



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

Reducción del sesgo en estimación de máxima verosimilitud en modelos asimétricos

Por

Jaime Roberto Arrué Álvarez

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

Orientador : **Reinaldo Boris Arellano Valle**

Co-orientador : **Héctor Wladimir Gómez Geraldo**

Julio, 2016

Santiago, Chile

Comisión Informante:

- Dra. María Dolores Jiménez Gamero, U. de Sevilla, España.
- Dr. Pedro Jodrá Esteban, U. de Zaragoza, España.
- Dr. Jorge González Burgos, PUC.
- Dr. Manuel Galea Rojas, PUC.
- Dr. Reinaldo Arellano Valle, PUC.

Índice general

Resumen	1
1. Introducción	2
1.1. Formulación del problema	2
1.1.1. Descripción del método de Firth	4
1.2. Definición de los objetivos	9
1.3. Resumen de los contenidos	9
2. Distribución skew-normal modificada	13
2.1. Introducción	13
2.2. Definición y propiedades del modelo MSN	17
2.2.1. Momentos	19
2.2.2. Coeficiente de asimetría y curtosis	20
2.3. Estimación de máxima verosimilitud	22
2.3.1. Caso parámetro de forma	22
2.3.2. Caso localización, escala y forma	26
2.3.3. Matriz información	28
2.4. Score modificada para la familia MSN	35

2.4.1.	Caso parámetro de forma	36
2.4.2.	Caso localización, escala y forma	45
2.5.	Aplicación:	48
2.6.	Apéndice	52
3.	I transformación de variable aleatoria MSN: Distribución log-MSN	54
3.1.	Introducción	54
3.2.	Definición y propiedades del modelo LMSN	56
3.2.1.	Coeficiente de asimetría y curtosis	60
3.3.	Estimación de máxima verosimilitud	62
3.3.1.	Perfil de verosmilitud	62
3.3.2.	Matriz información	63
3.4.	Reducción de sesgo del EMV	63
3.4.1.	Estudio de simulación: Localización, escala y forma	64
3.5.	Aplicación	65
4.	II transformación de variable aleatoria MSN: Distribución Birnbaum-Saunders skew-normal modificada	71
4.1.	Introducción	71
4.2.	Definición y propiedades del modelo BSMSN	73
4.2.1.	Función de Supervivencia	76
4.2.2.	Momentos	77
4.2.3.	Coeficiente de asimetría y curtosis	78
4.3.	Estimación de máxima verosimilitud	79

4.3.1. Matriz información	79
4.4. Reducción de sesgo del EMV	82
4.4.1. Estudio de simulación	84
4.5. Aplicación	87
4.6. Apéndice	92
5. Distribución skew t-normal modificada	95
5.1. Introducción	95
5.2. Definición y propiedades del modelo MStN	96
5.2.1. Momentos	99
5.2.2. Coeficiente de asimetría y curtosis	100
5.3. Estimación de máxima verosimilitud	103
5.3.1. Matriz información	103
5.4. Reducción de sesgo del EMV de λ	107
5.4.1. Caso: parámetro de forma	111
5.4.2. Estudio de simulación: parámetro de forma	116
5.4.3. Caso: Localización, escala y forma	118
5.4.4. Estudio de simulación: Localización, escala y forma	119
5.5. Aplicación	121
5.6. Apéndice	127
6. Conclusiones y futuras extensiones	130
Bibliografía	133

Índice de figuras

2.1. Gráficas de MSN(0) (línea continua), MSN(1) (línea punteada), SN(1) (línea segmentada) y MSN(3) (línea continua gruesa).	18
2.2. Coeficientes de asimetría y curtosis para distribución MSN(λ) (línea continua) y SN(λ) (línea punteada).	22
2.3. Perfil de las funciones log-verosimilitud para los modelos MSN (panel superior) y SN (panel inferior), para muestras de tamaños $n = 20$ (izquierda), $n = 50$ (centro) y $n = 100$ (derecha) simulada de la distribución $MSN(0) \equiv SN(0) \equiv N(0,1)$	28
2.4. Score modificada (panel izquierdo) y su integral (panel derecho) para los modelos SN(λ) (línea punteada) y MSN(λ) (línea continua).	37
2.5. Funciones score (panel izquierdo) y log-verosimilitud relativa (panel derecho), modificada (línea continua) y no modificada (línea segmentada). La línea de puntos horizontal en el panel de la derecha representa el intervalo de confianza de 0.95.	42

2.6.	Funciones score (panel izquierdo) y log-verosimilitud relativa (panel derecho), modificada (línea continua) y no modificada (línea segmentada) de $ z $. La línea de puntos horizontales en el panel de la derecha representa el intervalo de confianza de 0.95.	43
2.7.	Versiones relativas modificada, $l_M^P(\lambda)$ (línea continua) y no modificada, $l^P(\lambda)$ (línea segmentada). La línea horizontal punteada corresponde al intervalo de confianza de 0.95.	47
2.8.	Datos ajustados	50
3.1.	Gráfica de LMSN.	56
3.2.	Función de riesgo para el modelo LMSN.	58
3.3.	Gráfica de asimetría y curtosis para LMSN (línea continua) y LSN (línea punteada).	61
3.4.	Perfil de verosimilitud de LMSN.	63
3.5.	Ajuste de datos.	67
3.6.	Gráfica de QQ y CDF para LMSN y LN.	68
3.7.	Datos reajustados.	70
4.1.	Gráfica de BSMSN.	74
4.2.	Gráfica de función de riesgo de BSMSN.	76
4.3.	Gráfica de asimetría (γ_1) y curtosis (γ_2) del modelo BSMSN.	78
4.4.	Ajuste datos.	88
4.5.	QQ plot de BSMSN, BS y EXPBSn.	89
4.6.	CDF empírica y teórica del modelo BSMSN.	89

4.7. Datos ajustados.	91
5.1. Gráfica de MStN para a) $\nu = 1$ y b) $\nu = 5$	97
5.2. Gráfica de Asimetría y Curtosis para el modelo MStN.	103
5.3. Gráfica de función $h(\nu)$	106
5.4. Probabilidad de divergencia para el modelo MStN con (a) $\nu = 1$ y (b) $\nu = 5$	109
5.5. Función modificada (panel izquierdo) y función modificada integrada (panel derecho) para el modelo SN (línea segmentada), MStN con $\nu = 1$ (línea continua) y $\nu = 3$ (línea continua gruesa) y MSN (línea de puntos).	112
5.6. Perfil de log-verosimilitud para ν	123
5.7. Ajuste de datos Níquel.	124
5.8. QQ plot para MStN y MSN.	124
5.9. CDF empírica y teórica para el modelo MStN y MSN.	125
5.10. Datos ajustados.	126

Índice de tablas

2.1. Probabilidad de divergencia de EMV $p(\lambda, n)$	24
2.2. Simulación de 5,000 iteraciones del modelo MSN(λ).	26
2.3. Simulación de 5,000 iteraciones del modelo MSN(0, 1, λ)	33
2.4. Simulación de 5,000 iteraciones del modelo MSN(-1, 1, λ).	34
2.5. Simulación de 5,000 iteraciones del modelo MSN(1, 1, λ).	35
2.6. Estimaciones no modificada y modificada para λ	43
2.7. Estimaciones no modificada y modificada para λ basada en los datos $ z $. . .	44
2.8. Simulación de 5,000 iteraciones del modelo MSN(λ).	44
2.9. Estimaciones de μ , σ y λ	47
2.10. Simulación de 5,000 iteraciones del modelo MSN(0, 1, λ).	48
2.11. Estadística descriptiva de los datos.	49
2.12. EMV de μ , σ y λ	49
2.13. Intervalo confianza para λ	50
3.1. Rango de valores para la Asimetría y Curtosis.	60
3.2. Simulación de 5,000 iteraciones del modelo LMSN(0, 1, λ).	65
3.3. Estadística descriptiva de los datos.	66

3.4.	EMV de los modelos LMSN, LN y LSN.	66
3.5.	EMV de μ , σ y λ	69
3.6.	Intervalo de confianza para λ	69
4.1.	Simulación de 5,000 iteraciones del modelo BSMSN(α, β, λ).	85
4.2.	Resumen estadístico de los datos.	87
4.3.	EMV de los modelos BS, BSMSN y EXPBSn.	87
4.4.	EMV de α , β y λ	90
4.5.	Intervalos de confianza para λ	90
5.1.	Rango de Asimetría, Curtosis para distintos valores de ν	102
5.2.	Simulación de 5,000 iteraciones del modelo MStN(λ, ν), ν conocido.	108
5.3.	Simulación de 5,000 iteraciones del modelo MStN(0,1, λ, ν), ν conocido.	110
5.4.	Simulación de 5,000 iteraciones del modelo MStN(λ, ν), ν conocido.	117
5.5.	Simulación de 5,000 iteraciones del modelo MStN(0, 1, λ, ν), ν conocido.	120
5.6.	Estadística descriptiva de los datos de Níquel.	121
5.7.	EMV del modelo MStN que ajusta los datos, ν fijo.	122
5.8.	EMV del modelo SN, MSN y MStN.	123
5.9.	EMV de μ , σ y λ	125
5.10.	Intervalos de confianza para λ	126

Agradecimientos

Quiero agradecer ante todo a Dios por darme la oportunidad de alcanzar este gran objetivo trazado en mi vida, por darme fuerza y perseverancia en los momentos difíciles, por encontrar a las personas indicadas en mi camino, ayudándome incondicionalmente a lograrlo. En especial, mi sincero agradecimiento:

- A mi gran amigo y Profesor Dr. Héctor Gómez Geraldo, por su apoyo y confianza incondicional hacia mi persona. Por sus valiosas sugerencias y orientaciones que permitieron llevar a buen término esta tesis.
- Al Profesor Dr. Reinaldo Arellano Valle, por el honor de tenerlo como profesor y ser testigo de su gran calidad humana y profesional. También, agradecer su fundamental orientación, sugerencia y contribuciones en el desarrollo de este trabajo.
- A los Profesor Héctor Varela y René Maluenda por la confianza depositada en mí, dándome el apoyo y respaldo para poder llegar a buen término a mi objetivo trazado.
- A amiga Carla Leal con quien compartimos gratos momentos de estudio y de reflexión sobre la vida.
- A mi amigo Juan Olivares de quien siempre recibí apoyo, disposición desinteresada a ayudar y palabras de aliento.
- A Conicyt por financiar mis estudios, pues sin ello no hubiese sido posible cumplir este sueño.

...Todo el esfuerzo e inspiración depositado en esta Tesis se la dedicó a mi esposa e hija, por su apoyo incondicional y paciencia infinita, sobretodo en lo momentos más difíciles de mi vida.....las amo.....como una vez dije: *Un buen fruto no depende sólo del árbol que los sostiene, sino también de quien lo cuida y protege.*

Reducción del sesgo en estimación de máxima verosimilitud en modelos asimétricos

Jaime Roberto Arrué Álvarez

Orientador: Reinaldo Boris Arellano Valle

Co-orientador: Héctor Wladimir Gómez Geraldo

Resumen

En la presente tesis, se presenta el método de reducción de sesgo de Firth [28] para ser aplicado a modelos asimétricos y así, dar solución al problema de sobrestimación del estimador máximo verosímil de los parámetros que definen a estos modelos. Este método, consiste en sustituir la ecuación de verosimilitud estándar de la distribución, por una nueva ecuación denominada ecuación de verosimilitud modificada. La solución a esta nueva ecuación se le denomina estimador máximo verosímil modificado, el cual por construcción tiene un menor sesgo que el estimador máximo verosímil no modificado ó estándar. Aplicaremos este método, en particular, a una familia de distribuciones que compite con la familia skew-normal, denominada skew-normal modificada, puesto que ambas tienen características similares. Se estudia algunas propiedades probabilísticas del modelo propuesto, se realiza inferencia estadística, se presentan simulaciones para ilustrar la reducción de sesgo y por último, se aplica el modelo y el método de reducción a datos reales. También, se analiza el comportamiento

del sesgo de los estimadores máximo verosímiles de modelos que se deducen de la distribución skew-normal modificada. Estas son las distribuciones log-skew-normal modificada y skew-normal modificada Birnbaum Saunders, se obtienen por medio de una transformación de una variable que se distribuye skew-normal modificada. Ambas, resultan ser una extensión de las distribuciones log-normal y Birnbaum Saunders clásica, respectivamente, y tienen la ventaja de tener un rango más amplio de asimetría y curtosis, es decir, son más flexibles comparados con los respectivos modelos clásicos. La distribución skew t-normal modificada, se obtiene de una mezcla de escala del modelo skew-normal modificada. Esta nueva distribución incorpora en su densidad a la distribución Student- t , aumentando de esta manera la flexibilidad de la estructura paramétrica, así, este modelo es más flexible que el modelo skew-normal modificada. Estos nuevos modelos, derivados de la distribución propuesta, se estudian siguiendo la estructura anteriormente señalada. Para finalizar, se entregan resultados y conclusiones derivados del desarrollo de esta tesis, además se entregan algunos temas propuestos para desarrollarlas en trabajos futuros.

Capítulo 1

Introducción

1.1. Formulación del problema

Podemos distinguir dos clases de métodos para reducir el sesgo, haciendo la diferencia entre corrección de sesgo y reducción de sesgo. Los métodos de corrección de sesgo se basan en los siguientes pasos:

1. Obtener el término de sesgo de primer orden del estimador máximo verosímil.
2. Restarlo del estimador máximo verosímil.

De lo anterior, las estimaciones de sesgo corregido no están definidas cuando los estimadores máximo verosímil son infinitos, entonces estos métodos dependen de la finitud de estos estimadores. Además, para tamaños de muestra pequeños el método de corrección de sesgo tiende a corregir más allá del verdadero valor del parámetro. Esto se ilustra a través de los estudios empíricos en Bull et al. [23], ellos comparan el método sesgo corregido con la reducción de sesgo en regresiones logísticas. Quenouille [57] fue el primero en desarrollar un

método de reducción de sesgo que es aplicable a familias de distribuciones generales. Este procedimiento, denominado de Jackknife, tiene como objetivo la eliminación de términos del sesgo hasta un orden específico. Sin embargo, si el estimador máximo verosímil no tiene forma cerrada, este método puede llegar a ser computacionalmente costoso, debido a que las estimaciones de máxima verosimilitud tienen que ser obtenidas de forma iterativa, para cada uno de los posibles subconjuntos de la muestra, de acuerdo con cierto esquema de particiones.

Método de reducción de sesgo:

Consiste básicamente en la modificación de las funciones score de modo que las raíces de las ecuaciones score modificada resultante es un estimador insesgado de primer orden. Firth [28], fue el primero en dar una base formal de argumentos asintóticos para familias regulares. Él desarrolla un método general para la eliminación del término de primer orden en la expansión asintótica del sesgo del estimador máximo verosímil. Demostró que para la familia exponencial de parametrización canónica, el método de penalización de la función verosimilitud se reduce a la priori invariante de Jeffreys [38].

Las ventajas del método de reducción de sesgo son:

1. Es independiente del estimador máximo verosímil, por lo tanto no depende de su finitud.
2. Los nuevos estimadores obtenidos tienen el término de primer orden más pequeño, o incluso cero en la expansión asintótica de su sesgo.
3. Estos estimadores, de sesgo reducido, al ser obtenidos por estimación de máxima verosimilitud, tienen todas las propiedades asintóticas deseadas, esto es: normalidad asintótica, suficiencia asintótica, insesgamiento y eficiencia.

Este enfoque de las funciones score modificado para reducir el sesgo está ganando popularidad (por ejemplo: Mehrabi y Matthews [52], Pettitt et al. [55], Heinze y Schemper [36], Bull et al. [24], Zorn [67], Sartori [60], Bull et al. [25], Kosmidis y Firth [40], debido a las propiedades superiores del estimador de sesgo reducido sobre el tradicional estimador de máxima verosimilitud, generalmente en modelos para respuestas categóricas. En particular, es aplicado en regresión logística, donde el método de reducción de sesgo coincide con la log-verosimilitud penalizada por la priori de Jeffreys y las estimaciones de sesgo reducido son siempre finitas.

1.1.1. Descripción del método de Firth

A continuación, se entrega una descripción breve del método general de reducción del sesgo propuesto por Firth [28] y aplicada por Sartori [60] en el contexto skew-normal (SN); ver también otros trabajos de Kosmidis y Firth [40], Lagos-Álvarez y Jiménez-Gamero [41] y Azzalini y Arellano-Valle [17]. Se conserva la notación considerada por Sartori [60] y se considera sólo el caso de un parámetro.

Sea $l(\theta)$ sea la función de log-verosimilitud de una familia paramétrica regular. Denotemos por $U(\theta) = l'(\theta)$ la función score y $j(\theta) = -l''(\theta)$ la cantidad información observada, donde l' y l'' son la primera y segunda derivada de l , respectivamente. También, consideremos las funciones esperadas

$$I(\theta) = E_{\theta}\{j(\theta)\}, \quad \nu_{\theta,\theta,\theta}(\theta) = E_{\theta}\{l'(\theta)^3\}, \quad \nu_{\theta,\theta\theta}(\theta) = E_{\theta}\{l'(\theta)l''(\theta)\}.$$

En muestras aleatorias de tamaño n , $j(\theta)$ es de orden $O_P(n)$ y los valores esperados son de orden $O(n)$. Por otra parte, si $\hat{\theta}$ es el EMV de θ , Cox y Snell [27] mostraron que el sesgo de

$\widehat{\theta}$ es (véase también Firth [28])

$$E_{\theta}(\widehat{\theta} - \theta) = b(\theta) + O(n^{-2}), \text{ con } b(\theta) = -\frac{1}{2}I(\theta)^{-2}(\nu_{\theta,\theta,\theta} + \nu_{\theta,\theta\theta}) = O(n^{-1}).$$

Esto produce un estimador estándar con sesgo corregido de θ dado por $\widehat{\theta}_{BC} = \widehat{\theta} - b(\widehat{\theta})$, el cual reduce el sesgo a $E_{\theta}(\widehat{\theta}_{BC} - \theta) = O(n^{-2})$. Como ha señalado por Firth [28], este enfoque es correctivo, en lugar de reductivo, ya que se aplica una vez que $\widehat{\theta}$ se ha calculado. Además, este enfoque tiene un inconveniente práctico, ya que su aplicación depende de la finitud de $\widehat{\theta}$. Para superar estos inconvenientes, Firth [28] propone un método para reducir el sesgo de $b(\theta)$ antes de calcular $\widehat{\theta}$. En concreto, su enfoque consiste en modificar la función score sin modificar $U(\theta)$ como

$$U_M(\theta) = U(\theta) + M(\theta), \quad (1.1)$$

donde $M(\theta)$ es $O(1)$ cuando $n \rightarrow \infty$. La solución a la ecuación de verosimilitud modificada $U_M(\theta) = 0$ produce el EMV modificado, digamos $\widehat{\theta}_M$; y por la imposición de la condición $E_{\theta}(\widehat{\theta}_M - \theta) = O(n^{-2})$, la modificación $M(\theta)$ es entonces elegida como

$$M(\theta) = -I(\theta)b(\theta) = \frac{1}{2}I(\theta)^{-1}(\nu_{\theta,\theta,\theta} + \nu_{\theta,\theta\theta}). \quad (1.2)$$

Además, la existencia de $\widehat{\theta}_M$ no depende de la existencia de $\widehat{\theta}$, $\widehat{\theta}_M$ es de segundo orden equivalente a $\widehat{\theta}_{BC}$, y de primer orden equivalente a $\widehat{\theta}$. En particular, $\widehat{\theta}_M$ es un estimador consistente y su distribución asintótica de primer orden coincide con la de $\widehat{\theta}$, es decir,

$$\sqrt{n}(\widehat{\theta}_M - \theta) \xrightarrow{d} N(0, I^{-1}(\theta)), \quad \text{as } n \rightarrow \infty.$$

A consecuencia de este enfoque, se obtiene a partir de la función score modificada, la función cuasi-log-verosimilitud se define como

$$l_M(\theta) = \int_c^{\theta} U_M(t)dt = l(\theta) - l(c) + \int_c^{\theta} M(t)dt, \quad (1.3)$$

donde c es una constante arbitraria. $l_M(\theta)$ es una función de log-verosimilitud penalizada, con penalización de orden $O(1)$. También, obtenemos la razón de log-verosimilitud modificada dada por

$$W_M(\theta) = 2\{l_M(\hat{\theta}_M) - l_M(\theta)\}, \quad (1.4)$$

que es asintóticamente distribuida χ_1^2 , es útil para realizar pruebas de hipótesis y obtener intervalos de confianza para θ .

Debido a las ventajas del método de reducción de sesgo sobre el método de corrección de sesgo, se aplicará a modelos asimétricos que presentan problemas de sobrestimación en la EMV de sus parámetros. Uno de los modelos asimétricos más importante de la literatura, es la distribución skew-normal univariada, introducida por Azzalini [12]. En este sentido, diremos que una variable aleatoria Z se distribuye skew-normal con parámetro λ , denotada por $Z \sim SN(\lambda)$, si su densidad viene dada por:

$$f_Z(z; \lambda) = 2\phi(z)\Phi(\lambda z), \quad (1.5)$$

donde z y $\lambda \in \mathbb{R}$, ϕ y Φ son la función de distribución de probabilidad (pdf) y función de distribución acumulada (cdf) de la distribución normal estándar, respectivamente. Cuando $\lambda = 0$ en (1.5), se obtiene la distribución normal estándar. Para valores distintos de cero, el parámetro λ controla la asimetría del modelo, haciendo de éste un modelo más flexible que el modelo normal.

Azzalini [12] estudia las propiedades del modelo SN, estimadores de máxima verosimilitud y matriz información de Fisher. Esta matriz, es singular cuando el parámetro de asimetría toma el valor cero, por lo que recurre a una reparametrización para obtener una nueva ma-

triz de información no singular y así, poder realizar inferencia asintótica sobre la nulidad del parámetro de forma. Henze [37], da una representación estocástica para este modelo y con ella deriva los momentos impares. Azzalini [13], realiza un estudio a modelos más generales que el modelo skew-normal. Arnold et al. [8] realiza truncaciones al modelo normal utilizando la distribución skew-normal. Los problemas de inferencia de este modelo son estudiados por Pewsey [56]. Posteriormente, Gupta y Chen [33] presentan un test de bondad de ajuste. Nadarajah y Kotz [53], extienden el modelo skew-normal manteniendo como kernel la distribución normal. Arellano-Valle et al. [2] introduce la distribución skew-normal-generalizada y contiene al modelo skew-normal como un caso particular. Liseo [46] realiza inferencia bayesiana y Liseo y Loperfido [48], analizan de un punto de vista bayesiano el parámetro de asimetría del modelo skew-normal.

En el caso multivariado del modelo skew-normal; es estudiado por Azzalini y Dalla-Valle [14], sus propiedades y aplicación son analizadas por Azzalini y Capitanio [15]. Branco y Dey [22], extienden el resultado de Azzalini y Capitanio [14] a distribuciones skew-elípticas multivariadas, a través de un método de condicionamiento. El modelo skew Cauchy es estudiado por Arnold y Beaver [9]. Genton He y Liu [29], obtienen los momentos de los vectores aleatorios del modelo skew-normal multivariado. Gupta et al. [34] definen y estudian las propiedades de una nueva clase de distribuciones skew-normal. Arellano-Valle, del Pino y San Martín [1] muestran que algunas propiedades del modelo skew-normal se cumplen para una clase general de distribuciones, definidas en términos de independencia condicional en el signo y el valor absoluto. Este trabajo es extendido, por Arellano-Valle y Genton [3], dando origen a una nueva clase de distribuciones skew-normal multivariada denominada skew-normal fundamental. Sartori [60] reduce el sesgo del estimador máximo verosímil del

parámetro de asimetría en la distribución skew-normal y skew-t-normal, aplicando el método de Firth [28]. Arellano y Azzalini [4] extienden la definición de parametrización centrada al caso multivariado de la distribución skew-normal, estudiando su matriz de información. Azzalini y Arellano-Valle [17] proponen una verosimilitud penalizada para reducir el sesgo de la estimación máximo verosímil del parámetro de asimetría, y hacen este estudio para el caso univariado y multivariado de las distribuciones skew-normal y skew-t.

Debido a que en la naturaleza se encuentran diversos conjuntos de datos, cuya función de distribución desconocida que los gobierna puede ser asimétrica ó simétrica, se ha generado una incesante búsqueda para encontrar modelos paramétricos que sean lo suficientemente flexibles, en lo que se refiere a poseer un cierto rango de asimetría y curtosis. Estos modelos más flexibles, evitan el gran problema de encontrar una reparametrización adecuada y en consecuencia perder interpretabilidad de los parámetros originales. También, en lo posible contengan a la distribución normal en forma estricta y no como un caso límite. Uno de los modelos más importante y conocido, en este sentido, es la distribución skew-normal univariada introducida por Azzalini [12]. Este modelo tiene un problema de sobrestimación en el parámetro de asimetría, cuando se estima por el método de máxima verosimilitud. Sobretudo, ocurre, cuando estimamos valores grandes del verdadero parámetro y el tamaño de muestra es pequeño o moderado. Incluso, estas estimaciones pueden ser infinitas. Por consiguiente, genera evidentemente un sesgo en la estimación y esto es una preocupación para los profesionales e investigadores de la Estadística. La magnitud de este sesgo juega un papel importante en esta disciplina, puesto que si éste es grande se puede realizar inferencias potencialmente erróneas.

En resumen, es de gran importancia aplicar el método de reducción de sesgo de Firth [28], por

su efectividad, a modelos asimétricos flexibles que tengan características similares al modelo skew-normal. En particular, en esta tesis, se propone el modelo skew-normal modificada.

1.2. Definición de los objetivos

Objetivo General: Aplicar el método de reducción de sesgo a distribuciones asimétricas.

En particular, al modelo skew-normal modificada y a nuevos modelos derivados de ella.

Aplicar a datos reales los modelos y el método mencionado anteriormente.

Objetivos Específicos:

- Definir la distribución skew-normal modificada y aplicar el método de reducción de sesgo de Firth [28].
- Derivar los nuevos modelos a partir de la distribución skew-normal modificada.
- Estudiar el comportamiento de los estimadores máximo verosímil de los nuevos modelos y aplicar el método mencionado previamente.
- Ajustar los modelos propuestos a conjuntos de datos reales y compararlos con otros de su misma clase.
- Ilustrar el método de reducción en el ejemplo práctico.

1.3. Resumen de los contenidos

Con el fin de cumplir con los objetivos trazados, esta tesis ha sido organizada de la siguiente manera: En el Capítulo 2 se entrega una breve descripción para generar modelos skew

simétricos. Se define la familia skew-normal modificada, ésta incluye en forma estricta a la normal cuando su parámetro de forma toma el valor cero. Se muestran algunas propiedades del modelo, representación estocástica y rangos de los coeficientes de asimetría y curtosis. Se realiza inferencia estadística por el método de máxima verosimilitud. Se obtiene la matriz de información observada y esperada. El método de Firth [28], se ilustra utilizando simulaciones del modelo en estudio, tanto para el caso de parámetro de forma, como para el caso de presencia de parámetros de localización, escala y forma. Por último, se ajusta el modelo skew-normal modificada a un conjunto de datos reales y se aplica el método de reducción de sesgo al parámetro en forma.

El Capítulo 3, consiste en presentar una nueva distribución, denominada log-skew-normal modificada, obtenida por medio de una transformación de la variable aleatoria de una distribución skew-normal modificada. Esto es, la variable aleatoria $Y = e^Z$, se distribuye log-skew-normal modificada si Z se distribuye skew-normal modificada. El nuevo modelo propuesto contiene a la distribución log-normal clásica, cuando el parámetro de forma es cero. Se presenta una forma general para los momentos y se demuestra la no existencia de la función generadora de momentos. Se compara los rangos de coeficiente de asimetría y curtosis del modelo log-skew-normal modificada, log-normal y log-skew-normal, para el caso particular $\mu = 0$ y $\sigma = 1$. Se entrega una aplicación del modelo propuesto a datos reales, mostrando una mayor flexibilidad comparado con otros modelos. Se aplica el método de reducción de sesgo a este ejemplo práctico.

En el Capítulo 4 se muestra el origen de una nueva distribución obtenida del modelo skew-normal modificada, denominada Birnbaum-Saunders skew normal modificada, obtenida por transformación de la variable aleatoria que se distribuye skew-normal modificada. Es decir,

una variable aleatoria T tiene distribución Birnbaum-Saunders skew normal modificada de parámetros α , β y λ , denotada por $T \sim BSMSN(\alpha, \beta, \lambda)$, si T corresponde a la transformación:

$$T = \beta \left(\frac{\alpha Z}{2} + \sqrt{\left(\frac{\alpha Z}{2}\right)^2 + 1} \right)^2,$$

donde $Z \sim MSN(\lambda)$. Ésta, incluye en forma estricta a la distribución Birnbaum-Saunders cuando el parámetro λ es nulo. Además, se entregan algunas propiedades relevantes y se definen algunos conceptos de análisis de sobrevivencia. Se muestra una expresión general para los momentos y se explora los rangos de asimetría y curtosis. El desarrollo y obtención de la matriz información observada y esperada son entregados, analizando la singularidad de esta última. Se realizan simulaciones que ilustran la efectividad de la reducción de sesgo del estimador sobrestimado respectivo. Se presenta un ejemplo práctico de este modelo, mostrando su capacidad de buen ajuste, comparado con otros modelos, y se ilustra la aplicación del método de reducción.

El Capítulo 5 presenta al modelo denominado skew t-normal modificada, el cual se obtiene a través de una mezcla en el parámetro de escala del modelo MSN. Algunas propiedades importantes de esta familia son estudiadas. Una forma general de los momentos es dada y se entregan los rangos de valores de los coeficientes de asimetría y curtosis. Se obtiene una expresión general de la matriz información observada y esperada, siendo esta última no singular bajo la hipótesis de simetría. Se aplica el método de reducción de sesgo, para obtener un nuevo estimador con menor sesgo que el estimador de máxima verosimilitud, previa demostración de su existencia. Por lo anterior, se realizan simulaciones para mostrar esta reducción. Se ilustra una aplicación a un conjunto de datos, de concentración de Níquel,

en relación al ajuste del modelo y a la reducción de sesgo.

Capítulo 2

Distribución skew-normal modificada

2.1. Introducción

Azzalini [12], presenta una forma general para obtener distribuciones asimétricas a través del siguiente Lema:

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

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

es una función densidad de la variable X para cualquier $\lambda \in \mathbb{R}$. Se denota por $X \sim Sf(\lambda)$.

Algunos autores, en base al Lema 2.1, han generado nuevas distribuciones asimétricas, denominadas skew simétricas, esto es: Gupta y otros [35] consideran f como la función densidad normal, Laplace y Logistic, y G la función de distribución acumulada respectiva. Nadarajah y Kotz [53], reemplazan f por la densidad normal y G por la distribución acumulada normal, Student- t , Cauchy y Laplace, respectivamente. Gómez, Venegas y Bolfarine [31] consideran f

como la densidad normal, Student- t , Laplace y Logistic, y G como la función de distribución normal acumulada.

Arellano-Valle et al. [2] introducen otra alternativa de distribución skew simétrica, denominada skew-normal-generalizada (en adelante SGN), la cual es una mezcla en el parámetro de asimetría de las distribuciones skew-normal y normal. Más específicamente, si $Z|X = x \sim SN(x)$ y $X \sim N(\lambda_1, \lambda_2)$ entonces la distribución marginal, $Z \sim SGN(\lambda_1, \lambda_2)$, tiene como función densidad

$$f_Z(z; \lambda_1, \lambda_2) = 2\phi(z)\Phi\left(\frac{\lambda_1 z}{\sqrt{1 + \lambda_2 z^2}}\right), \quad z \in \mathbb{R}, \quad (2.1)$$

donde $\lambda_1 \in \mathbb{R}$, $\lambda_2 \geq 0$, ϕ y Φ son la pdf y función de distribución acumulada (cdf) de la distribución normal estándar, respectivamente. Para $\lambda_1 = 0$, o el caso límite $\lambda_2 = \infty$, la pdf del modelo SGN dada en (2.1) tiende a la pdf normal estándar $\phi(z)$, $z \in \mathbb{R}$, y para $\lambda_2 = 0$ a la pdf skew-normal de Azzalini [12] (en adelante SN) dada por $2\phi(z)\Phi(\lambda_1 z)$, $z \in \mathbb{R}$.

La distribución SGN ha sido considerada en diferentes trabajos, por ejemplo, Arellano-Valle et al. [5] considera este modelo en análisis de regresión Bayesiano, mientras que Sever et al. [61] lo usa en análisis discriminante; Arnold et al. [10] y Arellano-Valle et al. [6] obtienen una extensión bivariada y multivariada de SGN respectivamente; Gómez et al. [30] estudia el caso especial llamado skew-normal-curvada correspondiente al parámetro restringido $\lambda_2 = \lambda_1^2$; y Arellano-Valle et al. [7] estudia la matriz información de Fisher para el modelo SGN de localización y escala, mostrando que ésta es singular en $\lambda_1 = 0$ independiente del valor de λ_2 . Específicamente, para $\lambda_1 = 0$, estos autores encuentran que la función score de λ_2 es igual a cero, también que la función score del parámetro de localización es proporcional a la función score de λ_1 . Este hecho viola las condiciones usuales de regularidad estudiadas para la

distribución asintótica normal de los estimadores máximo verosímiles (EMVs). Por lo tanto, como fue mostrado por Arellano-Valle et al. [7], bajo estas circunstancias la metodología no estándar propuesta por Rotnitzky et al. [58] no puede ser aplicada para estudiar las propiedades asintóticas de los EMVs en el caso normal $\lambda_1 = 0$.

La familia SGN puede ser reparametrizada como $\lambda_1 = \lambda/\sqrt{\nu}$ y $\lambda_2 = 1/\nu$, donde $\lambda \in \mathbb{R}$ y $\nu > 0$. En términos de los nuevos parámetros de forma (λ, ν) , la pdf de SGN dada en (2.1) es modificada de la siguiente manera

$$f_Z(z; \lambda, \nu) = 2\phi(z)\Phi\left(\frac{\lambda z}{\sqrt{\nu + z^2}}\right), \quad z \in \mathbb{R}. \quad (2.2)$$

La correspondiente distribución es denominada skew-normal-generalizada modificada y es denotada por $Z \sim MSGN(\lambda, \nu)$. Nuevamente la pdf de MSGN (2.2) se reduce a la pdf $N(0,1)$ para $\lambda = 0$ o en el caso límite cuando $\nu = \infty$, y para la pdf signo-skew-normal (en adelante SSN) $2\phi(z)\Phi\{\lambda \text{sing}(z)\}$, $z \in \mathbb{R}$, donde $\text{sing}(z) = \pm 1$, cuando $\nu = 0$. Sin embargo, la familia MSGN no incluye al modelo SN en ningún caso. No es difícil mostrar que la función score de ν en el modelo MSGN es también cero en $\lambda = 0$; pero en este caso, cuando el modelo incluye el parámetro de localización μ , la dependencia lineal entre las funciones score de μ y λ en $\lambda = 0$ desaparece. Por lo tanto, la metodología no estándar propuesta por Rotnitzky et al. [58] ahora no es necesaria para estudiar las propiedades asintóticas de los EMVs en el caso normal $\lambda = 0$.

Casos particulares interesantes son obtenidos fijando λ_2 en (2.1) o ν en (2.2); por ejemplo, en este trabajo consideraremos el caso $\lambda_2 = \nu = 1$, bajo el cual las distribuciones SGN y MSGN son completamente equivalentes. Para este caso en especial nos referiremos simplemente como distribución skew-normal modificada (en adelante MSN). El interés en este modelo es

que es un buen competidor para el modelo SN, ya que ambos controlan la asimetría con un simple parámetro escalar, digamos $\lambda \in \mathbb{R}$, tal que si $\lambda = 0$ entonces se obtiene el modelo normal simétrico. Sin embargo, como se muestra en este trabajo, una importante ventaja del modelo MSN sobre el SN es que, en presencia de parámetros de localización y escala, la matriz de información de Fisher del modelo MSN es no singular en $\lambda = 0$. Por lo tanto, para el modelo MSN, las condiciones de regularidad estándar que conducen a una distribución asintótica normal de los EMVs también se mantiene bajo la hipótesis nula de normalidad dada por $H_0 : \lambda = 0$. Por otro lado, las funciones de verosimilitud de los modelos $\text{SN}(\lambda)$ y $\text{MSN}(\lambda)$ tienen algunas propiedades similares.

En particular, en muestras aleatorias con elementos de diferente signo, las correspondientes ecuaciones de verosimilitud tienen solución única, obteniéndose en ambos modelos un EMV finito para λ . Este hecho fue probado para el caso SN en Martínez et al. [50]. En este trabajo, se demuestra que este resultado también es válido para el modelo MSN. En caso contrario, cuando los elementos de la muestra tienen todos el mismo signo, las funciones de verosimilitud de los modelos $\text{SN}(\lambda)$ y $\text{MSN}(\lambda)$ son monótonas creciente, produciendo en ambos casos un EMV infinito para λ . Se reduce la probabilidad de este evento cuando aumenta el tamaño de las muestras, pero puede ocurrir con probabilidad no despreciable en muestras finitas, especialmente si λ está lejos de cero.

