal independientes y dependientes. Las distribuciones resultantes pueden ser interpretadas desde el
punto de vista Bayesiano como si se considerase una priori normal para el parámetro de
asimetría. Por lo tanto, y como se establecerá en las proposiciones siguientes, la función de
densidad predictiva y posteriori asociada al parámetro de asimetría define la densidad de la
distribución skew-normal-generalizada multivariada.

Para cualquier vector n-dimensional \mathbf{w} , se denota por $D(\mathbf{w})$ la matriz diagonal de dimensión $n \times n$ formada por los componentes w_1, \ldots, w_n de \mathbf{w} . Entonces, para dos vectores cualesquiera de dimensión n, \mathbf{s} y \mathbf{z} , se tiene $D(\mathbf{s})\mathbf{z} = D(\mathbf{z})\mathbf{s} = (s_1z_1, \ldots, s_nz_n)^T$. La demostración de las proposiciones siguientes está basada en el Lema 3.8.1, Apéndice B, Capítulo 3 del presente documento.

Proposición 4.3.1. Supóngase que dado S, Z_i , con $i=1,\ldots,n$, son variables aleatorias independientes con distribución skew-normal $SN(s_i)$, donde el vector aleatorio $S = (S_1, \ldots, S_n)^T$ tiene distribución normal multivariada $N_n(\eta, \Omega)$. Entonces, la densidad marginal de $Z = (Z_1, \ldots, Z_n)^T$ está dada por

$$f(\mathbf{z}|\boldsymbol{\eta}, \boldsymbol{\Omega}) = 2^n \phi_n(\mathbf{z}) \Phi_n(D(\boldsymbol{\eta})\mathbf{z}; \mathbf{0}, \mathbf{I}_n + D(\mathbf{z})\boldsymbol{\Omega}D(\mathbf{z})),$$

la cual contiene a la densidad $\mathcal{N}_n(\mathbf{0}, \mathbf{I}_k)$ para $\boldsymbol{\eta} = \mathbf{0}$ y la densidad skew-normal multivariada independiente dada por $2^n \phi_n(\mathbf{z}) \Phi_n(D(\boldsymbol{\eta})\mathbf{z})$ para $\Omega = \mathbf{0}$, la matriz cero. Más aún, la densidad condicional de \mathbf{S} dado \mathbf{Z} está dada por

$$f(\mathbf{s}|\mathbf{z}, \boldsymbol{\eta}, \boldsymbol{\Omega}) = \frac{\phi_n(\mathbf{s}; \boldsymbol{\eta}, \boldsymbol{\Omega}) \Phi_n(D(\mathbf{z})\mathbf{s})}{\Phi_n(D(\boldsymbol{\eta})\mathbf{z}; \mathbf{0}, \mathbf{I}_n + D(\mathbf{z})\boldsymbol{\Omega}D(\mathbf{z}))}.$$

Proposición 4.3.2. Supóngase que dado S, el vector aleatorio Z tiene distribución skewnormal $SN_n(s)$, con densidad $2\phi_n(z)\Phi(s^Tz)$, $z \in \mathbb{R}^n$, donde $S = (S_1, \ldots, S_n)^T$ tiene distribución normal multivariada $N_n(\eta, \Omega)$. Entonces, la densidad marginal de $Z = (Z_1, \ldots, Z_n)^T$ es dada por

$$f(\mathbf{z}|\boldsymbol{\eta}, \boldsymbol{\Omega}) = 2\phi_n(\mathbf{z})\Phi\left(\frac{\boldsymbol{\eta}^T\mathbf{z}}{\sqrt{1+\mathbf{z}^T\boldsymbol{\Omega}\mathbf{z}}}\right),$$

la cual contiene a la densidad $\mathcal{N}_n(0, \mathbf{I}_k)$ para $\eta = 0$ y a la densidad skew-normal multivariada dependiente $2\phi_n(\mathbf{z})\Phi(\eta^T\mathbf{z})$ para $\Omega = \mathbf{O}$. Más aún, la densidad condicional de \mathbf{S} dado Z está dada por

$$f(\mathbf{s}|\mathbf{z}, \boldsymbol{\eta}, \boldsymbol{\Omega}) = \phi_n(\mathbf{s}; \boldsymbol{\eta}, \boldsymbol{\Omega}) \frac{\Phi(\mathbf{z}^T \mathbf{s})}{\Phi\left(\frac{\boldsymbol{\eta}^T \mathbf{z}}{\sqrt{1 + \mathbf{z}^T \boldsymbol{\Omega} \mathbf{z}}}\right)}.$$

Algunos casos particulares de la distribución skew-normal-generalizada multivariada se obtienen asumiendo que:

- I) S_1, \ldots, S_n son variables aleatorias independientes con distribución idéntica normal de parámetros η y τ^2 , es decir, $\eta = \eta \mathbf{1}_n$ y $\Omega = \tau^2 \mathbf{I}_n$; o
- II) S_i son variables aleatorias independientes con distribución normal $\mathcal{N}(\eta_i, \tau_i^2)$, para $i=1,\ldots,n$, por lo que $\Omega=\mathrm{diag}\{\tau_1^2,\ldots,\tau_n^2\}$; o
- III) S_1, \ldots, S_n son variables aleatorias normales permutables, lo cual es equivalente a considerar $\eta = \eta \mathbf{1}_n$ y $\Omega = \tau^2 \{ (1 \rho) \mathbf{I}_n + \rho \mathbf{1}_n \mathbf{1}_n^T \}$, donde $\rho \in [0, 1)$.

Nótese que i) es un caso particular de ii) y iii).

Todos los modelos hasta aquí mencionados se derivan asumiendo observaciones skewnormal con parámetros de asimetría diferentes. La situación en la que el parámetro de asimetría es común para todas las observaciones se considera a continuación.

Proposición 4.3.3. Supóngase que dado S, Z_1, \ldots, Z_n son variables aleatorias independientes con distribución idéntica skew-normal SN(s), donde S tiene distribución normal con parámetros $\eta y \tau^2$. Entonces, la densidad marginal de $\mathbf{Z} = (Z_1, \ldots, Z_n)^T$ está dada por

$$f(\mathbf{z}|\boldsymbol{\eta},\tau^2) = 2^n \phi_n(\mathbf{z}) \Phi_n(\boldsymbol{\eta}\mathbf{z};\mathbf{0},\mathbf{I}_n + \tau^2 \mathbf{z}\mathbf{z}^T),$$

y la densidad condicional de S dado ${f Z}$ es dada por

$$f(s|\mathbf{z}, \eta, \tau) = \frac{\phi(s; \eta, \tau^2)\Phi_n(\mathbf{z}s)}{\Phi_n(\eta \mathbf{z}; \mathbf{0}, \mathbf{I}_n + \tau^2 \mathbf{z}\mathbf{z}^T)}.$$

Proposición 4.3.4. Sea \mathbb{Z} dado S=s un vector aleatorio con distribución skew-normal $S\mathcal{N}_n(s\mathbf{1}_n)$, es decir, su densidad es $2\phi_n(\mathbf{z})\Phi(ns\bar{z})$, donde $\bar{z}=n^{-1}\sum_{i=1}^n z_i\,y\,S$ tiene distribución normal $\mathcal{N}(\eta,\tau^2)$. Entonces, la densidad marginal de $\mathbb{Z}=(Z_1,\ldots,Z_n)^T$ está dada por

$$f(\mathbf{z}|\eta,\tau) = 2\phi_n(\mathbf{z})\Phi\left(\frac{n\eta\bar{z}}{\sqrt{1+n^2\tau^2\bar{z}^2}}\right),$$

y la densidad condicional de S dado ${f Z}$ está dada por

$$f(s|\mathbf{z}, \eta, \tau) = \frac{\phi(s; \eta, \tau^2)\Phi(n\bar{z}s)}{\Phi\left(\frac{n\eta\bar{z}}{\sqrt{1+n^2\tau^2\bar{z}^2}}\right)}.$$

Un hecho interesante es que los resultados en las Proposiciones 4.3.3 y 4.3.4 pueden ser obtenidas como casos particulares de las Proposiciones 4.3.1 y 4.3.2, respectivamente, cuando se considera la distribución normal permutable descrita en (iii) con $\rho = 1$ para las variables de asimetría S_1, \ldots, S_n .

En todos los casos anteriores, la distribución skew-normal-generalizada de localización-escala multivariada para \mathbf{Z} puede ser obtenida a través de la transformación lineal afin $\mathbf{Y} = \boldsymbol{\xi} + \omega \mathbf{Z}$, para un vector de localización $\boldsymbol{\xi} \in \mathbb{R}^n$ y una matriz de escala no singular $\omega \in \mathcal{C}_{n \times n}$. Sin embargo, en la Proposición 4.3.1, otra posibilidad es incorporar una matriz de escala ω directamente en la distribución condicional de \mathbf{Z} dado \mathbf{S} . Finalmente, otras familias de distribuciones asimétricas multivariadas pueden ser obtenidas utilizando el mismo procedimiento.

4.4. Inferencia Bayesiana en modelos de regresión skewnormal

La inferencia y problemas relacionados a modelos de regresión han sido estudiados desde el punto de vista Bayesiano bajo el supuesto de que los términos de error son simétricamente distribuidos. La mayoría de la investigación ha sido desarrollada bajo distribuciones multivariadas normales esféricas para el vector de errores. Sin embargo, hay algunas extensiones del modelo lineal normal al modelo lineal esférico (véase, por ejemplo, Arellano-Valle *et al*- [6] y las referencias dentro de este artículo), donde la atención es centrada en la robustez de la teoría normal. Sahu *et al*. [75] implementaron el análisis a posteriori del modelo de regresión considerando distribuciones asimétricas para los términos de error. En esta sección, se considera el análisis Bayesiano del modelo de regresión lineal con observaciones $(Y_i, \mathbf{x}_i), Y_i \in \mathbb{R}, \mathbf{x}_i \in \mathbb{R}^k, i = 1, \dots, n$, cuando los términos de error son independientes con distribución idéntica skew-normal.