Se han propuesto diferentes métodos para resolver este problema en el contexto del modelo SN. Por ejemplo, basado en el método de reducción del sesgo de Firth [28], Sartori [60] propuso sustituir la ecuación de verosimilitud SN habitual de λ por una ecuación de verosimilitud modificada, y él prueba que para cualquier muestra la ecuación de verosimilitud modificada tiene al menos una solución finita. A esta solución, se le denomina estimador máximo

verosímil modificado. Resultados similares también son considerados para la distribución skew- t por Sartori [60] y Lagos-Álvarez y Jiménez-Gamero [41].

En este trabajo se aplica este enfoque al modelo MSN(λ). Como Sartori [60], se obtiene una ecuación de verosimilitud modificada para estimar el parámetro de forma del modelo MSN, y se muestra que el EMV modificado de λ es siempre finito. Para obtener el intervalo de confianza para λ , se considera el enfoque de cuasi-verosimilitud dado por Sartori [60] y es aplicado al modelo MSN. Cuando el modelo incluye parámetros de localización y escala, los EMVs no modificados de estos parámetros se combinan con el EMV modificado para el parámetro de forma.

2.2. Definición y propiedades del modelo MSN

En esta sección, se define formalmente la distribución de MSN y son examinados algunos de sus propiedades principales. También, se revisa la inferencia básica basada en el enfoque de máxima verosimilitud.

Diremos que una variable aleatoria Z tiene distribución skew-normal modificada, denotada por $Z \sim MSN(\lambda)$, si su densidad tiene la forma

$$f_Z(z; \lambda) = 2\phi(z)\Phi(\lambda u(z)), \quad z \in \mathbb{R}, \quad (2.3)$$

donde $u(z) = \frac{z}{\sqrt{1+z^2}}$ y $\lambda \in \mathbb{R}$. Aquí ϕ y Φ denotan las funciones de densidad y de distribución de la $N(0,1)$, respectivamente. Si $\lambda = 0$, entonces la densidad MSN en (2.3) se reduce a la densidad normal estándar, de lo contrario valores no nulos del parámetro λ afectarán directamente en la asimetría del modelo. En particular, cuando $\lambda \rightarrow \infty$ el modelo tiende a la distribución Half-Normal. La relevancia de esta clase de modelo skew-normal es que en

presencia de parámetros de localización y escala, la matriz de información de Fisher asociada correspondiente es no singular cuando $\lambda = 0$. La Figura 2.1, muestra la forma de la densidad para distintos valores de λ . En ella se puede apreciar la sutil diferencia entre el modelo MSN y SN para un mismo valor del parámetro, digamos $\lambda = 1$, indicado con línea punteada y segmentada.

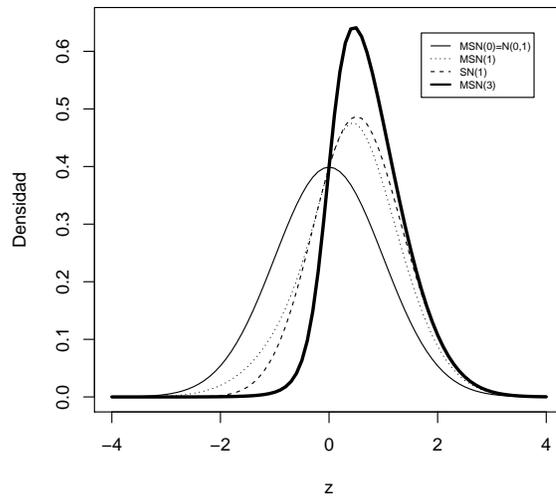


Figura 2.1: Gráficas de MSN(0) (línea continua), MSN(1) (línea punteada), SN(1) (línea segmentada) y MSN(3) (línea continua gruesa).

Las siguientes propiedades se obtiene de Arellano-Valle et al. [2]:

Propiedades : Sea $Z \sim MSN(\lambda)$, entonces:

1. $Z \sim MSN(\lambda) \Rightarrow -Z \sim MSN(-\lambda)$.
2. $Z \sim MSN(\lambda)$, entonces $|Z| \sim HN(0, 1)$.
3. $Z|S = s \sim SN(s)$ y $S \sim N(\lambda, 1)$ entonces $Z \sim MSN(\lambda)$.

4. $MSN(0) \equiv N(0, 1)$, y $MSN(\lambda) \equiv MSGN(\lambda, 1) \equiv SGN(\lambda, 1)$.

La Propiedad 3 presenta a la distribución MSN como una mezcla en el parámetro de asimetría de una distribución SN con una distribución normal. Para una discusión más detallada sobre esta propiedad y sus implicaciones, ver Arellano-Valle et al. [6].

2.2.1. Momentos

Los momentos, por Propiedad 3, se pueden obtener como resultado de una mezcla en el parámetro de asimetría entre la distribución SN y la distribución normal estándar. Luego, podemos escribir $E(Z^k) = E(E(Z^k|S = s))$, donde $E(Z^k|S = s)$ con $k = 0, 1, \dots$, son los momentos del modelo SN con parámetro de asimetría S . Luego, cuando k es par, los momentos de la distribución SN son constantes y por lo tanto coincidirán con los momentos del modelo MSN. Para el caso de momentos impares, consideraremos la expresión dada por Henze [37], así

$$\begin{aligned} E(Z^{2k+1}) &= E(E(Z^{2k+1}|S = s)) \\ &= E\left(\frac{b(2k+1)!}{2^k} \sum_{\nu=0}^k \frac{\nu!2^{2\nu}}{(2\nu+1)!(k-\nu)!} \frac{S^{2\nu+1}}{(1+S^2)^{(2k+1)/2}}\right) \\ &= \frac{b(2k+1)!}{2^k} \sum_{\nu=0}^k \frac{\nu!2^{2\nu}}{(2\nu+1)!(k-\nu)!} E\left(\frac{S^{2\nu+1}}{(1+S^2)^{(2k+1)/2}}\right), \end{aligned}$$

donde $k = 0, 1, 2, \dots$, $b = \sqrt{\frac{2}{\pi}}$. Considerando

$$\psi_{k,\nu} = \psi_{k,\nu}(\lambda) = E\left(\frac{S^{2\nu+1}}{(1+S^2)^{(2k+1)/2}}\right) = \int_{-\infty}^{\infty} \frac{s^{2\nu+1}}{(1+s^2)^{(2k+1)/2}} \phi(s-\lambda) ds,$$

y $\psi_k = \psi_{k,k}$, tenemos que

$$E(Z^{2k+1}) = \frac{b(2k+1)!}{2^k} \sum_{\nu=0}^k \frac{\nu!2^{2\nu}}{(2\nu+1)!(k-\nu)!} \psi_{k,\nu}.$$

Luego, los 4 primeros momentos son

$$E(Z) = b\psi_0,$$

$$E(Z^2) = 1,$$

$$E(Z^3) = b(3\psi_0 - \psi_1),$$

$$E(Z^4) = 3.$$

Los momentos impares también pueden ser obtenidos de Arellano [2], a través de la fórmula

$$E(Z^{2k+1}) = 2c_k - 2^k k! b,$$

para $k = 0, 1, 2, \dots$ y donde $c_k := c_k(\lambda) = \int_0^\infty x^k \phi(\sqrt{x}) \Phi(\lambda u(\sqrt{x})) dx$.

Generatriz de Momentos Como se mencionó anteriormente, en Propiedad 3, si $Z|S = s \sim SN(S)$ y $S \sim N(\lambda, 1)$ entonces $Z \sim MSN(\lambda)$, luego podemos escribir

$$M_Z(t) = E(E(e^{Zt}|S = s)) = E(M_{Z|S=s}(t)) = E\left(2e^{t^2/2} \Phi\left(\frac{S}{\sqrt{1+S^2}}t\right)\right). \quad (2.4)$$

Como $M_{Z|S=s}(t)$ corresponde a la generatriz de momentos del modelo skew-normal, sus derivadas existen y son finitas, entonces todos los momentos de Z se obtienen como sigue

$$E(Z^n) = \frac{\partial^n M_Z(t)}{\partial^n t} \Big|_{t=0} = E\left(\frac{\partial^n}{\partial^n t} \left[2e^{t^2/2} \Phi\left(\frac{S}{\sqrt{1+S^2}}t\right)\right] \Big|_{t=0}\right).$$

2.2.2. Coeficiente de asimetría y curtosis

Para el cálculo de los momentos centrales, utilizamos la siguiente relación

$$\mu'_2 = \mu_2 - \mu_1^2, \quad \mu'_3 = \mu_3 - 3\mu_1\mu_2 + 2\mu_1^3 \quad y \quad \mu'_4 = \mu_4 - 4\mu_1\mu_3 + 6\mu_1^2\mu_2 - 3\mu_1^4, \quad (2.5)$$

donde $\mu'_k = E(x^k)$. Entonces, los coeficientes de asimetría y de curtosis, en función de estos momentos, tienen las siguientes expresiones

$$\gamma_1 = \frac{\mu'_3}{(\mu'_2)^{3/2}}, \quad \gamma_2 = \frac{\mu'_4}{(\mu'_2)^2} \quad (2.6)$$

Luego, el coeficiente de asimetría, en función de λ , viene dado por la expresión

$$\gamma_1(\lambda) = \frac{2b^3\psi_0^3 - b\psi_1}{(1 - (b\psi_0)^2)^{3/2}},$$

El mínimo y máximo del coeficiente se obtiene de

$$\lim_{\lambda \rightarrow \pm\infty} \gamma_1(\lambda) = \pm \frac{\sqrt{2}(4 - \pi)}{(\pi - 2)^{3/2}} = \pm 0.995,$$

y la expresión para la curtosis es

$$\gamma_2(\lambda) = \frac{3 - 6(b\psi_0)^2 + 4b^2\psi_0\psi_1 - 3(b\psi_0)^4}{(1 - (b\psi_0)^2)^2}.$$

Entonces, el mínimo y máximo para este coeficiente esta dado por

$$\lim_{\lambda \rightarrow 0} \gamma_2(\lambda) = 3, \quad \lim_{\lambda \rightarrow \pm\infty} \gamma_2(\lambda) = \frac{3\pi^2 - 4\pi - 12}{(\pi - 2)^2} = 3.869.$$

Las gráficas para la simetría y curtosis, en función de λ , son respectivamente

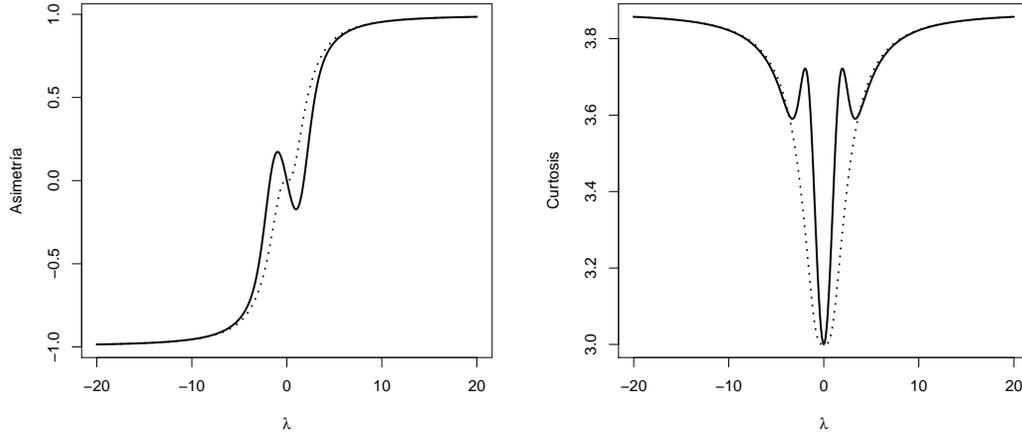


Figura 2.2: Coeficientes de asimetría y curtosis para distribución $MSN(\lambda)$ (línea continua) y $SN(\lambda)$ (línea punteada).

La Figura (2.2), muestra la asimetría como una función impar de λ , en donde las asíntotas horizontales son ± 0.995 , cuando $\lambda \rightarrow \pm\infty$, respectivamente. La curtosis es una función par de λ y crece asintóticamente hasta el valor 3.869, cuando $\lambda \rightarrow \pm\infty$. Se observa que ambos modelos tienen el mismo rango de asimetría y curtosis.

2.3. Estimación de máxima verosimilitud

2.3.1. Caso parámetro de forma

EMV de λ : Para una muestra aleatoria de tamaño n de $Z \sim MSN(\lambda)$, la función de log-verosimilitud de λ es

$$l(\lambda) = \frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^n z_i^2 + \sum_{i=1}^n \log(\Phi(\lambda u(z_i))). \quad (2.7)$$

Es fácil ver que la función $u(x)$, $x \in \mathbb{R}$, es monótona creciente. Es claro que de (2.7) el máximo valor de $l(\lambda)$ es en $\lambda = \infty$, si $z_i > 0$ para todo $i = 1, \dots, n$, y es en $\lambda = -\infty$, si $z_i < 0$ para todo $i = 1, \dots, n$. Un resultado análogo, fue notado por Liseo [46] para el modelo SN (ver también Sartori [60], Azzalini y Arellano-Valle [17]). Por lo tanto, en ambos casos el EMV de λ diverge cuando todas las observaciones tienen el mismo signo. Este comportamiento es explicado por el hecho que ambas distribuciones SN y MSN, el valor límite $\lambda = \infty$ corresponde precisamente a la distribución half-normal o χ , mientras que el valor límite para $\lambda = -\infty$ es la distribución $-\chi$.

Una consecuencia práctica de este resultado indeseable es que podría ocurrir con probabilidad no nula en tamaños de muestra moderados, incluso si el verdadero valor de λ es finito. En efecto, si Z_1, \dots, Z_n es una muestra aleatoria de distribución SN(λ) o MSN(λ), entonces la probabilidad de diverger del EMV de λ es $p(\lambda, n) = \prod_{i=1}^n P(Z_i < 0) + \prod_{i=1}^n P(Z_i > 0)$. En el caso SN(λ), la probabilidad es (Liseo [46]; Azzalini y Arellano-Valle [17])

$$p_{\text{SN}}(\lambda, n) = \left(\frac{1}{2} + \frac{1}{\pi} \arctan \lambda \right)^n + \left(\frac{1}{2} - \frac{1}{\pi} \arctan \lambda \right)^n.$$

Por lo tanto, aplicando Propiedad 3 al caso SN(λ), esta probabilidad es

$$p_{\text{MSN}}(\lambda, n) = \left[\frac{1}{2} + \frac{1}{\pi} E \{ \arctan(Z_0 + \lambda) \} \right]^n + \left[\frac{1}{2} - \frac{1}{\pi} E \{ \arctan(Z_0 + \lambda) \} \right]^n,$$

donde $Z_0 \sim N(0, 1)$. De $\arctan x \in (-\pi/2, \pi/2)$ para $x \in (-\infty, \infty)$, se sigue que para valores finitos de λ esta probabilidad converge rápidamente a 0 cuando $n \rightarrow \infty$. Sin embargo, para un tamaño de muestra pequeño o moderado, estas probabilidades pueden ser no despreciables, incluso para valores moderados de λ . La Tabla 2.1, reporta valores de $p_{\text{SN}}(\lambda, n)$ y $p_{\text{MSN}}(\lambda, n)$ para diferentes valores de λ y n .

Tabla 2.1: Probabilidad de divergencia de EMV $p(\lambda, n)$.

Modelo	$n = 20$	$n = 50$	$n = 100$	γ_1	γ_2
SN(3)	0.1152	0.0045	0.0000	0.6670	3.5097
SN(5)	0.2731	0.0389	0.0015	0.8509	3.7053
SN(10)	0.5219	0.1994	0.0397	0.9555	3.8232
MSN (3)	0.0871	0.0022	0.0000	0.5178	3.6000
MSN (5)	0.2580	0.0338	0.0011	0.8318	3.6918
MSN (10)	0.5212	0.1962	0.0385	0.9543	3.8219

Por otra parte, la ecuación de verosimilitud inducida por (2.7) es

$$U(\lambda) = \frac{\partial l(\lambda)}{\partial \lambda} = \sum_{i=1}^n u(z_i) \zeta\{\lambda u(z_i)\} = 0, \quad (2.8)$$

donde $\zeta(x) = \phi(x)/\Phi(x)$. Nótese que $\zeta(x) > 0$ es una función diferenciable y monótonamente decreciente desde ∞ a 0 cuando el rango x es $-\infty$ a ∞ (Sampford [59]). Como se ha visto anteriormente, si todas las observaciones en la muestra tienen el mismo signo, entonces el EMV de λ diverge. Afortunadamente, si al menos uno de los elementos de la muestra tiene un signo diferente al resto, la ecuación de estimación (2.8) tiene una raíz única. Lo anterior, se puede formalizar a través de la siguiente proposición:

Proposición 2.1. *Dada una muestra z_1, \dots, z_n iid de $MSN(\lambda)$, supongamos que existen i y j tal que $z_i > 0$ y $z_j < 0$, entonces la función dada en (2.8), tiene una única raíz.*

Demostración: Supongamos una muestra z_1, \dots, z_n son tales que $z_i > 0$ y $z_j < 0$ y así $u(z_i) > 0$ y $u(z_j) < 0$ para algún $i \neq j$, entonces (2.8) se puede separar en elemen-

tos positivos y negativos de la muestra, esto es

$$U(\lambda) = \sum_{i: z_i > 0} u(z_i) \zeta\{\lambda u(z_i)\} + \sum_{j: z_j < 0} u(z_j) \zeta\{\lambda u(z_j)\} = U_+(\lambda) + U_-(\lambda) = 0.$$

Claramente se sigue que la función score de $U(\lambda)$ es la suma de una función decreciente positivo, digamos $U_+(\lambda)$, y una función creciente negativa, digamos $U_-(\lambda)$, tal que $U(-\infty) = \infty + 0$ y $U(\infty) = 0 - \infty$. Esto implica que $U(\lambda)$ tiene una raíz. La monotonía de las funciones $U_+(\lambda) = \sum_{i: z_i < 0} u(z_i) \zeta\{\lambda u(z_i)\}$ y $U_-(\lambda) = \sum_{i: z_i < 0} u(z_i) \zeta\{\lambda u(z_i)\}$ implica la unicidad de esta raíz. \square

Un resultado similar para el modelo SN fue probada en Martínez et al. [50].

Primer estudio de simulación : Sea $\hat{\lambda}$ el EMV de λ . Con el fin de estudiar el sesgo de $\hat{\lambda}$ en muestras finitas, se presenta a continuación un pequeño estudio de simulación. Para generar las muestras de la distribución MSN se utilizan la Propiedad 3 y el método de composición (ver Tanner [63]).

La Tabla 2.2, informa de los resultados de la simulación de 5,000 iteraciones para diferentes valores del tamaño de muestra n y diferentes valores del parámetro de forma λ . Aunque el sesgo se calcula utilizando sólo aquellas muestras donde $\hat{\lambda}$ es finito, es bastante evidente el problema de la sobrestimación de λ , en particular para valores grandes de este parámetro y tamaño de muestra pequeño.

Tabla 2.2: Simulación de 5,000 iteraciones del modelo MSN(λ).

λ	n=20		n=50		n=100	
	$sesgo(\hat{\lambda})^a$	$\%(\hat{\lambda} < \infty)$	$sesgo(\hat{\lambda})^a$	$\%(\hat{\lambda} < \infty)$	$sesgo(\hat{\lambda})^a$	$\%(\hat{\lambda} < \infty)$
1	0.12883	99.92	0.03422	100.00	0.01632	100.00
3	1.30481	90.70	0.43527	99.84	0.16066	100.00
5	1.89707	74.84	1.71885	96.36	0.72847	99.92
7	2.19043	60.66	2.87950	90.14	2.83049	98.94
10	2.43177	48.20	5.15723	80.26	3.69982	96.04

Sesgo y porcentajes de casos en los que $\hat{\lambda}$ existe. ^aCalculado cuando $\hat{\lambda} < \infty$.

Como se mencionó anteriormente, la divergencia de $\hat{\lambda}$ puede también ocurrir cuando todas las observaciones son negativas, el cual se asocia a valores negativos del parámetro de asimetría λ . De acuerdo con la Propiedad 1, el estudio de simulación se desarrolla considerando sólo los valores positivos de λ .

2.3.2. Caso localización, escala y forma

La extensión a localización y escala de la distribución MSN se introduce considerando la variable aleatoria $X = \mu + \sigma Z$, con $\mu \in \mathbb{R}$ y $\sigma > 0$, cuyo pdf se define como $f_X(x; \mu, \sigma, \lambda) = \frac{1}{\sigma} f_Z\left(\frac{x-\mu}{\sigma}\right)$. Así, de (2.3) se sigue que

$$f_X(x; \mu, \sigma, \lambda) = \frac{2}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right) \Phi\left(\lambda u\left(\frac{x-\mu}{\sigma}\right)\right), \quad y \in \mathbb{R}. \quad (2.9)$$

En este caso de tres parámetros, se utiliza la notación $X \sim MSN(\mu, \sigma, \lambda)$, y bajo esto se explora el comportamiento de los EMV de (μ, σ, λ) .

EMV de $\theta = (\mu, \sigma, \lambda)^\top$: La log-verosimilitud para una variable aleatoria X que se distribuye MSN(θ) basada en (2.9), con $\theta = (\mu, \sigma, \lambda)$, está dada por

$$l(\theta) = -\log(\sqrt{2\pi}\sigma) - \frac{1}{2}z^2 + \log(\Phi(\lambda u(z))), \quad (2.10)$$

donde $z = \frac{x-\mu}{\sigma}$, $x \in \mathbb{R}$ y $\lambda \in \mathbb{R}$.

Incluso en el caso de tres parámetros, el EMV de λ puede diverger, el cual, al menos empíricamente, parece ocurrir con más frecuencia que en el caso de un sólo parámetro de forma.

Como en el modelo SN(μ, σ, λ), este hecho también ocurre en el modelo MSN(μ, σ, λ) cuando los elementos de la muestra y_1, \dots, y_n son tales que, para todo $i = 1, \dots, n$, $y_i - \hat{\mu} > 0$ o $y_i - \hat{\mu} < 0$, donde $\hat{\mu}$ es el EMV del parámetro de localización μ . En tales situaciones, la función perfil de verosimilitud definida por $l^P(\lambda) = l(\hat{\mu}(\lambda), \hat{\sigma}(\lambda), \lambda)$, donde $(\hat{\mu}(\lambda), \hat{\sigma}(\lambda))$ es el EMV de (μ, σ) para λ fijo, es monótona creciente y converge a un valor constante igual a $l(\hat{\mu}, \hat{\sigma}, \infty)$ cuando λ va a ∞ , lo que indica que el EMV $(\hat{\mu}, \hat{\sigma})$ de (μ, σ) siempre existe. Hechos similares fueron observados por Sartori [60] para el modelo de tres parámetros SN(μ, σ, λ).

El panel superior de la Figura 2.3 muestra el perfil de las funciones de log-verosimilitud de λ de una distribución MSN, para diferentes tamaños de muestra ($n = 20, 50, 100$), simulada de una distribución N(0,1). Es claro, que estos gráficos no tienen un punto de inflexión en $\lambda = 0$, el cual se relaciona con el hecho de que la matriz de información es no singular en ese punto. El panel inferior muestra el perfil de las funciones de log-verosimilitud de SN sobre la base de las mismas muestras, todos tienen un punto de inflexión en $\lambda = 0$; en este caso, la matriz de información es singular en $\lambda = 0$.

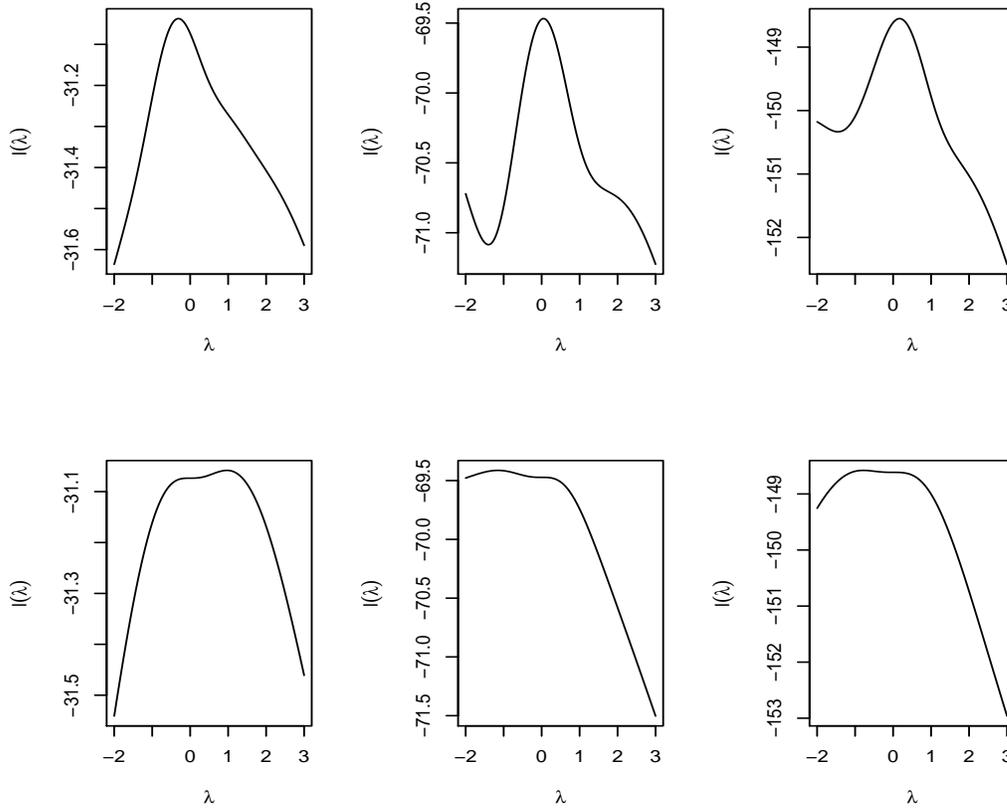


Figura 2.3: Perfil de las funciones log-verosimilitud para los modelos MSN (panel superior) y SN (panel inferior), para muestras de tamaños $n = 20$ (izquierda), $n = 50$ (centro) y $n = 100$ (derecha) simulada de la distribución $MSN(0) \equiv SN(0) \equiv N(0,1)$.

2.3.3. Matriz información

Matriz información observada: Sea una muestra aleatoria x_1, x_2, \dots, x_n de una variable aleatoria $X \sim MSN(\theta)$, con $\theta = (\mu, \sigma, \lambda)$, entonces la log-verosimilitud en asociada a dicha muestra está dada por

$$l(\theta) = \frac{n}{2} \log\left(\frac{2}{\pi}\right) - n \log(\sigma) - \frac{1}{2} \sum_{i=1}^n z_i^2 + \sum_{i=1}^n \log(\Phi(\lambda u(z_i))), \quad (2.11)$$

donde $z_i = \frac{x_i - \mu}{\sigma}$, $u(x) = \frac{x}{\sqrt{1+x^2}}$, $x_i \in \mathbb{R}$ y $\lambda \in \mathbb{R}$. Tenemos que el vector score asociado a (2.11) es

$$\begin{aligned} S_\mu &= \frac{n}{\sigma} (\bar{z} - \lambda \bar{\rho}_{03} \zeta(\lambda u(z_i))), \\ S_\sigma &= \frac{n}{\sigma} \left(-1 + \bar{z}^2 - \lambda \bar{\rho}_{13} \right), \\ S_\lambda &= n \lambda \bar{\rho}_{11}, \end{aligned}$$

donde

$$\begin{aligned} \rho_{nmi} &= \frac{z_i^n}{(1+z_i^2)^{m/2}} \zeta(\lambda u(z_i)), \\ \eta_{nmi} &= \frac{z_i^n}{(1+z_i^2)^m} \zeta^2(\lambda u(z_i)), \\ \bar{\eta}_{nm} &= \frac{1}{n} \sum_{i=1}^n \eta_{nmi}, \\ \bar{\rho}_{nm} &= \frac{1}{n} \sum_{i=1}^n \rho_{nmi}, \\ \zeta(x) &= \frac{\phi(x)}{\Phi(x)}. \end{aligned}$$

Las entradas de la matriz información observada de la distribución MSN, asociada a la log-verosimilitud dada en (2.11), según la fórmula

$$J_{\theta_i \theta_j} = -\frac{\partial^2 l(\theta)}{\partial \theta_i \partial \theta_j}, \quad i, j = 1, 2, 3 \quad (2.12)$$

están dadas por

$$\begin{aligned}
J_{\mu\mu} &= -\frac{n}{\sigma^2} (1 + 3\lambda\bar{\rho}_{15} + \lambda^3\bar{\rho}_{17} + \lambda^2\bar{\eta}_{03}), \\
J_{\mu\sigma} &= \frac{n}{\sigma^2} \left(-2\bar{z} + \lambda\bar{\rho}_{05} - 2\lambda\bar{\rho}_{25} - \widehat{\lambda}^3\bar{\rho}_{27} - \lambda^2\bar{\eta}_{13} \right), \\
J_{\mu\lambda} &= -\frac{n}{\sigma} (\bar{\rho}_{03} - \lambda^2\bar{\rho}_{25} - \lambda\bar{\eta}_{12}), \\
J_{\sigma\sigma} &= \frac{n}{\sigma^2} \left(1 - 3\bar{z}^2 + \lambda\bar{\rho}_{13} + \lambda\bar{\rho}_{15} - 2\lambda\bar{\rho}_{35} - \lambda^3\bar{\rho}_{37} - \lambda^2\bar{\eta}_{23} \right), \\
J_{\sigma\lambda} &= \frac{n}{\sigma} (-\bar{\rho}_{13} + \lambda^2\bar{\rho}_{35} + \lambda\bar{\eta}_{22}), \\
J_{\lambda\lambda} &= -n(\lambda\bar{\rho}_{33} + \bar{\eta}_{21}).
\end{aligned}$$

Para más detalle ver Apéndice 2.6.

Matriz información de Fisher La log-verosimilitud para una variable aleatoria X que se distribuye MSN(θ), con $\theta = (\mu, \sigma, \lambda)$, dada en (2.11), tiene asociado el siguiente vector score

$$\begin{aligned}
S_\mu &= \frac{1}{\sigma} \left(Z - \frac{\lambda}{(1+Z^2)^{3/2}} \zeta(\lambda u(Z)) \right), \\
S_\sigma &= \frac{1}{\sigma} \left(-1 + Z^2 - \frac{\lambda Z}{(1+Z^2)^{3/2}} \zeta(\lambda u(Z)) \right), \\
S_\lambda &= u(Z) \zeta(\lambda u(Z)),
\end{aligned}$$

donde $Z = \frac{X-\mu}{\sigma}$.

Las entradas de la matriz información de Fisher, están dadas por la fórmula

$$I_{\theta_i\theta_j} = -E \left(\frac{\partial^2 l(\theta)}{\partial \theta_i \partial \theta_j} \right), \quad i, j = 1, 2, 3. \quad (2.13)$$

Así, para el modelo en estudio se tienen las siguientes expresiones (Ver Apéndice 2.6)

$$\begin{aligned}
I_{\mu\mu} &= \frac{1}{\sigma^2} (1 + \lambda^2 \eta_{03}), \\
I_{\sigma\mu} &= -\frac{1}{\sigma^2} (-2E(Z) + \lambda\rho_{05} - 2\lambda\rho_{25} - \lambda^3\rho_{27} - \lambda^2\eta_{13}), \\
I_{\mu\lambda} &= \frac{1}{\sigma} (\rho_{03} - \lambda^2\rho_{25} - \lambda\eta_{12}), \\
I_{\sigma\sigma} &= \frac{1}{\sigma^2} (2 + \lambda^2\eta_{23}), \\
I_{\lambda\sigma} &= -\frac{\lambda}{\sigma}\eta_{22}, \\
I_{\lambda\lambda} &= \eta_{21},
\end{aligned}$$

donde

$$\begin{aligned}
\rho_{nm} &= E\left(\frac{Z^n}{(1+Z^2)^{m/2}}\zeta(\lambda u(Z))\right), \quad \rho_{nm} = 0 \quad \text{si } n = \text{impar}, \\
\eta_{nm} &= E\left(\frac{Z^n}{(1+Z^2)^m}\zeta^2(\lambda u(Z))\right), \\
\zeta = \zeta(x) &= \frac{\phi(x)}{\Phi(x)}.
\end{aligned}$$

Considerando el caso $\lambda = 0$, tenemos que la matriz información de Fisher de MSN es

$$I_{\text{MSN}}(\mu, \sigma, 0) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 & \frac{d_1}{\sigma} \\ 0 & \frac{2}{\sigma^2} & 0 \\ \frac{d_1}{\sigma} & 0 & d_2 \end{pmatrix},$$

donde $d_1 = \rho_{03}(0) = 2(2/\pi)^{1/2} \int_0^\infty \frac{\phi(x)}{(1+x^2)^{3/2}} dx$ y $d_2 = \eta_{21}(0) = \frac{2}{\pi}[1 - (2\pi)^{1/2}e^{1/2}\Phi(-1)]$ deben ser calculados numéricamente. Se deduce que la matriz de información $I_{\text{MSN}}(\mu, \sigma, \lambda)$ es no singular en $\lambda = 0$, con inversa

$$I_{\text{MSN}}^{-1}(\mu, \sigma, 0) = \begin{pmatrix} \frac{d_2}{d_2-d_1^2} \sigma^2 & 0 & -\frac{d_1}{d_2-d_1^2} \sigma \\ 0 & \frac{1}{2} \sigma^2 & 0 \\ -\frac{d_1}{d_2-d_1^2} \sigma & 0 & \frac{1}{d_2-d_1^2} \end{pmatrix}.$$

En particular, si $\hat{\theta} = (\hat{\mu}, \hat{\sigma}, \hat{\lambda})^\top$ es el EMV de $\theta = (\mu, \sigma, \lambda)^\top$ basada en una muestra aleatoria de tamaño n , entonces bajo la hipótesis de simetría normal $\lambda = 0$, $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N_3(0, I_{\text{MSN}}^{-1}(\theta_0))$, cuando $n \rightarrow \infty$, y donde $\theta_0 = (\mu, \sigma, 0)^\top$. Por lo tanto, se puede utilizar la teoría de la distribución normal inmediatamente para probar $H_0 : \lambda = 0$ versus $H_1 : \lambda \neq 0$. Por ejemplo, H_0 se rechaza si $|\hat{\lambda}|$ es "bastante grande". De acuerdo con la teoría, se puede calcular el estadístico estandarizado $\hat{z} = \sqrt{n(d_2 - d_1^2)} \hat{\lambda}$ y p-valor = $P(|\hat{Z}| > |\hat{z}|)$, donde $\hat{Z} \sim N(0, 1)$.

Nótese que estos últimos resultados no son aplicables en el contexto SN, ya que la matriz de información $I_{\text{SN}}(\mu, \sigma, \lambda)$ es singular en $\lambda = 0$. Para superar este problema, Azzalini [12] propuso la llamada parametrización centrada (véase también Chiogna [26], y Arellano-Valle y Azzalini [4]).

Segundo estudio de simulación: Las Tablas 2.3, 2.4 y 2.5 informan los resultados sobre el sesgo de los EMV $(\hat{\mu}, \hat{\sigma}, \hat{\lambda})$ de (μ, σ, λ) , que se basan en 5,000 simulaciones del modelo $\text{MSN}(\mu, \sigma, \lambda)$ para diferentes valores de sus parámetros y diferentes tamaños de muestra. Específicamente, son considerados los valores de $\mu = \pm 1$, $\sigma = 1$, $\lambda = 3, 5, 7, 10$ y $n = 50, 100, 200$. Estos resultados muestran que el sesgo de $\hat{\mu}$ y $\hat{\sigma}$ son bastante pequeños, sin importar el tamaño de la muestra y el verdadero valor de λ . Lo contrario ocurre con el sesgo de $\hat{\lambda}$, que muestra una clara sobrestimación de λ . También, en este caso la divergencia de $\hat{\lambda}$ se produce con más frecuencia que en el caso parámetro de forma. De acuerdo con estos resultados, es suficiente para aplicar el método de Firth [28] y reducir sólo el sesgo de $\hat{\lambda}$, por lo tanto implementar un método de estimación de dos etapas para los tres parámetros del modelo $\text{MSN}(\mu, \sigma, \lambda)$, como lo sugiere Sartori [60] para el modelo $\text{SN}(\mu, \sigma, \lambda)$.

Tabla 2.3: Simulación de 5,000 iteraciones del modelo MSN(0, 1, λ)

λ	n	$sesgo(\hat{\mu})$	$sesgo(\hat{\sigma})$	$sesgo(\hat{\lambda})^a$	$\%(\hat{\lambda} < \infty)$
3	50	0.02172	-0.01474	0.82271	96.80
	100	0.00919	-0.00619	0.39387	99.88
	200	0.00144	-0.00104	0.14659	100.00
5	50	0.00603	-0.00726	1.81600	87.54
	100	0.00122	-0.00504	1.50431	98.28
	200	0.00118	-0.00286	0.52377	99.96
7	50	0.00951	-0.01406	1.84417	77.44
	100	0.00146	-0.00490	2.50426	94.84
	200	0.00092	-0.00136	1.26313	99.70
10	50	0.01562	-0.01497	1.38927	65.68
	100	0.00228	-0.00428	3.72836	87.66
	200	0.00041	-0.00236	2.56727	98.36

Sesgo de $\hat{\mu}$, $\hat{\sigma}$ y $\hat{\lambda}$ y porcentajes de casos en los que $\hat{\lambda}$ existe. ^aCalculado cuando $\hat{\lambda} < \infty$.

Tabla 2.4: Simulación de 5,000 iteraciones del modelo MSN($-1, 1, \lambda$).

λ	n	$sesgo(\hat{\mu})$	$sesgo(\hat{\sigma})$	$sesgo(\hat{\lambda})^a$	$\%(\hat{\lambda} < \infty)$
3	50	0.01893	-0.00991	0.87821	96.38
	100	0.00640	-0.00403	0.36423	99.90
	200	0.00210	-0.00162	0.12564	100.00
5	50	0.00860	-0.00978	1.77914	86.88
	100	0.00085	-0.00308	1.45500	97.90
	200	-0.00005	0.00009	0.57297	99.94
7	50	0.01051	-0.01273	1.97267	77.28
	100	0.00375	-0.00387	2.22588	94.72
	200	0.00022	-0.00051	1.24983	99.58
10	50	0.01504	-0.01986	1.43642	65.48
	100	0.00399	-0.00617	3.72827	88.68
	200	0.00060	-0.00258	2.79012	98.36

Sesgo de $\hat{\mu}$, $\hat{\sigma}$ y $\hat{\lambda}$ y porcentajes de casos en los que $\hat{\lambda}$ existe. ^aCalculado cuando $\hat{\lambda} < \infty$.

Tabla 2.5: Simulación de 5,000 iteraciones del modelo MSN(1, 1, λ).

λ	n	$sesgo(\hat{\mu})$	$sesgo(\hat{\sigma})$	$sesgo(\hat{\lambda})^a$	$\%(\hat{\lambda} < \infty)$
3	50	0.02264	-0.01151	0.96230	95.50
	100	0.00672	-0.00537	0.41651	98.96
	200	0.00471	-0.00353	0.12541	99.86
5	50	0.00779	-0.00876	1.81196	87.14
	100	0.00013	-0.00247	1.40356	98.30
	200	-0.00060	-0.00120	0.54906	99.96
7	50	0.01338	-0.01424	1.92785	78.30
	100	0.00026	-0.00248	2.55455	94.64
	200	-0.00018	-0.00069	1.28891	99.70
10	50	0.01476	-0.01837	1.59102	65.52
	100	0.00371	-0.00570	3.73348	88.34
	200	0.00194	-0.00275	2.65022	98.74

Sesgo de $\hat{\mu}$, $\hat{\sigma}$ y $\hat{\lambda}$ y porcentajes de casos en los que $\hat{\lambda}$ existe. ^aCalculado cuando $\hat{\lambda} < \infty$.

2.4. Score modificada para la familia MSN

En esta sección, la metodología descrita anteriormente se aplica a la familia MSN asumiendo $\theta = \lambda$, y así reducir el sesgo en el EMV del parámetro que controla la asimetría.

2.4.1. Caso parámetro de forma

La log-verosimilitud $l(\lambda)$ y la función score $U(\lambda) = l'(\lambda)$ basada en n observaciones independientes de $Z \sim MSN(\lambda)$ están definidas en (2.7) y (2.8), respectivamente. La cantidad de información observada, derivada de (2.8), es

$$j(\lambda) = -l''(\lambda) = \lambda \sum_{i=1}^n u^3(z_i) \zeta\{\lambda u(z_i)\} + \sum_{i=1}^n u^2(z_i) \zeta^2\{\lambda u(z_i)\}.$$

Ahora, consideremos las funciones

$$a_{km}(\lambda) = E_{\lambda}[u^k(Z) \zeta^m\{\lambda u(Z)\}],$$

y como $a_{k1}(\lambda) = 0$ cuando k es impar, se sigue que

$$I(\lambda) = na_{22}(\lambda), \quad \nu_{\lambda,\lambda,\lambda}(\lambda) = na_{33}(\lambda), \quad \nu_{\lambda,\lambda\lambda}(\lambda) = -n\{a_{42}(\lambda) + a_{33}(\lambda)\}.$$

Nótese que si $u(z)$ es reemplazado por z , entonces este último resultado corresponde al modelo $SN(\lambda)$. Por lo tanto, reemplazando en (1.2) la modificación se convierte en

$$M(\lambda) = -\frac{\lambda a_{42}(\lambda)}{2 a_{22}(\lambda)} = -\frac{\lambda A_4(\lambda)}{2 A_2(\lambda)}, \quad (2.14)$$

donde $A_{2k}(\lambda) = \int_0^{\infty} u^{2k}(z) \phi(z) h\{\lambda u(z)\} dz$, $k = 1, 2$ y $h(x) = \zeta(x) \zeta(-x)$. Como $h(-x) = h(x)$, se sigue que las funciones $A_{2k}(\lambda)$, con $k = 1, 2$, son pares. Por lo tanto, en (2.14) la modificación $M(\lambda)$ es claramente una función impar. Además, como se muestra en el panel izquierdo de la Figura 2.4, las modificaciones $M_{SN}(\lambda)$ y $M_{MSN}(\lambda)$ asociado con los modelos $SN(\lambda)$ y $MSN(\lambda)$, son funciones acotadas, con valores máximos y mínimos dados por $M_{SN}(\pm 1.07) = \pm 0.83$ y $M_{MSN}(\pm 2.96) = \pm 0.55$, y tienden a cero cuando λ diverge. El panel de la derecha de la Figura 2.4, muestra la forma de las funciones obtenidas mediante la integración de estas modificaciones.

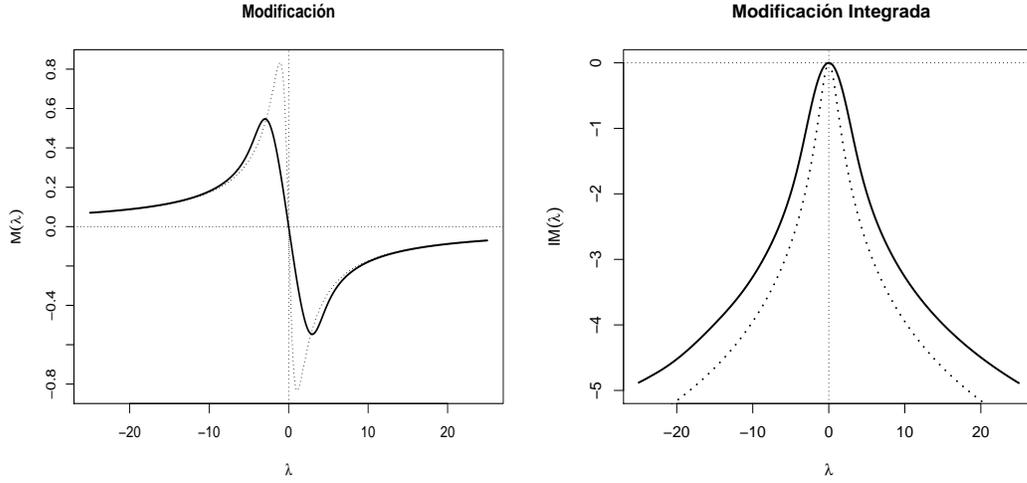


Figura 2.4: Score modificada (panel izquierdo) y su integral (panel derecho) para los modelos $SN(\lambda)$ (línea punteada) y $MSN(\lambda)$ (línea continua).

Como lo garantiza la siguiente proposición, en el modelo $MSN(\lambda)$ la ecuación de verosimilitud $U_M(\lambda) = U(\lambda) + M(\lambda) = 0$, con $U(\lambda)$ y $M(\lambda)$ definido en (2.8) y (2.14), respectivamente, siempre ofrece una solución finita, aunque el EMV no modificado de λ pueda diverger. Esto está garantizado por la Proposición 2.2 que se da a continuación, debido a que $M(\lambda) = \Omega(\lambda^{-1})$. La notación $g = \Omega(h)$ es equivalente a $g = O(h)$ y $h = O(g)$, y significa que g/h está acotado por arriba y por abajo. Dado que la función $M(\lambda)$ es impar, este resultado puede probarse, sin perder generalidad, para $\lambda > 0$. Por ejemplo, si todas las observaciones son positivas, entonces la función score en (2.8) es siempre positiva, esto implica que la ecuación de verosimilitud $U(\lambda) = 0$ no tiene una solución finita. También, a partir de (2.8) se desprende que $U(\lambda) = O_p(e^{-\lambda^2})$ y de la Proposición 2.2 que $M(\lambda) = O(\lambda^{-1})$. Así, como $M(\lambda)$ es negativo para $\lambda > 0$, $U_M(\lambda) = U(\lambda) + M(\lambda) \rightarrow 0^-$ cuando $\lambda \rightarrow \infty$. Pero, como $U_M(0) = U(0) + M(0) = U(0) > 0$, debe existir al menos una solución finita $\hat{\lambda}_M$ tal que

$$U_M(\widehat{\lambda}_M) = 0.$$

Proposición 2.2. *Sea $M(\lambda)$ la función modificación para el modelo $MSN(\lambda)$. Entonces, $M(\lambda) = \Omega(\lambda^{-1})$, es decir, ambas colas convergen a razón λ^{-1} .*

Demostración:

Primeramente, la prueba de que $a_{22}(\lambda)$ es una función simétrica y decreciente respecto de λ , se puede obtener en forma similar a la realizada por Sartori [60]. Esto es

$$\begin{aligned} a_{22}(\lambda) &= 2 \int_0^\infty u^2(z)\phi(z) \frac{\phi^2(\lambda u(z))}{\Phi(\lambda u(z))} dz + 2 \int_0^\infty u^2(-z)\phi(-z) \frac{\phi^2(-\lambda u(z))}{1 - \Phi(\lambda u(z))} dz \\ &= 2 \int_0^\infty u^2(z)\phi(z) \frac{\phi^2(-\lambda u(z))}{\Phi(\lambda u(z))[1 - \Phi(\lambda u(z))]} dz \\ &= a_{22}(-\lambda). \end{aligned}$$

Es decir, $a_{22}(\lambda)$ es simétrica respecto de λ .

Ahora, se analiza el signo de la derivada de $a_{22}(\lambda)$ para $\lambda > 0$

$$\frac{\partial a_{22}(\lambda)}{\partial \lambda} = -2 \int_0^\infty u^3(z)\phi(z) \frac{\phi^2(\lambda u(z))}{\Phi^2(\lambda u(z))[1 - \Phi(\lambda u(z))]^2} H(\lambda u(z)) dz,$$

donde $H(s) = 2s\Phi(s)[1 - \Phi(s)] + \phi(s)[1 - 2\Phi(s)]$. $\forall s > 0$, $s\Phi(-s) < \phi(s)$ (Sartori [60]), entonces $H(s) > 0$. Por lo tanto, $a_{22}(\lambda)$ es una función decreciente de λ .

Por la simetría de $a_{22}(\lambda)$, sólo se prueba la convergencia de la cola derecha, considerando $\lambda > 1$.

Es conveniente tener las siguientes consideraciones para la demostración:

$\forall z > 0$ se tiene que:

1i) Para $\lambda > 1$, $\frac{1}{\lambda}u(\lambda z) < u(z) < u(\lambda z) < z < \lambda z$.

2i) Las funciones $\Phi(z)$ y $u(z)$ son crecientes. La función $\phi(z)$ es decreciente.

$$3i) \frac{1}{1-\Phi(\lambda u(z))} < \frac{1}{1-\Phi(\lambda z)} < \frac{1}{1-\Phi(1)} = c = \text{cte para } 0 < z < 1/\lambda \text{ y } \lambda > 0.$$

$$4i) \frac{\phi(z)}{1-\Phi(z)} < \frac{z^2+1}{z}.$$

$$5i) 1 < z < \lambda \Rightarrow \frac{z}{\sqrt{2}} < \lambda u(z/\lambda) \Rightarrow \phi(\lambda u(z/\lambda)) < \phi\left(\frac{z}{\sqrt{2}}\right).$$

$$6i) 1 < \lambda < z \Rightarrow \frac{\lambda}{\sqrt{2}} < \lambda u(z/\lambda) \Rightarrow \phi(\lambda u(z/\lambda)) < \phi\left(\frac{\lambda}{\sqrt{2}}\right).$$

Descomponiendo $a_{22}(\lambda)$ se obtiene que

$$\begin{aligned} a_{22}(\lambda) &= A(\lambda) + B(\lambda) \\ &= 2 \int_0^\infty u^2(z)\phi(z) \frac{\phi^2(\lambda u(z))}{\Phi(\lambda u(z))} dz + 2 \int_0^\infty u^2(z)\phi(z) \frac{\phi^2(\lambda u(z))}{1-\Phi(\lambda u(z))} dz. \end{aligned}$$

Resolviendo $A(\lambda)$:

$\forall z > 0$ y $\lambda > 1$, se tiene que $1 < 1/\Phi(\lambda u(z)) < 2$, entonces $A^*(\lambda) < A(\lambda) < 2A^*(\lambda)$ donde

$$A^*(\lambda) = 2 \int_0^\infty u^2(z)\phi(z)\phi^2(\lambda u(z))dz = I_3 + I_4,$$

con

$$\begin{aligned} I_3 &= 2 \int_0^{1/\lambda} u^2(z)\phi(z)\phi^2(\lambda u(z))dz, \\ I_4 &= 2 \int_{1/\lambda}^\infty u^2(z)\phi(z)\phi^2(\lambda u(z))dz. \end{aligned}$$

Tenemos la siguiente desigualdad

$$O(\lambda^{-3}) = I_1 < A^*(\lambda) < 2(I_3 + I_4) < O(\lambda^{-3}).$$

Las cotas para I_1 , I_3 y I_4 se obtienen utilizando 1i), 2i) y haciendo el cambio de variable

$r = \lambda z$, esto es

$$\begin{aligned}
I_1 &= \frac{2}{\lambda^2} \int_0^\infty u^2(\lambda z) \phi(\lambda z) \phi^2(\lambda z) dz = \frac{1}{\lambda^3} I_2 = O(\lambda^{-3}), \\
I_2 &= 2 \int_0^\infty u^2(r) \phi^3(r) dr < \infty, \\
I_3 &< 2 \int_0^{1/\lambda} u^2(z) \phi^3(u(\lambda z)) dz < 2 \int_0^{1/\lambda} z^2 \phi^3(u(\lambda z)) dz = \frac{1}{\lambda^3} I_6 = O(\lambda^{-3}), \\
I_6 &= 2 \int_0^1 r^2 \phi^3(u(r)) dr < \infty, \\
I_4 &< 2 \int_{1/\lambda}^\infty z^2 \phi(z) \phi^2(\lambda u(z)) dz = \frac{2}{\lambda^3} \int_1^\infty r^2 \phi(r/\lambda) \phi^2(\lambda u(r/\lambda)) dz = \frac{1}{\lambda^3} (I_7^{2,2} + I_8^{2,2}),
\end{aligned}$$

donde $I_7^{2,2}$ y $I_8^{2,2}$ se obtienen de la expresión general que sigue, usando 5i) y 6i), entonces

$$\begin{aligned}
I_7^{k,l} &= 2 \int_1^\lambda r^k \phi(r/\lambda) \phi^l(\lambda u(r/\lambda)) dr < 2 \int_1^\lambda r^k \phi(r/\lambda) \phi^l(r/\sqrt{2}) dr < I_9, \\
I_9 &= 2 \phi(0) \int_0^\infty r^k \phi^l(r/\sqrt{2}) dr < \infty, \\
I_8^{k,l} &= 2 \int_\lambda^\infty r^k \phi(r/\lambda) \phi^l(\lambda u(r/\lambda)) dr < 2 \phi^l(\lambda/\sqrt{2}) \int_\lambda^\infty r^k \phi(r/\lambda) dr = I_{12} \\
I_{12} &= \lambda^{k+1} \phi^l(\lambda/\sqrt{2}) I_{10} = O(\lambda^{k+1} e^{-\lambda^2}), \\
I_{10} &= 2 \int_1^\infty v^k \phi(v) dv < \infty.
\end{aligned}$$

Luego

$$\begin{aligned}
I_4 &< O(\lambda^{-3})(I_7^{2,2} + I_8^{2,2}) \\
&< O(\lambda^{-3})[O(1) + O(\lambda^3 e^{-\lambda^2})] \\
&< O(\lambda^{-3})O(\text{máx}\{1, \lambda^3 e^{-\lambda^2}\}) \\
&< O(\lambda^{-3})O(1) \\
&< O(\lambda^{-3}).
\end{aligned}$$

Entonces, $A^*(\lambda) = \Omega(\lambda^{-3})$, por lo tanto $A(\lambda) = \Omega(\lambda^{-3})$.

Resolviendo $B(\lambda)$: Se tiene que $A^*(\lambda) = O(\lambda^{-3}) < B(\lambda)$, ya que $1 - \Phi(\lambda u(z)) < 1$. Ahora, acotamos por arriba a $B(\lambda)$,

$$\begin{aligned} B(\lambda) &= B_1(\lambda) + B_2(\lambda) \\ &= 2 \int_0^{1/\lambda} u^2(z)\phi(z) \frac{\phi^2(\lambda u(z))}{1 - \Phi(\lambda u(z))} dz + 2 \int_{1/\lambda}^{\infty} u^2(z)\phi(z) \frac{\phi^2(\lambda u(z))}{1 - \Phi(\lambda u(z))} dz. \end{aligned}$$

De 3i) se tiene

$$B_1(\lambda) < 2c \int_0^{1/\lambda} u^2(z)\phi(z)\phi^2(\lambda u(z))dz = cI_3 < O(\lambda^{-3}).$$

Usando 4i), 1i) y haciendo el cambio de variable $r = \lambda z$, tenemos que

$$\begin{aligned} B_2(\lambda) &< 2 \int_{1/\lambda}^{\infty} u^2(z)\phi(z)\phi(\lambda u(z)) \frac{(\lambda u(z))^2 + 1}{(\lambda u(z))} dz < I_{11}, \\ I_{11} &= \frac{2}{\lambda} \int_{1/\lambda}^{\infty} z\phi(z)\phi(\lambda u(z))(2(\lambda z)^2)dz = 4\lambda \int_{1/\lambda}^{\infty} z^3\phi(z)\phi(\lambda u(z))dz \\ &= \frac{4}{\lambda^3} \int_1^{\infty} r^3\phi(r/\lambda)\phi(\lambda u(r/\lambda))dr = O(\lambda^{-3})(2I_7^{3,1} + 2I_8^{3,1}). \end{aligned}$$

Luego

$$\begin{aligned} I_{11} &= O(\lambda^{-3})(2I_7^{3,1} + 2I_8^{3,1}) \\ &< O(\lambda^{-3})[O(1) + O(\lambda^4 e^{-\lambda^2})] \\ &< O(\lambda^{-3})O(\text{máx}\{1, \lambda^4 e^{-\lambda^2}\}) \\ &< O(\lambda^{-3})O(1) \\ &< O(\lambda^{-3}) \end{aligned}$$

Entonces, $B(\lambda) = \Omega(\lambda^{-3})$ y en consecuencia $a_{22}(\lambda) = \Omega(\lambda^{-3})$. Similarmente al desarrollo anterior, se demuestra la simetría, monotonía y el orden de convergencia respecto de λ para $a_{42}(\lambda)$, en este caso $a_{42}(\lambda) = \Omega(\lambda^{-5})$. Por lo tanto, se ha demostrado que $M(\lambda) = \Omega(\lambda^{-1})$. \square

Primer ejemplo. Consideremos la distribución MSN(λ) con $\lambda = 5$ de la que generamos una muestra aleatoria de tamaño $n = 20$: $z = (0.350, -0.046, -0.061, 0.869, 0.356, 1.192, 0.108, 0.308, 1.335, 1.660, 0.527, 0.446, 0.605, 0.311, 0.301, 0.396, 0.359, 1.304, 0.525, 0.558)$. La Figura 2.5, muestra en el panel izquierdo, las funciones score no modificada y modificada de $U(\lambda)$ y $U_M(\lambda)$, respectivamente, y en el panel derecho, la correspondientes funciones de log-verosimilitud relativas de $l(\lambda)$ y $l_M(\lambda)$. Para esta muestra, son calculados los EMV no modificado y modificado $\hat{\lambda}$ y $\hat{\lambda}_M$, respectivamente. También, se calcula el intervalo de confianza de 0.95 basado en $W_M(\lambda) = 2\{l_M(\hat{\lambda}_M) - l_M(\lambda)\}$. Estos resultados son resumidos en la Tabla 2.6.

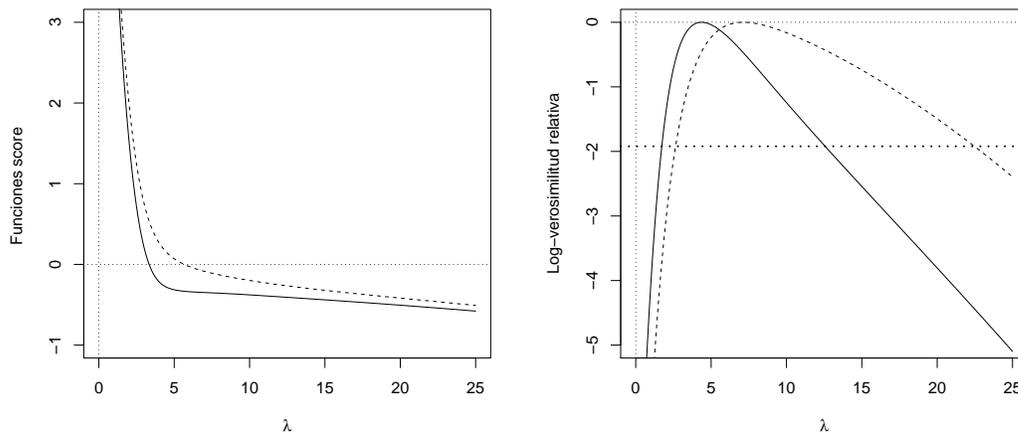


Figura 2.5: Funciones score (panel izquierdo) y log-verosimilitud relativa (panel derecho), modificada (línea continua) y no modificada (línea segmentada). La línea de puntos horizontal en el panel de la derecha representa el intervalo de confianza de 0.95.

Tabla 2.6: Estimaciones no modificada y modificada para λ .

$\hat{\lambda}$	$W(\lambda)IC$	$\hat{\lambda}_M$	$W_M(\lambda)IC$
7.143	(2.643, 22.497)	4.360	(1.728, 12.564)

Segundo ejemplo. Ahora, consideremos el valor absoluto de los datos de la muestra z , en la que sólo la segunda y tercera de las observaciones cambia. Las funciones score y log-verosimilitud relativa (modificada y no modificada) se presentan en los paneles izquierdo y derecho de la Figura 2.6, respectivamente. Ambas funciones no modificadas indican claramente que el EMV no modificado de λ diverge. Las versiones modificadas de estas funciones, muestran claramente la existencia y la finitud del EMV modificado de λ . Los resultados numéricos obtenidos se resumen en la Tabla 2.7.

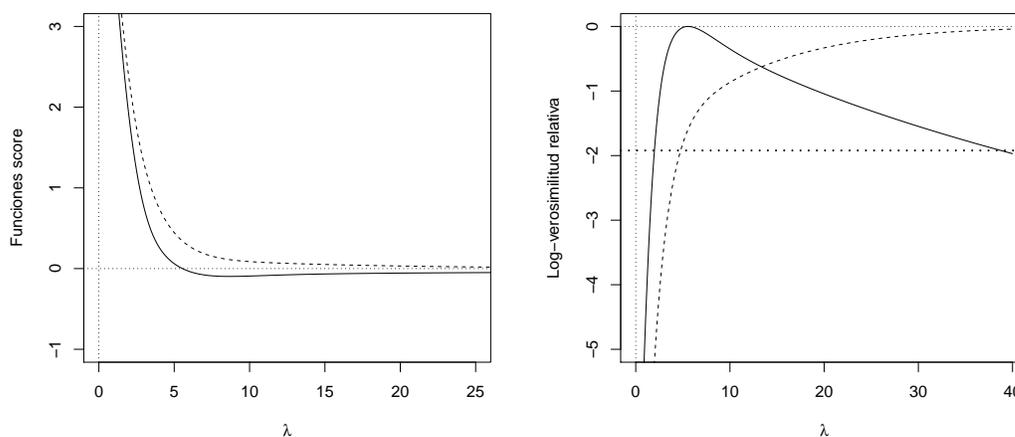


Figura 2.6: Funciones score (panel izquierdo) y log-verosimilitud relativa (panel derecho), modificada (línea continua) y no modificada (línea segmentada) de $|z|$. La línea de puntos horizontales en el panel de la derecha representa el intervalo de confianza de 0.95.

Tabla 2.7: Estimaciones no modificada y modificada para λ basada en los datos $|z|$.

$\hat{\lambda}$	$W(\lambda)IC$	$\hat{\lambda}_M$	$W_M(\lambda)IC$
∞	—	5.548	(1.981, 38.741)

Tabla 2.8: Simulación de 5,000 iteraciones del modelo MSN(λ).

λ	n	$sesgo(\hat{\lambda})^a$	$sesgo(\hat{\lambda}_M)$	$W_M(\lambda)IC$	$\%(\hat{\lambda} < \infty)$
5	20	2.201	-1.423	0.947	75.08 (74.19)
	50	1.743	-0.189	0.957	96.60 (96.62)
	100	0.677	-0.011	0.954	99.90 (99.84)
10	20	2.216	-5.768	0.904	48.46 (47.52)
	50	4.789	-2.507	0.941	81.04 (80.05)
	100	3.930	-0.478	0.958	96.00 (96.00)

Sesgo de $\hat{\lambda}$ y $\hat{\lambda}_M$, cobertura empírica de los intervalos de confianza de 0.95 basado en $W_M(\lambda)$ y el porcentaje empírico (teórico) para casos en que $\hat{\lambda}$ existe. ^a Calculado cuando $\hat{\lambda} < \infty$.

Cuarto estudio de simulación: Finalmente, para obtener una evaluación empírica del sesgo del estimador modificado, se realiza un estudio de simulación con 5,000 iteraciones para $\lambda = 5, 10$ y $n = 20, 50, 100$. Los resultados obtenidos se resumen en la Tabla 2.8 e indican que el sesgo de $\hat{\lambda}_M$ tiende a ser mayor cuando el verdadero valor de λ es grande y el tamaño de la muestra n es pequeña. En este análisis, se debe considerar también la proporción de no existencia de EMV no modificada de λ . Por ejemplo, por $\lambda = 10$ y $n = 20$, el sesgo de $\hat{\lambda}_M$ es

-5.768, en este caso más de la mitad de las muestras dan un valor infinito para $\hat{\lambda}$. Además, excluyendo el caso cuando λ es grande y n es pequeña, los porcentajes de cobertura empírica tienden a ser similar al nivel nominal.

2.4.2. Caso localización, escala y forma

Consideremos la función de log-verosimilitud del modelo $\text{MSN}(\mu, \sigma, \lambda)$ dado en (2.11). Sea $\hat{\mu}$, $\hat{\sigma}$ y $\hat{\lambda}$ los EMVs no modificadas de μ , σ y λ , respectivamente. Como se ha expuesto anteriormente en 2.3.3, en este caso el $\hat{\mu}$ y $\hat{\sigma}$ son siempre finitos, incluso si $\hat{\lambda}$ diverge. Además, los resultados de simulación (algunos de los cuales se resumen en las Tablas 2.3, 2.4 y 2.5) muestran que las propiedades de $\hat{\mu}$ y $\hat{\sigma}$ son bastante buenas en términos de sesgo y error medio cuadrático, y tienen el comportamiento asintótico usual. Esto nos indica que no es necesario investigar la reducción de sesgo para estos estimadores. Por lo tanto, la misma metodología utilizada por Sartori [60] para estimar los tres parámetros en los casos del modelo skew-normal y skew- t , también se utilizan para el caso $\text{MSN}(\mu, \sigma, \lambda)$, es decir, primero se calcula el EMV de $\hat{\mu}(\lambda)$ y $\hat{\sigma}(\lambda)$ para λ fijo, y luego aplicar el método de Firth al perfil de la función score de λ dada por

$$U^P(\lambda) = \frac{\partial}{\partial \lambda} l^P(\lambda) = \sum_{i=1}^n u(\hat{z}_i) \zeta\{\lambda u(\hat{z}_i)\}, \quad (2.15)$$

donde $\hat{z}_i = (y_i - \hat{\mu}(\lambda))/\hat{\sigma}(\lambda)$ y como se definió previamente $l^P(\lambda) = l(\hat{\mu}(\lambda), \hat{\sigma}(\lambda), \lambda)$. El perfil de la función score modificada se define como

$$U_M^P(\lambda) = U^P(\theta) + M(\lambda), \quad (2.16)$$

con $M(\lambda)$ dado por (2.14). Si $\hat{\lambda}$ diverge, entonces $\hat{\mu}(\lambda) \rightarrow \hat{\mu}$ cuando $\lambda \rightarrow \infty$, donde $\hat{\mu}$ es tal que $y_i - \hat{\mu} > 0$ para todo $i = 1, \dots, n$. Esto implica que $U^P(\lambda) \rightarrow U(\lambda)$ cuando $\lambda \rightarrow \infty$, y

por lo tanto la existencia $\widehat{\lambda}_M$ está garantizada. Finalmente, la razón de verosimilitud puede expresarse como

$$W_M^P(\lambda) = 2\{l_M^P(\widehat{\lambda}_M) - l_M^P(\lambda)\}, \quad (2.17)$$

también puede se usado para obtener los intervalos de confianza para λ , donde

$$l_M^P(\lambda) = \int_c^\lambda U_M^P(t)dt = l^P(\lambda) - l^P(c) + \int_c^\lambda M(t)dt, \quad (2.18)$$

donde c es una constante arbitraria.

Tercer ejemplo. La siguiente muestra aleatoria de tamaño $n = 50$ fue generada de la población MSN(0,1,5): $y = (0.754, 0.112, 1.619, -0.085, 0.514, 1.090, 1.083, 0.470, 1.854, -0.034, 0.592, 0.649, 0.110, 0.170, 0.716, 1.725, 0.587, -0.141, 0.306, 0.489, 1.330, -0.135, -0.006, -0.186, 2.204, -0.117, 0.793, 0.253, 0.266, 0.746, 0.143, 1.334, 0.746, 1.214, 0.488, 0.586, 0.234, 1.212, 2.655, 0.113, 3.240, -0.060, 0.449, 0.932, 0.323, 1.316, 0.694, 1.078, 0.165, 1.730)$. La Figura 2.7, muestra el comportamiento del perfil cuasi-verosimilitud relativa sin modificar y modificado, $l^P(\lambda)$ y $l_M^P(\lambda)$, respectivamente para estos datos. Se ve claramente que mientras el EMV no modificado de λ diverge, el EMV modificado de λ existe y es finito. La Tabla 2.9, reporta los resultados de estas estimaciones y el intervalo de confianza de 0.95 basado en $W_M^P(\lambda)$.

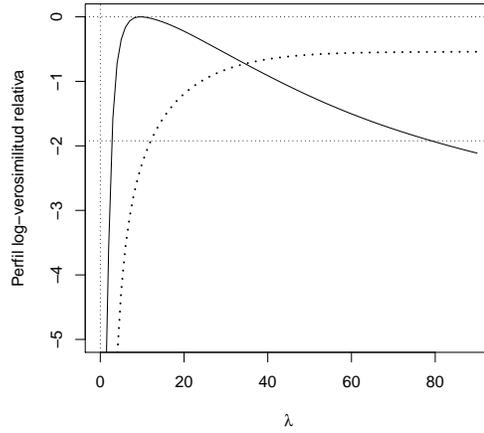


Figura 2.7: Versiones relativas modificada, $l_M^P(\lambda)$ (línea continua) y no modificada, $l^P(\lambda)$ (línea segmentada). La línea horizontal punteada corresponde al intervalo de confianza de 0.95.

Tabla 2.9: Estimaciones de μ , σ y λ .

$\hat{\mu}$	$\hat{\sigma}$	$\hat{\lambda}$	$W(\lambda)IC$	$\hat{\lambda}_M$	$W_M^P(\lambda)IC$
-0.187	1.173	∞	—	9.554	(2.755, 79.040)

Estudio de simulación final : Por último, se lleva a cabo un estudio de simulación de 5,000 iteraciones para $\mu = 0$, $\sigma = 1$, $\lambda = 5, 10$ y distintos tamaños de muestras. La Tabla 2.10, informa el sesgo de los EMV no modificados y modificados, la cobertura empírica de los intervalos de confianza basado en $W_M^P(\lambda)$ y el porcentaje de casos en que el EMV no modificada de λ es finito. Se observa que los EMV $\hat{\mu}$ y $\hat{\sigma}$ tienen un buen comportamiento con respecto al sesgo, que no se ve afectado por el comportamiento de $\hat{\lambda}$. Las coberturas empíricas de probabilidades son en general, bastante similares al valor nominal, excepto en

el caso de $\lambda = 10$ y $n = 50$, donde la proporción de casos con $\hat{\lambda} = \infty$ es significativa.

Tabla 2.10: Simulación de 5,000 iteraciones del modelo MSN(0, 1, λ).

λ	n	$sesgo(\hat{\mu})$	$sesgo(\hat{\sigma})$	$sesgo(\hat{\lambda})^a$	$sesgo(\hat{\lambda}_M)$	$W_M^P(\lambda)IC$	$\%(\hat{\lambda} < \infty)$
5	50	-0.00196	-0.00033	2.16050	-0.64363	0.952	87.36
	100	0.00032	0.00007	1.40534	-0.22914	0.945	98.66
	200	-0.00005	-0.00122	0.54437	-0.08913	0.945	99.98
10	50	0.01455	-0.01811	1.46304	-3.46331	0.892	64.52
	100	0.00362	-0.00693	3.67886	-1.38844	0.923	87.90
	200	0.00047	-0.00207	2.98198	-0.24559	0.935	98.04

Sesgo de $\hat{\mu}$, $\hat{\sigma}$, $\hat{\lambda}$ y $\hat{\lambda}_M$, cobertura empírica del intervalo de confianza de 0.95 basado en $W_M^P(\lambda)$ y porcentaje empírico (teórico) de casos cuando $\hat{\lambda}$ existe. ^a calculado cuando $\hat{\lambda} < \infty$.

2.5. Aplicación:

La existencia de sesgo en la MLEs, ha sido una preocupación constante para los profesionales e investigadores en Estadística. La magnitud de este sesgo es de vital importancia, puesto que si éste es relativamente grande las conclusiones y las decisiones pueden ser totalmente alejadas de la realidad. En este ejemplo, consideraremos la base de datos correspondiente a la concentración de níquel en 86 muestras de suelo analizado en el Departamento de Minería en la Universidad de Atacama - Chile. Las estadísticas básicas de los datos se presentan en Tabla 2.11.

Tabla 2.11: Estadística descriptiva de los datos.

<i>Datos</i>	<i>n</i>	<i>media</i>	<i>d.e.</i>	$\sqrt{\gamma_1}$	γ_2
<i>Nickel</i>	86	21.337	16.639	2.355	11.191

Tabla 2.12: EMV de μ , σ y λ .

$\hat{\mu}$	$\hat{\sigma}$	$\hat{\lambda}$	$\hat{\lambda}_M$	$l(\hat{\mu}, \hat{\sigma}, \hat{\lambda})$	$l(\hat{\mu}, \hat{\sigma}, \hat{\lambda}_M)$
2.571(1.260)	25.027(2.154)	10.261(5.236)	5.882(2.185)	-344.769	-345.719

Ajustando el modelo $MSN(\mu, \sigma, \lambda)$ a los datos dados, obtenemos el EMV no modificado $\hat{\theta} = (\hat{\mu}, \hat{\sigma}, \hat{\lambda})^T$ de $\theta = (\mu, \sigma, \lambda)^T$. A continuación, a partir de la ecuación score modificado para el parámetro de forma de este modelo, se calcula el EMV modificado $\hat{\lambda}_M$. Los resultados de las estimaciones y los respectivos errores estándar, dados entre paréntesis, se presentan en Tabla 2.12. Los errores se obtuvieron a partir de la matriz de información de Fisher del modelo $MSN(\mu, \sigma, \lambda)$. Debido a que esta matriz de información de Fisher no es singular, el estimador no modificado de $\hat{\theta} = (\hat{\mu}, \hat{\sigma}, \hat{\lambda})^T$ y el estimador modificado $\hat{\theta} = (\hat{\mu}, \hat{\sigma}, \hat{\lambda}_M)^T$, satisfacen las propiedades asintóticas regulares de los EMV. En particular, tienen distribución asintótica normal $N_3(\theta, I_{MSN}^{-1}(\theta)/n)$, donde $I_{MSN}(\theta)$ es la matriz de información Fisher del modelo MSN dada en 2.3.3. Por construcción el EMV modificado tiene siempre menor sesgo que el EMV estándar.