Considérese el modelo de regresión lineal

$$Y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \omega \varepsilon_i, \quad i = 1, \dots, n,$$

donde $\varepsilon_1, \ldots, \varepsilon_n$ son errores aleatorios independientes con distribución idéntica skewnormal $\mathcal{SN}(\lambda)$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)^T$, ω y λ son parámetros desconocidos. Esto es, se considera el modelo de regresión lineal skew-normal

$$[Y_i|\beta,\omega^2,\lambda] \stackrel{ind.}{\sim} S\mathcal{N}(\mathbf{x}_i^T\beta,\omega^2,\lambda), \quad i=1,\ldots,n,$$
 (4.4.1)

cuya función de verosimilitud es

$$f(\mathbf{y}|\boldsymbol{\beta}, \omega^2, \lambda) = \frac{2^n}{\omega^n} \phi_n \left(\frac{\mathbf{y} - \mathbf{X}\boldsymbol{\beta}}{\omega} \right) \Phi_n \left(\lambda \left(\frac{\mathbf{y} - \mathbf{X}\boldsymbol{\beta}}{\omega} \right) \right), \tag{4.4.2}$$

donde $\mathbf{y} = (y_1, \dots, y_n)^T \mathbf{y} \mathbf{X}$ es la matriz de $n \times k$ cuyas filas son $\mathbf{x}_1^T, \dots, \mathbf{x}_n^T$

Con el objeto de realizar inferencia a posteriori para $(\beta, \omega^2, \lambda)$, se asume que

$$\lambda \perp \!\!\! \perp (\beta, \omega^2),$$
 (4.4.3)

y se considera

$$\pi(\boldsymbol{\beta}, \omega^2) \propto \frac{1}{\omega^2}$$
 y $\lambda \sim \mathcal{N}(a, b^2)$. (4.4.4)

4.4.1. Función de verosimilitud marginal de $(m{eta},\omega^2)$ cuando $\lambda\sim\mathcal{N}(a,b^2)$

El objetivo principal de esta sección es obtener la verosimilitud marginal de (β, ω^2) bajo las distribuciones a priori antes señaladas. Como consecuencia de este resultado, se muestra que bajo tales especificaciones a priori se obtiene una función predictiva propia, lo cual garantiza que la distribución a posteriori de $(\beta, \omega^2, \lambda)$ es también propia.

Proposición 4.4.1. Considere el modelo de regresión lineal skew-normal en (4.4.1). Entonces, si $\pi(\beta, \omega^2)$ es arbitraria y λ tiene distribución normal $\mathcal{N}(a, b^2)$, la función de verosimilitud marginal de (β, ω^2) es

$$f(\mathbf{y}|\boldsymbol{\beta},\omega^2) = \frac{2^n}{\omega^n} \,\phi_n \left(\frac{\mathbf{y} - \mathbf{X}\boldsymbol{\beta}}{\omega} \right) \Phi_n \left(a \left(\frac{\mathbf{y} - \mathbf{X}\boldsymbol{\beta}}{\omega} \right) \, \middle| \, \mathbf{I}_n + b^2 \left(\frac{\mathbf{y} - \mathbf{X}\boldsymbol{\beta}}{\omega} \right) \left(\frac{\mathbf{y} - \mathbf{X}\boldsymbol{\beta}}{\omega} \right)^T \right).$$

Dem. Sea $Z = \frac{\alpha - a}{b}$ y $\mathbf{t} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})/\omega$. Dado que λ tiene distribución normal $\mathcal{N}(a, b^2)$, se tiene de (4.4.2) que $f(\mathbf{y}|\boldsymbol{\beta}, \omega^2) = \frac{2^n}{\omega^n} \phi_n(\mathbf{t}) E[\Phi_n(\mathbf{b}Z + \mathbf{a})]$, donde $\mathbf{a} = a\mathbf{t}$, $\mathbf{b} = b\mathbf{t}$ y Z

tiene distribución normal estándar. Luego, la demostración sigue del Lema 3.8.1.

Nota 4.4.1. Una verosimilitud marginal alternativa es obtenida cuando en (4.4.1) el parámetro común de asimetría λ es reemplazado por λ_i y estos son independientes con distribución normal $\mathcal{N}(a_i,b_i^2)$, para $i=1,\ldots,n$, con $(\lambda_1,\ldots,\lambda_n)$ independiente de $(\boldsymbol{\beta},\omega^2)$. En este caso, la Proposición 4.3.1 origina $[Y_i|\boldsymbol{\beta},\omega^2] \stackrel{ind.}{\sim} \mathcal{SGN}(\mathbf{x}_i^T\boldsymbol{\beta},\omega^2,a_i,b_i^2), i=1,\ldots,n$, cuya densidad conjunta es

$$f(\mathbf{y}|\boldsymbol{\beta}, \omega^2) = \frac{2^n}{\omega^n} \phi_n \left(\frac{\mathbf{y} - \mathbf{X}\boldsymbol{\beta}}{\omega} \right) \prod_{i=1}^n \Phi \left(\frac{a_i \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\omega} \right)}{\sqrt{1 + b_i^2 \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\omega} \right)^2}} \right).$$

Entonces, la especificación a priori $\lambda_i \stackrel{ind}{\sim} \mathcal{N}(0,b_i^2), i=1,\ldots,n$, es equivalente a considerar el modelo de regresión lineal normal estandar para la distribución condicional de Y_1,\ldots,Y_n dado (β,ω^2) . Un resultado análogo es obtenido cuando $b_i^2\to\infty, i=1,\ldots,n$, lo cual puede ser interpretado como una priori conjunta difusa para $\lambda_1,\ldots,\lambda_n$.

Una consecuencia de la Proposición 4.4.1 es dada en el siguiente corolario, el cual establece que la densidad predictiva y la posteriori de β , ω^2 y λ , es propia cuando se considera una priori normal para λ , y una priori impropia para (β, ω^2) .

Corolario 4.4.2. Si en la Proposición 4.4.1 se considera las especificaciones a priori en (4.4.3) y (4.4.4), entonces la distribución a posteriori de $(\beta, \omega^2, \lambda)$ es propia.

Dem. Por la Proposición 4.4.1, se tiene que bajo las especificaciones a priori antes

señaladas, la densidad predictiva es

$$f(\mathbf{y}) = \int_{\mathbb{R}^{k}} \int_{0}^{\infty} \frac{2^{n}}{\omega^{n+2}} \phi_{n} \left(\frac{\mathbf{y} - \mathbf{X}\boldsymbol{\beta}}{\omega} \right) \times \Phi_{n} \left(a \left(\frac{\mathbf{y} - \mathbf{X}\boldsymbol{\beta}}{\omega} \right) \middle| \mathbf{I}_{n} + b^{2} \left(\frac{\mathbf{y} - \mathbf{X}\boldsymbol{\beta}}{\omega} \right) \left(\frac{\mathbf{y} - \mathbf{X}\boldsymbol{\beta}}{\omega} \right)^{T} \right) d\omega^{2} d\boldsymbol{\beta}$$

$$\leq 2^{n} \int_{\mathbb{R}^{k}} \int_{0}^{\infty} \frac{1}{\omega^{n+2}} \phi_{n} \left(\frac{\mathbf{y} - \mathbf{X}\boldsymbol{\beta}}{\omega} \right) d\omega^{2} d\boldsymbol{\beta}$$

$$< \infty,$$

donde se utiliza el hecho de que la última integral corresponde a la densidad predictiva correspondiente al modelo de regresión lineal normal estándar $[\mathbf{y}|\boldsymbol{\beta},\omega^2] \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta},\omega^2\mathbf{I}_n)$ bajo la priori no informativa usual $\pi(\boldsymbol{\beta},\omega^2) = \omega^{-2}$, la cual es bien sabido que es propia. \square Nota 4.4.2. Una generalización del Corolario 4.4.2 propuesto recientemente por Arellano-Valle et al. [5] establece que cuando se considera $\pi(\boldsymbol{\beta},\omega^2) = \omega^{-2}$ y $\pi(\lambda)$ general y propia, entonces la distribución a posteriori $\pi(\boldsymbol{\beta},\omega^2,\lambda)$ también es propia.