Tabla 2.13: Intervalo confianza para λ .

	95 %	98 %	99 %
IC	(2.006,19.231)	(-0.134,21.372)	(-1.561,22.800)
IC^*	(2.288, 7.527)	(1.395,10.370)	(0.799,10.965)

La Tabla 2.13, presenta los intervalos de confianza para λ obtenidos de su distribución asintótica de $\hat{\lambda}$ y $\hat{\lambda}_M$. Notamos que mientras el intervalo de confianza estándar (IC) modifica la hipótesis de normalidad, $H_0 : \lambda = 0$, para niveles de confianza superiores al 95 % los intervalos de confianza modificados, obtenidos del EMV modificado (IC^*), permanecen desfavorables a la hipótesis en los tres niveles de confianza considerados. En resumen, los intervalos de confianza modificados son más pequeños y precisos.

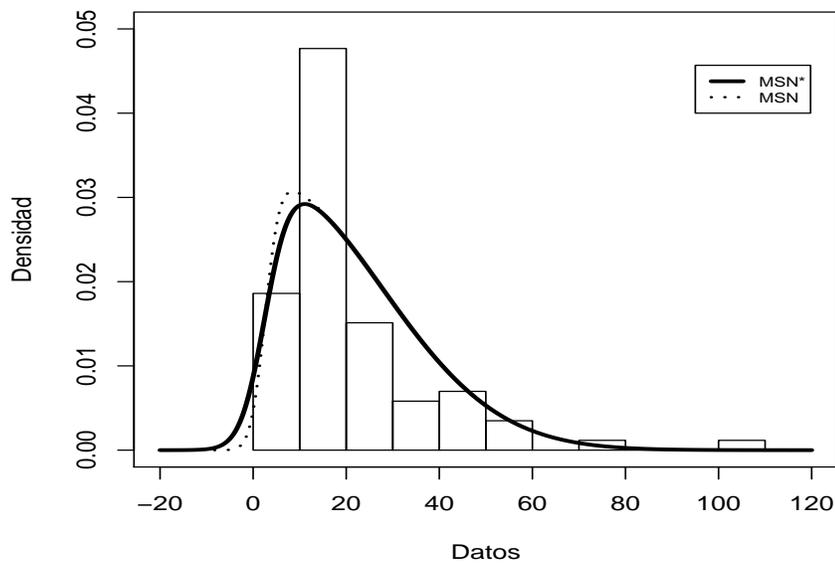


Figura 2.8: Datos ajustados

La figura 2.8 muestra el ajuste de los datos con $MSN(\hat{\mu}, \hat{\sigma}, \hat{\lambda})$ y el reajuste $MSN^*(\hat{\mu}, \hat{\sigma}, \hat{\lambda}_M)$ utilizando el EMV de sesgo reducido $\hat{\lambda}_M$.

2.6. Apéndice

Sea $X \sim MSN(\mu, \sigma, \lambda)$. Consideremos la notación:

$$\tau = (\mu, \sigma), \quad Z = \frac{X - \mu}{\sigma}, \quad u(x) = \frac{x}{(1 + x^2)^{1/2}}, \quad \zeta = \zeta(\lambda u(Z)) = \frac{\phi(\lambda u(Z))}{\Phi(\lambda u(Z))}$$

y las siguientes derivadas

$$\begin{aligned} \frac{\partial Z}{\partial \mu} &= -\frac{1}{\sigma}, & \frac{\partial Z}{\partial \sigma} &= -\frac{Z}{\sigma}, & \frac{\partial u(Z)}{\partial \tau} &= \frac{1}{(1 + Z^2)^{3/2}} \frac{\partial Z}{\partial \tau} \\ \frac{\partial \zeta}{\partial \tau} &= \left(-\frac{\lambda^2 Z}{(1 + Z^2)^2} \zeta - \frac{\lambda}{(1 + Z^2)^{3/2}} \zeta^2 \right) \frac{\partial Z}{\partial \tau} \\ \frac{\partial}{\partial \tau} \left(\frac{Z}{(1 + Z^2)^{3/2}} \right) &= \frac{1 - 2Z^2}{(1 + Z^2)^{5/2}} \frac{\partial Z}{\partial \tau} \\ \frac{\partial \zeta}{\partial \lambda} &= -\frac{\lambda Z^2}{(1 + Z^2)} \zeta - \frac{Z}{(1 + Z^2)^{1/2}} \zeta^2 \end{aligned}$$

tenemos que las funciones score asociados al modelo $MSN(\mu, \sigma, \lambda)$ son

$$\begin{aligned} S_\mu &= \frac{1}{\sigma} \left(Z - \frac{\lambda}{(1 + Z^2)^{3/2}} \zeta \right) \\ S_\sigma &= \frac{1}{\sigma} \left(-1 + Z^2 - \frac{\lambda Z}{(1 + Z^2)^{3/2}} \zeta \right) \\ S_\lambda &= \frac{Z}{(1 + Z^2)^{1/2}} \zeta \end{aligned}$$

y las entradas de la matriz información de Fisher son

$$\begin{aligned}
I_{\mu\mu} &= -E(S_{\mu\mu}) \\
&= -\frac{1}{\sigma^2} E \left(-1 - \frac{3\lambda Z}{(1+Z^2)^{5/2}} \zeta - \frac{\lambda^3 Z}{(1+Z^2)^{7/2}} \zeta - \frac{\lambda^2}{(1+Z^2)^3} \zeta^2 \right) \\
&= \frac{1}{\sigma^2} (1 + \lambda^2 \eta_{03})
\end{aligned}$$

$$\begin{aligned}
I_{\mu\sigma} &= -E(S_{\sigma\mu}) \\
&= -\frac{1}{\sigma^2} E \left(-2Z + \frac{\lambda(1-2Z^2)}{(1+Z^2)^{5/2}} \zeta - \frac{\lambda^3 Z^2}{(1+Z^2)^{7/2}} \zeta - \frac{\lambda^2 Z}{(1+Z^2)^3} \zeta^2 \right) \\
&= -\frac{1}{\sigma^2} (-2E(Z) + \lambda\rho_{05} - 2\lambda\rho_{25} - \lambda^3\rho_{27} - \lambda^2\eta_{13})
\end{aligned}$$

$$\begin{aligned}
I_{\mu\lambda} &= -E(S_{\mu\lambda}) \\
&= -\frac{1}{\sigma} E \left(-\frac{1}{(1+Z^2)^{3/2}} \left\{ \zeta - \frac{\lambda^2 Z^2}{(1+Z^2)} \zeta - \frac{\lambda Z}{(1+Z^2)^{1/2}} \zeta^2 \right\} \right) \\
&= \frac{1}{\sigma} (\rho_{03} - \lambda^2\rho_{25} - \lambda\eta_{12})
\end{aligned}$$

$$\begin{aligned}
I_{\sigma\sigma} &= -E(S_{\sigma\sigma}) \\
&= -\frac{1}{\sigma^2} E \left(1 - 3Z^2 + \frac{\lambda Z}{(1+Z^2)^{3/2}} \zeta + \frac{\lambda(Z-2Z^3)}{(1+Z^2)^{5/2}} \zeta - \frac{\lambda^3 Z^3}{(1+Z^2)^{7/2}} \zeta - \frac{\lambda^2 Z^2}{(1+Z^2)^3} \zeta^2 \right) \\
&= \frac{1}{\sigma^2} (2 + \lambda^2 \eta_{23})
\end{aligned}$$

$$\begin{aligned}
I_{\sigma\lambda} &= -E(S_{\lambda\sigma}) \\
&= -\frac{1}{\sigma} E \left(-\frac{Z}{(1+Z^2)^{3/2}} \zeta + \frac{\lambda^2 Z^3}{(1+Z^2)^{5/2}} \zeta + \frac{\lambda Z^2}{(1+Z^2)^2} \zeta^2 \right) \\
&= -\frac{\lambda}{\sigma} \eta_{22}
\end{aligned}$$

$$\begin{aligned}
I_{\lambda\lambda} &= -E(S_{\lambda\lambda}) \\
&= -E \left(-\frac{\lambda Z^3}{(1+Z^2)^{3/2}} \zeta - \frac{Z^2}{1+Z^2} \zeta^2 \right) \\
&= \eta_{21}
\end{aligned}$$

Capítulo 3

I transformación de variable aleatoria

MSN: Distribución log-MSN

3.1. Introducción

En este Capítulo, estudiaremos el comportamiento del sesgo de los EMV de los parámetros de un nuevo modelo, denominada log-skew-normal modificada (LMSN), que nace de una transformación de una variable aleatoria que tiene distribución MSN. En forma más específica, $Y = e^Z \sim LMSN(\lambda)$, donde $Z \sim MSN(\lambda)$. Según esta transformación el Jacobiano no depende del parámetro de forma (λ), luego las verosimilitudes del modelo MSN y LMSN son proporcionales entre sí, entonces las inferencias que se deriven de ambas verosimilitudes serán análogas.

El modelo LMSN adquiere gran importancia, puesto que se puede comparar naturalmente con modelos de su misma clase, por ejemplo, el modelo log-normal y log-skew-normal, los cuales son muy utilizados en aplicaciones a datos reales. A continuación, entregamos la

definición de estos modelos que utilizaremos más adelante: El modelo log-normal es una distribución de probabilidad muy utilizada para modelar el comportamiento de observaciones con asimetría positiva, por lo cual la mayoría de los valores de estas observaciones ocurren en las proximidades de su valor mínimo. Algunas de las aplicaciones que tiene este modelo son: patrones de abundancia de especies, concentraciones ambientales, precio de las acciones, distribución del peso molecular de los polímeros, producción de nanopartículas de cobre, etc. La distribución log-normal se obtiene a través de una transformación de una variable aleatoria $Z \sim N(0, 1)$, a saber $Y = e^Z$. Entonces, diremos que una variable aleatoria Y se distribuye log-normal, en adelante LN, si su densidad está dada por

$$f_Y(y) = \frac{1}{y} \phi(\log(y)),$$

donde $y \in \mathbb{R}^+$ y ϕ es la densidad normal estándar.

Estas observaciones por su naturaleza, requieren ser modeladas por distribuciones con soporte positivo y, además tengan un cierto grado de flexibilidad en lo que refiere a asimetría y curtosis. Sin embargo, existen datos que no son modelados adecuadamente por la distribución log-normal, ya que ellos presentan cierta asimetría y curtosis la cuales están fuera de su rango natural. Un modelo que presenta esta característica es la distribución log-skew-normal (LSN), introducida y estudiada por Azzalini [16]. Esta es una versión con soporte positivo de la distribución skew-normal (SN), definida de la siguiente forma: Diremos que una variable aleatoria Y tiene distribución log-skew-normal con parámetros μ , σ y λ , denotada por $Y \sim LSN(\mu, \sigma, \lambda)$, sí y solo si, $\log(Y)$ tiene distribución skew-normal con parámetro μ , σ y λ , es decir, $Y = e^Z$ donde $Z \sim SN(\mu, \sigma, \lambda)$, de modo que su pdf esta dada por

$$f_Y(y; \mu, \sigma, \lambda) = \frac{2}{\sigma y} \phi(z) \Phi(\lambda z),$$

con $z = \frac{\log(y) - \mu}{\sigma}$, $y \in \mathbb{R}^+$, $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}^+$ y λ , ϕ y Φ son la pdf y función de distribución acumulada (cdf) de la distribución normal estándar, respectivamente. Una aplicación de este modelo a datos reales se puede encontrar en Azzalini [16], para un conjunto multivariado en Marchenko y Genton [49] y un caso bimodal en Bolfarine et al. [21].

3.2. Definición y propiedades del modelo LMSN

Diremos que una variable aleatoria Y tiene distribución log-skew-normal modificada con parámetro λ , denotada por $Y \sim LMSN(\lambda)$, sí y solo si, $\log(Y)$ tiene distribución skew-normal modificada. Esto es, $Y = e^Z$, donde $Z \sim MSN(\lambda)$, su densidad viene dada por

$$f_Y(y; \lambda) = \frac{2}{y} \phi(\log(y)) \Phi(\lambda u(\log(y))),$$

donde $u(x) = \frac{x}{\sqrt{1+x^2}}$, $y \in \mathbb{R}^+$ y $\lambda \in \mathbb{R}$, ϕ y Φ son la pdf y función de distribución acumulada (cdf) de la distribución normal estándar, respectivamente.

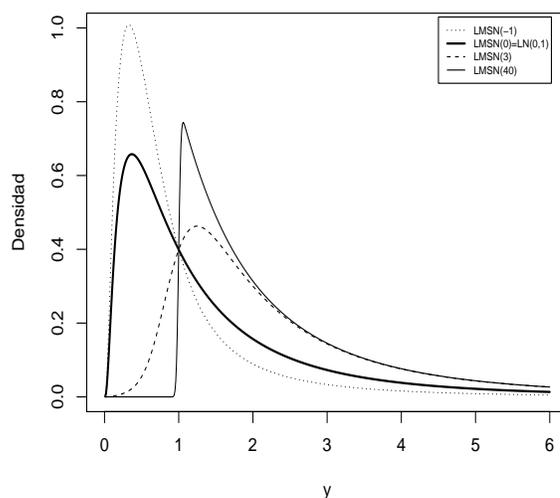


Figura 3.1: Gráfica de LMSN.

La Figura 3.1, muestra el comportamiento del modelo $MLSN(\lambda)$ para distintos valores de λ . Cuando $\lambda = 0$, ésta coincide con la distribución LN (línea continua gruesa), luego el modelo LMSN es una extensión del modelo LN.

Podemos deducir que el modelo LMSN puede modelar datos de tiempos de falla, dentro de un período dado, que no logra capturar la distribución LN. La ventaja de esta nueva distribución, es que permite flexibilizar el rango de variación de la asimetría y curtosis de la distribución LN, puesto que posee un parámetro adicional que contribuye a esta variación.

Si $Y \sim LMSN(\mu, \sigma, \lambda)$, su densidad viene expresada por

$$f_Y(y; \mu, \sigma, \lambda) = \frac{2}{y\sigma} \phi(z) \Phi(\lambda u(z)), \quad (3.1)$$

donde $z = \frac{\log(y) - \mu}{\sigma}$, $u(x) = \frac{x}{\sqrt{1+x^2}}$, y y $\sigma \in \mathbb{R}^+$, λ y $\mu \in \mathbb{R}$, ϕ y Φ son la pdf y función de distribución acumulada (cdf) de la distribución normal estándar, respectivamente.

Función de sobrevivencia: Es de gran importancia, en análisis de sobrevivencia definir la función de riesgo asociada a la distribución que modela los datos. La función de riesgo definida como $r(t) = \frac{f(t)}{S(t)}$, para el modelo LMSN, tiene la siguiente expresión

$$r(t) = \frac{\frac{2}{t} \phi(\log(t)) \Phi(\lambda u(\log(t)))}{1 - \int_0^t \frac{2}{w} \phi(\log(w)) \Phi(\lambda u(\log(w))) dw}. \quad (3.2)$$

Haciendo una simple operación matemática en (3.2), se tiene que $\lim_{t \rightarrow \infty} r(t) = 0$.

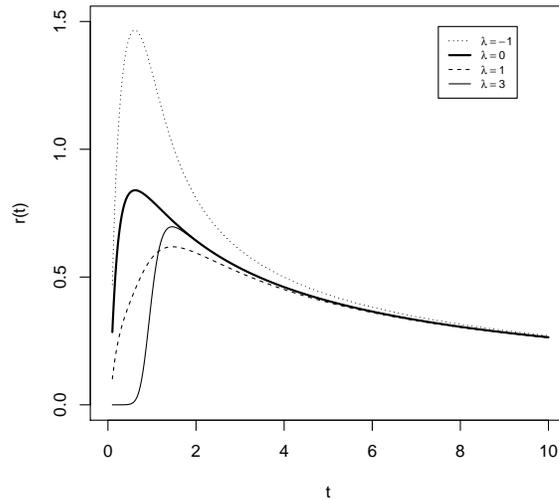


Figura 3.2: Función de riesgo para el modelo LMSN.

La Figura 3.2, muestra la forma que toma la función de riesgo $r(t)$ para distintos valores de λ . En la práctica, la distribución LMSN podría modelar el comportamiento de una población que es una mezcla de individuos que tienden a tener tiempos cortos o largos de vida. Por ejemplo, el tiempo de supervivencia después de un tratamiento para algunos pacientes de cáncer, donde las personas que son exitosamente tratadas, se convierten en sobrevivientes con un período vida más largo. También, otro ejemplo, puede ser la duración de los matrimonios, los cuales después de cierto número de años, el riesgo de disolución del matrimonio por divorcio decrece, etc. Los valores máximos de la función de riesgo $r(t)$, para cada λ , son decrecientes cuando $\lambda \in (-\infty, 1.448)$ y fuera de ese intervalo es creciente. Cuando $\lambda = -\infty$, $r(t)$ es una función estrictamente creciente definida en el intervalo $(0, 1]$ y fuera de éste toma el valor cero. Ahora, sí $\lambda = \infty$, $r(t)$ es una función estrictamente decreciente definida en el intervalo $[1, \infty)$, coincidiendo con la función de riesgo del modelo LN en dicho intervalo, y

toma el valor cero en el intervalo $(0, 1)$.

Momentos

Sea una variable aleatoria $Y = e^Z \sim LMSN(\lambda)$, donde $Z \sim MSN(\lambda)$. El momento r -ésimo puede ser expresado por

$$E(Y^r) = 2e^{r^2/2} E\left(\Phi\left(\frac{S}{\sqrt{1+S^2}}r\right)\right),$$

para $r = 0, 1, 2, \dots$ y $S \sim N(\lambda, 1)$. Esta expresión, se obtiene directamente de la generatriz de momentos del modelo MSN, dada en (2.4), puesto que es válida para todo $t > 0$, en particular cuando $t = r$. Estos momentos, no son explícitos y deben ser calculados numéricamente.

No existe generadora de momentos para la distribución LMSN.

Proposición 1. *Para todo $\lambda \in \mathbb{R}$, la variable $Y \sim LMSN(\lambda)$ no tiene función generadora de momentos.*

Demostración: Utilizando parte de la demostración de Lin and Stoyanov [45], tenemos que para cada $t > 0$

$$E(e^{tY}) = \int_0^\infty h_\lambda(y) dy,$$

donde

$$h_\lambda(y) = \frac{e^{ty}}{y} \phi(\log(y)) \Phi(\lambda u(\log(y))) > 0, \quad \forall y > 0,$$

entonces para

$$\lambda \geq 0, \quad \lim_{y \rightarrow \infty} \Phi(\lambda u(\log(y))) \geq \frac{1}{2}$$

y

$$\lambda < 0, \quad \lim_{y \rightarrow \infty} \Phi(\lambda u(\log(y))) \geq \Phi(\lambda),$$

en ambos casos $h_\lambda(y) \rightarrow \infty$ cuando $y \rightarrow \infty$. Por lo tanto, dado $t > 0$, $E(e^{tY}) = \infty$, para cualquier λ . \square

3.2.1. Coeficiente de asimetría y curtosis

Los coeficientes de asimetría y de curtosis pueden ser obtenidas por las fórmulas dadas en (2.5) y (2.6). Estos deben ser calculados numéricamente, puesto que no tienen forma explícita.

Tabla 3.1: Rango de valores para la Asimetría y Curtosis.

	MSN		LSN	
Límites	Asimetría	Curtosis	Asimetría	Curtosis
Mínimo	0.110	1.952	0.110	1.952
Máximo	10.985	524.351	6.184	113.936

La Tabla 3.1, muestra que el rango de valores que puede tomar el coeficiente de asimetría y curtosis para los modelos LMSN y LSN. Estos cálculos, consideran el caso $\mu = 0$ y $\sigma = 1$.

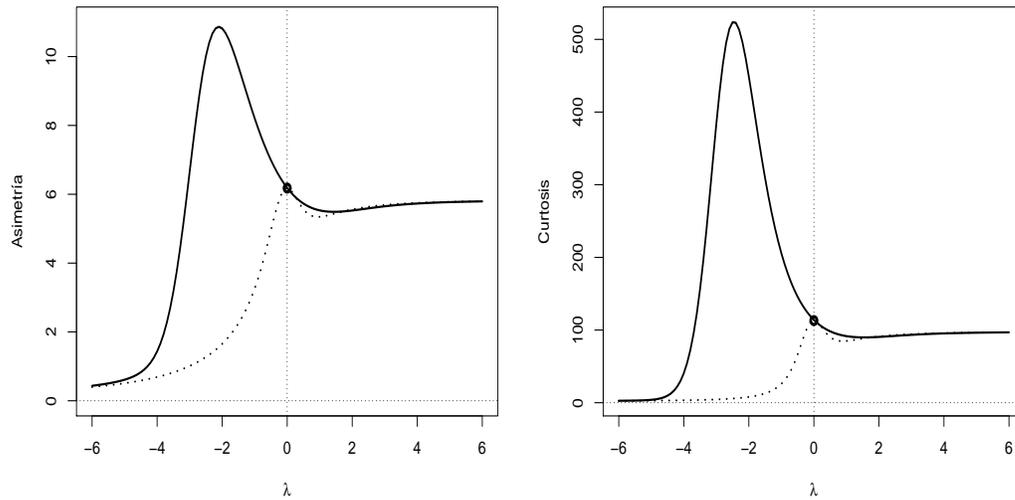


Figura 3.3: Gráfica de asimetría y curtosis para LMSN (línea continua) y LSN (línea punteada).

La Figura 3.3, muestra el comportamiento del coeficiente de asimetría y curtosis, en función de λ , correspondiente a las distribuciones LMSN y LSN. Cuando $\lambda = 0$, el punto representa los valores de los coeficientes correspondientes a la distribución LN. Esto demuestra la gran flexibilidad del nuevo modelo, con respecto al modelo LN y LSN. También, se puede observar que los valores mínimos corresponden a las asíntotas de la cola izquierda. Por otra parte, a partir de valores relativamente grandes de λ , digamos $\lambda > 2$, las colas que están al lado derecho de las gráficas se estabilizan alrededor de las asíntotas horizontales 5.83 y 97.93, respectivamente.

3.3. Estimación de máxima verosimilitud

Sea una muestra aleatoria y_1, y_2, \dots, y_n de una variable aleatoria $Y \sim LMSN(\theta)$, con $\theta = (\mu, \sigma, \lambda)$. Entonces, la log-verosimilitud correspondiente a dicha muestra, está dada por

$$l(\theta) = \frac{n}{2} \log\left(\frac{2}{\pi}\right) - n \log(\sigma) - \sum_{i=1}^n \log(y_i) - \frac{1}{2} \sum_{i=1}^n z_i^2 + \sum_{i=1}^n \log(\Phi(\lambda u(z_i))), \quad (3.3)$$

donde $z_i = \frac{\log(y_i) - \mu}{\sigma}$, $u(x) = \frac{x}{\sqrt{1+x^2}}$, $y_i \in \mathbb{R}^+$ y $\lambda \in \mathbb{R}$. Luego, el vector score asociado a (3.3)

es

$$\begin{aligned} S_\mu &= \frac{n}{\sigma} \left(\bar{z} - \frac{\lambda}{n} \sum_{i=1}^n \frac{1}{(1+z_i^2)^{3/2}} \zeta(\lambda u(z_i)) \right), \\ S_\sigma &= \frac{n}{\sigma} \left(-1 + \bar{z}^2 - \frac{\lambda}{n} \sum_{i=1}^n \frac{z_i}{(1+z_i^2)^{3/2}} \zeta(\lambda u(z_i)) \right), \\ S_\lambda &= \sum_{i=1}^n \frac{z_i}{(1+z_i^2)^{1/2}} \zeta(\lambda u(z_i)). \end{aligned}$$

Igualando a cero las funciones score, se sigue que

$$\begin{aligned} \bar{z} &= \frac{\lambda}{n} \sum_{i=1}^n \frac{1}{(1+z_i^2)^{3/2}} \zeta(\lambda u(z_i)), \\ \bar{z}^2 - 1 &= \frac{\lambda}{n} \sum_{i=1}^n \frac{z_i}{(1+z_i^2)^{3/2}} \zeta(\lambda u(z_i)), \\ 0 &= \sum_{i=1}^n \frac{z_i}{(1+z_i^2)^{1/2}} \zeta(\lambda u(z_i)). \end{aligned}$$

Resolviendo el sistema numéricamente obtenemos los EMV de μ , σ y λ .

3.3.1. Perfil de verosimilitud

El perfil de verosimilitud de la distribución LMSN, respecto del parámetro λ , para una muestra aleatoria de tamaño $n = 50, 100$ y 200 , obtenidas de una distribución $LN(0,1)=LMSN(0)$

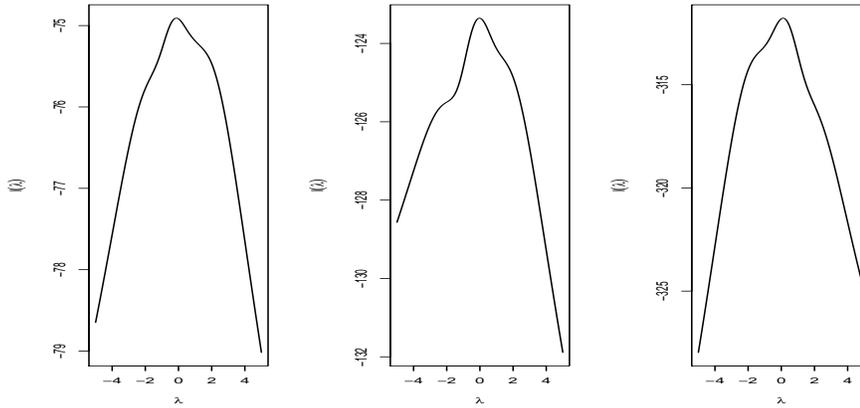


Figura 3.4: Perfil de verosimilitud de LMSN.

3.3.2. Matriz información

Debido a que el modelo LMSN se obtiene por transformación de una variable aleatoria que se distribuye MSN, el Jacobiano de la transformación no depende de los parámetros, en particular de λ . Esto implica, que la matriz información del modelo LMSN es similar a la matriz de información de la distribución MSN. Esto es, la matriz información observada y esperada para el modelo LMSN son las dadas en 2.3.3. De lo anterior, se desprende que el modelo propuesto, LMSN, hereda la no singularidad de la matriz información de Fisher del modelo MSN, cuando el parámetro λ toma el valor 0.

3.4. Reducción de sesgo del EMV

Las verosimilitudes del modelo MSN y LMSN se diferencian sólo en el Jacobiano, el cual no depende de los parámetros, luego son proporcionales entre sí. Además, el comportamiento del factor $\Phi(\lambda y)$, respecto de λ , es similar al comportamiento del factor $\Phi(\lambda u(y))$. Esto es, si

la verosimilitud del modelo MSN es divergente, también lo será la verosimilitud del modelo LMSN, entonces las inferencias que se deriven de ellas serán análogas. Por lo argumentado anteriormente, respecto al Jacobiano, la función modificada (M) del modelo MSN, dada en (1.2), es la misma para el modelo LMSN. Luego, la reducción del sesgo será similar en ambos modelos. En resumen, el modelo LMSN hereda el problema de sobrestimación en la estimación máximo verosímil del parámetro de forma λ y también la función modificada (M) del modelo MSN. Esto último, hace que la metodología y resultados de las simulaciones para reducir el sesgo, en el modelo LMSN, sean similares a los obtenidos en el Capítulo 2.

3.4.1. Estudio de simulación: Localización, escala y forma

A continuación se presenta un estudio de simulación de 5,000 iteraciones de muestras de variables aleatorias iid de $LMSN(0, 1, \lambda)$, $\lambda = 5, 10$ y distintos tamaños de muestras. La Tabla 3.2, informa el sesgo de los EMV no modificados y modificados, la cobertura empírica de los intervalos de confianza basado en $W_M^P(\lambda)$ y el porcentaje de casos en que el EMV no modificada de λ es finito. Se observa que los EMV $\hat{\mu}$ y $\hat{\sigma}$ tienen un buen comportamiento con respecto al sesgo, el cual no se ve afectado por el comportamiento de $\hat{\lambda}$. Esto permite aplicar el método de Firth [28] sólo al parámetro λ . Las coberturas empíricas de probabilidades son, en general, bastante similares al valor nominal, excepto en el caso de $\lambda = 10$ y $n = 50$, donde la proporción de casos con $\hat{\lambda} = \infty$ es significativa.

Tabla 3.2: Simulación de 5,000 iteraciones del modelo LMSN(0, 1, λ).

λ	n	$sesgo(\hat{\mu})$	$sesgo(\hat{\sigma})$	$sesgo(\hat{\lambda})^a$	$sesgo(\hat{\lambda}_M)$	$W_M^P(\lambda)IC$	$\%(\hat{\lambda} < \infty)$
5	50	0.00027	-0.00317	1.97804	-0.68632	94.82	87.06
	100	0.00066	-0.00238	1.35883	-0.23525	94.20	98.22
	200	0.00001	-0.00048	0.57867	-0.08691	94.68	99.98
10	50	0.01697	-0.01970	1.41342	-3.48979	87.74	66.32
	100	0.00324	-0.00489	3.68726	-1.34379	92.58	88.32
	200	0.00035	-0.00132	2.87639	-0.35524	93.74	98.56

Sesgo de $\hat{\mu}$, $\hat{\sigma}$, $\hat{\lambda}$ y $\hat{\lambda}_M$, cobertura empírica del intervalo de confianza de 0.95 basado en $W_M^P(\lambda)$ y porcentaje empírico (teórico) de casos cuando $\hat{\lambda}$ existe. ^a calculado cuando $\hat{\lambda} < \infty$.

Donde $W_M^P(\lambda) = 2\{l_M^P(\hat{\lambda}_M) - l_M^P(\lambda)\}$ es el estadístico razón de verosimilitud basado en la función perfil de cuasi-verosimilitud $l^P(\lambda) = l(\hat{\mu}(\lambda), \hat{\sigma}(\lambda), \lambda)$, $\hat{\mu}$ y $\hat{\sigma}$ son los EMV de μ y σ , respectivamente, para un valor de λ fijo.

3.5. Aplicación

Un modelo que presenta problema de sobrestimación en su estimación máximo verosímil genera automáticamente un sesgo. Dependiendo de la magnitud de éste es cuanto podemos errar en las inferencias estadísticas y por tanto obtener conclusiones que no se ajustan a la realidad. Por ello, es imprescindible disponer de metodologías de reducción de sesgo que permitan superar este problema, como por ejemplo el método de Firth [28]. El conjunto de

datos reales fueron analizados por Nadarajah [54] y Leiva et al. [42]. Para las concentraciones de contaminantes atmosféricos, se asume generalmente que los datos son independientes y no están correlacionados, por lo tanto no requieren el análisis de tendencias diurna o cíclica (ver Gokhale y Khare [32]). Esta información corresponde a mediciones de nivel de ozono tomados en forma diaria (en $ppb = ppm \times 1000$) en la ciudad de Nueva York entre mayo y septiembre de 1973, provenientes del Departamento de Conservación del Estado de Nueva York. La Tabla 3.3 muestra las estadísticas básicas que representan a los datos involucrados e indican un cierto grado de asimetría (b_1) y curtosis (b_2) muestral, esto es:

Tabla 3.3: Estadística descriptiva de los datos.

<i>Datos</i>	<i>n</i>	<i>media</i>	<i>d.e.</i>	$\sqrt{b_1}$	b_2
<i>Ozono</i>	116	42.129	32.987	1.209	4.112

Tabla 3.4: EMV de los modelos LMSN, LN y LSN.

EMVs.	LMSN	LN	LSN
$\hat{\mu}$	4.532(0.135)	3.418(0.080)	4.374(0.237)
$\hat{\sigma}$	1.411(0.137)	0.861(0.056)	1.287(0.195)
$\hat{\lambda}$	-4.095(1.498)	-	-2.420(1.430)
log-likelihood	-542.105	-543.883	-541.655
AIC	1090.211	1091.766	1089.310

La Tabla 3.4 muestra los EMV de los parámetros de las distribuciones LMSN, LN y LSN

y los valores entre paréntesis corresponden a los errores estándar. A pesar que el valor de AIC del modelo LSN es levemente menor que el modelo LMSN, no podemos testear simetría ($H_0 : \lambda = 0$) para el modelo LSN, puesto que su matriz de información de Fisher es singular. Luego, el modelo que ajusta mejor los datos es la distribución LMSN.

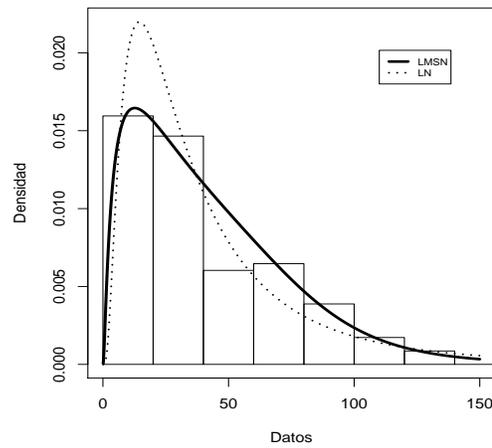


Figura 3.5: Ajuste de datos.

La Figura 3.5 presenta el histograma de los datos con las correspondientes distribuciones ajustadas, $LMSN(\hat{\mu}, \hat{\sigma}, \hat{\lambda})$ (línea continua) y $LN(\hat{\mu}, \hat{\sigma}, \hat{\lambda})$ (línea punteada).

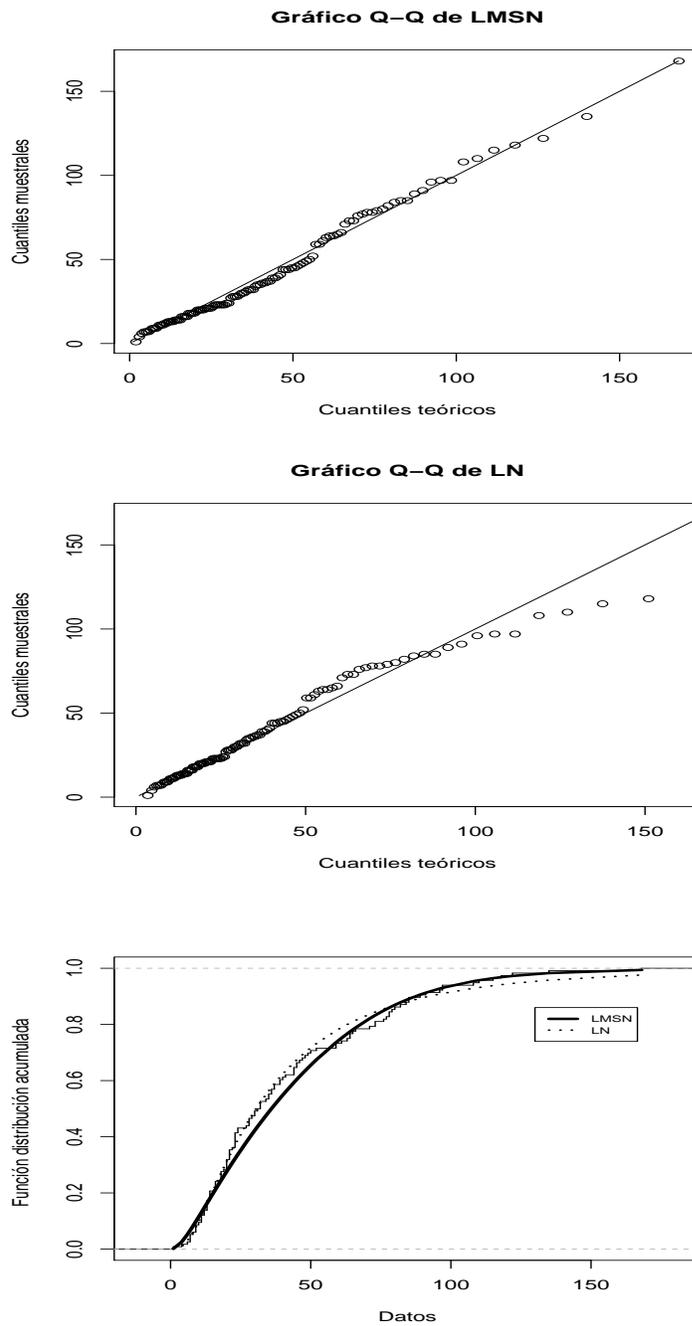


Figura 3.6: Gráfica de QQ y CDF para LMSN y LN.