4.4.2. Muestreo de Gibbs para el modelo de regresión skew-normal

En esta sección, se proporcionan las distribuciones condicionales necesarias para implementar el muestreo de Gibbs con el objetivo de obtener el análisis a posteriori del modelo de regresión lineal skew-normal (4.4.1) cuando se consideran las distribuciones a priori (4.4.3) y (4.4.4). Para tal fin, se utiliza una conveniente representación estocástica en (4.2.8), condicionalmente en $S = \lambda$, obteniéndose la siguiente especificación equivalente

de(4.4.1):

$$(i) \quad [Y_i|\boldsymbol{\beta},\omega^2,\lambda,\tau_i] \overset{ind.}{\sim} \mathcal{N}\left(\frac{\lambda\tau_i}{\sqrt{1+\lambda^2}} + \mathbf{x}_i^T\boldsymbol{\beta},\frac{\omega^2}{1+\lambda^2}\right), \ i=1,\ldots,n,$$

(ii)
$$\tau_i \stackrel{i.i.d.}{\sim} \mathcal{HN}(0,1) \text{ y } \tau_i \perp \!\!\! \perp \!\!\! (\boldsymbol{\beta}, \omega^2, \lambda), i = 1, \ldots, n.$$

Entonces, directamente se obtiene las distribuciones condicionales requeridas para implementar el muestreo de Gibbs. De hecho, considerando la transformación $w_i = \psi \tau_i$, $i = 1, \ldots, n$, donde ψ es un nuevo parámetro de escala definido por

$$\psi = \frac{\omega}{\sqrt{1+\lambda^2}},\tag{4.4.5}$$

la representación del modelo anterior puede ser reescrita como

$$[Y_i|\boldsymbol{\beta}, \psi^2, \lambda, w_i] \stackrel{ind.}{\sim} \mathcal{N}(\lambda w_i + \mathbf{x}_i^T \boldsymbol{\beta}, \psi^2), i = 1, \dots, n,$$
 (4.4.6)

$$[w_i|\psi^2] \stackrel{i.i.d.}{\sim} \mathcal{HN}(0,\psi^2) \text{ y } w_i \perp \!\!\! \perp \!\!\! (\boldsymbol{\beta},\lambda), i=1,\ldots,n.$$
 (4.4.7)

Mas aún, es fácil mostrar que de (4.4.3), (4.4.4) y (4.4.5) las distribuciones a priori asociadas a los parámetros β , ψ^2 y λ son tales que

$$\pi(\beta, \psi^2 | \lambda) \propto \frac{1}{\psi^2} \quad \text{y} \quad \lambda \sim N(a, b^2).$$
 (4.4.8)

Por lo tanto, en términos de la nueva parametrización $(\beta_1, \ldots, \beta_k, \psi^2, \lambda, w_1, \ldots, w_n)$, las siguientes distribuciones condicionales son necesarias para implementar el procedimiento de muestreo de Gibbs:

$$[\boldsymbol{\beta}|\psi^2, \lambda, \mathbf{w}, \mathbf{y}]; \quad [\psi^2|\boldsymbol{\beta}, \lambda, \mathbf{w}, \mathbf{y}]; \quad [\lambda|\boldsymbol{\beta}, \psi^2, \mathbf{w}, \mathbf{y}]; \quad [w_i|\boldsymbol{\beta}, \psi^2, \lambda, w_{-i}, \mathbf{y}];$$

donde $\mathbf{w} = (w_1, \dots, w_n)^T$ y para cualquier vector $\mathbf{u} = (u_1, \dots, u_p)^T$, el vector \mathbf{u}_{-i} es definido por $(u_1, \dots, u_{i-1}, u_{i+1}, \dots, u_p)^T$.

Dado que bajo (4.4.8) se garantiza que las distribuciones a posteriori marginales son propias (Corolario 4.4.2), las distribuciones condicionales necesarias para implementar el Muestreo de Gibbs son establecidas en la siguiente proposición, cuya demostración es omitida pues sigue de manipulaciones algebráicas estándares.

Proposición 4.4.3. Considérese la representación condicional (4.4.6)-(4.4.7) del modelo de regresión lineal skew-normal (4.4.1), con distribuciones a priori (4.4.8). Entonces,

$$\begin{aligned} & [\boldsymbol{\beta}|\psi^2, \lambda, \mathbf{w}, \mathbf{y}] &\sim & \mathcal{N}_k \left(\hat{\boldsymbol{\beta}}(\mathbf{y}) - \lambda \hat{\boldsymbol{\beta}}(\mathbf{w}), \psi^2(\mathbf{X}^T \mathbf{X})^{-1} \right), \\ & [\mathbf{w}|\boldsymbol{\beta}, \psi^2, \lambda, \mathbf{y}] &\sim & \mathcal{T} \mathcal{N}_n \left(\mathbf{0}; \frac{\lambda}{1 + \lambda^2} \boldsymbol{\epsilon}, \frac{\psi^2}{1 + \lambda^2} \mathbf{I}_n \right), \\ & [\psi^2|\boldsymbol{\beta}, \lambda, \mathbf{w}, \mathbf{y}] &\sim & \mathcal{G} \mathcal{I} \left(n, \frac{\|\boldsymbol{\epsilon} - \lambda \mathbf{w}\|^2 + \|\mathbf{w}\|^2}{2} \right), \\ & [\lambda|\boldsymbol{\beta}, \psi^2, \mathbf{w}, \mathbf{y}] &\sim & \mathcal{N} \left(\frac{b^2 \mathbf{w}^T \boldsymbol{\epsilon} + a \psi^2}{b^2 \|\mathbf{w}\|^2 + \psi^2}, \frac{b^2 \psi^2}{b^2 \|\mathbf{w}\|^2 + \psi^2} \right), \end{aligned}$$

donde $\hat{\boldsymbol{\beta}}(\mathbf{z}) = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{z}$, $\epsilon = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$, $\mathcal{GI}(\alpha, \gamma)$ es la distribución Gamma Inversa y $\mathcal{TN}_n(\mathbf{c}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denota la distribución $\mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ truncada sobre \mathbf{c} .

En la Sección 5 se muestra una aplicación basada en los resultados de la Proposición 4.4.3.

4.5. Inferencia Bayesiana para mezclas de asimetría en modelos de regresión lineal

4.5.1. Distribución skew-normal-generalizada

Considerése el modelo de regresión lineal skew-normal-generalizado

$$[Y_i|\boldsymbol{\beta}, \omega^2, \alpha_1, \alpha_2] \stackrel{ind.}{\sim} SGN(\mathbf{x}_i^T \boldsymbol{\beta}, \omega^2, \alpha_1, \alpha_2), \quad i = 1, \dots, n,$$
 (4.5.1)

cuya función de verosimilitud es

$$f(\mathbf{y}|\boldsymbol{\beta}, \omega^2, \alpha_1, \alpha_2) = (2/\omega)^n \ \phi_n(\mathbf{z}) \Phi_n(\alpha_1 \mathbf{z}; I_n + \alpha_2 D^2(\mathbf{z})),$$

donde $\mathbf{z} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})/\omega$. De los resultados dados en la Sección 4.2.4 (ver, por ejemplo, (4.2.8)), (4.5.1) se puede escribir como

$$[Y_i|\boldsymbol{\beta}, \omega^2, \lambda_i] \stackrel{ind.}{\sim} \mathcal{SN}(\mathbf{x}_i^T \boldsymbol{\beta}, \omega^2, \lambda_i), \quad i = 1, \dots, n,$$
 (4.5.2)

$$[\lambda_i | \alpha_1, \alpha_2] \stackrel{i.i.d.}{\sim} \mathcal{N}(\alpha_1, \alpha_2), \quad i = 1, \dots, n.$$
 (4.5.3)

Por lo tanto, haciendo $\psi_i^2 = \omega^2/(1+\lambda_i^2)$, $i=1,\ldots,n$, y asumiendo que $[\xi_i \perp \!\!\! \perp \lambda_i | \psi_i^2]$, una conveniente representación jerárquica de (4.5.1) es

$$[Y_i|\boldsymbol{\beta}, \omega^2, \xi_i, \lambda_i] \stackrel{ind.}{\sim} \mathcal{N}(\mathbf{x}_i^T \boldsymbol{\beta} + \lambda_i \xi_i, \psi_i^2), \quad i = 1, \dots, n,$$
(4.5.4)

$$[\xi_i|\psi_i^2] \qquad \stackrel{ind.}{\sim} \mathcal{HN}(0,\psi_i^2), \quad i = 1,\dots, n, \tag{4.5.5}$$

$$[\psi_i^2|\lambda_i] \stackrel{ind.}{\sim} G_i, \qquad [\lambda_i|\alpha_1,\alpha_2] \stackrel{i.i.d.}{\sim} \mathcal{N}(\alpha_1,\alpha_2), \quad i = 1,\dots, n,$$
 (4.5.6)

donde G_i son las distribuciones condicionales en λ_i , las cuales son determinadas por la distribución a priori de (β, ω^2) .

Considérese además las siguientes especificaciones a priori:

$$\pi(\boldsymbol{\beta}, \omega^2, \alpha_1, \alpha_2) \propto \frac{1}{\omega^2} \pi(\alpha_1 | \alpha_2) \pi(\alpha_2),$$
 (4.5.7)

con $\alpha_1 \sim \mathcal{N}(a_1, \alpha_2)$ y $\alpha_2 \sim \mathcal{IG}(a_2/2, b_2/2)$. Nótese que para $a_1 = 0$ se sigue de (4.5.2)-(4.5.3) y considerando el resultado de la Sección 3 (véase Proposición 4.3.1) que $[Y_i|\beta,\omega^2] \stackrel{ind.}{\sim} \mathcal{N}(\mathbf{x}_i^T\beta,\omega^2), \ i=1,\ldots,n$. Entonces, en esta situación las inferencias a posteriori de (β,ω^2) se basan en el modelo de regresión normal estándar. Más aún, dado que se considera una priori impropia $\pi(\beta,\omega^2) = \omega^{-2}$, se tiene entonces en (4.5.6) que las distribuciones G_i son tales que

$$\pi(\boldsymbol{\beta}, \psi_i^2 | \lambda_i) \propto \frac{1}{\psi_i^2}, \quad i = 1, \dots, n.$$
 (4.5.8)

Las distribuciones condicionales necesarias para implementar el Muestreo de Gibbs son establecidas en la siguiente proposición, cuya demostración es omitida.