El QQ-plot y la función de distribución acumulada empírica y teórica del modelo LMSN y LN se muestran en la Figura 3.6, obtenido de las estimaciones no modificadas de los parámetros

del modelo, ratifican visualmente el buen ajuste del modelo en estudio.

Tabla 3.5: EMV de μ , σ y λ .

$\hat{\mu}$	$\hat{\sigma}$	$\hat{\lambda}$	$\hat{\lambda}_M$	$l(\hat{\mu}, \hat{\sigma}, \hat{\lambda})$	$l(\hat{\mu}, \hat{\sigma}, \hat{\lambda}_M)$
4.532(0.105)	1.411(0.121)	-4.095(1.057)	-	-542.105	
4.532(0.123)	1.411(0.130)	-	-3.342(0.791)	-	-542.663

La Tabla 3.5 muestra los EMV $\hat{\mu}$, $\hat{\sigma}$, $\hat{\lambda}$ y el EMV modificado $\hat{\lambda}_M$ con los respectivos errores estándar, dadas entre paréntesis. Éstos, fueron obtenidos de la matriz información de Fisher de MLSN, definida en 2.3.3, puesto que la distribución asintótica $\hat{\theta}$ y $\hat{\theta}_M$ es $N_3(\theta, I^{-1}(\theta)/n)$, donde $\theta = (\mu, \sigma, \lambda)$. Además, claramente se aprecia que el valor del EMV modificado $\hat{\lambda}_M$ es mayor que el valor de EMV $\hat{\lambda}$ y por construcción tiene menor sesgo asintótico.

Tabla 3.6: Intervalo de confianza para λ

	95 %	98 %	99 %
IC	(-5.834,-2.356)	(-6.266,-1.924)	(-6.554,-1.636)
IC^*	(-4.644,-2.040)	(-4.968,-1.716)	(-5.183,-1.500)

La Tabla 3.6 presenta los intervalos de confianza de λ para distintos porcentajes, obtenidos de la distribución asintótica de $\hat{\theta}$ y $\hat{\theta}_M$. Si comparamos la longitud de los intervalos de confianza, obtenidos utilizando el EMV estándar (IC), con los intervalos de confianza, obtenidos utilizando el EMV modificado (IC^*), para los respectivos porcentajes, estos últimos, son más pequeños y en consecuencia más precisos.

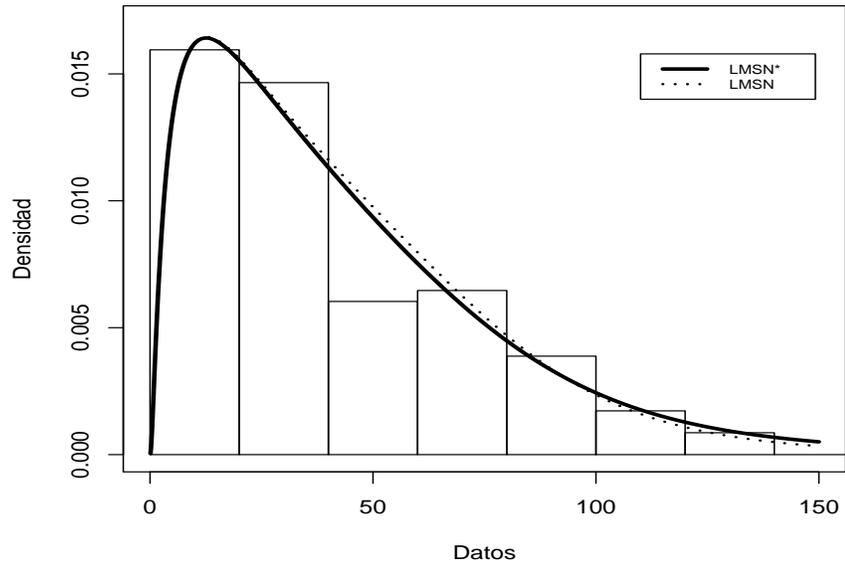


Figura 3.7: Datos reajustados.

La figura 3.7 muestra el ajuste de los datos con $LMSN(\hat{\mu}, \hat{\sigma}, \hat{\lambda})$ y el reajuste $LMSN^*(\hat{\mu}, \hat{\sigma}, \hat{\lambda}_M)$ utilizando el EMV de sesgo reducido $\hat{\lambda}_M$.

Capítulo 4

II transformación de variable aleatoria

MSN: Distribución

Birnbaum-Saunders skew-normal modificada

4.1. Introducción

A continuación, se analiza el comportamiento de la reducción de sesgo a un nuevo modelo, denominado Birnbaum-Saunders skew-normal modificada (BSMSN), que se origina por la transformación de una variable aleatoria con distribución MSN. Más específicamente, diremos que una variable aleatoria T tiene distribución BSMSN, denotada por $T \sim BSMSN(\alpha, \beta, \lambda)$,

si T corresponde a la transformación dada por

$$T = \beta \left(\frac{\alpha Z}{2} + \sqrt{\left(\frac{\alpha Z}{2}\right)^2 + 1} \right)^2, \quad (4.1)$$

donde $Z \sim MSN(\lambda)$. De acuerdo a esta transformación dada en (4.1), el Jacobiano no depende de parámetro de forma (λ), pero sí de los nuevos parámetros. Esta situación es diferente en el modelo LMSN, puesto que, el Jacobiano es independiente de los parámetros. Sin embargo, dado α y β , la log-verosimilitud de la distribución BSMSN, en función de λ , es proporcional a la log-verosimilitud del modelo MSN, por lo tanto, el comportamiento de ambas verosimilitudes, respecto de λ , es similar.

La importancia del nuevo modelo propuesto, BSMSN, nace del hecho que es una extensión del modelo Birbaum-Sanders clásico, siendo este último un modelo con bastante aplicación en temas de confiabilidad. Ahora, se presenta este modelo que citaremos más adelante:

En 1969, nace la distribución Birbaum-Sanders, conocida como la distribución que modela datos de fatiga de materiales. Birbaum y Saunders [19] presentaron un modelo probabilístico ingenioso, que da solución a problemas de vibración encontrados en los nuevos aviones comerciales y en los problemas de fatiga de materiales. Esta interesante distribución, tiene la siguiente representación estocástica, dada en (4.1), donde

$$Z = \frac{1}{\alpha} \left[\sqrt{\frac{T}{\beta}} - \sqrt{\frac{\beta}{T}} \right] \sim N(0, 1).$$

Diremos que una variable aleatoria T se distribuye Birbaum-Sanders de parámetros α y β , denotada por $T \sim BS(\alpha, \beta)$, si su densidad viene dada por

$$f_T(t) = = \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{1}{2\alpha^2} \left[\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}} \right]^2 \right) \frac{t^{-3/2}}{2\alpha\sqrt{\beta}} (t + \beta), \quad (4.2)$$

donde $t > 0$, $\alpha > 0$ es el parámetro de forma y $\beta > 0$ es el parámetro de escala y mediana de la distribución. Algunas de sus propiedades se presentan a continuación. Sea $T \sim BS(\alpha, \beta)$, entonces:

1. $cT \sim BS(\alpha, c\beta)$, $c > 0$.
2. $1/T \sim BS(\alpha, 1/\beta)$.

La primera propiedad indica que la distribución BS pertenece a la familia de escala. La segunda propiedad nos dice que la distribución BS es cerrada bajo recíprocos.

4.2. Definición y propiedades del modelo BSMSN

Diremos que una variable aleatoria T tiene distribución Birnbaum-Saunders skew-normal modificada, denotada por $T \sim BSMSN(\alpha, \beta, \lambda)$, si T corresponde a la transformación dada en 4.1, donde $Z \sim MSN(\lambda)$ y su densidad está definida por

$$f_T(t; \alpha, \beta, \lambda) = 2\phi(a_t(\alpha, \beta))\Phi(\lambda u(a_t(\alpha, \beta))) \frac{t^{-3/2}}{2\alpha\sqrt{\beta}}(t + \beta), \quad (4.3)$$

donde $a_t = a_t(\alpha, \beta) = \frac{1}{\alpha} \left(\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}} \right)$, $u(z) = \frac{z}{\sqrt{1+z^2}}$, $t > 0$, $\alpha > 0$, $\beta > 0$ y $\lambda \in \mathbb{R}$. Aquí ϕ y Φ denotan las funciones de densidad y distribución de la $N(0,1)$, respectivamente. Si $\lambda = 0$, entonces la densidad BSMSN en (4.3) se reduce a la densidad Birnbaum-Saunders clásica (BS) dada en (4.2).

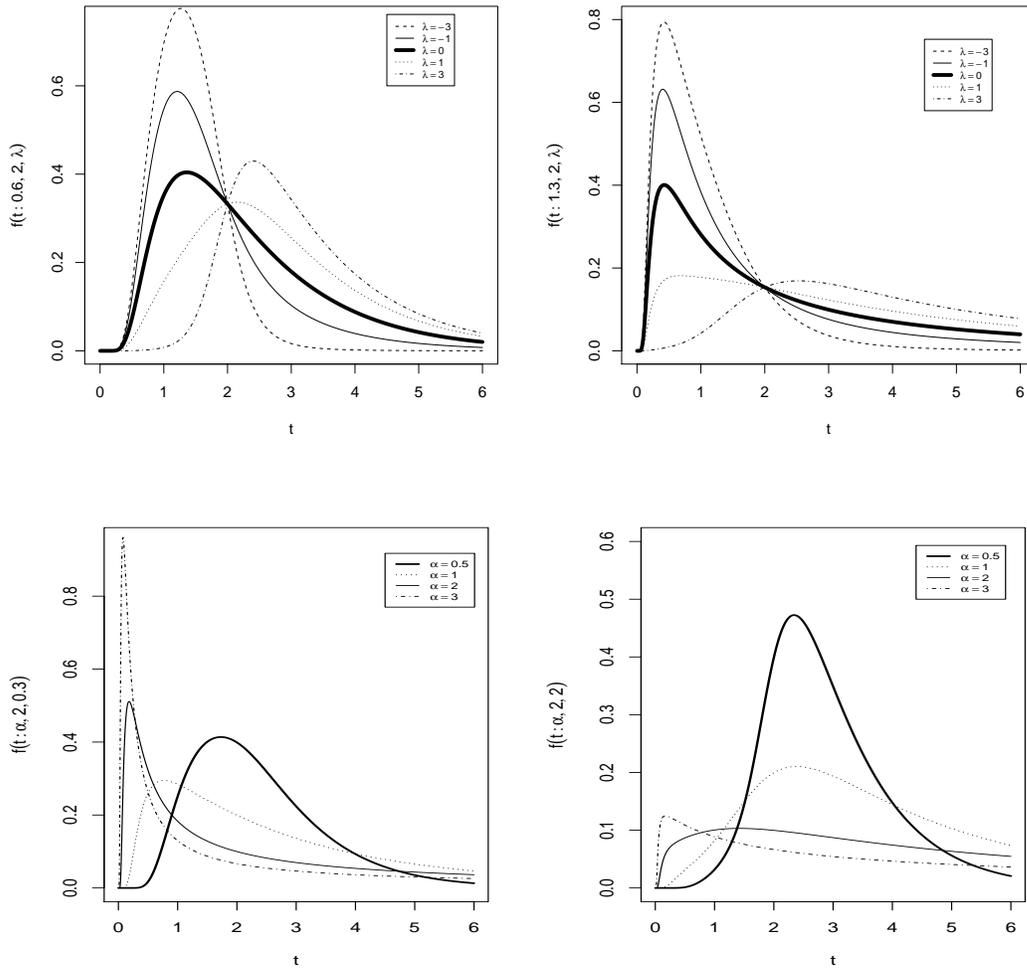


Figura 4.1: Gráfica de BSMSN.

La Figura 4.1 muestra el comportamiento de la densidad para distintos valores de α , β y λ , reflejando una gran flexibilidad respecto de la asimetría y curtosis, las cuales varían en la medida que cambian los valores de α y λ . Notamos que en las primeras dos Figuras superiores, la línea continua gruesa (sólida) representa a la distribución BS. También, se puede observar la concurrencia de todas las densidades en $t = \beta = 2$, al variar λ y manteniendo fijo los parámetros α y β . Por otra parte, cuando $\lambda \rightarrow +\infty$ la densidad coincide con dos veces la

densidad BS, para $t > b$, y es 0 en otro caso. Sin embargo, para $\lambda \rightarrow -\infty$ la densidad coincide con dos veces la densidad BS, para $t < b$ y es 0 en otro caso.

Propiedades

Proposición 4.2.1. *Sea $T \sim BSMSN(\alpha, \beta, \lambda)$, entonces:*

1. $bT \sim BSMSN(\alpha, b\beta, \lambda)$.
2. $T^{-1} \sim BSMSN(\alpha, 1/\beta, -\lambda)$.

Demostración:

1) Sea $v = bt$ entonces $dv = bdt$, luego reemplazando en (4.3)

$$f_V(v; \alpha, \beta, \lambda) = 2\phi(a_v(\alpha, b\beta))\Phi(\lambda u(a_v(\alpha, b\beta))) \frac{v^{-3/2}}{2\alpha\sqrt{b\beta}}(v + b\beta).$$

2) Sea $v = t^{-1}$ entonces $dv = -t^{-2}dt$, luego reemplazando en (4.3)

$$f_V(v; \alpha, \beta, \lambda) = 2\phi(a_v(\alpha, \beta^{-1}))\Phi(-\lambda u(a_v(\alpha, \beta^{-1}))) \frac{\sqrt{\beta}v^{-3/2}}{2\alpha}(v + \frac{1}{\beta}).$$

Proposición 4.2.2. *Sea $T \sim BSMSN(\alpha, \beta, \lambda)$ y $Z \sim MSN(\lambda)$ entonces $F_T(t; \alpha, \beta, \lambda) = F_Z(a_t(\alpha, \beta); \lambda)$.*

Demostración:

$$\begin{aligned} F_T(t; \alpha, \beta, \lambda) &= \int_0^t 2\phi(a_v(\alpha, \beta))\Phi(\lambda u(a_v(\alpha, \beta))) \frac{d}{dv}a_v(\alpha, \beta)dv \\ &= \int_0^{a_t(\alpha, \beta)} 2\phi(z)\Phi(\lambda u(z))dz \\ &= F_Z(a_t(\alpha, \beta); \lambda). \end{aligned}$$

Proposición 4.2.3. El percentil p -ésimo de la distribución BSMSN(α, β, λ), $t_p = F_T^{-1}(p; \alpha, \beta, \lambda)$

está dada por

$$t_p = \beta \left(\frac{\alpha z_p}{2} + \sqrt{\left(\frac{\alpha z_p}{2} \right)^2 + 1} \right)^2,$$

donde z_p es el percentil p -ésimo de la distribución MSN(λ), dado por $z_p = F_Z^{-1}(p; \lambda)$.

4.2.1. Función de Supervivencia

Algunas funciones de importancia en análisis de supervivencia se entregan a continuación.

Las funciones de supervivencia (S), riesgo (r), riesgo acumulativo (H) y riesgo inversa (R) para la distribución BSMSN están dadas por

$$S(t) = 1 - F_T(t), \quad H(t) = -\log(S(t)),$$

$$r(t) = \frac{f_T(t)}{1 - F_T(t)}, \quad R(t) = \frac{f_T(t)}{F_T(t)},$$

donde $F_T(t) = \int_0^{a_t(\alpha, \beta)} 2\phi(x)\Phi(\lambda u(x))dx$ y $f_T(t) = F_T'(t)$.

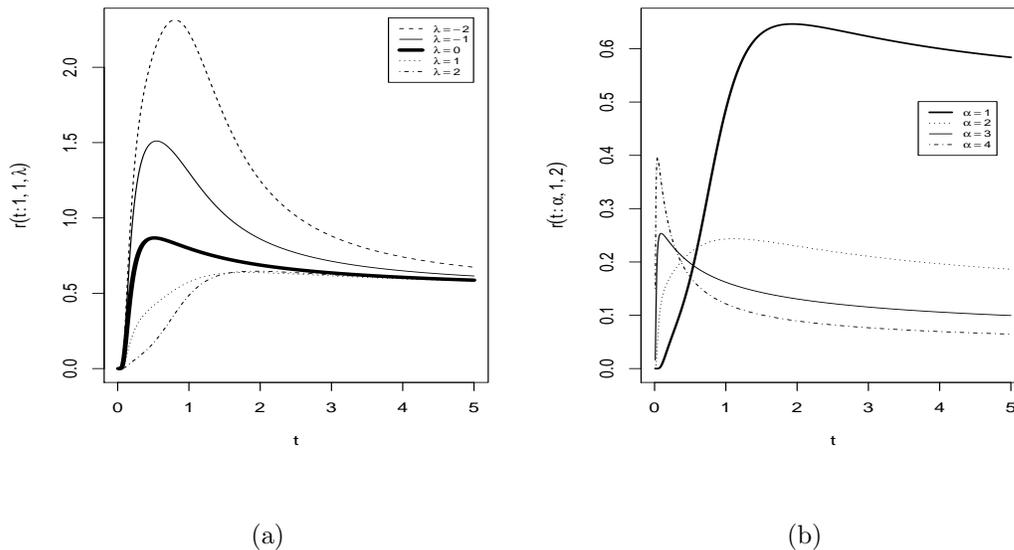


Figura 4.2: Gráfica de función de riesgo de BSMSN.

La Figura 4.2, muestra el comportamiento de la función de riesgo y la forma unimodal que posee para distintos valores de λ y α . En ambas gráficas, se puede apreciar claramente la convergencia de $r(t)$ dada en Proposición 4.2.4. En la Figura 4.2 (a), se visualiza que el valor máximo de $r(t)$ es una función decreciente respecto de λ .

Proposición 4.2.4. $\lim_{t \rightarrow \infty} r(t) = \frac{1}{2\alpha^2\beta}$.

4.2.2. Momentos

Proposición 4.2.5. Sea $T \sim BSMSN(\alpha, \beta, \lambda)$ y $Z \sim MSN(\lambda)$, entonces los momentos de orden r están dados por

$$E(T^r) = 2\beta^r \sum_{i=0}^r C_{2i}^{2r} \sum_{k=0}^i C_k^i \left(\frac{\alpha}{2}\right)^{2(r-k)} (2(r-k) - 1)!! + 2\beta^r \sum_{i=0}^{r-1} C_{2i+1}^{2r} H_i,$$

donde $H_i = \int_{-\infty}^{\infty} h_i(z) \phi(z) \Phi(\lambda u(z)) dz$ y $h_i(z) = \left(\frac{\alpha z}{2}\right)^{2r-(2i+1)} \left(\left(\frac{\alpha z}{2}\right)^2 + 1\right)^{(2i+1)/2}$.

Demostración:

$$\begin{aligned} E(T^r) &= \int_0^{\infty} t^r 2\phi(a_t(\alpha, \beta)) \Phi(\lambda u(a_t(\alpha, \beta))) \frac{t^{-3/2}}{2\alpha\sqrt{\beta}} (t + \beta) dt \\ &= \int_{-\infty}^{\infty} 2\beta^r \left(\frac{\alpha z}{2} + \sqrt{\left(\frac{\alpha z}{2}\right)^2 + 1}\right)^{2r} \phi(z) \Phi(\lambda u(z)) dz \\ &= 2\beta^r \int_{-\infty}^{\infty} \left\{ \sum_{i=0}^r C_{2i}^{2r} \sum_{k=0}^i C_k^i \left(\frac{\alpha z}{2}\right)^{2(r-k)} + \sum_{i=0}^{r-1} C_{2i+1}^{2r} h_i(z) \right\} \phi(z) \Phi(\lambda u(z)) dz \\ &= 2\beta^r \sum_{i=0}^r C_{2i}^{2r} \sum_{k=0}^i C_k^i \left(\frac{\alpha}{2}\right)^{2(r-k)} E(z^{2(r-k)}) + 2\beta^r \sum_{i=0}^{r-1} C_{2i+1}^{2r} \int_{-\infty}^{\infty} h_i(z) \phi(z) \Phi(\lambda u(z)) dz \\ &= 2\beta^r \sum_{i=0}^r C_{2i}^{2r} \sum_{k=0}^i C_k^i \left(\frac{\alpha}{2}\right)^{2(r-k)} (2(r-k) - 1)!! + 2\beta^r \sum_{i=0}^{r-1} C_{2i+1}^{2r} H_i, \end{aligned}$$

donde $h_i(z) = \left(\frac{\alpha z}{2}\right)^{2r-(2i+1)} \left(\left(\frac{\alpha z}{2}\right)^2 + 1\right)^{(2i+1)/2}$, $H_i = \int_{-\infty}^{\infty} h_i(z) \phi(z) \Phi(\lambda u(z)) dz$ debe ser calculada numéricamente y $E(z^{2(r-k)}) = (2(r-k) - 1)!!$ coincide con los correspondientes momentos pares de la distribución normal. \square

Nota. $(2n - 1)!! = (2n - 1) * (2n - 3) * .. * 1$.

4.2.3. Coeficiente de asimetría y curtosis

Las expresiones de los coeficientes de asimetría y curtosis se obtiene de las fórmulas dadas en (2.5) y (2.6). Éstos no son explícitos y deben ser calculados numéricamente. La Figura 4.3, muestra y nos da una idea del comportamiento de estos coeficientes.

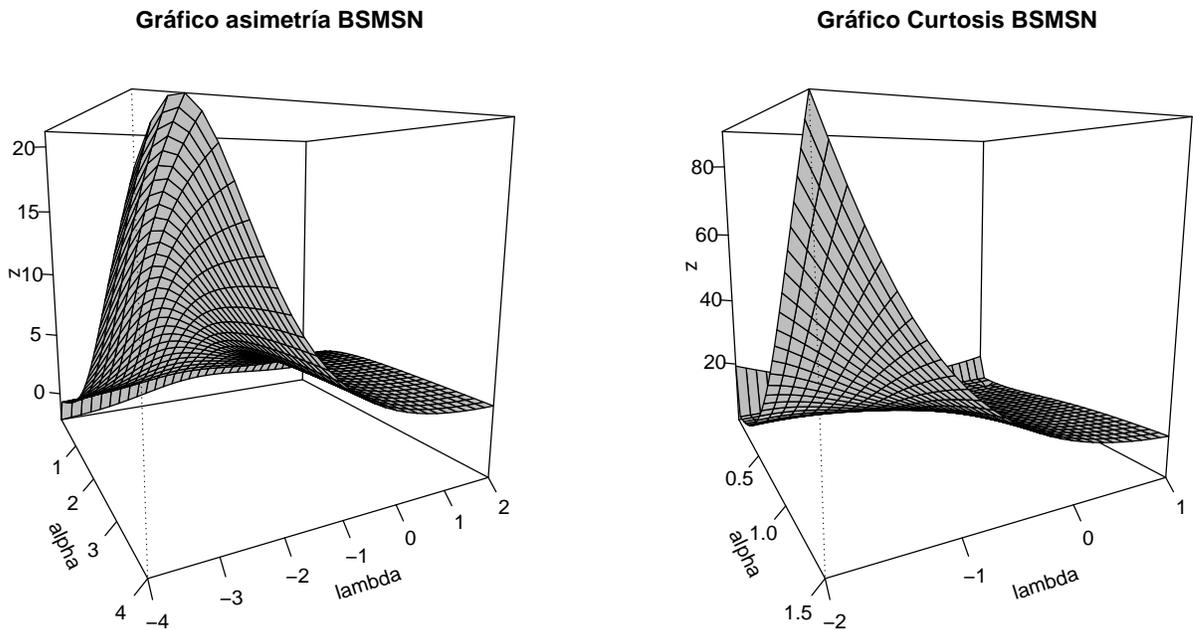


Figura 4.3: Gráfica de asimetría (γ_1) y curtosis (γ_2) del modelo BSMSN.

Respecto a la Figura 4.3, panel izquierdo, γ_1 converge a $2\sqrt{2}$ cuando $\lambda \rightarrow +\infty$ y $a \rightarrow +\infty$, sin embargo, cuando $\lambda \rightarrow -\infty$ y $a \rightarrow +\infty$, γ_1 crece indefinidamente. Por otra parte, si $\lambda \rightarrow \pm\infty$ y $a \rightarrow 0$ entonces $\gamma_1 \rightarrow -\infty$. La Figura 4.3, panel derecho, muestra que γ_2 converge a 15 cuando $\lambda \rightarrow +\infty$ y $a \rightarrow \infty$, sin embargo, cuando $\lambda \rightarrow -\infty$ y $a \rightarrow \infty$ ésta

crece indefinidamente, lo mismo ocurre cuando $\lambda \rightarrow \pm\infty$ y $a \rightarrow 0$. Por otra parte, el valor mínimo de la curtosis es $\gamma_2 = 1.95$. En ambas gráficas se consideró $\beta = 1$.

4.3. Estimación de máxima verosimilitud

4.3.1. Matriz información

Matriz información observada: Dada una muestra aleatoria de tamaño n , digamos t_1, \dots, t_n cuya distribución es BSMSN (θ) , con $\theta = (\alpha, \beta, \lambda)$. La log-verosimilitud del modelo está dada por

$$l(\theta) = c_0 - n \log(\alpha) - \frac{n}{2} \log(\beta) - \frac{1}{2} \sum_{i=1}^n a_i^2 - \frac{3}{2} \sum_{i=1}^n \log(t_i) + \sum_{i=1}^n \log(t_i + \beta) + \sum_{i=1}^n \log(\Phi(\lambda u(a_i))), \quad (4.4)$$

donde $c_0 = cte$, $a_i = a_{t_i}(\alpha, \beta) = \frac{1}{\alpha} \left(\sqrt{\frac{t_i}{\beta}} - \sqrt{\frac{\beta}{t_i}} \right)$, $t_i > 0$ para $i = 1, \dots, n$, $\alpha > 0$, $\beta > 0$ y $\lambda \in \mathbb{R}$. El vector score asociado a (4.4) es

$$\begin{aligned} S_\alpha &= \frac{n}{\alpha} \left(-1 + \bar{a}^2 - \lambda \bar{\rho}_{103} \right), \\ S_\beta &= n \left(-\frac{1}{2\beta} + \frac{1}{2n\alpha^2} \sum_{i=1}^n \left(\frac{t_i}{\beta^2} - \frac{1}{t_i} \right) + \frac{1}{n} \sum_{i=1}^n \frac{1}{t_i + \beta} - \frac{\lambda}{2\alpha\beta} \bar{\rho}_{013} \right), \\ S_\lambda &= n \bar{\rho}_{101}. \end{aligned}$$

Las entradas de la matriz información observada, según fórmula (2.12), son (Ver Apéndice 4.6)

$$\begin{aligned}
J_{\alpha\alpha} &= \frac{n}{\alpha^2} \left(1 - 3\bar{a}^2 + \lambda(\bar{\rho}_{103} + \bar{\rho}_{105} - 2\bar{\rho}_{305}) - \lambda^3\bar{\rho}_{307} - \lambda^2\bar{\eta}_{203} \right), \\
J_{\beta\alpha} &= \frac{n}{2\alpha^2\beta} \left(-\frac{2}{n\alpha} \sum_{i=1}^n \left(\frac{t_i}{\beta} - \frac{\beta}{t_i} \right) + \lambda(\bar{\rho}_{015} - 2\bar{\rho}_{215}) - \lambda^3\bar{\rho}_{217} - \lambda^2\bar{\eta}_{113} \right), \\
J_{\alpha\lambda} &= \frac{n}{\alpha} (-\bar{\rho}_{103} + \lambda^2\bar{\rho}_{305} + \lambda\bar{\eta}_{202}), \\
J_{\beta\beta} &= n \left\{ \frac{1}{2\beta^2} - \frac{1}{n\alpha^2\beta^3} \sum_{i=1}^n t_i - \frac{1}{n} \sum_{i=1}^n \frac{1}{(t_i + \beta)^2} + \frac{1}{4\alpha^2\beta^2} [\lambda(2\alpha\bar{\rho}_{013} + \alpha^2\bar{\rho}_{103}) \right. \\
&\quad \left. - \lambda^3(\alpha^2\bar{\rho}_{307} + 4\bar{\rho}_{107}) - \lambda^2(\alpha^2\bar{\eta}_{203} + 4\bar{\eta}_{003}) - 3\lambda(\alpha^2\bar{\rho}_{305} + 4\bar{\rho}_{105})] \right\}, \\
J_{\lambda\beta} &= \frac{n}{2\alpha\beta} \{ -\bar{\rho}_{013} + \lambda^2\bar{\rho}_{215} + \lambda\bar{\eta}_{112} \}, \\
J_{\lambda\lambda} &= -n(\lambda\bar{\rho}_{303} + \bar{\eta}_{201}),
\end{aligned}$$

donde

$$\begin{aligned}
u(x) &= \frac{x}{(1+x^2)^{1/2}}, \\
\zeta_i = \zeta(\lambda u(a_i)) &= \frac{\phi(\lambda u(a_i))}{\Phi(\lambda u(a_i))}, \\
\tilde{w}_i &= \sqrt{\frac{t_i}{\beta}} + \sqrt{\frac{\beta}{t_i}}, \\
\bar{\eta}_{npm} &= \frac{1}{n} \sum_{i=1}^n \frac{a_i^n \tilde{w}_i^p \zeta_i^2}{(1+a_i^2)^m}, \\
\bar{\rho}_{npm} &= \frac{1}{n} \sum_{i=1}^n \frac{a_i^n \tilde{w}_i^p \zeta_i}{(1+a_i^2)^{m/2}}, \\
\bar{a}^2 &= \frac{1}{n} \sum_{i=1}^n a_i^2.
\end{aligned}$$

Matriz información de Fisher: La log-verosimilitud para una variable aleatoria T que se distribuye BSMSN(θ), con $\theta = (\alpha, \beta, \lambda)$, está dada por

$$l(\theta) = -\frac{1}{2} \log\left(\frac{2}{\pi}\right) - \log(\alpha) - \frac{1}{2} \log(\beta) - \frac{\alpha^2}{2} - \frac{3}{2} \log(T) + \log(T + \beta) + \log(\Phi(\lambda u(a_T))) \quad (4.5)$$

donde $a = a_T(\alpha, \beta) = \frac{1}{\alpha} \left(\sqrt{\frac{T}{\beta}} - \sqrt{\frac{\beta}{T}} \right)$, $T > 0$, $\alpha > 0$, $\beta > 0$ y $\lambda \in \mathbb{R}$. El vector función score asociado a (4.5) es

$$\begin{aligned} S_\alpha &= \frac{1}{\alpha} \left(-1 + a^2 - \frac{\lambda a \zeta}{(1 + a^2)^{3/2}} \right), \\ S_\beta &= -\frac{1}{2\beta} + \frac{1}{2\alpha^2} \left(\frac{T}{\beta^2} - \frac{1}{T} \right) + \frac{1}{T + \beta} - \frac{\lambda}{2\alpha\beta} \frac{\tilde{w}\zeta}{(1 + a^2)^{3/2}}, \\ S_\lambda &= \frac{a\zeta}{(1 + a^2)^{1/2}}. \end{aligned}$$

Resolviendo el sistema dado por las funciones score, iguales a 0, obtenemos el siguiente sistema de ecuaciones simplificado $-\frac{1}{2\beta} + \frac{1}{2n\alpha^2} \sum_{i=1}^n \left(\frac{t_i}{\beta^2} - \frac{1}{t_i} \right) + \frac{1}{n} \sum_{i=1}^n \frac{1}{t_i + \beta} = \frac{\hat{\lambda}}{2\alpha\beta} \bar{\rho}_{013}$, $\bar{a}^2 - 1 = \lambda \bar{\rho}_{103}$ y $\bar{\rho}_{101} = 0$, la solución basada en métodos numéricos, nos entregará los EMV de los parámetros del modelo.

Las entradas de la matriz información de Fisher, según fórmula (2.13), tienen las siguientes expresiones (Ver Apéndice 4.6)

$$\begin{aligned} I_{\alpha\alpha} &= \frac{1}{\alpha^2} (2 + \lambda^2 \eta_{203}), \\ I_{\beta\alpha} &= -\frac{1}{2\alpha^2\beta} \left\{ -\frac{2}{\alpha} \left[\frac{E(T)}{\beta} - \beta E\left(\frac{1}{T}\right) \right] + \lambda(\rho_{015} - 2\rho_{215}) - \lambda^3 \rho_{217} - \lambda^2 \eta_{113} \right\}, \\ I_{\alpha\lambda} &= -\frac{1}{\alpha} \lambda \eta_{202}, \\ I_{\beta\beta} &= -\frac{1}{2\beta^2} + \frac{E(T)}{\alpha^2\beta^3} + E\left(\frac{1}{(T + \beta)^2}\right) + \frac{1}{4\alpha^2\beta^2} \{-2\lambda\alpha\rho_{013} + \lambda^2(\alpha^2\eta_{203} + 4\eta_{003})\}, \\ I_{\lambda\beta} &= -\frac{1}{2\alpha\beta} (-\rho_{013} + \lambda^2\rho_{215} + \lambda\eta_{112}), \\ I_{\lambda\lambda} &= \eta_{201}, \end{aligned}$$

donde

$$\begin{aligned}
u(x) &= \frac{x}{(1+x^2)^{1/2}}, \\
\zeta = \zeta(\lambda u(a)) &= \frac{\phi(\lambda u(a))}{\Phi(\lambda u(a))}, \\
\tilde{w} &= \sqrt{\frac{T}{\beta}} + \sqrt{\frac{\beta}{T}}, \\
\rho_{npm} &= E\left(\frac{a^n \tilde{w}^p \zeta}{(1+a^2)^{m/2}}\right), \\
\rho_{npm} &= 0 \quad \text{si } n = \text{impar} \quad \text{y } p = 0, \\
\eta_{npm} &= E\left(\frac{a^n \tilde{w}^p \zeta^2}{(1+a^2)^m}\right).
\end{aligned}$$

Para el caso $\lambda = 0$, tenemos que

$$I(\alpha, \beta, 0) = \begin{pmatrix} \frac{2}{\alpha^2} & 0 & 0 \\ 0 & d_{22} & d_{23} \\ 0 & d_{23} & d_{33} \end{pmatrix},$$

donde

$$\begin{aligned}
d_{23} &= 2\sqrt{\frac{2}{\pi}} \int_0^\infty \frac{\sqrt{1 + (\frac{\alpha z}{2})^2}}{(1+z^2)^{3/2}} \phi(z) dz, \\
d_{33} &= \frac{2}{\pi} [1 - (2\pi)^{1/2} e^{1/2} \Phi(-1)], \\
d_{22} &= \frac{1}{\alpha^2 \beta^2} + \frac{1}{\sqrt{2\pi} \beta^2} \left[\sqrt{\frac{\pi}{2}} - \frac{\pi e^{2/\alpha^2} (1 - \Phi(\frac{2}{\alpha}))}{\alpha} \right].
\end{aligned}$$

Así, la matriz I es no singular.

4.4. Reducción de sesgo del EMV

En simulaciones previas, observamos que el EMV del parámetro λ del modelo BSMSN en muchos casos sobrestima el valor del verdadero parámetro y en otros puede ser infinito, con

cierta probabilidad. En particular, cuando todos los términos $a_i > 0$, es decir, $\min(t_i) > \hat{\beta}$, el perfil de log-verosimilitud definido por $l_P(\lambda) = l(\hat{\alpha}(\lambda), \hat{\beta}(\lambda), \lambda)$, donde $\hat{\alpha}$ y $\hat{\beta}$ son los EMV de α y β para un valor de λ fijo, es monótona creciente y por lo tanto obtenemos un EMV infinito (Liseo [46]). Sin embargo, la probabilidad de estimaciones infinitas decrece rápidamente a medida que aumenta el tamaño de la muestra. Por otra parte, $\min(t_i) > \hat{\beta}$ es equivalente a $z_i = a_{t_i} > 0$, con $i = 1, \dots, n$, y $Z \sim MSN(\lambda)$ implica que $-Z \sim MSN(-\lambda)$, entonces centraremos el estudio sólo a valores positivos de λ . En las simulaciones señaladas anteriormente, también observamos que los EMV para los parámetros α y β tienen sesgo muy pequeño, por ello aplicaremos el método propuesto por Firth [28] para reducir el sesgo sólo al parámetro λ . Como mencionamos en 1.1.1, el sesgo del EMV de λ es de orden $O(n^{-1})$ y el sesgo del estimador modificado es de orden $O(n^{-2})$. Éste último, es solución de la ecuación de perfil de verosimilitud modificada U_M^P , dada en (2.16), donde el perfil de la función score U^P , dada por (2.15), se transforma en

$$U^P(\lambda) = \sum_{i=1}^n u(\tilde{a}_i) \zeta(\lambda u(\tilde{a}_i)),$$

con $\tilde{a}_i = \frac{1}{\hat{\alpha}} \left(\sqrt{\frac{t_i}{\hat{\beta}}} - \sqrt{\frac{\hat{\beta}}{t_i}} \right)$ y $\zeta(x) = \frac{\phi(x)}{\Phi(x)}$. Las expresiones de las funciones perfil de cuasi-verosimilitud $l_M^P(\lambda)$, asociada a $U_M^P(\lambda)$, y el estadístico razón de verosimilitud modificada $W_M^P(\lambda)$, asociada a $l_M^P(\lambda)$, son similares a los presentados en (2.17) y (2.18), pero utilizando el modelo BSMSN.

La siguiente proposición garantiza la existencia del EMV modificado λ_M .

Proposición 4.1. *Sea $M(\lambda)$ la modificación en (2.14) para la función score del modelo BSMSN(λ). Entonces, $M(\lambda) = \Omega(\lambda^{-1})$, es decir, ambas colas de la función M convergen a tasa λ^{-1} .*

La demostración se puede obtener en forma similar a la Proposición 2.2.

4.4.1. Estudio de simulación

Se realizó simulaciones de una variable aleatoria T correspondiente a la distribución $BSMSN(\alpha, \beta, \lambda)$, para distintos tamaños de muestra y distintos valores de los parámetros. Las muestras fueron obtenidas de su representación estocástica de la variable aleatoria T dada en (4.1), donde $Z \sim MSN(\lambda)$.

Tabla 4.1: Simulación de 5,000 iteraciones del modelo BSMSN(α, β, λ).

n	α	β	λ	$sesgo(\hat{\alpha})$	$sesgo(\hat{\beta})$	$sesgo(\hat{\lambda})^a$	$sesgo(\hat{\lambda}_M)$	$W_M^P(\lambda)IC$	$\%(\hat{\lambda} < \infty)$
50	0.5	1	5	-0.0037	0.0035	1.8412	-0.7553	94.5	87.2
100	0.5	1	5	-0.0020	0.0017	1.2481	-0.3145	94.0	98.1
200	0.5	1	5	-0.0010	0.0005	0.5644	-0.0940	94.5	99.9
50	0.5	1	10	-0.0087	0.0082	1.6201	-3.4943	88.3	65.6
100	0.5	1	10	-0.0039	0.0026	3.4054	-1.4599	91.7	88.2
200	0.5	1	10	-0.0010	0.0006	2.8928	-0.3672	93.3	98.3
50	1	1	5	-0.0065	0.0145	2.1261	-0.8364	93.5	87.6
100	1	1	5	-0.0001	0.0033	1.5285	-0.2849	94.2	97.5
200	1	1	5	0.0012	0.0014	0.6865	-0.0948	94.7	100.0
50	1	1	10	-0.0169	0.0173	1.4553	-3.6286	87.9	64.1
100	1	1	10	-0.0076	0.0050	3.3726	-1.5270	92.1	88.3
200	1	1	10	-0.0032	0.0019	2.9408	-0.4791	93.4	98.4
50	2	1	5	-0.0092	0.0788	2.0509	-1.2465	85.2	84.6
100	2	1	5	0.0085	0.0274	1.8715	-0.4878	92.1	96.9
200	2	1	5	0.0089	0.0045	0.8351	-0.1385	94.6	99.9
50	2	1	10	-0.0484	0.0550	1.4590	-4.1786	85.7	63.0
100	2	1	10	-0.0060	0.0134	3.7671	-1.6031	91.7	86.4
200	2	1	10	-0.0042	0.0069	2.9325	-0.6352	92.8	98.0

Sesgo de $\hat{\alpha}$, $\hat{\beta}$, $\hat{\lambda}$ y $\hat{\lambda}_M$, cobertura empírica con intervalo de confianza de 0.95 basado en $W_M^P(\lambda)$ y porcentaje empírico de casos que $\hat{\lambda}$ es finito. ^a Calculado cuando $\hat{\lambda} < \infty$.

En la Tabla 4.1, se observa que existe una sobrestimación importante del parámetro λ y también casos en que la estimación es ∞ . Ésta, depende del tamaño de la muestra y del

verdadero valor del parámetro de λ , pero no de los parámetros α y β . Sin embargo, la estimación para los parámetros α y β es bastante buena, ya que siempre son finitos y el sesgo es bastante pequeño. Por lo anterior, se aplica el método de Firth [28] sólo al parámetro λ . El estimador $\hat{\lambda}_M$ siempre existe y es finito. La reducción del sesgo de λ es bastante buena, considerando el hecho que la reducción se aplica cuando $\hat{\lambda}$ es finito y/o infinito. La cobertura empírica es cercana al valor nominal (95 %) y levemente menor, cuando el tamaño de muestra es pequeño, lo cual se debe al mayor porcentaje de $\hat{\lambda}$ infinitos.

4.5. Aplicación

Consideraremos un conjunto de datos, analizados anteriormente por Birnbaum y Saunders [20] y Martínez-Flórez et al. [51], correspondiente a tiempos de vida en ciclos $\times 10^{-3}$ de $n = 101$ piezas de aluminio 6061-T6, cortadas en paralelo a la dirección de rotación, a una velocidad de 18 ciclos por segundo y un esfuerzo máximo de 21.000 psi. La Tabla 4.2, muestra las estadísticas básicas que representan a los datos involucrados, esto es:

Tabla 4.2: Resumen estadístico de los datos.

n	<i>media</i>	<i>d.e.</i>	$\sqrt{\gamma_1}$	γ_2
101	1400.84	391.01	0.14	2.72

Tabla 4.3: EMV de los modelos BS, BSMSN y EXPBSn.

EMVs.	BS	BSMSN	EXPBSn
$\hat{\alpha}$	0.310(0.035)	0.498(0.049)	0.099(0.0001)
$\hat{\beta}$	1336.563(93.860)	1951.042(84.172)	2135.990(18.590)
$\hat{\lambda}$	-	-4.046(1.190)	0.052(0.005)
log-verosimilitud	-751.332	-746.574	-747.548
AIC	1510.664	1505.149	1507.096

La Tabla 4.3, muestra los EMV de los parámetros de las distribuciones BS, BSMSN y EXPBSn (Exponencial Birnbaum-Saunders Normal) con pdf

$$f_T(t; \alpha, \beta, \lambda) = \lambda \phi(a_t) \Phi(a_t)^{\lambda-1} \frac{t^{-3/2}}{2\alpha\sqrt{\beta}} (t + \beta),$$

donde $a_t = a_t(\alpha, \beta) = \frac{1}{\alpha} \left(\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}} \right)$, $t > 0$, $\alpha > 0$, $\beta > 0$ y $\lambda > 0$, y los valores entre paréntesis corresponden a los errores estándar. También, muestra los valores de log-verosimilitud y AIC, los que indican el mejor ajuste del modelo propuesto. Lo anterior, se puede ver en Figura 4.4. Además, si consideramos la prueba de hipótesis de no diferencia entre las distribuciones BS y BSMSN para los datos dados, esto es

$$H_0 : \lambda = 0 \quad \text{v/s} \quad H_1 : \lambda \neq 0.$$

Entonces, usando el estadístico de razón de verosimilitud

$$\Lambda = \frac{l_{BS(\hat{\alpha}, \hat{\beta})}}{l_{BSMSN(\hat{\alpha}, \hat{\beta}, \hat{\lambda})}},$$

tenemos que $-2 \log \Lambda = -2(-751.332 + 746.574) = 9.516$, cuyo valor es mayor que $\chi_1^2 = 3.84$, por lo que la hipótesis nula se rechaza. De esta manera, podemos decir que el modelo BSMSN se ajusta significativamente mejor a los datos dados.

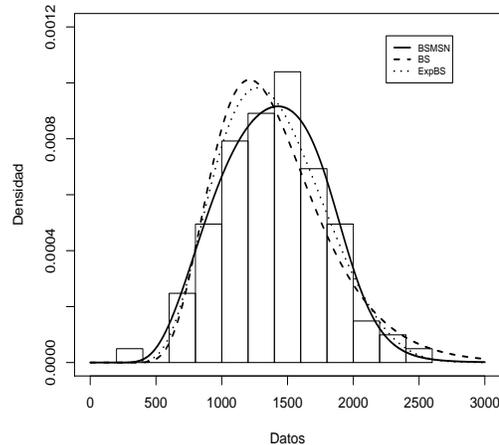


Figura 4.4: Ajuste datos.

La Figura 4.4, presenta el histograma de los datos con las correspondientes distribuciones ajustadas, $BSMSN(\hat{\alpha}, \hat{\beta}, \hat{\lambda})$ (línea sólida), $BS(\hat{\alpha}, \hat{\beta})$ (línea segmentada) y $EXPBSn(\hat{\alpha}, \hat{\beta}, \hat{\lambda})$ (línea de puntos).

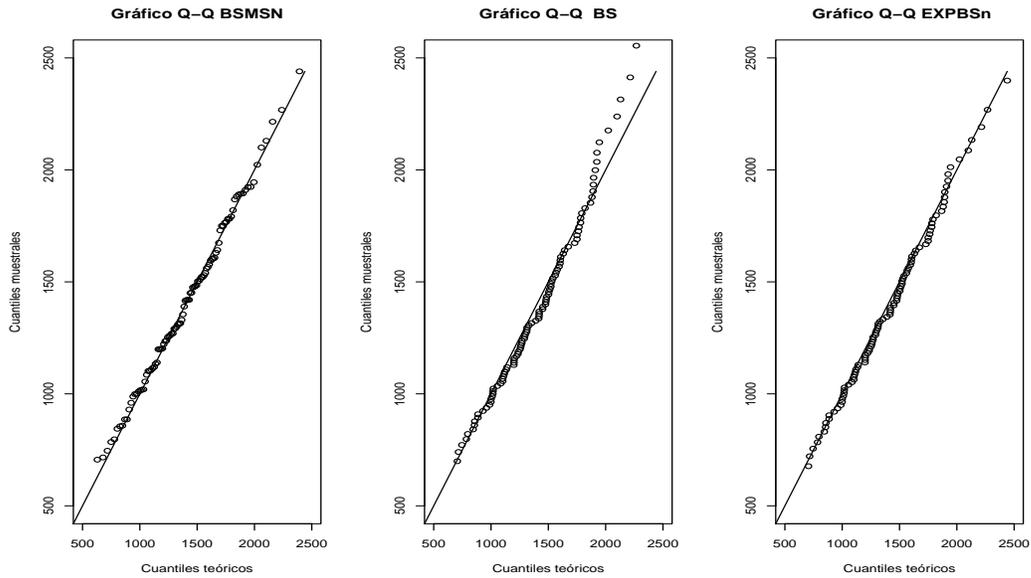


Figura 4.5: QQ plot de BSMSN, BS y EXPBSn.

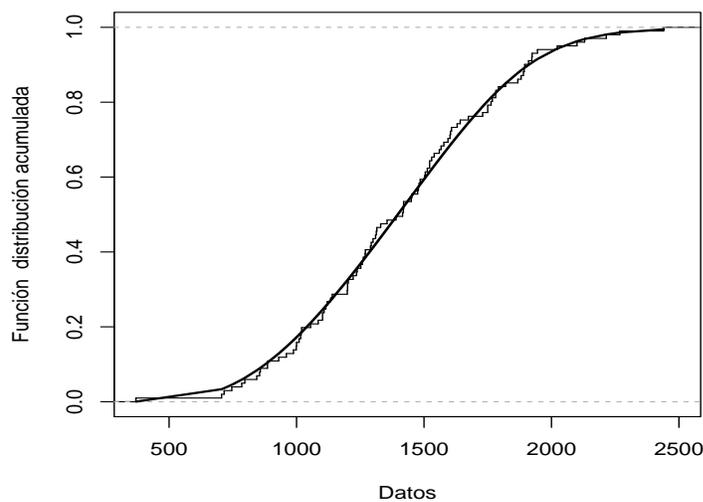


Figura 4.6: CDF empírica y teórica del modelo BSMSN.

Los QQ-plot de los modelos BSMSN, BS y EXPBSn y la función de distribución acumulada empírica y teórica de BSMSN se muestran en las Figuras 4.5 y 4.6. Estas gráficas, ratifican visualmente el mejor ajuste del modelo en estudio.

Tabla 4.4: EMV de α , β y λ .

$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\lambda}_M$	$l(\hat{\alpha}, \hat{\beta}, \hat{\lambda})$	$l(\hat{\alpha}, \hat{\beta}, \hat{\lambda}_M)$
0.498(0.047)	1951.236(80.390)	-4.043(1.131)	-	- 746.574	
0.498(0.050)	1951.236(90.789)	-	-3.342(0.913)	-	-746.880

La Tabla 4.4 muestra los EMV $\hat{\alpha}$, $\hat{\beta}$, $\hat{\lambda}$ y el EMV modificado $\hat{\lambda}_M$ y los respectivos errores estándar, entre paréntesis, fueron obtenidos de la matriz información de Fisher, dada en 4.3.1, puesto que la distribución asintótica $\hat{\theta}$ y $\hat{\theta}_M$ es $N_3(\theta, I^{-1}(\theta)/n)$, siendo $\theta = (\alpha, \beta, \lambda)$. Claramente, el valor del EMV modificado $\hat{\lambda}_M$ es mayor que el valor de EMV $\hat{\lambda}$ y por construcción $\hat{\lambda}_M$ tiene menor sesgo asintótico que $\hat{\lambda}$.

Tabla 4.5: Intervalos de confianza para λ .

	95 %	98 %	99 %
IC	(-5.905,-2.184)	(-6.367,-1.721)	(-6.676,-1.413)
IC^*	(-4.982,-1.976)	(-5.356,-1.602)	(-5.605,-1.353)

La Tabla 4.5, presenta los intervalos de confianza de λ para tres valores de niveles diferentes. Los intervalos de confianza obtenidos utilizando el EMV modificado $\hat{\lambda}_M$ (IC^*) son más precisos comparados con el intervalo de confianza basado en el EMV no modificado $\hat{\lambda}$ (IC)

debido a que son de menor longitud y consecuentemente sus estimaciones son más exactas.

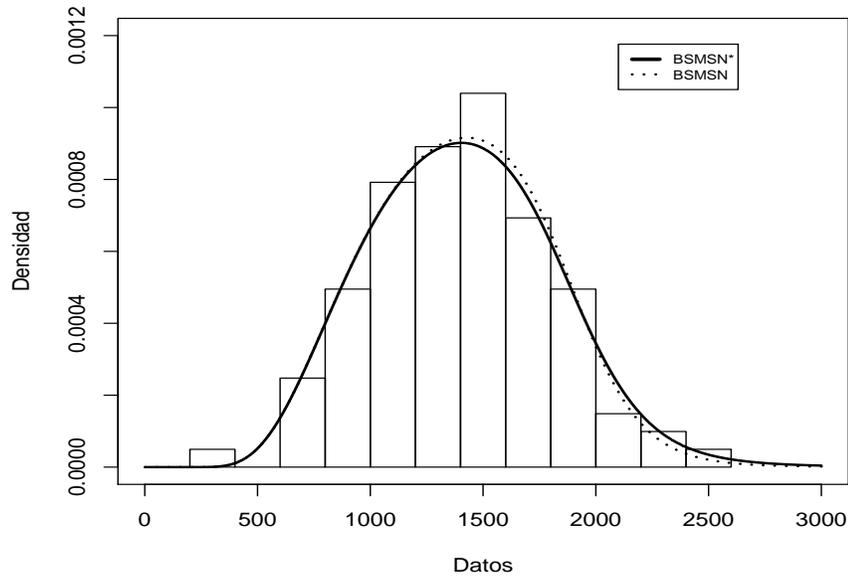


Figura 4.7: Datos ajustados.

La Figura 4.7, muestra el ajuste de los datos con $BSMSN(\hat{\alpha}, \hat{\beta}, \hat{\lambda})$ y el nuevo ajuste utilizando el EMV modificado de sesgo reducido $\hat{\lambda}_M$, esto es, $BSMSN^*(\hat{\alpha}, \hat{\beta}, \hat{\lambda}_M)$.

4.6. Apéndice

Sea $T \sim BSMSN(\alpha, \beta, \lambda)$. Consideremos la notación:

$$\tau = (\alpha, \beta), \quad u = u(a) = \frac{a}{(1+a^2)^{1/2}}, \quad \zeta = \zeta(\lambda u(a)) = \frac{\phi(\lambda u(a))}{\Phi(\lambda u(a))}$$

$$a = a_T(\alpha, \beta) = \frac{1}{\alpha} w, \quad w = \sqrt{\frac{T}{\beta}} - \sqrt{\frac{\beta}{T}}, \quad \tilde{w} = \sqrt{\frac{T}{\beta}} + \sqrt{\frac{\beta}{T}}$$

y las siguientes derivadas

$$\frac{\partial a}{\partial \alpha} = -\frac{1}{\alpha} a, \quad \frac{\partial a}{\partial \beta} = -\frac{1}{2\alpha\beta} \tilde{w}, \quad \frac{\partial u(a)}{\partial \tau} = \frac{1}{(1+a^2)^{3/2}} \frac{\partial a}{\partial \tau}$$

$$\begin{aligned} \frac{\partial}{\partial \tau} \left(\frac{a}{(1+a^2)^{3/2}} \right) &= \frac{1-2a^2}{(1+a^2)^{5/2}} \frac{\partial a}{\partial \tau} \\ \frac{\partial \zeta}{\partial \tau} &= \left(-\frac{\lambda^2 a \zeta}{(1+a^2)^2} - \frac{\lambda \zeta^2}{(1+a^2)^{3/2}} \right) \frac{\partial a}{\partial \tau} \\ \frac{\partial \zeta}{\partial \lambda} &= -\frac{\lambda a^2 \zeta}{(1+a^2)} - \frac{a \zeta^2}{(1+a^2)^{1/2}} \end{aligned}$$

tenemos que las funciones score de α , β y λ del modelo BSMSN son

$$\begin{aligned} S_\alpha &= \frac{1}{\alpha} \left(-1 + a^2 - \frac{\lambda a \zeta}{(1+a^2)^{3/2}} \right) \\ S_\beta &= -\frac{1}{2\beta} + \frac{1}{2\alpha^2} \left(\frac{T}{\beta^2} - \frac{1}{T} \right) + \frac{1}{T+\beta} - \frac{\lambda}{2\alpha\beta} \frac{\tilde{w}\zeta}{(1+a^2)^{3/2}} \\ S_\lambda &= \frac{a\zeta}{(1+a^2)^{1/2}} \end{aligned}$$

y las entradas de la matriz información de Fisher son:

$$\begin{aligned}
I_{\alpha\alpha} &= -E(S_{\alpha\alpha}) \\
&= -\frac{1}{\alpha^2} E \left(1 - 3a^2 + \lambda \frac{a\zeta}{(1+a^2)^{3/2}} + \lambda \frac{(a-2a^3)\zeta}{(1+a^2)^{5/2}} - \lambda^3 \frac{a^3\zeta}{(1+a^2)^{7/2}} - \lambda^2 \frac{a^2\zeta^2}{(1+a^2)^3} \right) \\
&= -\frac{1}{\alpha^2} (-2 + \lambda(\rho_{103} + \rho_{105} - 2\rho_{305}) - \lambda^3\rho_{307} - \lambda^2\eta_{203}) \\
&= \frac{1}{\alpha^2} (2 + \lambda^2\eta_{203})
\end{aligned}$$

$$\begin{aligned}
I_{\beta\alpha} &= -E(S_{\beta\alpha}) \\
&= -E \left(-\frac{1}{\alpha^3\beta} w\tilde{w} + \frac{1}{2\alpha^2\beta} \left[\frac{\lambda(1-2a^2)\tilde{w}\zeta}{(1+a^2)^{5/2}} - \frac{\lambda^3 a^2 \tilde{w}\zeta}{(1+a^2)^{7/2}} - \frac{\lambda^2 a \tilde{w}\zeta^2}{(1+a^2)^3} \right] \right) \\
&= -\frac{1}{2\alpha^2\beta} \left(-\frac{2}{\alpha} \left[\frac{E(T)}{\beta} - \beta E \left(\frac{1}{T} \right) \right] + \lambda(\rho_{015} - 2\rho_{215}) - \lambda^3\rho_{217} - \lambda^2\eta_{113} \right)
\end{aligned}$$

$$\begin{aligned}
I_{\alpha\lambda} &= -E(S_{\alpha\lambda}) \\
&= -E \left(-\frac{1}{\alpha} \frac{a}{(1+a^2)^{3/2}} \left\{ \zeta - \frac{\lambda^2 a^2 \zeta}{(1+a^2)} - \frac{\lambda a \zeta^2}{(1+a^2)^{1/2}} \right\} \right) \\
&= -\frac{1}{\alpha} (-\rho_{103} + \lambda^2\rho_{305} + \lambda\eta_{202}) \\
&= -\frac{1}{\alpha} \lambda\eta_{202}
\end{aligned}$$

$$\begin{aligned}
I_{\beta\beta} &= -E(S_{\beta\beta}) \\
&= -E \left(\frac{1}{2\beta^2} - \frac{T}{\alpha^2\beta^3} - \frac{1}{(T+\beta)^2} + \frac{\lambda(2\alpha\tilde{w} + \alpha^2 a)\zeta}{4\alpha^2\beta^2(1+a^2)^{3/2}} - \frac{\lambda^3(\alpha^2 a^3 + 4a)\zeta}{4\alpha^2\beta^2(1+a^2)^{7/2}} - \frac{\lambda^2(\alpha^2 a^2 + 4)\zeta^2}{4\alpha^2\beta^2(1+a^2)^3} \right. \\
&\quad \left. - \frac{3\lambda(\alpha^2 a^3 + 4a)\zeta}{4\alpha^2\beta^2(1+a^2)^{5/2}} \right) \\
&= -\frac{1}{2\beta^2} + \frac{E(T)}{\alpha^2\beta^3} + E \left(\frac{1}{(T+\beta)^2} \right) + \frac{1}{4\alpha^2\beta^2} \{ -\lambda(2\alpha\rho_{013} + \alpha^2\rho_{103}) + \lambda^3(\alpha^2\rho_{307} + 4\rho_{107}) \\
&\quad + \lambda^2(\alpha^2\eta_{203} + 4\eta_{003}) + 3\lambda(\alpha^2\rho_{305} + 4\rho_{105}) \} \\
&= -\frac{1}{2\beta^2} + \frac{E(T)}{\alpha^2\beta^3} + E \left(\frac{1}{(T+\beta)^2} \right) + \frac{1}{4\alpha^2\beta^2} \{ -2\lambda\alpha\rho_{013} + \lambda^2(\alpha^2\eta_{203} + 4\eta_{003}) \}
\end{aligned}$$

$$\begin{aligned}
I_{\lambda\beta} &= -E(S_{\lambda\beta}) \\
&= -\frac{1}{2\alpha\beta}E\left(-\frac{\tilde{w}\zeta}{(1+a^2)^{3/2}} + \frac{\lambda^2 a^2 \tilde{w}\zeta}{(1+a^2)^{5/2}} + \frac{\lambda a \tilde{w}\zeta^2}{(1+a^2)^2}\right) \\
&= -\frac{1}{2\alpha\beta}(-\rho_{013} + \lambda^2 \rho_{215} + \lambda \eta_{112})
\end{aligned}$$

$$\begin{aligned}
I_{\lambda\lambda} &= -E(S_{\lambda\lambda}) \\
&= -E\left(-\frac{\lambda a^3 \zeta}{(1+a^2)^{3/2}} - \frac{a^2 \zeta^2}{1+a^2}\right) \\
&= \lambda \rho_{303} + \eta_{201} \\
&= \eta_{201}
\end{aligned}$$

Capítulo 5

Distribución skew t-normal modificada

5.1. Introducción

La idea en este Capítulo, es estudiar el efecto de la reducción del sesgo en otros modelo ligados al modelo MSN, dada en (2.3). En particular, el modelo skew t-normal modificada, en adelante MStN, se puede representar una como mezcla en el parámetro de escala del modelo MSN, esto es:

$$Z|V = v \sim MSN(0, V^{-1}, \lambda), \quad \text{con } V \sim \text{Gamma}(\nu/2, \nu/2).$$

El objetivo de introducir este modelo, es disponer de una distribución más robusta para realizar inferencia en datos que poseen colas más pesadas y a la vez, rescatar la aplicación de la metodología de reducción de sesgo planteada para el modelo MSN. Esta nueva distribución se obtiene reemplazando la densidad normal en (2.3) por la densidad Student- t , con ν grados

de libertad y tiene la ventaja que la matriz información de Fisher, bajo la hipótesis de simetría, es no singular. Esta distribución pertenece a la familia de distribuciones skew simétricas introducidas por Wang, Boyer and Genton [65]. Kheradmandi et al. [39] estudia las propiedades de la distribución skew t-normal generalizada, siendo el modelo MStN un caso particular de tipo I y $\lambda_2 = 1$. El modelo MStN contiene como caso límite a la distribución MSN y tiene una mayor flexibilidad, respecto de la asimetría y curtosis, al incorporar los grados de libertad de la Student- t .

5.2. Definición y propiedades del modelo MStN

Diremos que una variable aleatoria Z tiene distribución skew t-normal modificada con parámetro de asimetría λ y ν grados de libertad, denotada por $Z \sim MStN(\lambda, \nu)$, si su densidad tiene la forma

$$f_Z(z; \lambda, \nu) = 2t_\nu(z)\Phi(\lambda u(z)), \quad (5.1)$$

donde $u(z) = \frac{z}{\sqrt{1+z^2}}$, $z \in \mathbb{R}$, $\nu > 0$ y $\lambda \in \mathbb{R}$. Aquí, t_ν y Φ denotan la función densidad Student- t con ν grados de libertad y la función de distribución $N(0,1)$, respectivamente. Si $\lambda = 0$, entonces la densidad MStN en (5.1) se reduce a la densidad Student- t . En particular, cuando $\lambda \rightarrow \infty$ el modelo tiende a dos veces la densidad Student- t con soporte positivo. Además, cuando $\nu \rightarrow \infty$ el modelo converge a la distribución skew-normal modificada. La siguiente Figura muestra la forma de la densidad para distintos valores de λ .

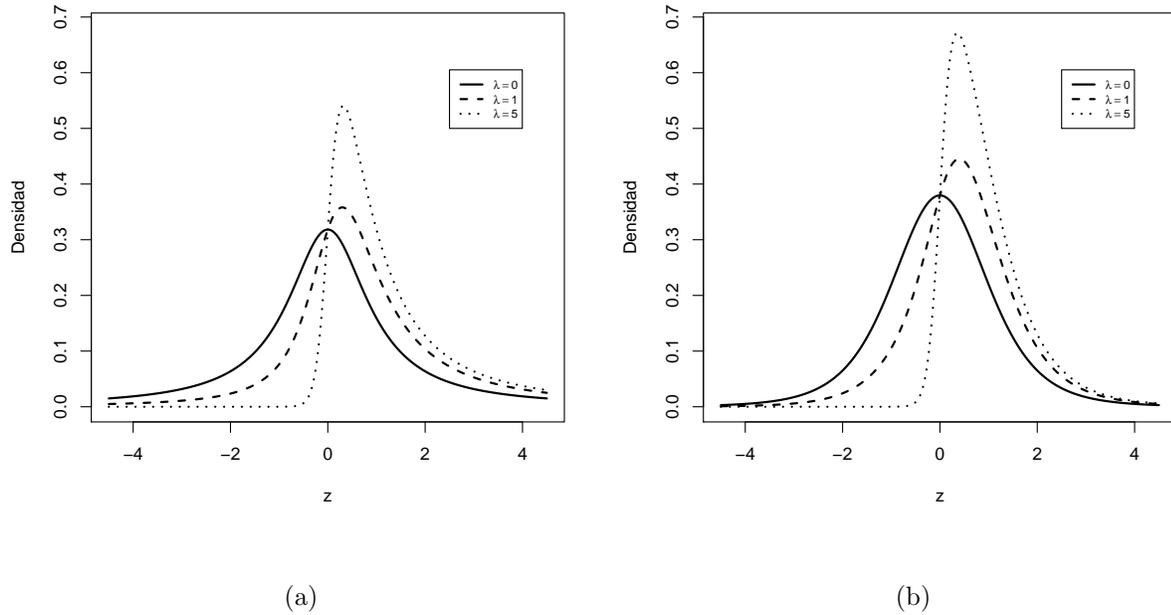


Figura 5.1: Gráfica de MStN para a) $\nu = 1$ y b) $\nu = 5$.

En la Figura 5.1, se puede apreciar el comportamiento de la densidad para distintos valores de ν y λ . La línea continua gruesa considera el caso $\lambda = 0$, es decir, cuando el modelo MStN coincide con la densidad Student- t . En ambas figuras, se observa como la curtosis y asimetría se ven afectadas a medida que ν y λ varían.

Para el caso localización y escala, es decir, $X = \mu + \sigma Z$, donde $\mu \in \mathbb{R}$, $\sigma > 0$, $\lambda \in \mathbb{R}$ y $Z \sim MStN(\lambda, \nu)$, con $\nu > 0$, denotada por $X \sim MStN(\mu, \sigma, \lambda, \nu)$, entonces tenemos que su densidad es

$$f_X(x; \mu, \sigma, \lambda, \nu) = \frac{2}{\sigma} t_\nu \left(\frac{x - \mu}{\sigma} \right) \Phi \left(\lambda u \left(\frac{x - \mu}{\sigma} \right) \right). \quad (5.2)$$

La relevancia del modelo skew t-normal modificada con localización y escala, es que la matriz de información de Fisher asociada es no singular cuando $\lambda = 0$.

Propiedades Sea $Z \sim MStN(\lambda, \nu)$, entonces:

1. $Z \sim MStN(\lambda, \nu) \Rightarrow -Z \sim MStN(-\lambda, \nu)$.
2. $Z|S = s \sim StN(S, \nu)$ y $S \sim N(\lambda, 1)$ entonces $Z \sim MStN(\lambda, \nu)$.
3. $Z|S = s, V = \nu \sim SN(0, \nu^{-1}, \nu^{-1/2}s)$, $V \sim Gamma(\nu/2, \nu/2)$ y $S \sim N(\lambda, 1)$ entonces $Z \sim MStN(\lambda, \nu)$.
4. $Z|V = \nu \sim MSN(0, \nu^{-1}, \lambda)$, con $V \sim Gamma(\nu/2, \nu/2)$ entonces $Z \sim MStN(\lambda, \nu)$.
5. $Z \sim MStN(\lambda, \nu)$ entonces $|Z| \sim 2t_\nu$ (con soporte positivo). En particular , cuando $\nu \rightarrow \infty$, entonces $|Z| \sim HN(0, 1)$.
6. $MStN(0, \nu) = t_\nu$.
7. $MStN(\lambda, 1) = MSCN$ (skew Cauchy normal Modificada).
8. Si $\nu \rightarrow \infty$, entonces $MStN(\lambda, \nu) \rightarrow MSN(\lambda)$. En particular si $\lambda = 0$, entonces $MStN(\lambda, \nu) \rightarrow MSN(0) = N(0, 1)$.

La Propiedad 2, nos muestra cómo nace la distribución MStN. Esto es, la mezcla en el parámetro de forma entre la distribución skew t-normal y la distribución normal. La Propiedad 3, presenta a la distribución MStN como una mezcla de forma y escala de la distribución skew-normal. La Propiedad 4, representa a la distribución MStN como una mezcla de forma y escala de la distribución MSN. La demostración de la Propiedad 2, 3 y 4 se obtienen utilizando el concepto de distribución condicional y aplicando el Lemma 2, dado por Azzalini [12]. Las restantes propiedades se demuestran en forma directa.

5.2.1. Momentos

Los momentos los podemos obtener utilizando la propiedad (2), es decir, podemos escribir $E(Z^k) = E(E(Z^k|S = s))$, donde $E(Z^k|S = s)$ con $k = 0, 1, \dots$, son los momentos condicionales del modelo StN dado el parámetro de forma S y $S \sim N(\lambda, 1)$, $\nu > k$. Luego, cuando k es par los momentos de la distribución StN coinciden con los momentos de la distribución Student- t . Por lo tanto, los momentos pares para el modelo MStN, están dados por

$$E(Z^k) = \frac{\Gamma\left(\frac{k+1}{2}\right)\Gamma\left(\frac{\nu-k}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}\nu^{k/2}.$$

En particular, $d_2 = E(Z^2) = \frac{\nu}{\nu-2}$, $\nu > 2$ y $d_4 = E(Z^4) = \frac{3\nu^2}{(\nu-2)(\nu-4)}$, $\nu > 4$.

Cuando k y ν son impares, consideramos la expresión dada por Gómez et al. [31] correspondiente a los momentos impares del modelo StN, entonces

$$\begin{aligned} E(Z^k) &= E(E(Z^k|S = s)) \\ &= \frac{\sqrt{2}\nu^{(k+1)/2}\Gamma\left(\frac{\nu-k-1}{2}\right)\Gamma\left(\frac{k+2}{2}\right)}{\pi\Gamma\left(\frac{k}{2}\right)}E\left[S {}_2F_2\left(\frac{1}{2}, \frac{2+k}{2}; \frac{3}{2}, \frac{3-\nu+k}{2}; \frac{\nu S^2}{2}\right)\right] \\ &+ \frac{2^{(k-\nu+4)/2}\nu^{\nu/2}\Gamma\left(\frac{\nu+1}{2}\right)}{(\nu-k)(\nu-k-1)\cos\left(\frac{(k-\nu)\pi}{2}\right)\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{\nu-k-1}{2}\right)} \\ &\times E\left[\text{sign}(S)S^{\nu-k} {}_2F_2\left(\frac{\nu+1}{2}, \frac{\nu-k}{2}; \frac{1+\nu-k}{2}, \frac{2+\nu-k}{2}; \frac{\nu S^2}{2}\right)\right], \end{aligned}$$

con $S \sim N(\lambda, 1)$, sign es la función signo y pFq es la función hipergeométrica dada por

$$pFq(u, g, z) = \sum_{r=1}^{\infty} \left(\prod_{i=1}^p \frac{\Gamma(u_i + r)}{\Gamma(u_i)z^r} \left(\prod_{i=1}^q \frac{\Gamma(g_i + r)}{\Gamma(g_i)} r! \right)^{-1} \right),$$

donde z es una variable, $u = [u_1, u_2, \dots, u_p]$ y $g = [g_1, g_2, \dots, g_q]$.

Para el caso, k impar y ν par, los momentos deben ser calculados en forma numérica. Los

momentos impares también pueden ser obtenidos a través de la fórmula

$$E(Z^k) = b_k - c_k,$$

con $k = 1, 3, \dots$, donde $b_k := b_k(\lambda, \nu) = \int_0^\infty 4x^k t_\nu(x) \Phi(\lambda u(x)) dx$ y $c_k := c_k(\nu) = c_1 \nu^{\frac{k-1}{2}} \prod_{i=1}^{\frac{k-1}{2}} \frac{(2i)!!}{\nu - (2i+1)}$, $k > 1$; con $c_1 := c_1(\nu) = \frac{2\nu\Gamma(\frac{\nu+1}{2})}{(\nu-1)\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}$ y $(2k)!! = 2k * (2k-2) * \dots * 2$.

5.2.2. Coeficiente de asimetría y curtosis

Para obtener los coeficientes de asimetría y curtosis en función de λ y ν dado, se utilizan las fórmulas entregadas en (2.6) y (2.5). Así, el coeficiente de asimetría, tiene la forma

$$\begin{aligned} \gamma_1(\lambda) &= \frac{E(Z^3) - 3E(Z^2)E(Z) + 2E^3(Z)}{(E(Z^2) - E^2(Z))^{3/2}} \\ &= \frac{b_3 - c_3 - 3d_2(b_1 - c_1) + 2(b_1 - c_1)^3}{(d_2 - (b_1 - c_1)^2)^{3/2}}. \end{aligned}$$

Dado que $E(Z^k; -\lambda) = -E(Z^k; \lambda)$, con k impar, se tiene que $\gamma_1(\lambda)$ es una función impar de λ , esto se observa en el panel izquierdo de la Figura 5.2. El mínimo y máximo de este coeficiente nos permite conocer el rango de asimetría para cada valor de ν . Éstos, se pueden observar en Tabla 5.1 y son obtenidos de la siguiente expresión

$$\lim_{\lambda \rightarrow \pm\infty} \gamma_1 = \pm c_1 \frac{\frac{2\nu}{\nu-3} - 3d_2 + 2c_1^2}{(d_2 - c_1^2)^{3/2}}.$$

Por otra parte, la expresión para el coeficiente de curtosis es

$$\begin{aligned} \gamma_2(\lambda) &= \frac{E(Z^4) - 4E(Z^3)E(Z) + 6E(Z^2)E^2(Z) - 3E^4(Z)}{(E(Z^2) - E^2(Z))^4} \\ &= \frac{d_4 - 4(b_3 - c_3)(b_1 - c_1) + 6d_2(b_1 - c_1)^2 - 3(b_1 - c_1)^4}{(d_2 - (b_1 - c_1)^2)^2}. \end{aligned}$$

Tenemos que $E(Z^k; -\lambda) = E(Z^k; \lambda)$, con k par, entonces $\gamma_2(\lambda)$ es una función par de λ como se muestra en el panel derecho de la Figura 5.2. El mínimo y máximo de este coeficiente,

para cada valor de ν , nos permite conocer su rango. Éstos, se pueden observar en Tabla 5.1 y son obtenidos de la siguientes fórmulas

$$\gamma_{min} = \lim_{\lambda \rightarrow 0} \gamma_2 = \frac{3(\nu - 2)}{\nu - 4}, \quad \nu > 4,$$

y

$$\gamma_{max} = \lim_{\lambda \rightarrow \pm\infty} \gamma_2 = \frac{d_4 - 4\frac{2\nu}{\nu-3}c_1^2 + 6d_2c_1^2 - 3c_1^4}{(d_2 - c_1^2)^2}.$$

Los coeficientes de asimetría y curtosis debe ser calculados numéricamente, debido a que no tienen forma explícita. Además, el rango de éstos, coinciden con el rango del modelo skew t-normal, dado por Gómez [31], con $\nu = 5, \dots, 20$. Estos rangos se presentan en Tabla 5.1.