Proposición 4.5.1. Considérese la representación condicional (4.5.4)-(4.5.6) del modelo de regresión lineal skew-normal-generalizado (4.5.1), con distribuciones a priori para los

parámetros del modelo dadas en (4.5.7)-(4.5.8). Entonces,

$$\begin{split} [\boldsymbol{\beta}|\boldsymbol{\psi},\alpha_{1},\alpha_{2},\boldsymbol{\xi},\boldsymbol{\lambda},\mathbf{y}] &\sim \mathcal{N}_{k}\left(\hat{\boldsymbol{\beta}}(\mathbf{y})-\hat{\boldsymbol{\beta}}(D(\boldsymbol{\lambda})\boldsymbol{\xi}),(\mathbf{X}^{T}D(\boldsymbol{\psi})^{-1}\mathbf{X})^{-1}\right), \\ [\boldsymbol{\xi}|\boldsymbol{\beta},\boldsymbol{\psi},\alpha_{1},\alpha_{2},\boldsymbol{\lambda},\mathbf{y}] &\sim \mathcal{T}\mathcal{N}_{n}\left(\mathbf{0};[I_{n}+D(\boldsymbol{\lambda})D(\boldsymbol{\lambda})]^{-1}D(\boldsymbol{\lambda})\boldsymbol{\epsilon},[I_{n}+D(\boldsymbol{\lambda})D(\boldsymbol{\lambda})]^{-1}D(\boldsymbol{\psi})\right), \\ [\boldsymbol{\psi}|\boldsymbol{\beta},\alpha_{1},\alpha_{2},\boldsymbol{\xi},\boldsymbol{\lambda},\mathbf{y}] &\sim \prod_{i=1}^{n}\mathcal{G}\mathcal{I}\left(1,\frac{(\epsilon_{i}-\lambda_{i}\xi_{i})^{2}+\xi_{i}^{2}}{2}\right), \\ [\boldsymbol{\lambda}|\boldsymbol{\beta},\boldsymbol{\psi},\alpha_{1},\alpha_{2},\boldsymbol{\xi},\mathbf{y}] &\sim \mathcal{N}_{n}\left(D(\boldsymbol{\tau})[D(\boldsymbol{\xi})D(\boldsymbol{\psi})^{-1}\boldsymbol{\epsilon}+\alpha_{1}\alpha_{2}^{-1}I_{n}\mathbf{1}_{n}],D(\boldsymbol{\tau})\right), \\ [\alpha_{1}|\boldsymbol{\beta},\boldsymbol{\psi},\alpha_{2},\boldsymbol{\xi},\boldsymbol{\lambda},\mathbf{y}] &\sim \mathcal{N}\left(\frac{\boldsymbol{\lambda}^{T}\mathbf{1}_{n}+a_{1}}{n+1},\frac{\alpha_{2}}{n+1}\right), \\ [\alpha_{2}|\boldsymbol{\beta},\boldsymbol{\psi},\alpha_{1},\boldsymbol{\xi},\boldsymbol{\lambda},\mathbf{y}] &\sim \mathcal{G}\mathcal{I}\left(\frac{a_{2}+n+1}{2},\frac{||\boldsymbol{\lambda}-\alpha_{1}\mathbf{1}_{n}||^{2}+(\alpha_{1}-a_{1})^{2}+b_{2}}{2}\right), \\ donde\ \hat{\boldsymbol{\beta}}(\mathbf{z}) &= (\mathbf{X}^{T}D(\boldsymbol{\psi})^{-1}\mathbf{X})^{-1}\mathbf{X}^{T}D(\boldsymbol{\psi})^{-1}\mathbf{z},\ \boldsymbol{\epsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta},\ \boldsymbol{\psi} = (\boldsymbol{\psi}_{1}^{2},\ldots,\boldsymbol{\psi}_{n}^{2})^{T},\ \boldsymbol{\tau} = \left(\frac{\alpha_{2}\boldsymbol{\psi}_{1}^{2}}{\alpha_{2}\boldsymbol{\xi}_{1}^{2}+\boldsymbol{\psi}_{1}^{2}},\ldots,\frac{\alpha_{2}\boldsymbol{\psi}_{n}^{2}}{\alpha_{2}\boldsymbol{\xi}_{n}^{2}+\boldsymbol{\psi}_{n}^{2}}\right)^{T}. \end{split}$$

4.5.2. Distribución skew-normal-t

Sea el siguiente resultado.

Proposición 4.5.2. Si Z dado S tiene distribución skew-normal $SN(\alpha_1\sqrt{s})$, donde S tiene distribución gamma $G(\nu/2,\nu/2)$, entonces marginalmente, la función de densidad de Z está dada por

$$f_Z(z) = 2\phi(z)T(\alpha_1 z; \nu).$$

Dem. La función de densidad marginal de Z está dada por

$$f_{Z}(z) = 2\phi(z) \int_{0}^{\infty} \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)} s^{\nu/2-1} e^{-\nu s/2} \Phi(\alpha_{1}\sqrt{s}z) ds$$
$$= 2\phi(z) \int_{0}^{\infty} \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)} s^{\nu/2-1} e^{-\nu s/2} \int_{t < z} \alpha_{1} \sqrt{s} \phi(\alpha_{1}\sqrt{s}t) dt ds.$$

Haciendo w = t - z, se tiene

$$f_Z(z) = 2\phi(z) \int_0^\infty \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)} s^{\nu/2 - 1} e^{-\nu s/2} \int_{w \le 0} \alpha_1 \sqrt{s} \phi(\alpha_1 \sqrt{s}(w + z)) dw ds, \quad (4.5.9)$$

Aplicando el Teorema de Fubini y luego de algunas manipulaciones algebráicas, se obtiene de (4.5.9) que

$$f_Z(z) = 2\phi(z) \int_{w<0} \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)} \frac{\alpha_1}{\sqrt{\pi\nu}} \left[1 + \frac{(\alpha_1(w+z))^2}{\nu} \right]^{-(\nu+1)/2} dw,$$

lo cual concluye la demostración.

Usando la Proposición 4.5.2, se puede definir el modelo de regresión lineal skew-normal-t (\mathcal{SNT})

$$[Y_i|\boldsymbol{\beta}, \omega^2, \alpha_1, \nu] \stackrel{ind.}{\sim} \mathcal{SNT}(\mathbf{x}_i^T \boldsymbol{\beta}, \omega^2, \alpha_1, \nu), \quad i = 1, \dots, n,$$
(4.5.10)

cuya función de verosimilitud está dada por

$$f(\mathbf{y}|\boldsymbol{\beta},\omega^2,\alpha_1,\nu) = (2/\omega)^n \ \phi_n(\mathbf{z}) \prod_{i=1}^n T(\alpha_1 z_i;\nu), \tag{4.5.11}$$

donde $z = (y - X\beta)/\omega$. De manera análoga al caso del modelo skew-normal-generalizado, (4.5.11) se puede reescribir como

$$[Y_i|\boldsymbol{\beta},\omega^2,\lambda_i,\alpha_1] \overset{ind.}{\sim} \mathcal{SN}(\mathbf{x}_i^T\boldsymbol{\beta},\omega^2,\alpha_1\sqrt{\lambda_i}), \quad i=1,\ldots,n,$$
$$[\lambda_i|\nu] \overset{i.i.d.}{\sim} \mathcal{G}(\nu/2,\nu/2), \quad i=1,\ldots,n.$$

Entonces, haciendo $\psi_i^2=\omega^2/(1+\alpha_1^2\lambda_i),\,i=1,\ldots,n,$ y asumiendo $[\xi_i\!\perp\!\!\!\perp (\lambda_i,\alpha_1)|\psi_i^2]$ se

propone la siguiente representación jerárquica de (4.5.11):

$$[Y_i|\boldsymbol{\beta},\omega^2,\alpha_1,\xi_i,\lambda_i] \stackrel{ind.}{\sim} \mathcal{N}(\mathbf{x}_i^T\boldsymbol{\beta} + \alpha_1\sqrt{\lambda_i}\xi_i,\psi_i^2), \quad i=1,\ldots,n, \quad (4.5.12)$$

$$[\xi_i|\psi_i^2] \qquad \stackrel{ind.}{\sim} \mathcal{HN}(0,\psi_i^2), \quad i = 1,\dots, n, \tag{4.5.13}$$

$$[\psi_i^2|\alpha_1,\lambda_i] \stackrel{ind.}{\sim} G_i, \qquad [\lambda_i|\nu] \stackrel{i.i.d.}{\sim} \mathcal{G}(\nu/2,\nu/2), \quad i=1,\ldots,n, \tag{4.5.14}$$

donde G_i es la distribución condicional en (α_1, λ_i) y es determinada por la distribución a priori de (β, ω^2) . Considerando el mismo esquema para las distribuciones a priori que en el caso del modelo skew-normal-generalizado con

$$\alpha_1 \sim \mathcal{N}(a_1, b_1^2), \quad \text{y} \quad \nu \sim Exponencial(c_1/2), \quad (4.5.15)$$

se obtienen las siguientes distribuciones condicionales.

Proposición 4.5.3. Considérese la representación condicional (4.5.12)-(4.5.14) del modelo de regresión lineal skew-normal-t (4.5.10), con distribuciones a priori (4.5.15). Entonces,

$$\begin{split} [\boldsymbol{\beta}|\boldsymbol{\psi},\alpha_{1},\boldsymbol{\nu},\boldsymbol{\xi},\boldsymbol{\lambda},\mathbf{y}] &\sim \mathcal{N}_{k}\left(\hat{\boldsymbol{\beta}}(\mathbf{y})-\hat{\boldsymbol{\beta}}(\alpha_{1}\boldsymbol{\kappa}),(\mathbf{X}^{T}D(\boldsymbol{\psi})^{-1}\mathbf{X})^{-1}\right), \\ [\boldsymbol{\xi}|\boldsymbol{\beta},\boldsymbol{\psi},\alpha_{1},\boldsymbol{\nu},\boldsymbol{\lambda},\mathbf{y}] &\sim \mathcal{T}\mathcal{N}_{n}\left(\mathbf{0};\alpha_{1}[I_{n}+\alpha_{1}^{2}D(\boldsymbol{\lambda})]^{-1}D(\boldsymbol{\lambda})^{1/2}\boldsymbol{\epsilon},[I_{n}+\alpha_{1}^{2}D(\boldsymbol{\lambda})]^{-1}D(\boldsymbol{\psi})\right), \\ [\boldsymbol{\psi}|\boldsymbol{\beta},\alpha_{1},\boldsymbol{\nu},\boldsymbol{\xi},\boldsymbol{\lambda},\mathbf{y}] &\sim \prod_{i=1}^{n}IG\left(1,\frac{(\epsilon_{i}-\alpha_{1}\kappa_{i})^{2}+\xi_{i}^{2}}{2}\right), \\ [\boldsymbol{\lambda}|\boldsymbol{\beta},\boldsymbol{\psi},\alpha_{1},\boldsymbol{\nu},\boldsymbol{\xi},\mathbf{y}] &\sim \prod_{i=1}^{n}\mathcal{G}\left(\frac{\boldsymbol{\nu}}{2},\frac{\alpha_{1}^{2}\xi_{i}^{2}+\boldsymbol{\psi}_{i}^{2}\boldsymbol{\nu}}{2\boldsymbol{\psi}_{i}^{2}}\right)\exp\left\{\frac{\alpha_{1}\kappa_{i}\epsilon_{i}}{\boldsymbol{\psi}_{i}^{2}}\right\}, \\ [\alpha_{1}|\boldsymbol{\beta},\boldsymbol{\psi},\boldsymbol{\nu},\boldsymbol{\xi},\boldsymbol{\lambda},\mathbf{y}] &\sim \mathcal{N}\left(\frac{a_{1}+b_{1}^{2}\boldsymbol{\kappa}^{T}D(\boldsymbol{\psi})^{-1}\boldsymbol{\epsilon}}{1+b_{1}^{2}\boldsymbol{\kappa}^{T}D(\boldsymbol{\psi})^{-1}\boldsymbol{\kappa}},\frac{b_{1}^{2}}{1+b_{1}^{2}\boldsymbol{\kappa}^{T}D(\boldsymbol{\psi})^{-1}\boldsymbol{\kappa}}\right), \\ [\boldsymbol{\nu}|\boldsymbol{\beta},\boldsymbol{\psi},\alpha_{1},\boldsymbol{\xi},\boldsymbol{\lambda},\mathbf{y}] &\sim \exp\left\{-\boldsymbol{\nu}\left(\frac{\boldsymbol{\lambda}^{T}\mathbf{1}_{n}+c_{1}}{2}\right)\right\}\prod_{i=1}^{n}\boldsymbol{\lambda}_{i}^{\boldsymbol{\nu}/2-1}, \\ donde\ \hat{\boldsymbol{\beta}}(\mathbf{z}) &= (\mathbf{X}^{T}D(\boldsymbol{\psi})^{-1}\mathbf{X})^{-1}\mathbf{X}^{T}D(\boldsymbol{\psi})^{-1}\mathbf{z},\ \boldsymbol{\epsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta},\ \boldsymbol{\psi} = (\boldsymbol{\psi}_{1}^{2},\ldots,\boldsymbol{\psi}_{n}^{2})^{T}\ \boldsymbol{y}\ \boldsymbol{\kappa} = (\boldsymbol{\lambda}_{1}^{1/2}\boldsymbol{\xi}_{1},\ldots,\boldsymbol{\lambda}_{n}^{1/2}\boldsymbol{\xi}_{n})^{T}. \end{split}$$

4.5.3. Distribución skew-t-normal

Sea el siguiente resultado.

Proposición 4.5.4. Si Z dado S tiene distribución skew-normal $SN(0, s^{-1}, \alpha_1 s^{-1/2})$, donde S distribuye gamma $G(\nu/2, \nu/2)$, entonces la densidad marginal de Z está dada por

$$f_Z(z) = 2t(z; \nu)\Phi(\alpha_1 z).$$

Dem. La demostración es directa utilizando el bien conocido resultado sobre la distribución predictiva del modelo Bayesiano normal-gamma (véase Bernardo y Smith [25]), es decir,

$$f_Z(z) = 2\Phi(\alpha_1 z) \int_0^\infty \mathcal{N}(z; 0, s^{-1}) Gamma(s; \nu/2, \nu/2) ds.$$

Usando la Proposición 4.5.4, el modelo regresión lineal skew-t-normal (STN) es definido como sigue:

$$[Y_i|\boldsymbol{\beta}, \sigma^2, \alpha_1, \nu] \stackrel{ind.}{\sim} \mathcal{STN}(\mathbf{x}_i^T \boldsymbol{\beta}, \omega^2, \alpha_1, \nu), \quad i = 1, \dots, n,$$
(4.5.16)

con función de verosimilitud dada por

$$f(\mathbf{y}|\boldsymbol{\beta}, \omega^2, \alpha_1, \nu) = (2/\sigma)^n \left\{ \prod_{i=1}^n t(z_i; \nu) \right\} \Phi_n(\alpha_1 \mathbf{z}), \tag{4.5.17}$$

y z = $(y - X\beta)/\omega$. El modelo (4.5.17) puede además ser reescrito como

$$[Y_i|\boldsymbol{\beta}, \sigma^2, \lambda_i, \alpha_1] \stackrel{ind.}{\sim} \mathcal{SN}(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma^2 \lambda_i^{-1}, \alpha_1 \lambda_i^{-1/2}), \quad i = 1, \dots, n,$$
$$[\lambda_i|\nu] \stackrel{iid}{\sim} \mathcal{G}(\nu/2, \nu/2), \quad i = 1, \dots, n,$$

Haciendo $\psi_i^2 = \omega^2/(\lambda_i + \alpha_1^2)$, i = 1, ..., n, y asumiendo $[\xi_i \perp \!\!\! \perp (\lambda_i, \alpha_1) | \psi_i^2]$ se propone la siguiente representación jerárquica de (4.5.17):

$$[Y_i|\boldsymbol{\beta}, \sigma^2, \alpha_1, \xi_i, \lambda_i] \stackrel{ind.}{\sim} \mathcal{N}(\mathbf{x}_i^T \boldsymbol{\beta} + \alpha_1 \xi_i, \psi_i^2), \quad i = 1, \dots, n,$$
 (4.5.18)

$$[\xi_i|\psi_i^2] \stackrel{ind.}{\sim} \mathcal{HN}(0,\psi_i^2), \quad i=1,\ldots,n,$$
(4.5.19)

$$[\psi_i^2|\alpha_1,\lambda_i]$$
 $\stackrel{ind.}{\sim} G_i, \quad [\lambda_i|\nu] \stackrel{iid}{\sim} \mathcal{G}(\nu/2,\nu/2), \quad i=1,\ldots,n,.$ (4.5.20)

Nuevamente, G_i es la distribución condicional en (α_1, λ_i) y es determinada por la distribución a priori de (β, ω^2) . El esquema de prioris para este modelo es el mismo que para el modelo de regresión skew-normal-t. Nótese que este modelo de regresión considera una mezcla en el parámetro de escala y asimetría del modelo skew-normal.

Proposición 4.5.5. Considérese la representación condicional (4.5.18)-(4.5.20) del modelo de regresión lineal skew-t-normal (4.5.16), con distribuciones a priori (4.5.15). Entonces,

$$[\boldsymbol{\beta}|\boldsymbol{\psi},\alpha_{1},\nu,\boldsymbol{\xi},\boldsymbol{\lambda},\mathbf{y}] \sim \mathcal{N}_{k}\left(\hat{\boldsymbol{\beta}}(\mathbf{y})-\alpha_{1}\hat{\boldsymbol{\beta}}(\boldsymbol{\xi}),(\mathbf{X}^{T}D(\boldsymbol{\psi})^{-1}\mathbf{X})^{-1}\right),$$

$$[\boldsymbol{\xi}|\boldsymbol{\beta},\boldsymbol{\psi},\alpha_{1},\nu,\boldsymbol{\lambda},\mathbf{y}] \sim \mathcal{T}\mathcal{N}_{n}\left(\mathbf{0};\frac{\alpha_{1}}{1+\alpha_{1}^{2}}\boldsymbol{\epsilon},\frac{1}{1+\alpha_{1}^{2}}D(\boldsymbol{\psi})\right),$$

$$[\boldsymbol{\psi}|\boldsymbol{\beta},\alpha_{1},\nu,\boldsymbol{\xi},\boldsymbol{\lambda},\mathbf{y}] \sim \prod_{i=1}^{n}\mathcal{I}\mathcal{G}\left(1,\frac{(\epsilon_{i}-\alpha_{1}\xi_{i})^{2}+\xi_{i}^{2}}{2}\right),$$

$$[\boldsymbol{\lambda}|\boldsymbol{\beta},\boldsymbol{\psi},\alpha_{1},\nu,\boldsymbol{\xi},\mathbf{y}] \sim \prod_{i=1}^{n}\mathcal{G}\left(\frac{\nu}{2},\frac{\nu}{2}\right),$$

$$[\alpha_{1}|\boldsymbol{\beta},\boldsymbol{\psi},\nu,\boldsymbol{\xi},\boldsymbol{\lambda},\mathbf{y}] \sim \mathcal{N}\left(\frac{a_{1}+b_{1}^{2}\boldsymbol{\xi}^{T}D(\boldsymbol{\psi})^{-1}\boldsymbol{\epsilon}}{1+b_{1}^{2}\boldsymbol{\xi}^{T}D(\boldsymbol{\psi})^{-1}\boldsymbol{\xi}},\frac{b_{1}^{2}}{1+b_{1}^{2}\boldsymbol{\xi}^{T}D(\boldsymbol{\psi})^{-1}\boldsymbol{\xi}}\right),$$

$$[\nu|\boldsymbol{\beta},\boldsymbol{\psi},\alpha_{1},\boldsymbol{\xi},\boldsymbol{\lambda},\mathbf{y}] \sim \exp\left\{-\nu\left(\frac{\boldsymbol{\lambda}^{T}\mathbf{1}_{n}+c_{1}}{2}\right)\right\}\prod_{i=1}^{n}\lambda_{i}^{\nu/2-1},$$

donde $\hat{\boldsymbol{\beta}}(\mathbf{z}) = (\mathbf{X}^T D(\boldsymbol{\psi})^{-1} \mathbf{X})^{-1} \mathbf{X}^T D(\boldsymbol{\psi})^{-1} \mathbf{z}, \ \epsilon = \mathbf{y} - \mathbf{X} \boldsymbol{\beta} \ \mathbf{y} \ \boldsymbol{\psi} = (\psi_1^2, \dots, \psi_n^2)^T$