Tabla 5.1: Rango de Asimetría, Curtosis para distintos valores de ν .

ν	Rango de Asimetría	Rango de Curtosis
5	(-2.550, 2.550)	(9.00, 23.109)
6	(-2.052, 2.052)	(6.00, 12.673)
7	(-1.798, 1.798)	(5.000, 9.461)
8	(-1.643, 1.643)	(4.500, 7.936)
9	(-1.539, 1.539)	(4.200, 7.054)
10	(-1.463, 1.463)	(4.000, 6.482)
11	(-1.407, 1.407)	(3.857, 6.082)
12	(-1.362, 1.362)	(3.750, 5.787)
13	(-1.326, 1.326)	(3.667, 5.561)
14	(-1.297, 1.297)	(3.600, 5.382)
15	(-1.272, 1.272)	(3.545, 5.237)
16	(-1.251, 1.251)	(3.500, 5.117)
17	(-1.233, 1.233)	(3.462, 5.017)
18	(-1.218, 1.218)	(3.429, 4.931)
19	(-1.204, 1.204)	(3.400, 4.857)
20	(-1.192, 1.192)	(3.375, 4.793)
∞	(-0.995, 0.995)	(3.000, 3.869)

La Figura 5.2, muestra que la asimetría es una función impar y cuyo rango disminuye a medida que aumenta ν . Cuando $\nu \rightarrow \infty$, se obtiene el rango de asimetría del modelo MSN y SN coinciden y su valor es ± 0.995 . Por otro lado, la curtosis es una función par cuyo límite inferior y superior de su rango disminuyen, así como también la amplitud del rango, conforme crece ν . El valor límite de la curtosis se obtiene cuando $\nu \rightarrow \infty$, coincidiendo el rango de curtosis del modelo MSN y SN, cuya variación es $(3, 3.869)$. Las Figuras respectivas de asimetría y curtosis para el modelo MStN son

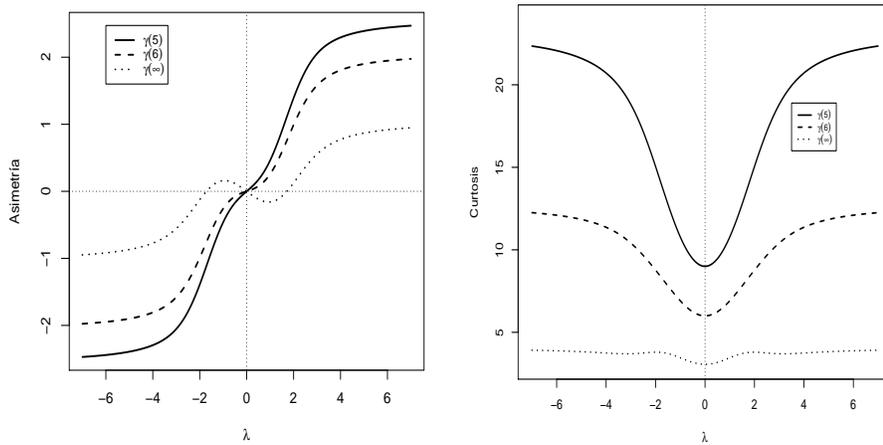


Figura 5.2: Gráfica de Asimetría y Curtosis para el modelo MStN.

5.3. Estimación de máxima verosimilitud

5.3.1. Matriz información

Matriz información de Fisher: La log-verosimilitud asociada a una muestra de tamaño $n = 1$ de una variable aleatoria X que se distribuye $MStN(\mu, \sigma, \lambda, \nu)$, ν conocido, está dada

por

$$l(\mu, \sigma, \lambda) = -\log(\sigma) - \frac{\nu+1}{2} \log\left(1 + \frac{Z^2}{\nu}\right) + \log(\Phi(\lambda u(Z))), \quad (5.3)$$

donde $Z = \frac{X-\mu}{\sigma}$, $X \in \mathbb{R}$, $\mu \in \mathbb{R}$, $\sigma > 0$, $\lambda \in \mathbb{R}$ y $\nu > 0$. El vector función score asociado a (5.3) es

$$\begin{aligned} S_\mu &= \frac{1}{\sigma} \left(-\frac{\lambda}{(1+Z^2)^{3/2}} \zeta(\lambda u(Z)) + \frac{\nu+1}{\nu} \frac{Z}{(1+\frac{Z^2}{\nu})} \right), \\ S_\sigma &= \frac{1}{\sigma} \left(-1 - \frac{\lambda Z}{(1+Z^2)^{3/2}} \zeta(\lambda u(Z)) + \frac{\nu+1}{\nu} \frac{Z^2}{(1+\frac{Z^2}{\nu})} \right), \\ S_\lambda &= u(Z) \zeta(\lambda u(Z)). \end{aligned}$$

Note que $S_\sigma = -\frac{1}{\sigma} + Z S_\mu$.

Las entradas de la matriz información de Fisher asociada a (5.3), de acuerdo a la fórmula dada en (2.13), serían las siguientes expresiones (Ver Apéndice 5.6):

$$\begin{aligned} I_{\mu\mu} &= \frac{1}{\sigma^2} \left(\lambda^2 \eta_{03} + \frac{\nu+2}{\nu+3} \right), \\ I_{\mu\sigma} &= -\frac{1}{\sigma^2} \left(\lambda \rho_{05} - 2\lambda \rho_{25} - \lambda^3 \rho_{27} - \lambda^2 \eta_{13} - \frac{2(\nu+1)}{\nu} \delta_2 \right), \\ I_{\mu\lambda} &= \frac{1}{\sigma} (\rho_{03} - \lambda^2 \rho_{25} - \lambda \eta_{12}), \\ I_{\sigma\sigma} &= \frac{1}{\sigma^2} \left(\lambda^2 \eta_{23} + \frac{2\nu}{\nu+3} \right), \\ I_{\sigma\lambda} &= -\frac{\lambda}{\sigma} \eta_{22}, \\ I_{\lambda\lambda} &= \eta_{21}, \end{aligned}$$

donde

$$\begin{aligned}\rho_{nm} &= E\left(\frac{Z^n}{(1+Z^2)^{m/2}}\zeta(\lambda u(Z))\right), \quad \rho_{nm} = 0 \quad \text{si } n = \text{impar}, \\ \eta_{nm} &= E\left(\frac{Z^n}{(1+Z^2)^m}\zeta^2(\lambda u(Z))\right), \\ \delta_n &= E\left(\frac{Z}{(1+\frac{Z^2}{\nu})^n}\right), \\ \zeta(x) &= \frac{\phi(x)}{\Phi(x)}.\end{aligned}$$

En particular, para $\lambda = 0$, tenemos que

$$I(\mu, \sigma, 0) = \begin{pmatrix} \frac{\nu+2}{\sigma^2(\nu+3)} & 0 & \frac{d_1}{\sigma} \\ 0 & \frac{2\nu}{\sigma^2(\nu+3)} & 0 \\ \frac{d_1}{\sigma} & 0 & d_2 \end{pmatrix},$$

donde

$$\begin{aligned}d_1 &= -\frac{2\sqrt{2}}{\pi\nu\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{\nu}{2}\right)} \left\{ \pi\Gamma\left(\frac{\nu+2}{2}\right) {}_2F_1\left[\frac{1}{2}, \frac{3}{2}, \frac{2-\nu}{2}, \nu\right] \right. \\ &\quad \left. - 2\nu^{\nu/2}\Gamma\left(\frac{2-\nu}{2}\right)\Gamma\left(\frac{\nu+1}{2}\right)\Gamma\left(\frac{\nu+3}{2}\right) {}_2F_1\left[\frac{\nu+1}{2}, \frac{\nu+3}{2}, \frac{\nu+2}{2}, \nu\right] \right\}, \\ d_2 &= -\frac{\sqrt{\nu}}{\sqrt{\pi}2^\nu(1-\nu)^{\frac{\nu+3}{2}}\nu\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{\nu}{2}\right)} \left\{ -2(\nu-1)\nu^{\frac{\nu+1}{2}}\Gamma\left(-\frac{\nu}{2}\right)\Gamma(\nu) + 2^\nu(1-\nu)^{\frac{\nu+1}{2}} \right. \\ &\quad \left. \times \sqrt{\nu}\Gamma\left(\frac{\nu-2}{2}\right) \left((\nu-3){}_2F_1\left[-\frac{1}{2}, 1, \frac{4-\nu}{2}, \nu\right] - 2(\nu-1){}_2F_1\left[\frac{1}{2}, 1, \frac{4-\nu}{2}, \nu\right] \right) \right\},\end{aligned}$$

donde ${}_pF_q$ es la función hipergeométrica y las expresiones d_1 y d_2 fueron obtenidas del software Matemática [66]. Esta matriz es no singular, puesto que el $|I| \neq 0$, en efecto,

$$|I| = \frac{2\nu(\nu+2)d_2}{\sigma^4(\nu+3)^2}(1-h(\nu)),$$

donde la función $h(\nu) = \frac{(\nu+3)d_1^2}{(\nu+2)d_2}$ es creciente y converge al valor 0.926, según muestra la

Figura 5.3.

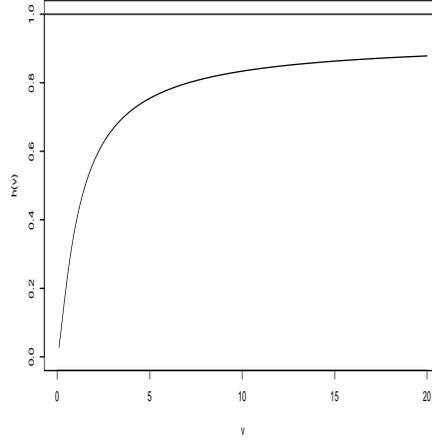


Figura 5.3: Gráfica de función $h(\nu)$.

Por otra parte, cuando $\nu \rightarrow \infty$, la matriz $I(\mu, \sigma, 0)$ coincide con la matriz información de Fisher del modelo MSN para el caso $\lambda = 0$.

Matriz información observada: La matriz información observada asociada a (5.3) tiene entradas, según la fórmula (2.12), como sigue (Ver Apéndice 5.6)

$$\begin{aligned}
 J_{\mu\mu} &= \frac{n}{\sigma^2} \left(3\lambda\ddot{\rho}_{15} + \lambda^3\ddot{\rho}_{17} + \lambda^2\ddot{\eta}_{03} - \frac{\nu+1}{\nu}\ddot{\delta}_{02} \right), \\
 J_{\mu\sigma} &= \frac{n}{\sigma^2} \left(\lambda\ddot{\rho}_{05} - 2\lambda\ddot{\rho}_{25} - \lambda^3\ddot{\rho}_{27} - \lambda^2\ddot{\eta}_{13} - 2\frac{\nu+1}{\nu}\ddot{\delta}_{12} \right), \\
 J_{\mu\lambda} &= \frac{n}{\sigma} \left(-\ddot{\rho}_{03} + \lambda^2\ddot{\rho}_{25} + \lambda\ddot{\eta}_{12} \right), \\
 J_{\sigma\sigma} &= \frac{n}{\sigma^2} \left(1 + \lambda\ddot{\rho}_{13} + \lambda\ddot{\rho}_{15} - 2\lambda\ddot{\rho}_{35} - \lambda^3\ddot{\rho}_{37} - \lambda^2\ddot{\eta}_{23} - 2\frac{\nu+1}{\nu}\ddot{\delta}_{22} - \frac{\nu+1}{\nu}\ddot{\delta}_{21} \right), \\
 J_{\sigma\lambda} &= -\frac{n}{\sigma} \left(-\ddot{\rho}_{13} + \lambda^2\ddot{\rho}_{35} + \lambda\ddot{\eta}_{22} \right), \\
 J_{\lambda\lambda} &= n(\lambda\ddot{\rho}_{33} + \ddot{\eta}_{21}),
 \end{aligned}$$

donde $\check{f} = \bar{f}(\mu, \sigma, \lambda, \nu) = \frac{1}{n} \sum_{i=1}^n f_i(\mu, \sigma, \lambda, \nu)$ y

$$\begin{aligned} u(x) &= \frac{x}{(1+x^2)^{1/2}}, \\ \zeta_i &= \frac{\phi(\lambda u(z_i))}{\Phi(\lambda u(z_i))}, \\ \eta_{pmi} &= \frac{z_i^p \zeta_i^2}{(1+z_i^2)^m}, \\ \rho_{pmi} &= \frac{z_i^p \zeta_i}{(1+z_i^2)^{m/2}}, \\ \delta_{pmi} &= \frac{z_i^p}{\left(1 + \frac{z_i^2}{\nu}\right)^m}. \end{aligned}$$

5.4. Reducción de sesgo del EMV de λ

El EMV del parámetro λ del modelo MStN, con ν fijo, sobrestima el valor del verdadero parámetro y puede ser infinito con probabilidad no despreciable, cuando toda la muestra es positiva, es decir, $\min(z_i) > 0$, $i = 1, \dots, n$. Lo anterior, se puede verificar con las simulaciones presentadas en la Tabla 5.2 que sigue

Tabla 5.2: Simulación de 5,000 iteraciones del modelo MStN(λ, ν), ν conocido.

n	λ	ν	$sesgo(\hat{\lambda}^a)$	$\%(\hat{\lambda} < \infty)$
20	5	3	1.99	70.68 (71.04)
50	5	3	2.00	95.10 (95.49)
100	5	3	0.82	99.74 (99.80)
20	10	3	1.52	45.62 (45.09)
50	10	3	4.58	77.32 (77.66)
100	10	3	3.73	95.18 (95.01)
20	5	5	2.17	71.48 (72.28)
50	5	5	2.00	95.36 (95.95)
100	5	5	0.86	99.90 (99.84)
20	10	5	2.31	46.44 (46.15)
50	10	5	5.00	79.08 (78.72)
100	10	5	3.64	96.16 (95.47)
20	5	10	2.06	74.52 (73.24)
50	5	10	1.97	96.24 (96.29)
100	5	10	0.83	99.84 (99.86)
20	10	10	2.67	47.26 (47.00)
50	10	10	4.38	80.40 (79.55)
100	10	10	3.88	95.54 (95.82)

$\%(\hat{\lambda} < \infty)$ porcentaje empírico (teórico) de casos cuando $\hat{\lambda}$ existe. ^a Calculado cuando $\hat{\lambda} < \infty$.

Esta probabilidad de divergencia en la EMV de λ , aumenta cuando los verdaderos valores de λ y ν aumentan, pero disminuye rápidamente conforme crece el tamaño de muestra. Esto, se observa en la Figura 5.4.

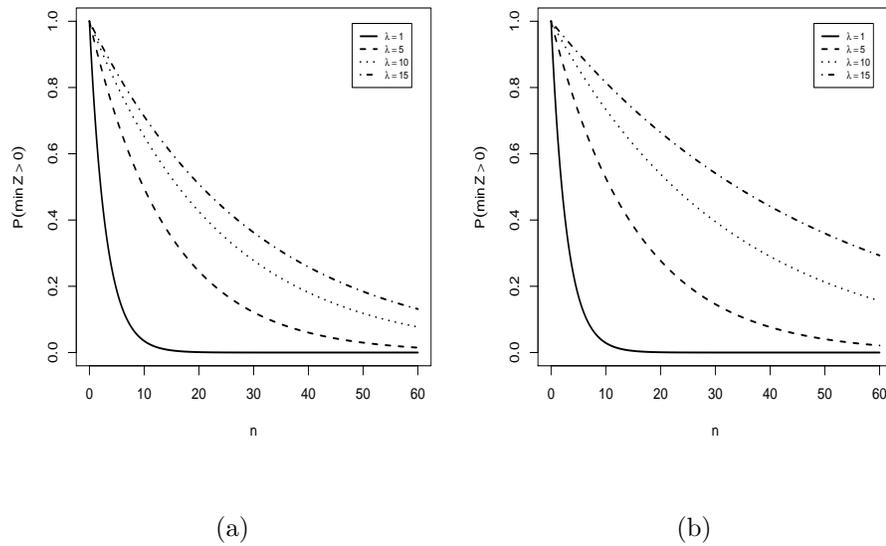


Figura 5.4: Probabilidad de divergencia para el modelo MStN con (a) $\nu = 1$ y (b) $\nu = 5$.

En el caso de parámetros de localización y escala, desconocidos, la sobrestimación de los parámetros ocurre sólo para el parámetro λ , esto se observa en la Tabla 5.3.

Tabla 5.3: Simulación de 5,000 iteraciones del modelo MStN(0,1, λ , ν), ν conocido.

n	λ	ν	$sesgo(\hat{\mu})$	$sesgo(\hat{\sigma})$	$sesgo(\hat{\lambda}^a)$	$\%(\hat{\lambda} < \infty)$
50	5	3	0.003	0.006	1.860	83.78
100	5	3	-0.005	0.010	1.797	96.98
200	5	3	-0.001	0.003	0.665	99.88
50	10	3	0.016	-0.010	1.052	61.80
100	10	3	0.004	0.001	3.499	85.24
200	10	3	0.000	0.002	3.211	97.60
50	5	5	0.003	0.004	1.840	85.54
100	5	5	-0.001	0.000	1.637	97.52
200	5	5	-0.001	0.003	0.666	99.96
50	10	5	0.016	-0.011	1.254	61.96
100	10	5	0.004	-0.003	3.645	86.08
200	10	5	0.001	-0.001	2.845	98.00
50	5	10	0.008	-0.006	1.798	87.10
100	5	10	0.000	0.001	1.525	98.22
200	5	10	0.000	0.002	0.641	99.94
50	10	10	0.017	-0.015	1.404	65.28
100	10	10	0.002	-0.001	3.621	87.80
200	10	10	-0.001	0.000	3.200	98.20

$\hat{\mu}$, $\hat{\sigma}$ y $\hat{\lambda}$ y porcentaje empírico de casos cuando $\hat{\lambda}$ existe. ^a Calculado cuando $\hat{\lambda} < \infty$.

Luego, como el sesgo de los EMV de μ y σ es prácticamente cero, se aplicará el método de Firth [28], para reducir el sesgo del EMV de λ de orden $O(n^{-1})$ (Cox and Snell [27]) y así obtener otro EMV $\widehat{\lambda}_M$ con sesgo de orden $O(n^{-2})$. Debido a que sí $Z \sim MStN(\lambda, \nu)$ entonces $-Z \sim MStN(-\lambda, \nu)$ centraremos el estudio para el caso $\lambda > 0$.

5.4.1. Caso: parámetro de forma

Consideramos el caso estándar, $\mu = 0$, $\sigma = 1$, ν conocido, y una muestra de n observaciones de z_1, \dots, z_n iid de $MStN(0, 1, \lambda, \nu)$, y su correspondiente log-verosimilitud dada en (5.3). Anteriormente, en 1.1.1 mencionamos que el sesgo del EMV de λ es de orden $O(n^{-1})$. Sin embargo, al aplicar el método de Firth al parámetro λ , obtenemos un nuevo estimador modificado cuyo sesgo es de orden $O(n^{-2})$ y es solución de la ecuación de verosimilitud modificada U_M dada en (1.1). Las expresiones de las funciones modificada (M), log-verosimilitud modificada (l_M) y el estadístico razón de verosimilitud (W_M), son similares a las presentados en (2.14), (1.3) y (1.4), respectivamente, pero utilizando el modelo $MStN$. En este caso,

$$a_{km}(\lambda) = \int_{-\infty}^{\infty} 2u^k(z)t_\nu(z)\Phi(\lambda u(z)) \left(\frac{\phi(\lambda u(z))}{\Phi(\lambda u(z))} \right)^m dz, \quad k = 2, 4 \quad y \quad m = 2.$$

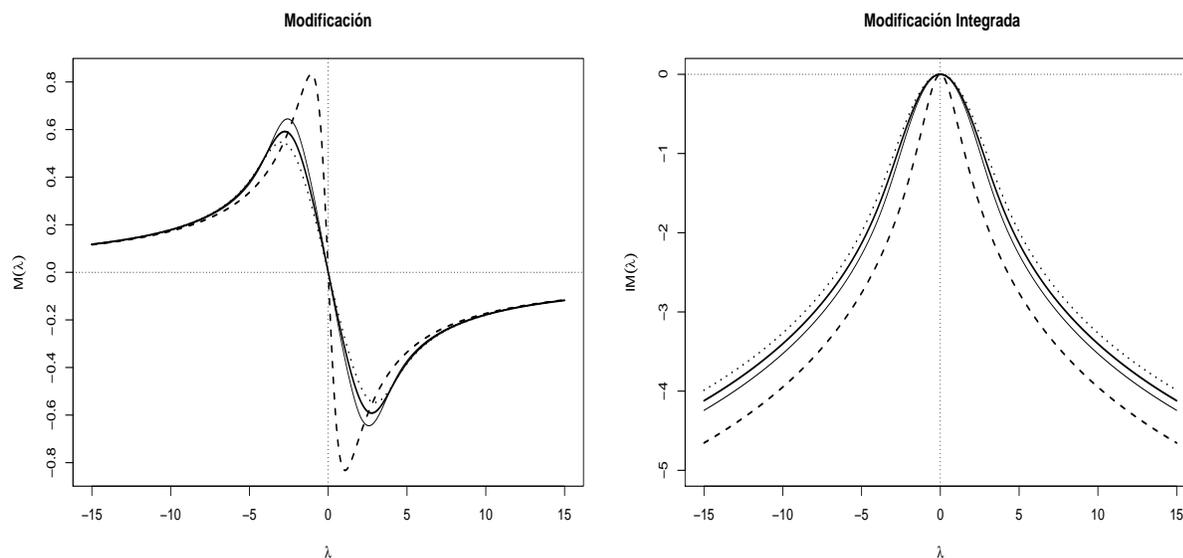


Figura 5.5: Función modificada (panel izquierdo) y función modificada integrada (panel derecho) para el modelo SN (línea segmentada), MStN con $\nu = 1$ (línea continua) y $\nu = 3$ (línea continua gruesa) y MSN (línea de puntos).

La Figura 5.4.1, panel izquierdo, muestra la forma de la función modificada $M(\lambda)$ para los modelos SN, MStN ,con $\nu = 1$ y $\nu = 3$, y MSN. Todas las funciones modificadas son impares y acotadas, para todo λ , con tasa de convergencia de orden $O(\lambda^{-1})$. El primer modelo tiene valor máximo $M_{SN}(1.07) = 0.83$, $M_{MStN}(2.58) = 0.64$ con $\nu = 1$, $M_{MStN}(2.75) = 0.59$ con $\nu = 3$ y $M_{MSN}(2.96) = 0.55$. Además, se observa que a medida que aumenta ν , la función modificada del modelo MStN tiende a la función modificada del modelo MSN. El panel derecho, muestra la forma de la función modificada integrada (IM), siendo ésta una función par y decreciente respecto de λ .

Para garantizar la existencia del estimador $\hat{\lambda}_M$ necesitamos la siguiente proposición

Proposición 5.1. *Sea $M(\lambda)$ la función modificación para el modelo $MStN(\lambda, \nu)$, entonces*

$M(\lambda) = \Omega(\lambda^{-1})$, $\forall \nu > 0$. Es decir, $M(\lambda)$ es acotada por arriba y por abajo.

Demostración:

Para demostrar que la función $a_{22}(\lambda)$ es simétrica y decreciente respecto de λ , se puede realizar en forma similar a la demostración dada por Sartori [60]. Esto es

$$\begin{aligned} a_{22}(\lambda) &= 2 \int_0^\infty u^2(z) t_\nu(z) \frac{\phi^2(\lambda u(z))}{\Phi(\lambda u(z))} dz + 2 \int_0^\infty u^2(-z) t_\nu(-z) \frac{\phi^2(-\lambda u(z))}{1 - \Phi(\lambda u(z))} dz \\ &= 2 \int_0^\infty u^2(z) t_\nu(z) \frac{\phi^2(-\lambda u(z))}{\Phi(\lambda u(z))[1 - \Phi(\lambda u(z))]} dz \\ &= a_{22}(-\lambda). \end{aligned}$$

Entonces, $a_{22}(\lambda)$ es simétrica respecto de λ .

Ahora, se analiza el signo de la derivada de $a_{22}(\lambda)$ para $\lambda > 0$, esto es

$$\frac{\partial a_{22}(\lambda)}{\partial \lambda} = -2 \int_0^\infty u^3(z) t_\nu(z) \frac{\phi^2(\lambda u(z))}{\Phi^2(\lambda u(z))[1 - \Phi(\lambda u(z))]^2} H(\lambda u(z)) dz,$$

donde $H(s) = 2s\Phi(s)[1 - \Phi(s)] + \phi(s)[1 - 2\Phi(s)]$. $\forall s > 0$, $s\Phi(-s) < \phi(s)$ (Sartori [60]),

entonces $H(s) > 0$. Luego, $a_{22}(\lambda)$ es una función decreciente de λ .

Por la simetría de $a_{22}(\lambda)$, analizamos sólo la convergencia de la cola derecha, considerando $\lambda > 1$.

Es conveniente tener las siguientes consideraciones para la demostración:

$\forall z > 0$ se tiene que:

1i) Para $\lambda > 1$, $\frac{1}{\lambda}u(\lambda z) < u(z) < u(\lambda z) < z < \lambda z$.

2i) Las funciones $\Phi(z)$ y $u(z)$ son crecientes. La función $t_\nu(z)$ es decreciente.

3i) $\frac{1}{1-\Phi(\lambda u(z))} < \frac{1}{1-\Phi(\lambda z)} < \frac{1}{1-\Phi(1)} = c = cte$ para $0 < z < 1/\lambda$ y $\lambda > 0$.

4i) $\frac{\phi(z)}{1-\Phi(z)} < \frac{z^2+1}{z}$.

$$5i) 1 < z < \lambda \Rightarrow \frac{z}{\sqrt{2}} < \lambda u(z/\lambda) \Rightarrow \phi(\lambda u(z/\lambda)) < \phi\left(\frac{z}{\sqrt{2}}\right).$$

$$6i) 1 < \lambda < z \Rightarrow \frac{\lambda}{\sqrt{2}} < \lambda u(z/\lambda) \Rightarrow \phi(\lambda u(z/\lambda)) < \phi\left(\frac{\lambda}{\sqrt{2}}\right).$$

Si descomponemos $a_{22}(\lambda)$ tenemos que

$$\begin{aligned} a_{22}(\lambda) &= A(\lambda) + B(\lambda) \\ &= 2 \int_0^\infty u^2(z)t_\nu(z) \frac{\phi^2(\lambda u(z))}{\Phi(\lambda u(z))} dz + 2 \int_0^\infty u^2(z)t_\nu(z) \frac{\phi^2(\lambda u(z))}{1 - \Phi(\lambda u(z))} dz. \end{aligned}$$

Resolviendo $A(\lambda)$:

$\forall z > 0$ y $\lambda > 1$, se tiene que $1 < 1/\Phi(\lambda u(z)) < 2$, entonces $A^*(\lambda) < A(\lambda) < 2A^*(\lambda)$ donde

$$A^*(\lambda) = 2 \int_0^\infty u^2(z)t_\nu(z)\phi^2(\lambda u(z))dz = I_3 + I_4,$$

con

$$\begin{aligned} I_3 &= 2 \int_0^{1/\lambda} u^2(z)t_\nu(z)\phi^2(\lambda u(z))dz, \\ I_4 &= 2 \int_{1/\lambda}^\infty u^2(z)t_\nu(z)\phi^2(\lambda u(z))dz. \end{aligned}$$

Se tiene la siguiente desigualdad

$$O(\lambda^{-3}) = I_1 < A^*(\lambda) < 2(I_3 + I_4) < O(\lambda^{-3}),$$

donde las cotas para I_1 , I_3 y I_4 se obtienen usando 1i), 2i) y haciendo el cambio de variable

$r = \lambda z$, es decir

$$I_1 = \frac{2}{\lambda^2} \int_0^\infty u^2(\lambda z)t_\nu(\lambda z)\phi^2(\lambda z)dz = \frac{1}{\lambda^3} I_2 = O(\lambda^{-3}),$$

$$I_2 = 2 \int_0^\infty u^2(r)t_\nu(r)\phi^2(r)dr < \infty,$$

$$I_3 < 2 \int_0^{1/\lambda} u^2(z)t_\nu(u(\lambda z))\phi^2(u(\lambda z))dz < 2 \int_0^{1/\lambda} z^2 t_\nu(u(\lambda z))\phi^2(u(\lambda z))dz = \frac{1}{\lambda^3} I_6 = O(\lambda^{-3}),$$

$$I_6 = 2 \int_0^1 r^2 t_\nu(u(r))\phi^2(u(r))dr < \infty,$$

$$I_4 = \frac{2}{\lambda} \int_1^\infty u^2(r/\lambda)t_\nu(r/\lambda)\phi^2(\lambda u(r/\lambda))dr = I_7^{2,2} + I_8^{2,2},$$

donde $I_7^{2,2}$ y $I_8^{2,2}$ son obtenidos de la expresión general que sigue, utilizando 5i) y 6i), entonces

$$\begin{aligned}
I_7^{k,l} &= \frac{2}{\lambda} \int_1^\lambda u^k(r/\lambda) t_\nu(r/\lambda) \phi^l(\lambda u(r/\lambda)) dr < \frac{2}{\lambda^{k+1}} \int_1^\lambda r^k t_\nu(r/\lambda) \phi^l(r/\sqrt{2}) dr = I_{12} \\
I_{12} &= \frac{1}{\lambda^{k+1}} I_9 = O(\lambda^{-(k+1)}), \\
I_9 &= 2t_\nu(0) \int_0^\infty r^k \phi^l(r/\sqrt{2}) dr < \infty, \\
I_8^{k,l} &= \frac{2}{\lambda} \int_\lambda^\infty u^k(r/\lambda) t_\nu(r/\lambda) \phi^l(\lambda u(r/\lambda)) dr < \frac{2}{\lambda} \phi^l(\lambda/\sqrt{2}) \int_\lambda^\infty u^k(r/\lambda) t_\nu(r/\lambda) dr = I_{13} \\
I_{13} &= \phi^l(\lambda/\sqrt{2}) I_{10} = O(e^{-\lambda^2}), \\
I_{10} &= 2 \int_1^\infty u(v)^k t_\nu(v) dv < \infty.
\end{aligned}$$

Luego

$$\begin{aligned}
I_4 &= I_7^{2,2} + I_8^{2,2} \\
&< O(\lambda^{-3}) + O(e^{-\lambda^2}) \\
&< O(\max\{\lambda^{-3}, e^{-\lambda^2}\}) \\
&< O(\lambda^{-3})
\end{aligned}$$

Entonces $A^*(\lambda) = \Omega(\lambda^{-3})$, esto implica que $A(\lambda) = \Omega(\lambda^{-3})$.

Resolviendo $B(\lambda)$:

Se tiene que $A^*(\lambda) = O(\lambda^{-3}) < B(\lambda)$, ya que $1 - \Phi(\lambda u(z)) < 1$. Ahora, se acota por arriba a $B(\lambda)$,

$$\begin{aligned}
B(\lambda) &= B_1(\lambda) + B_2(\lambda) \\
&= 2 \int_0^{1/\lambda} u^2(z) t_\nu(z) \frac{\phi^2(\lambda u(z))}{1 - \Phi(\lambda u(z))} dz + 2 \int_{1/\lambda}^\infty u^2(z) t_\nu(z) \frac{\phi^2(\lambda u(z))}{1 - \Phi(\lambda u(z))} dz.
\end{aligned}$$

Usando 3i), tenemos que

$$B_1(\lambda) < 2c \int_0^{1/\lambda} u^2(z) t_\nu(z) \phi^2(\lambda u(z)) dz = cI_3 < O(\lambda^{-3})$$

Ahora, para $B_2(\lambda)$ usamos 4i) y realizamos el cambio de variable $r = \lambda z$, se tiene

$$\begin{aligned}
B_2(\lambda) &< 2 \int_{1/\lambda}^{\infty} u^2(z)t_\nu(z)\phi(\lambda u(z))\frac{(\lambda u(z))^2 + 1}{(\lambda u(z))}dz = I_{11}, \\
I_{11} &= 2\lambda \int_{1/\lambda}^{\infty} u^3(z)t_\nu(z)\phi(\lambda u(z))dz + \frac{2}{\lambda} \int_{1/\lambda}^{\infty} u(z)t_\nu(z)\phi(\lambda u(z))dz \\
&= 2 \int_1^{\infty} u^3(r/\lambda)t_\nu(r/\lambda)\phi(\lambda u(r/\lambda))dr + \frac{2}{\lambda^2} \int_1^{\infty} u(r/\lambda)t_\nu(r/\lambda)\phi(\lambda u(r/\lambda))dr \\
&= \lambda(I_7^{3,1} + I_8^{3,1}) + \frac{1}{\lambda^2}\{\lambda(I_7^{1,1} + I_8^{1,1})\}.
\end{aligned}$$

Entonces

$$\begin{aligned}
I_{11} &= \lambda(I_7^{3,1} + I_8^{3,1}) + \frac{1}{\lambda^2}\{\lambda(I_7^{1,1} + I_8^{1,1})\} \\
&< \lambda(O(\lambda^{-4}) + O(e^{-\lambda^2})) + \frac{1}{\lambda}(O(\lambda^{-2}) + O(e^{-\lambda^2})). \\
&< \text{máx}\{\lambda^{-3}, \lambda e^{-\lambda^2}, \lambda^{-1}e^{-\lambda^2}\} \\
&< O(\lambda^{-3}).
\end{aligned}$$

Se concluye que $B(\lambda) = \Omega(\lambda^{-3})$ y por lo tanto $a_{22}(\lambda) = \Omega(\lambda^{-3})$. En forma similar al desarrollo anterior, se demuestra la simetría, monotonía y el orden de convergencia respecto de λ para $a_{42}(\lambda)$. En este caso, $a_{42}(\lambda) = \Omega(\lambda^{-5})$ y se concluye que $M(\lambda) = \Omega(\lambda^{-1})$. \square

La ecuación de verosimilitud modificada ($U_M(\lambda) = 0$) siempre tiene solución cuando la EMV de λ no existe, los argumentos son similares a los dados por Sartori [60].

5.4.2. Estudio de simulación: parámetro de forma

Se realizó simulaciones de una variable aleatoria Z correspondiente a la distribución MStN($\mu, \sigma, \lambda, \nu$), asumiendo $\mu = 0$, $\sigma = 1$ y ν conocido, para distintos tamaños de muestra; valores de λ y ν .

Tabla 5.4: Simulación de 5,000 iteraciones del modelo MStN(λ, ν), ν conocido.

n	λ	ν	$sesgo(\widehat{\lambda})^a$	$sesgo(\widehat{\lambda}_M)$	$W_M(\lambda)IC$	$\%(\widehat{\lambda} < \infty)$
20	5	3	1.867	-1.583	0.94	71.64 (71.04)
50	5	3	1.754	-0.298	0.95	95.06 (95.49)
100	5	3	0.788	-0.030	0.95	99.82 (99.80)
20	10	3	1.991	-6.034	0.90	44.78 (45.09)
50	10	3	4.299	-2.866	0.94	76.80 (77.66)
100	10	3	3.856	-0.694	0.94	94.84 (95.01)
20	5	5	2.148	-1.513	0.94	72.62 (72.28)
50	5	5	1.802	-0.293	0.95	96.48 (95.95)
100	5	5	0.815	-0.004	0.95	99.80 (99.84)
20	10	5	2.197	-5.949	0.90	46.46 (46.15)
50	10	5	4.116	-2.751	0.94	79.38 (78.72)
100	10	5	3.862	-0.626	0.95	95.38 (95.47)
20	5	10	2.177	-1.479	0.94	72.82 (73.24)
50	5	10	2.103	-0.236	0.96	96.64 (96.29)
100	5	10	0.776	0.018	0.95	99.90 (99.86)
20	10	10	2.274	-5.888	0.91	47.42 (47.00)
50	10	10	4.169	-2.626	0.94	79.18 (79.55)
100	10	10	4.338	-0.600	0.95	95.88 (95.82)

Sesgo de $\widehat{\lambda}$ y $\widehat{\lambda}_M$, cobertura empírica con intervalo de confianza de 0.95 basado en $W_M(\lambda)$ y porcentaje empírico de casos que $\widehat{\lambda}$ es finito. ^a Calculado cuando $\widehat{\lambda} < \infty$.