4.6. Aplicación a los datos de atletas australianos

Con el fin de ilustrar los resultados obtenidos en las secciones previas, se considera un set de datos de Cook y Weisberg [30] sobre características de atletas australianos provenientes del Instituto Australiano del Deporte (Australian Institute of Sport, AIS). Específicamente, se consideran las variables índice de masa corporal magra (Imc), altura (Al) y peso (Pe) asociadas a n=102 atletas austrialianos de género masculino. El Cuadro 4.1 presenta un resumen de las estadísticas descriptivas básicas para estas variables. Para este

Cuadro 4.1: Estadísticas descriptivas para el conjunto de datos de atletas australianos: media muestral \bar{z} , desviación estándar muestral s, y coeficientes muestrales de asimetría y curtosis $\sqrt{b_1}$ y b_2 , respectivamente.

Variable	\overline{z}	s	$\sqrt{b_1}$	b_2
Imc	74,66	9,89	0,28	0,71
Al	185,50	7,90	0,07	0,06
Pe	82,52	12,40	0,40	0,49

conjunto de datos, se propone estudiar el modelo de regresión lineal dado por

$$Imc_i = \beta_1 A l_i + \beta_2 P e_i + \omega \varepsilon_i, \quad i = 1, \dots, 102,$$

asumiendo que los términos de error ε_i son independientes y con distribución idéntica $SN(0, 1/v_i, \lambda_i), i = 1, ..., n$; de tal manera que,

- 1. para el modelo $\mathcal{SGN},$ $\lambda_i \sim \mathcal{N}(\alpha_1,\alpha_2)$ y $v_i=1,$ $i=1,\ldots,n.$
- 2. para el modelo $\mathcal{SNT}, \lambda_i = \alpha_1 \sqrt{w_i} \text{ con } w_i \sim \mathcal{G}(\nu/2, \nu/2), \text{ y } v_i = 1 \ i = 1, \dots, n.$

3. para el modelo STN, $\lambda_i = \alpha_1 \sqrt{w_i}$ y $v_i = w_i$ con $w_i \sim \mathcal{G}(\nu/2, \nu/2)$.

Con el objetivo de elicitar las distribuciones a priori para los modelos de mezclas de asimetría, primero se realizó el ajuste, desde el punto de vista Bayesiano del modelo de regresión skew-normal. Se consideró una distribución a priori impropia para los parámetros (β,ω^2) y una distribución normal centrada en 0 con varianza 10 como distribución a priori para el parámetro de asimetría. Esta especificación a priori refleja la creencia en favor de la normalidad de los datos. Utilizando la Proposición 4.4.3, se implementó el muestro de Gibbs. Para este algoritmo, se consideró 3 cadenas paralelas de largo 50,000 descartando las primeras 25,000 iteraciones como periodo de "quema" y con un salto entre iteraciones de tamaño 10 para evitar la autocorrelación. Las estimaciones a posteriori de los parámetros de este modelo se muestran en el Cuadro 4.2.

Cuadro 4.2: Estimaciones Bayesianas para el conjunto de datos de atletas australianos bajo el modelo de regresión lineal skew-normal.

	\mathcal{SN}					
Parametro	Media	SD	$P_{2,5}$	$P_{97,5}$		
eta_1	0,06	0,01	0,04	0,08		
eta_2	0,81	0,03	0,76	0,86		
σ	3,49	0,31	2,90	4,16		
α	-3,97	1,13	-6,51	-2,15		

En la Figura 4.1 se presenta el QQ-plot normal para los residuales estandarizados bajo el modelo de regresión lineal normal. Este plot, muestra fuerte evidencia en favor de la nonormalidad de los datos. Además, de acuerdo a las estimaciones Bayesianas reportadas en

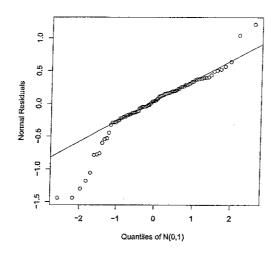


Figura 4.1: QQ-plot normal para los residuales estandarizados bajo el modelo de regresión lineal normal.

el Cuadro 4.2, los datos de atletas australianos muestran fuerte evidencia a favor de un comportamiento asimétrico. Esto se concluye al observar la media a posteriori del parámetro α , la cual es igual a -3,97 con un intervalo de credibilidad dado por [-6,51,-2,15]. También se constata que el el indice de masa magra esta relacionado linealmente y positivamente al peso y la altura de cada individuo. Usando esta información, se procedió al ajuste de los diferentes modelos de regresión de mezclas de asimetría, bajo la perspectiva Bayesiana. Para cada modelo, se consideró prioris impropias para los parámetros de regresión y de escala. Además, y con base en las estimaciones a posteriori del modelo de regresión skew-normal, se adoptó una priori normal centrada en -4 con varianza 1, para α_1 . En el caso del modelo SGN, y para el parámetro α_2 se consideró una priori $\mathcal{IG}(3,2)$. Esta elección a priori significa fijar, la media y la varianza a priori de este parámetro en 1. En el caso de los modelos

 \mathcal{SNT} y \mathcal{STN} , se consideró una priori exponencial Exp(0,10) truncada por encima de 2 para el parámetro ν .

La implementación del enfoque Bayesiano se realizó utilizando las condicionales completas dadas en las Proposiciones 4.5.1,4.5.3 y 4.5.5, através del muestreo de Gibbs utilizando R. Sin embargo, la conveniente representación jerárquica de los modelos SGN, SNT y STN permite además la implementación de éstos usando WinBUGS. En ambos casos, los resultados obtenidos son similares. El esquema MCMC utilizado para el ajuste de estos modelos es similar al del modelo de regresión skew-normal. Los resultados son reportados en el Cuadro 4.3. Para la comparación de los modelos, se utilizó el criterio de información de devianza (DIC): véase Spiegelhalter *et al.* [81]. Adicionalmente, se adoptó el criterio de validación cruzada, calculando pseudo factores de Bayes (PFB) (véase Geisser y Eddy [46]; Gelfand y Dey [47]), utilizando la pseudo log-verosimilitud marginal (PLVM) tal como sugiere Chen *et al.* [28].

De acuerdo a los valores del DIC reportados en el Cuadro 4.3, se concluye que el modelo skew-normal-generalizado es mejor que los otros modelos de mezclas de asimetría para
este conjunto de datos. Además, y usando la información proporcionada por el PLVM, se
comparó la especificación Bayesiana correspondiente al modelo SGN contra la especificación Bayesiana de los modelos SNT y STN através de PFB. Los valores respectivos de
las PLVM están dadas en el Cuadro 4.3 y originan un $2 \log PFB$ de 59,64 (modelo SGNversus modelo SNT) y de 50,76 (modelo SGN versus modelo STN) lo cual es interpretado como una fuerte evidencia a favor de la especificación Bayesiana correspondiente al
modelo skew-normal-generalizado. Adicionalmente, a pesar de que las estimaciones a posteriori de α_1 y ν son similiares en los modelos SNT y STN, el efecto en los coeficientes
de asimetría y curtosis es muy diferente; para más detalles véase Nadarajah y Kotz [67] y

Gómez et al. [50].

Finalmente, se concluye además que el indice de masa magra tiene una relación lineal y positiva con el peso y la altura del individuo.

4.7. Resumen y conclusiones

Los modelos estadísticos construidos a partir de mezclas de distribuciones skew-simétricas permiten generar una subclase de distribuciones que presenta mayor flexibilidad tanto en asimetría cuanto en curtosis, que otras distribuciones skew-simétricas.

Ejemplos de algunas distribuciones miembros de la clase antes mencionada son la distribución skew-normal-generalizada, skew-normal-t y skew-t-normal. Todas ellas se generan como mezclas de la distribución skew-normal en el parámetro de asimetría y bajo diferentes especificaciones de la distribución de mezclas.

Estos modelos skew-simétricos presentan propiedades interesantes, en mucho casos derivadas del supuesto de simetría en las funciones de distribución de la distribución de mezclas. Por ejemplo, muchas variantes de la versión multivariada del modelo estadístico skew-normal-generalizado (estudiado en detalle por Arellano-Valle *et l.* [10]) son obtenidas al considerar distribuciones normales como distribución de mezcla para el parámetro de asimetría.

La representación estocástica de la distribución skew-normal basada en la convolución de una variable aleatoria con distribución normal y otra con distribución normal truncada, además de la estructura de mezclas en el parámetro de asimetría, permite, desde el punto de vista Bayesiano, formular una conveniente estructura jerárquica para los modelos de mezclas de asimetría.

Los resultados obtenidos en el presente capítulo pueden ser aplicados a los modelos de

regresión, asumiendo que los términos de error provienen de distribuciones generadas por una mezcla de distribuciones skew-normal en el parámetro de asimetría. Específicamente, y de acuerdo a las estimaciones Bayesianas obtenidas con el conjunto de datos de atletas australianos, el modelo skew-normal-generalizado presenta un mejor ajuste que el resto de modelos de mezclas de asimetría.

4.8. Agradecimientos

Quiero expresar mi gratitud a Pilar Iglesias (Pontificia Universidad Católica de Chile) por sus comentarios y sugerencias en las versiones iniciales de este capítulo. También
a Reinaldo Arellano-Valle (Pontificia Universidad Católica de Chile), Marc Genton (University of Geneva) y a Héctor Gómez (Universidad de Concepción) por los comentarios,
sugerencias y aportes realizados en este capítulo. Además quiero agradecer los comentarios
y sugerencias de un referí anónimo.

Cuadro 4.3: Estimaciones Bayesianas para el conjunto de datos de atletas australianos bajo los modelos de regresión skew-normal-generalizado (\mathcal{SGN}), skew-normal-t (\mathcal{SNT}), y skew-t-normal (\mathcal{STN}).

`								
\mathcal{SGN}	β_1	β_2	σ	α_1	α_2	ν		
Media	0,05	0,81	1,14	-1,01	0,99	_		
SD	0,01	0,03	$0,\!16$	0,41	0,32	*****		
$P_{2,5}$	0,03	0,76	0,86	-1,88	0,50	_		
$P_{97,5}$	0,07	0,86	1,51	-0,26	1,75	-		
DIC	-52,88							
PLVM	-319,76							
SNT	β_1	β_2	σ	α_1	α_2	ν		
Media	0,05	0,81	0,86	-0,61	_	12,37		
SD	0,01	0,02	$0,\!13$	0,33		10,32		
$P_{2,5}$	0,03	0,76	0,67	-1,33	_	2,30		
$P_{97,5}$	0,07	0,85	1,16	-0.04	_	38,9		
DIC	145,16							
PLVM	-349,58							
STN	eta_1	eta_2	σ	α_1	α_2	ν		
Media	0,05	0,81	0,72	-0,59	_	11,87		
SD	0,01	0,02	0,07	0,31	_	9,90		
$P_{2,5}$	0,03	0,76	0,60	-1,29	_	2,26		
$P_{97,5}$	0,07	0,85	0,87	-0.06	_	38,29		
DIC	177,79							
PLVM	$-345,\!14$							

Conclusiones generales

La distribución skew-normal es típicamente definida como la reducción por condicionamiento de una distribución normal multivariada. En general, la distribución skew-normal de Azzalini se define considerando una matriz de correlación en dicha distribución normal.

La parametrización usada para indexar la distribución skew-normal está basada en la parametrización de la distribución normal original, antes de la reducción por condicionamiento. El problema que origina este hecho es que, tal parametrización no es identificada, aunque los parámetros de la distribución normal sean identificados. Es precisamente el proceso de reducción el que origina los problemas de identificación.

El significado de la parametrización utilizada para indexar el proceso de muestreo skewnormal debe estar basado en éste y no en un proceso normal no observable. El significado estadístico de tal parametrización puede ser establecido una vez que la identificabilidad de la misma es establecida.

Las distribuciones skew-normal propuestas por Azzalini [11] y Azzalini y Dalla Valle [15] pueden ser recuperadas introduciendo restricciones de identificación adicionales, a saber, considerar una matriz de correlación en la distribución normal multivariada que genera la reducción por condicionamiento.

Una vez adoptada una parametrización identificada, el modelo estadístico skew-normal puede ser utilizado en diferentes aplicaciones. Una de ellas es el enfoque de reducción probabilística propuesto por Spanos [78, 79, 80], el cual propone la especificación de un modelo estadístico que capture información relevante de los datos de manera confiable.

Bajo este enfoque, y tomando como reducción probabilística el modelo condicional skew-normal, el modelo de regresión que subyace a esta reducción es el modelo de regresión no-lineal skew-normal heterocedástico.

Sin embargo al imponer ciertas restricciones sobre el espacio paramétrico de los modelos marginal y condicional, la reducción probabilística genera el modelo de regresión lineal skew-normal homocedástico.

Desde el punto de vista Bayesiano, la distribución skew-normal parametrizada de manera identificada permite hacer inferencia para el parámetro de asimetría. Bajo ciertas especificaciones a priori de éste, tanto las densidades predictivas cuanto las densidades a posteriori del parámetro de asimetría pertenecen a diferentes clases de distribuciones asimétricas. De hecho, bajo distribuciones a priori skew-normal, se obtiene un modelo Bayesiano conjugado.

Además, bajo la familia de distribuciones skew-normal standard, se obtiene las distribuciones a posteriori exactas además de representaciones estocásticas para las mismas. Con tales representaciones, es fácil encontrar expresiones exactas para las medias y varianzas a posteriori del parámetro de asimetría, aplicando los resultados en la identificación de un punto de cambio en el parámetro de asimetría en una serie de retornos de mercados financieros emergentes.

Finalmente, los modelos estadísticos construidos a partir de mezclas de distribuciones skew-normales permiten generar una subclase de distribuciones que presenta mayor flexibilidad tanto en asimetría cuanto en curtosis.

Algunos ejemplos de estos modelos son la distribución skew-normal-generalizada, skew-normal-t y skew-t-normal. Todas ellas se generan como mezclas de la distribución skew-normal en el parámetro de asimetría y bajo diferentes especificaciones de la distribución de mezclas.

La representación estocástica de la distribución skew-normal basada en la convolución de una variable aleatoria con distribución normal y otra con distribución normal truncada, además de la estructura de mezclas en el parámetro de asimetría, permite, desde el punto de vista Bayesiano, formular una conveniente estructura jerárquica para los modelos de mezclas de asimetría.

Bibliografía

- [1] E. Andersen. Discrete statistical models with social science applications. North-Holland Publishing Company, Amsterdam, 1980.
- [2] R.B. Arellano-Valle and A. Azzalini. On unification of skew normal families. *Scandinavian Journal of Statistics*, 33:561–574, 2006.
- [3] R.B. Arellano-Valle and H. Bolfarine. On some characterizations of the t-distribution. Statistics & Probability Letters, 25:79–85, 1995.
- [4] R.B. Arellano-Valle, M. Branco, and M.G. Genton. A unified view on skewed distributions arising from selections. *Canadian Journal of Statistics*, 34:581–601, 2006.
- [5] R.B. Arellano-Valle, L.M. Castro, M.G. Genton, and H.W. Gómez. Bayesian inference for shape mixtures of skewed distributions, with application to regression analysis. *Bayesian Analysis*, 3:513–540, 2008.
- [6] R.B. Arellano-Valle, G. del Pino, and P. Iglesias. Bayesian inference in spherical linear models: robustness and conjugate analysis. *Journal of Multivariate Analysis*, 97:179–197, 2006.
- [7] R.B. Arellano-Valle, G. del Pino, and E. San Martín. Definition and probabilistic properties of skew-distributions. *Statistics & Probability Letters*, 58:111–121, 2002.
- [8] R.B. Arellano-Valle and M.G. Genton. On fundamental skew distributions. *Journal of Multivariate Analysis*, 96:93–116, 2005.

- [9] R.B Arellano-Valle, M.G. Genton, and R.H. Loschi. Shape mixtures of multivariate skew-normal distributions. *Journal of Multivariate Analysis (por aparecer)*.
- [10] R.B. Arellano-Valle, H.W. Gómez, and F.A. Quintana. A new class of skew-normal distributions. Communications in Statististics: Theory and Methods, 33:1465–1480, 2004.
- [11] A. Azzalini. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12:171–178, 1985.
- [12] A. Azzalini. The skew normal distribution and related multivariate families (with discussion). *Scandinavian Journal of Statistics*, 32:159–200, 2005.
- [13] A. Azzalini and A. Capitanio. Statistical applications of the multivariate skew normal distributions. *Journal of the Royal Statistical Society. Series B*, 61:579–602, 1999.
- [14] A. Azzalini and A. Capitanio. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. *Journal of the Royal Statistical Society. Series B*, 65:367–389, 2003.
- [15] A. Azzalini and A. Dalla Valle. The multivariate skew normal distribution. *Biometrika*, 83:715–726, 1996.
- [16] A. Azzalini and M.G. Genton. Robust likelihood methods based on the skew-t and related distributions. *International Statistical Review*, 76:106–129, 2008.
- [17] E.W. Barankin. Sufficient parameters: solution of the minimal dimensionality problem. *The Annals of the Institute of Statistical Mathematics*, 12:91–118, 1961.
- [18] E.W. Barankin, H. Kudo, and T. Kusama. Specification of statistical models by sufficiency, pages 9–19. In K. Matusita, editor. Recent Developments in Statistical Inference and Data Analysis. North-Holland, Amsterdam, 1980.

- [19] O. Barndorff-Nielsen. s-ancillarity in exponential families. Sankhy \bar{a} , 37:354–385, 1975.
- [20] O. Barndorff-Nielsen. Factorization of likelihood functions for full exponential families. *Journal of the Royal Statistical Society. Series B*, 38:37–44, 1976.
- [21] D. Barry and J.A. Hartigan. Product partition models for change point problems. *The Annals of Statistics*, 20:260–279, 1992.
- [22] D. Barry and J.A. Hartigan. A Bayesian analysis for change point problem. *Journal of the American Statistical Association*, 88:309–319, 1993.
- [23] D. Basu. Statistical information and likelihood (with discussion). *Sankhyā*, 37:1–71, 1975.
- [24] A. Berman and R.J. Plemmons. *Nonnegative matrices in the mathematical sciences*. SIAM, Philadelphia, 1994.
- [25] J.M. Bernardo and A.F.M. Smith. Bayesian theory. Wiley, New York, 1994.
- [26] H. Bunke and O. Bunke. Identifiability and estimability. *Mathematische Operations-forschung Statistik*, 5:223–233, 1974.
- [27] B.P. Carlin, A.E. Gelfand, and A.F.M. Smith. Hierarchical bayesian analysis of changepoint problems. Applied Statistics: Journal of the Royal Statistical Society, 41:389– 405, 1992.
- [28] M. Chen, Q. Shao, and J. Ibrahim. *Monte Carlo methods in Bayesian computation*. Springer-Verlag, New York, 2000.
- [29] B. Christensen and N. Kiefer. Local cuts and separate inference. *Scandinavian Journal of Statistics*, 21:389–401, 1993.

- [30] R.D. Cook and S. Weisberg. *An introduction to regression graphics*. Wiley, New York, 1994.
- [31] D. Cox and N. Reid. Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society. Series B*, 49:1–39, 1987.
- [32] D.R. Cox and D.V. Hinkley. *Theoretical statistics*. Chapman and Hall, London, 1974.
- [33] E.M. Crowley. Product partition models for normal means. *Journal of the American Statistical Association*, 92:192–198, 1997.
- [34] A. Domínguez-Molina, G. Gonzalez-Farías, and A. Gupta. The multivariate closed-skew normal distribution. Technical Report Nº 03-12, Department of Mathematics and Statistics, Bowling Green State University, 2003.
- [35] A. Domínguez-Molina, G. Gonzalez-Farías, and R. Ramos-Quiroga. *Skew-normality in stochastic frontier analysis*, pages xx–xx. Chapman and Hall/CRC, Boca Raton, 2004.
- [36] M.J. Dueker. Markov switching in GARCH processes and mean-reverting stock-market volatility. *Journal of Business & Economic Statistics*, 15:26–34, 1997.
- [37] B.E. Ellison. Two theorems for inferences about the normal distribution with applications in acceptance sampling. *Journal of the American Statistical Association*, 76:89–95, 1964.
- [38] R. Engle, D. Hendry, and J. Richard. Exogeneity. Econometrica, 51:277–304, 1983.
- [39] P. Fearnhead and Z. Liu. On-line inference for multiple changepoint problems. *Journal of the Royal Statistical Society. Series B*, 69:589–605, 2007.
- [40] R.A. Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London, Series A*, 222:309–368, 1922.

- [41] R.A. Fisher. The moments of the distribution for normal samples of measures of departure from normality. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 130:16–28, 1930.
- [42] J.P Florens, M. Mouchart, and J.M. Rolin. On two definitions of identification. *Statistics*, 16:213–218, 1985.
- [43] J.P Florens, M. Mouchart, and J.M. Rolin. *Elements of Bayesian statistics*. Marcel Dekker, New York, 1990.
- [44] J.P. Florens and J.M. Rolin. Asymptotic sufficiency and exact estimability in Bayesian experiments., pages 121–142. In J. P. Florens, M. Mouchart, J. P. Raoult and L. Simar, Alternative Approaches to Time Series Analysis. Publication des Facultés Universitaires Saint Louis, Bruxelles, 1984.
- [45] A. Gabrielsen. Consistency and identifiability. *Journal of Econometrics*, 8:261–263, 1978.
- [46] S. Geisser and W. Eddy. A predictive approach to model selection. *Journal of the American Statistical Association*, 74:153–160, 1979.
- [47] A. Gelfand and D. Dey. Bayesian model choice: asymptotics and exact calculations. Journal of the Royal Statistical Society. Series B, 56:501–514, 1994.
- [48] M.G. Genton. Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality. Chapman and Hall/CRC, Boca Raton, 2004.
- [49] M.G. Genton and N. Loperfido. Generalized skew-elliptical distributions and their quadratic forms. *Annals of the Institute of Statistical Mathematics*, 57:389–401, 2005.
- [50] H.W. Gómez, O. Venegas, and H. Bolfarine. Skew-symmetric distributions generated by the distribution function of the normal distribution. *Environmetrics*, 18:395–407, 2007.

- [51] G. Gonzáles-Farías, J. Domínguez-Molina, and A. Gupta. Additive properties of skew normal random vectors. *Journal of Statistical Planning and Inference*, 126:521–534, 2004.
- [52] D.M. Hawkins. Fitting multiple change-point models to data. *Computational Statistics & Data Analysis*, 37:323–341, 2001.
- [53] L. Horváth and Kokoszka P. Change-point detection with non-parametric regression. *Statistics*, 36:9–31, 2002.
- [54] V.K. Jandhyala, Fotopoulos S.B., and Hawkins DM. Detection and estimation of abrupt changes in the variability of a process. *Computational Statistics & Data Analysis*, 40:1–19, 2002.
- [55] J. Kadane. The role of identification in Bayesian theory, pages 175–191. North-Holland, Amsterdam, 1974.
- [56] T.C. Koopmans and O. Reiersøl. The identification of structural characteristics. *The Annals of Mathematical Statistics*, 21:165–181, 1950.
- [57] T. Lancaster. The incidental parameter problem since 1948. *Journal of Econometrics*, 95:391–413, 2000.
- [58] E.L. Lehmann. Theory of point estimation. Wiley, London, 1983.
- [59] B. Liseo and N. Loperfido. A Bayesian interpretation of the multivariate skew normal distribution. *Statistics & Probability Letters*, 61:395–401, 2003.
- [60] H.F. Lopes and H.S. Migon. Comovements and Contagion in Emergin Markets: Stock Indexes Volatilities. Lecture Notes in Statistics Series 167, 2002.
- [61] R. Loschi, F. Cruz, and R.B. Arellano-Valle. Multiple change point analysis for the regular exponential family using the product partition model. *Journal of Data Science*, 3:305–330, 2005.

- [62] R.H. Loschi and F.R.B. Cruz. Extension to the product partition model: Computing the probability of a change. *Computational Statistics & Data Analysis*, 48:255–268, 2005.
- [63] R.H. Loschi, P. Iglesias, R.B. Arellano-Valle, and F.R.B. Cruz. Full predictivisto modeling to stock market data: An application to change point problems. *European Journal of Operational Research*, 180:282–291, 2007.
- [64] Y. Ma and M. G. Genton. A flexible class of skew-symmetric distributions. *Scandinavian Journal of Statistics*, 31:459–468, 2004.
- [65] G. Maddala. Introduction to econometrics. Prentice Hall, UK, 1992.
- [66] P. McCullagh. What is a statistical model? (with discussion). *The Annals of Statistics*, 30:1225–1310, 2002.
- [67] S.Ñadarajah and S. Kotz. Skewed distributions generated by the normal kernel. *Statistics & Probability Letters*, 65:269–277, 2003.
- [68] A. Oulhaj and M. Mouchart. Partial sufficiency with connection to the identification problem. METRON - International Journal of Statistics, 61:267–283, 2003.
- [69] J. Pearl. Causality: models, reasoning and inference. Cambridge University Press, Cambridge, 2000.
- [70] G. Picci. Some connections between the theory of sufficient statistics and the identifiability problem. *METRON International Journal of Statistics*, 33:383–398, 1977.
- [71] F. Quintana and P. Iglesias. Nonparametric bayesian clustering and product partition models. *Journal of the Royal Statistical Society. Series B*, 65:557–574, 2003.
- [72] J.P. Raoult. Structures statistiques. Presses Universitaires de France, Paris, 1975.

- [73] T.J. Rothenberg. Identification in parametric models. *Econometrica*, 39:577–591, 1971.
- [74] F. Ruggeri and S. Sivaganesan. On modeling change points in non-homogeneous poisson processes. *Statistical Inference for Stochastic Processes*, 8:311–329, 2005.
- [75] S.K. Sahu, D.K. Dey, and M. Branco. A new class of multivariate skew distributions with application to bayesian regression models. *Canadian Journal of Statistics*, 31:129–150, 2003.
- [76] E. San Martín. Spécification du modèle lisrel d'un poin de vue du modèle lineaire. Master's thesis, Universite Catholique de Louvain, 1997.
- [77] E. San Martín and F. Quintana. Consistency and identifiability revisited. *Brazilian Journal of Probability and Statistics*, 16:99–106, 2002.
- [78] A. Spanos. On modelling heteroskedasticity: The student's t and elliptical linear regression models. *Econometric theory*, 10:286–315, 1994.
- [79] A. Spanos. On theory testing in econometrics modelling with noexperimental data. *Journal of Econometrics*, 67:189–226, 1995.
- [80] A. Spanos. Where do statistical models come frome? : Revisiting the problem of specification. The Second Erich L. Lehmann Symposium, Lecture Notes-Monograph Series, Institute of Mathematical Statistics, 49:98–119, 2006.
- [81] D.J. Spiegelhalter, N.G. Best, B.P. Carlin, and A. Van Der Linde. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society. Series B*, 64:583–640, 2002.
- [82] G.M. Tallis. The moment generating function of the truncated muilti-normal distribution. *Journal of the Royal Statistical Society. Series B*, 23:223–229, 1961.

[83] J. Wang, J. Boyer, and M.G. Genton. A skew-symmetric representation of multivariate distributions. *Statistica Sinica*, 14:1259–1270, 2004.