En la Tabla 5.4, se puede confirmar que existe una sobrestimación del parámetro λ y casos en que la estimación es ∞ , esto depende del tamaño de muestra, los grados de libertad (ν)

y del verdadero valor del parámetro (λ). Al aplicar el método de Firth [28] al parámetro de forma λ , se obtiene un nuevo estimador $\widehat{\lambda}_M$ el cual siempre existe y es finito, lo que es consistente con la Proposición 5.1 que lo garantiza. La reducción del sesgo de $\widehat{\lambda}_M$ es bastante buena, considerando el hecho que este método se aplica cuando $\widehat{\lambda}$ es finito y/o infinito. Además, existe una subestimación del parámetro λ cuando los valores de éste es grande y el tamaño de muestra es pequeño. La cobertura empírica es cercana al valor nominal (95 %) y levemente inferior cuando la muestra es pequeña, el cual se ve influenciado por el porcentaje de $\widehat{\lambda}$ divergentes.

5.4.3. Caso: Localización, escala y forma

Consideramos el caso μ, σ desconocidos, ν conocido y una muestra de n observaciones de z_1, \dots, z_n iid de $\text{MStN}(\mu, \sigma, \lambda, \nu)$, así la log-verosimilitud del modelo está dada por

$$l(\mu, \sigma, \lambda) = -n \log(\sigma) - \frac{\nu + 1}{2} \sum_{i=1}^n \log \left(1 + \frac{x_i^2}{\nu} \right) + \sum_{i=1}^n \log(\Phi(\lambda u(x_i))),$$

donde $x_i = \frac{z_i - \mu}{\sigma}$, $z_i \in \mathbb{R}$ con $i = 1, \dots, n$, $\mu \in \mathbb{R}$, $\sigma > 0$, $\lambda \in \mathbb{R}$ y $\nu > 0$. Similar al caso del parámetro de forma, la EMV puede ser infinita con probabilidad no despreciable, pero ahora cuando el $\min(z_i) > \widehat{\mu}$, $i = 1, \dots, n$. Sea $l_P(\lambda) = l(\widehat{\mu}(\lambda), \widehat{\sigma}(\lambda), \lambda)$ el perfil de verosimilitud para λ , donde $\widehat{\mu}(\lambda)$ y $\widehat{\sigma}(\lambda)$ son los EMV para un valor de λ fijo. Por lo anterior, el comportamiento de la log-verosimilitud del modelo MStN , respecto de λ , es similar al modelo MSN , entonces las expresiones de la función modificada (M), perfil cuasi-verosimilitud modificada l_M^P y el estadístico razón de verosimilitud modificado (W_M^P), son similares a los presentados en (2.14), (2.17) y (2.18), respectivamente, pero utilizando el modelo MStN .

5.4.4. Estudio de simulación: Localización, escala y forma

Se realizó simulaciones de una muestra aleatoria con distribución $MStN(\mu, \sigma, \lambda, \nu)$, asumiendo $\mu = 0$, $\sigma = 1$ desconocidos y ν conocido, para distintos tamaños de muestra, valores de λ y ν .

La Tabla 5.5, presenta en forma similar al caso de parámetro escalar, la existencia de una sobrestimación del parámetro λ y también hay casos en que esta estimación es ∞ , pero en mayor porcentaje. Sin embargo, la estimación para los parámetros de localización y escala, μ y σ respectivamente, son bastante favorables debido a que siempre existen, son finitos y el sesgo es muy cercano cero. Por lo anterior, sólo aplicamos el método de Firth [28], al parámetro λ . El nuevo estimador, $\hat{\lambda}_M$, siempre existe y es finito, ya que se logra el objetivo de reducir el sesgo, tanto en el caso cuando $\hat{\lambda}$ es finito y/o infinito. Al igual que el caso del parámetro de forma, también existe una subestimación del parámetro λ y esto ocurre, para valores grandes de éste y un tamaño de muestra pequeño. Sin embargo, en menor magnitud, la cobertura empírica es cercana al valor nominal (95%) y relativamente menor cuando λ es grande y n es pequeño. Esto, obviamente, está influenciado por el mayor porcentaje de estimaciones de λ infinito. Es importante notar que, para un valor fijo de ν y el mismo tamaño de muestra, el sesgo de $\hat{\lambda}_M$ aumenta significativamente, pero disminuye a medida que aumenta el tamaño de muestra. Sin embargo, para un valor fijo de λ y el mismo tamaño de muestra, el sesgo de $\hat{\lambda}_M$ disminuye lentamente conforme aumenta ν .

Tabla 5.5: Simulación de 5,000 iteraciones del modelo $MStN(0, 1, \lambda, \nu)$, ν conocido.

n	λ	ν	$sesgo(\hat{\mu})$	$sesgo(\hat{\sigma})$	$sesgo(\hat{\lambda})^a$	$sesgo(\hat{\lambda}_M)$	$W_M^P(\lambda)IC$	$\%(\hat{\lambda} < \infty)$
50	5	3	0.004	0.005	1.889	-0.898	0.93	84.24
100	5	3	-0.002	0.005	1.617	-0.306	0.94	97.42
200	5	3	-0.001	0.004	0.712	-0.106	0.95	99.84
50	10	3	0.017	-0.008	1.097	-3.971	0.87	61.86
100	10	3	0.004	-0.003	3.301	-1.699	0.91	85.40
200	10	3	0.001	0.000	3.188	-0.534	0.94	97.96
50	5	5	0.004	0.002	1.828	-0.855	0.94	86.70
100	5	5	-0.002	0.005	1.755	-0.255	0.94	97.58
200	5	5	-0.001	0.002	0.628	-0.098	0.95	99.78
50	10	5	0.016	-0.013	1.185	-3.842	0.87	63.74
100	10	5	0.005	-0.002	3.736	-1.508	0.92	86.72
200	10	5	0.000	0.000	3.053	-0.401	0.94	97.78
50	5	10	0.006	-0.003	1.832	-0.813	0.92	86.84
100	5	10	0.000	0.002	1.523	-0.245	0.95	98.00
200	5	10	0.000	0.000	0.578	-0.108	0.94	99.94
50	10	10	0.014	-0.013	1.530	-3.689	0.88	64.18
100	10	10	0.004	-0.002	3.622	-1.311	0.92	86.40
200	10	10	0.002	-0.002	2.700	-0.450	0.93	98.06

Sesgo de $\hat{\mu}$, $\hat{\sigma}$, $\hat{\lambda}$ y $\hat{\lambda}_M$, cobertura empírica con intervalo de confianza de 0.95 basado en $W_M^P(\lambda)$ y porcentaje empírico de casos que $\hat{\lambda}$ es finito. ^a Calculado cuando $\hat{\lambda} < \infty$.

5.5. Aplicación

Consideraremos un conjunto de datos correspondiente a la concentración de níquel en 86 muestras de suelo analizado en el Departamento de Minería en la Universidad de Atacama - Chile. La Tabla 5.6, muestra las estadísticas básicas que representan a los datos involucrados e indican el valor del coeficiente de asimetría (γ_1) y curtosis (γ_2) muestral, esto es:

Tabla 5.6: Estadística descriptiva de los datos de Níquel.

<i>Datos</i>	<i>n</i>	<i>media</i>	<i>d.e.</i>	γ_1	γ_2
<i>Nickel</i>	86	21.337	16.639	2.355	11.191

Se realiza una exploración de los EMV para el modelo MStN, considerando distintos valores de ν fijo, con el fin de estudiar el comportamiento de la log-verosimilitud. La Tabla 5.7, refleja este comportamiento y muestra que el valor máximo de la log-verosimilitud ocurre en $\nu = 3$.

Tabla 5.7: EMV del modelo MStN que ajusta los datos, ν fijo.

$\hat{\mu}$	$\hat{\sigma}$	$\hat{\lambda}$	ν	log-verosimilitud
8,202	11,852	2,429	2,2	-338,545
7,841	12,420	2,594	2,4	-338,397
7,549	12,921	2,739	2,6	-338,311
7,302	13,361	2,870	2,8	-338,269
7,086	13,771	2,994	3,0	-338,261
6,901	14,134	3,105	3,2	-338,277
6,736	14,474	3,211	3,4	-338,312
6,583	14,785	3,313	3,6	-338,360
6,452	15,077	3,406	3,8	-338,418
6,335	15,347	3,491	4,0	-338,484

El comportamiento del perfil de log-verosimilitud, respecto de ν , se muestra en la Figura 5.6.

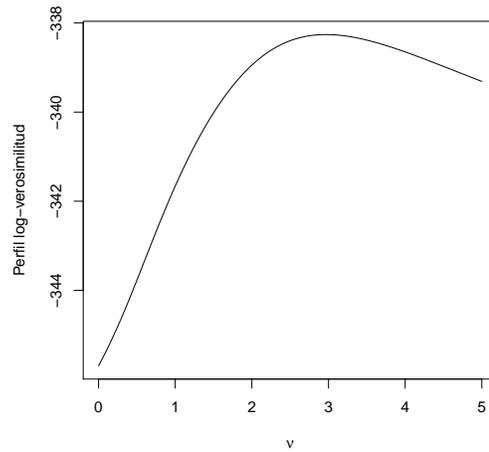


Figura 5.6: Perfil de log-verosimilitud para ν .

La Tabla 5.8, muestra el ajuste de los datos para diferentes modelos, esto es:

Tabla 5.8: EMV del modelo SN, MSN y MStN.

EMVs.	SN	MSN	MStN
$\hat{\mu}$	2.625 (2.068)	2.571 (1.959)	7.083(1.858)
$\hat{\sigma}$	24.968(2.461)	25.027 (2.405)	13.767(2.127)
$\hat{\lambda}$	10.261(9.615)	10.619 (9.427)	2.994(0.972)
ν	-	-	3
log-verosimilitud	-344.762	-344.769	-338.260
AIC	693.524	693.538	682.520

La Tabla 5.8, muestra los EMV de los parámetros de las distribuciones SN, MSN y MStN y los valores entre paréntesis corresponden a los errores estándar. Estos valores fueron obtenidos de la matriz información de Fisher de dichos modelos. También, se muestran los valores de

log-verosimilitud y de AIC, éstos indican claramente el buen ajuste del modelo propuesto comparado con las otras distribuciones.

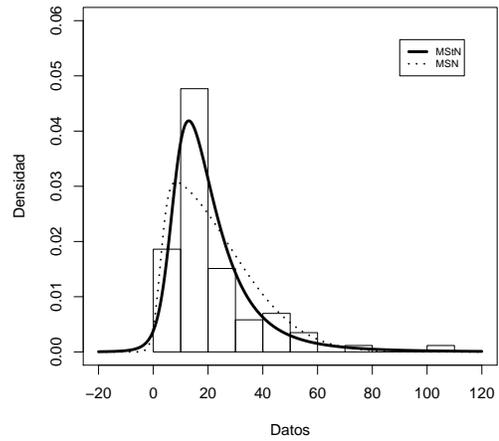


Figura 5.7: Ajuste de datos Níquel.

La Figura 5.7, presenta el histograma de los datos con las correspondientes distribuciones ajustadas, $MSN(\hat{\mu}, \hat{\sigma}, \hat{\lambda})$ (línea punteada) y $MStN(\hat{\mu}, \hat{\sigma}, \hat{\lambda}, 3)$ (línea continua).

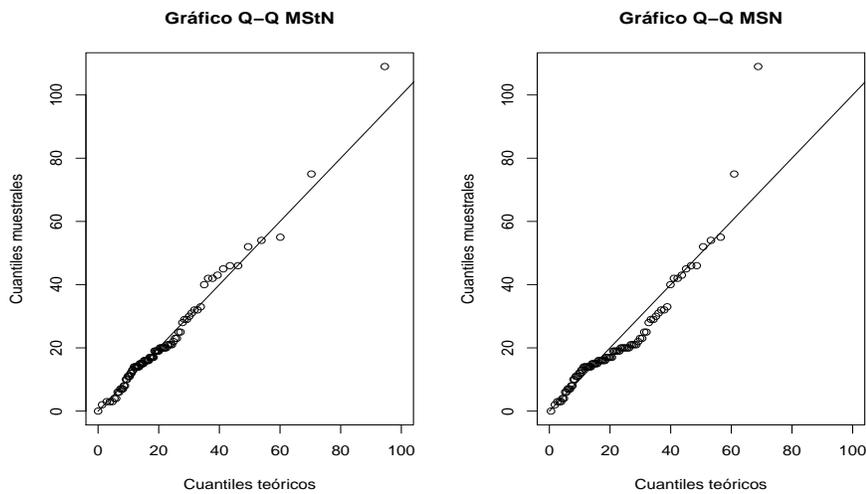


Figura 5.8: QQ plot para MStN y MSN.

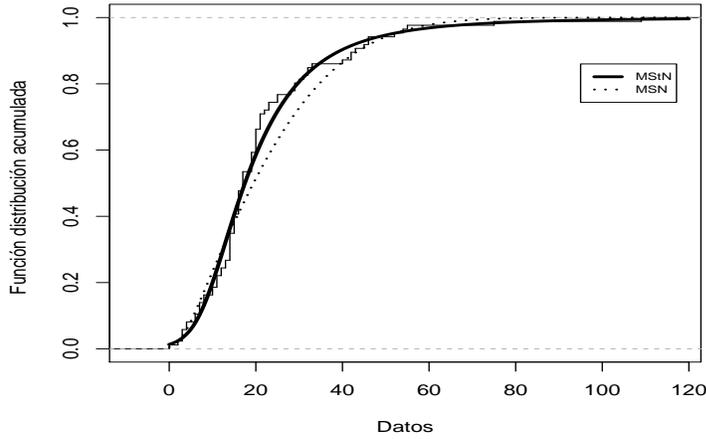


Figura 5.9: CDF empírica y teórica para el modelo MStN y MSN.

Los QQ-plot y la Función de distribución acumulada empírica y teórica de los modelos MSN y MStN que se muestra en las Figuras 5.8 y 5.9, son obtenidos de las estimaciones de los parámetros de cada distribución, confirman el buen ajuste del modelo de interés.

Tabla 5.9: EMV de μ , σ y λ .

$\hat{\mu}$	$\hat{\sigma}$	$\hat{\lambda}$	$\hat{\lambda}_M$	$l(\hat{\mu}, \hat{\sigma}, \hat{\lambda})$	$l(\hat{\mu}, \hat{\sigma}, \hat{\lambda}_M)$
7.083(1.402)	13.767(1.838)	2.994(0.789)	-	-338.260	-
7.083(1.545)	13.767(1.857)	-	2.528(0.638)	-	-338.657

La Tabla 5.9, muestra los EMV $\hat{\mu}$, $\hat{\sigma}$, $\hat{\lambda}$ y el EMV modificado $\hat{\lambda}_M$. Los errores estándar se presentan entre paréntesis y fueron obtenidos de la matriz información de Fisher, dada en 5.3.1, puesto que la distribución asintótica $\hat{\theta}$ y $\hat{\theta}_M$ es $N_3(\theta, I^{-1}(\theta)/n)$, donde $\theta = (\mu, \sigma, \lambda)$. El EMV modificado $\hat{\lambda}_M$ es de menor valor que el EMV $\hat{\lambda}$ y por construcción, tiene un menor sesgo asintótico.

Tabla 5.10: Intervalos de confianza para λ .

	95 %	98 %	99 %
IC	(1.696,4.292)	(1.373,4.615)	(1.158,4.830)
IC^*	(1.499,3.600)	(1.238,3.861)	(1.064,4.035)

La Tabla 5.10, presenta los intervalos de confianza de λ para tres valores de niveles diferentes. Los intervalos de confianza, obtenidos utilizando el EMV modificado (IC^*), son más precisos comparados con el intervalo de confianza, basado en el EMV no modificado (IC), puesto que tienen menor longitud, respectivamente.

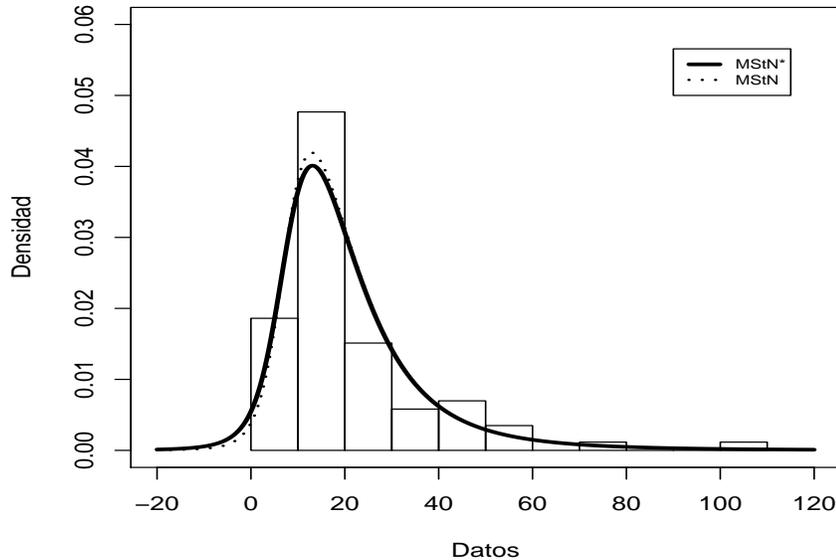


Figura 5.10: Datos ajustados.

La figura 5.10 muestra el reajuste de los datos con $MStN(\hat{\mu}, \hat{\sigma}, \hat{\lambda}, 3)$ y el reajuste $MStN^*(\hat{\mu}, \hat{\sigma}, \hat{\lambda}_M, 3)$ utilizando el EMV de sesgo reducido $\hat{\lambda}_M$.

5.6. Apéndice

Sea $X \sim MStN(\mu, \sigma, \lambda)$. Consideremos la notación:

$$\begin{aligned}\tau &= (\mu, \sigma), \quad Z = \frac{X - \mu}{\sigma}, \quad u(x) = \frac{x}{(1 + x^2)^{1/2}} \\ r(Z) &= \frac{Z^k}{1 + \frac{Z^2}{\nu}}, \quad w(\nu) = \frac{\nu + 1}{\nu^k(1 + \frac{Z^2}{\nu})} \\ \rho_{nm} &= E \left(\frac{Z^n}{(1 + Z^2)^{m/2}} \zeta(\lambda u(Z)) \right), \quad \rho_{nm} = 0 \quad \text{si } n = \text{impar} \\ \eta_{nm} &= E \left(\frac{Z^n}{(1 + Z^2)^m} \zeta^2(\lambda u(Z)) \right) \\ \delta_n &= E \left(\frac{Z}{(1 + \frac{Z^2}{\nu})^n} \right) \\ \zeta(x) &= \frac{\phi(x)}{\Phi(x)}\end{aligned}$$

y las siguientes derivadas

$$\begin{aligned}\frac{\partial Z}{\partial \mu} &= -\frac{1}{\sigma}, \quad \frac{\partial Z}{\partial \sigma} = -\frac{Z}{\sigma}, \quad \frac{\partial u(Z)}{\partial \tau} = \frac{1}{(1 + Z^2)^{3/2}} \frac{\partial Z}{\partial \tau}, \quad \frac{\partial r(Z)}{\partial \tau} = \frac{kz^{k-1}}{(1 + \frac{Z^2}{\nu})^2} \frac{\partial Z}{\partial \tau} \\ \frac{\partial \zeta}{\partial \tau} &= \left(-\frac{\lambda^2 Z}{(1 + Z^2)^2} \zeta - \frac{\lambda}{(1 + Z^2)^{3/2}} \zeta^2 \right) \frac{\partial Z}{\partial \tau} \\ \frac{\partial}{\partial \tau} \left(\frac{Z}{(1 + Z^2)^{3/2}} \right) &= \frac{1 - 2Z^2}{(1 + Z^2)^{5/2}} \frac{\partial Z}{\partial \tau} \\ \frac{\partial \zeta}{\partial \lambda} &= -\frac{\lambda Z^2}{(1 + Z^2)} \zeta - \frac{Z}{(1 + Z^2)^{1/2}} \zeta^2\end{aligned}$$

tenemos que las funciones score de μ , σ y λ del modelo $MStN(\lambda, \nu)$ pueden representarse

como:

$$\begin{aligned}S_\mu &= \frac{1}{\sigma} \left(-\frac{\lambda}{(1 + Z^2)^{3/2}} \zeta(\lambda u(Z)) + \frac{\nu + 1}{\nu} \frac{Z}{(1 + \frac{Z^2}{\nu})} \right) \\ S_\sigma &= \frac{1}{\sigma} \left(-1 - \frac{\lambda Z}{(1 + Z^2)^{3/2}} \zeta(\lambda u(Z)) + \frac{\nu + 1}{\nu} \frac{Z^2}{(1 + \frac{Z^2}{\nu})} \right) \\ S_\lambda &= u(Z) \zeta(\lambda u(Z))\end{aligned}$$

y en donde las entradas de la matriz información de Fisher son:

$$\begin{aligned}
I_{\mu\mu} &= -E(S_{\mu\mu}) \\
&= -\frac{1}{\sigma^2}E\left(-\frac{3\lambda Z}{(1+Z^2)^{5/2}}\zeta - \frac{\lambda^3 Z}{(1+Z^2)^{7/2}}\zeta - \frac{\lambda^2}{(1+Z^2)^3}\zeta^2 - \frac{\nu+1}{\nu}\frac{Z}{(1+\frac{Z^2}{\nu})^2}\right) \\
&= \frac{1}{\sigma^2}\left(\lambda^2\eta_{03} + \frac{\nu+2}{\nu+3}\right)
\end{aligned}$$

$$\begin{aligned}
I_{\mu\sigma} &= -E(S_{\sigma\mu}) \\
&= -\frac{1}{\sigma^2}E\left(\frac{\lambda(1-2Z^2)}{(1+Z^2)^{5/2}}\zeta - \frac{\lambda^3 Z^2}{(1+Z^2)^{7/2}}\zeta - \frac{\lambda^2 Z}{(1+Z^2)^3}\zeta^2 - \frac{2(\nu+1)}{\nu}\frac{Z}{(1+\frac{Z^2}{\nu})^2}\right) \\
&= -\frac{1}{\sigma^2}\left(\lambda\rho_{05} - 2\lambda\rho_{25} - \lambda^3\rho_{27} - \lambda^2\eta_{13} - \frac{2(\nu+1)}{\nu}\delta_2\right)
\end{aligned}$$

$$\begin{aligned}
I_{\mu\lambda} &= -E(S_{\mu\lambda}) \\
&= -\frac{1}{\sigma}E\left(-\frac{1}{(1+Z^2)^{3/2}}\left\{\zeta - \frac{\lambda^2 Z^2}{(1+Z^2)}\zeta - \frac{\lambda Z}{(1+Z^2)^{1/2}}\zeta^2\right\}\right) \\
&= \frac{1}{\sigma}(\rho_{03} - \lambda^2\rho_{25} - \lambda\eta_{12})
\end{aligned}$$

$$\begin{aligned}
&= -\frac{1}{\sigma} E \left(-\frac{\nu+1}{\nu} \frac{Z}{\left(1 + \frac{Z^2}{\nu}\right)^2} + \frac{1}{\nu} \frac{Z}{\left(1 + \frac{Z^2}{\nu}\right)} \right) \\
&= \frac{1}{\sigma} \left(\frac{(\nu+1)}{\nu^2} \delta_2 - \frac{1}{\nu} \delta_1 \right)
\end{aligned}$$

$$\begin{aligned}
I_{\sigma\sigma} &= -E(S_{\sigma\sigma}) \\
&= -\frac{1}{\sigma^2} E \left(1 + \frac{\lambda Z}{(1+Z^2)^{3/2}} \zeta + \frac{\lambda(Z-2Z^3)}{(1+Z^2)^{5/2}} \zeta - \frac{\lambda^3 Z^3}{(1+Z^2)^{7/2}} \zeta - \frac{\lambda^2 Z^2}{(1+Z^2)^3} \zeta^2 \right. \\
&\quad \left. - \frac{\nu+1}{\nu} \left[\frac{2Z^2}{\left(1 + \frac{Z^2}{\nu}\right)^2} + \frac{Z^2}{1 + \frac{Z^2}{\nu}} \right] \right) \\
&= \frac{1}{\sigma^2} \left(\lambda^2 \eta_{23} + \frac{2\nu}{\nu+3} \right)
\end{aligned}$$

$$\begin{aligned}
I_{\sigma\lambda} &= -E(S_{\lambda\sigma}) \\
&= -\frac{1}{\sigma} E \left(-\frac{Z}{(1+Z^2)^{3/2}} \zeta + \frac{\lambda^2 Z^3}{(1+Z^2)^{5/2}} \zeta + \frac{\lambda Z^2}{(1+Z^2)^2} \zeta^2 \right) \\
&= -\frac{\lambda}{\sigma} \eta_{22}
\end{aligned}$$

$$\begin{aligned}
I_{\lambda\lambda} &= -E(S_{\lambda\lambda}) \\
&= -E \left(-\frac{\lambda Z^3}{(1+Z^2)^{3/2}} \zeta - \frac{Z^2}{1+Z^2} \zeta^2 \right) \\
&= \eta_{21}
\end{aligned}$$

Capítulo 6

Conclusiones y futuras extensiones

- El modelo asimétrico estudiado, MSN, es un buen competidor para el modelo SN, puesto que ambos modelos controlan la asimetría con un parámetro escalar. Sin embargo, el modelo MSN, en presencia de parámetros de localización y escala, la matriz información esperada es no singular bajo la hipótesis de normalidad. Se realizó el estudio del comportamiento sesgo de los EMV del modelo MSN y se aplicó el método de reducción de sesgo, los resultados son similares a los obtenidos por Sartori [60] para el modelo SN. Lo anterior, se debe a que las verosimilitudes de ambos modelos son similares entre sí.
- A partir de la transformación de una variable que tiene distribución MSN, se obtuvieron las distribuciones LMSN y BSMSN. Estas distribuciones, resultaron ser más flexibles comparadas con otras de su misma clase. Además, estos modelos heredan la no singularidad de la matriz información esperada del modelo MSN, lo que nos permite realizar inferencia asintótica cuando el parámetro de forma es cero.
- La distribución MStN tiene una gran flexibilidad en su estructura paramétrica, por lo

tanto puede modelar datos que poseen colas más pesadas. Este modelo, está ligado al modelo MSN a través de su representación estocástica, de esta manera se pudo rescatar la aplicación del método de reducción de sesgo implementada en el modelo MSN.

- Los modelos presentados en esta tesis tienen problemas de sobrestimación en la estimación máximo verosímil del parámetro de forma λ , pudiendo ser finito ó infinito, esto se resuelve encontrando un nuevo EMV modificado ($\widehat{\lambda}_M$), a través del método de reducción de sesgo de Firth [28].
- Se demuestra que el EMV modificado del parámetro de forma, de los modelos MSN, LMSN, BSMSN y MStN obtenidos a través del método de Firth, siempre existe y es finito. Además, por construcción el EMV modificado tiene menor sesgo asintótico que el EMV no modificado. Los intervalos de confianza, en los estudios de simulación, fueron obtenidos a partir de la distribución asintótica χ^2 de la razón de cuasi-verosimilitud asociada a la función score modificada, puesto que captura mejor la asimetría de la log-verosimilitud y las coberturas empíricas de probabilidades son satisfactoriamente buenas. Para el caso mas general, cuando agregamos localización y escala al modelo, sólo se aplica el método de Firth al parámetro de forma, debido a que el sesgo de los restantes parámetros están muy próximo a cero. En general las coberturas empíricas son relativamente buenas, muy cercana al valor nominal (95 %), excepto para valores grandes de λ y tamaños de muestra pequeños, donde se produce un porcentaje relativamente alto de estimaciones infinitas.
- Se aplica el método de reducción de sesgo a los EMV obtenidos de cada modelo estudiado, pero ahora utilizando datos reales. El nuevo estimador modificado permite

obtener un intervalo de confianza (IC^*) más preciso, comparado con el intervalo de confianza estándar (IC), y por tanto las decisiones e inferencias son más confiables. Esto, se puede observar en el ejemplo aplicado al modelo MSN, el IC^* acepta la hipótesis de simetría para distintos niveles de confianza, mientras que el IC acepta la hipótesis y posteriormente la rechaza.

Algunos trabajos a futuros a realizar, son:

- Aplicar las distribuciones propuestas, en esta tesis, a modelos de regresión.
- Extender los modelos estudiados al caso multivariado.
- Estudiar el modelo MStN para el caso ν desconocido y aplicar el método de reducción de sesgo a este parámetro. Lo anterior, derivará en realizar el cálculo de 10 valores esperados, en vez de dos, como en el caso de ν conocido.

Bibliografía

- [1] Arellano-Valle, R.B., del Pino, G. and San Martín, 2002. *Definition and Probabilistic properties of skew-distributions*. Statistics and Probability Letters, **58(2)**, 111-121.
- [2] Arellano-Valle, R.B., Gómez, H.W., Quintana, F.A., 2004. *A New Class of Skew-Normal Distributions*. Communications in Statistics. Theory and Methods, **33(7)**, 1465-1480.
- [3] Arellano-Valle and Genton, M.G., 2005. *On fundamental skew distributions*. Journal of Multivariate Analysis, **96(1)**, 93-116.
- [4] Arellano-Valle, R.B., Azzalini, A., 2008. *The centered parametrization for the multivariate skew-normal distribution*. Journal of Multivariate Analysis, **99**, 1362-1382.
- [5] Arellano-Valle, R.B., Castro, L.M., Genton, M.G., Gómez, H.W., 2008. *Bayesian inference for shape mixtures of skewed distributions with application to regression analysis*. Bayesian Analysis, **3**, 513-540.
- [6] Arellano-Valle, R. B., Genton, M. G., Loschi, R. H., 2009. *Shape mixtures of multivariate skew-normal distributions*. Journal of Multivariate Analysis, **100(1)**, 91-1001.

- [7] Arellano-Valle, R.B., Gómez, H.W., Salinas, H.S., 2013. *A note on the Fisher information matrix for the skew-generalized-normal model*. Statistics and Operations Research Transactions. **37(1)**, 19-28.
- [8] Arnold, B.C., Beaver, R.J., Groeneveld, R.A., Meeker, W.Q., 1993. *The non-truncated marginal of a truncated bivariate normal distribution*. Psychometrika, **58**, 471-478.
- [9] Arnold, B.C. and Beaver, R.J., 2000. *The skew Cauchy distribution*. Statistics and Probability Letters, **49**, 285-290.
- [10] Arnold, B.C., Castillo, E., Sarabia, J.M., 2007. *Distributions with Generalized Skewed Conditionals and Mixtures of such Distributions*. Communications in Statistics. Theory and Methods, **36**, 1493-1504.
- [11] Astorga, J.M., Gómez, H.W., 2004. *Una Extensión del Modelo Log-Skew-Normal*. Revista de la Facultad de Ingeniería de la Universidad de Atacama, **17**, 4-7.
- [12] Azzalini, A., 1985. *A class of distributions which includes the normal ones*. Scandinavian Journal of Statistics, **12**, 171-178.
- [13] Azzalini, A., 1986. *Further results on a class of distributions which includes the normal ones*. Statistica, **46**,199-208.
- [14] Azzalini, A. and Dalla Valle,A, 1996. *The multivariate skew normal distribution*. Biometrika, **83(4)**,715-726.

- [15] Azzalini, A. and Capitanio, A., 1999. *Statistical applications of the multivariate skew normal distribution*. Journal of the Royal Statistical society, Series B, **61(3)**, 579-602.
- [16] Azzalini, A., Cappello, D., and Kotz, S., 2003. *Log-skew-normal and log-skew-t distributions as models for family income data*. Journal of Income Distribution, **11**, 12-20.
- [17] Azzalini, A., Arellano-Valle, R.B., 2013. *Maximum penalized likelihood estimation for skew-normal and skew-t distributions*. Journal of Statistical Planning and Inference, **143(2)**, 419-433.
- [18] Balakrishnan, N., Leiva, V., Sanhueza, A., Cabrera, E., 2009. *Mixture inverse Gaussian distribution and its transformations, moments and applications*. Statistics, **43**, 91-104.
- [19] Birnbaum ZW, Saunders SC, 1969. *A new family of life distributions*. Journal of Applied Probability, **6**, 319-327.
- [20] Birnbaum ZW, Saunders SC, 1969. *Estimation for a family distributions with applications to fatigue*. Journal of Applied Probability, **6**, 328-347.
- [21] Bolfarine H, Gómez HW, Rivas L., 2011. *The log-bimodal-skew-normal model. A geochemical application*. Journal of Chemometrics, **25(6)**, 329-332.
- [22] Branco, M. and Dey, D., 2001. *A general class of multivariate elliptical distributions*. Journal of Multivariate Analysis, **79(1)**, 99-113.

- [23] Bull, S. B., Greenwood, C. M. T. and Hauck, W. W., 1997. *Jackknife bias reduction for polychotomous logistic regression (Corr: 97V16 p2928)*. *Statistics in Medicine*, **16**, 545-560.
- [24] Bull, S. B., Mak, C. & Greenwood, C., 2002. *A modified score function estimator for multinomial logistic regression in small samples*. *Computational Statistics and Data Analysis*, **39**, 57-74.
- [25] Bull, S. B., Lewinger, J. B. & Lee, S. S. F., 2007. *Confidence intervals for multinomial logistic regression in sparse data*. *Statistics in Medicine*, **26**, 903-918.
- [26] Chiogna, M., 2005. *A note on the asymptotic distribution of the maximum likelihood estimator for the scalar skew-normal distribution*. *Statist. Methods Appl.*, **14**, 331-341.
- [27] Cox, D.R., Snell, E.J., 1968. *A general definition of residuals*. *Journal of the Royal Statistical Society, Series B* **30(2)**, 248-275.
- [28] Firth, D., 1993. *Bias reduction of maximum likelihood estimates*. *Biometrika*, **80**, 27-38. (Amendment : vol. 82, 667).
- [29] Genton, M. G., He, L. and Liu, X., 2001. *Moments of skew-normal random vectors and their quadratic forms*. *Statistics and Probability Letters*, **51(4)**, 319-325.
- [30] Gómez, H.W., Castro, L.M., Salinas, H.S., Bolfarine, H., 2010. *Properties and Inference on the Skew-curved-symmetric Family of Distributions*. *Communications in Statistics, Theory and Methods*, **39**, 884-898.

- [31] Gómez, H.W., Venegas, O., Bolfarine, H., 2007. *Skew-symmetric distributions generated by the distribution function of the normal distribution*. *Environmetrics*, **18**, 395-407.
- [32] Gokhale, S., Khare, M., 2007. *Statistical behavior of carbon monoxide from vehicular exhausts in urban environments*. *Environmental Modelling and Software*, **22**, 526-535.
- [33] Gupta, A.K. and Chen T., 2001 . *Goodness-of-fit test for the skew-normal distribution*. *Commun. Statist-Simula.*, **30(4)**, 907-930.
- [34] Gupta, A.K., Gonzáles-Farías and Domínguez-Molina, J.A., 2004 . *A multivariate skew normal distribution*. *Journal of Multivariate Analysis*, **89**, 181-190.
- [35] Gupta, A.K., Chang, F.C., and Huang, W.J., 2002 . *Some skew-symmetric models*. *Random Operators Stochastics Equations*, **10**, 113-140.
- [36] Heinze, G. & Schemper, M., 2002. *A solution to the problem of separation in logistic regression*. *Statistics in Medicine*, **21**, 2409-2419.
- [37] Henze, N.,1986. *A probabilistic representation of the skew-normal distribution*. *Scandinavian Journal of Statistics*, **13**, 271-275.
- [38] Jeffreys, H., 1946. *An invariant form for the prior probability in estimation problems*. *Proceedings of the Royal Society of London*, **186** 453-461.
- [39] Kheradmandi, A., Mohammadzadeh, M. and Sanjari Farsipour N., 2010. *Generalizations of the Skew t-Normal Distribution and their Properties*, *Journal of Sciences*, **21(4)**, 343-351.

- [40] Kosmidis, I., Firth, D., 2009. *Bias reduction in exponential family nonlinear models*. Biometrika, **96**, 793-904.
- [41] Lagos Alvarez, B., Jiménez Gamero, M.D., 2012, *A note on bias reduction of maximum likelihood estimates for the scalar skew t distribution*. Journal of Statistical Planning and Inference, **142(2)**, 608-612.
- [42] Leiva V., Vilca-Labra F., Balakrishnan N., Sanhueza A., 2010. *A Skewed Sinh-Normal Distribution and its Properties and Application to Air Pollution*. Communications in Statistics, Theory and Methods, **39**, 426-443.
- [43] Leiva, V., Barros, M., Paula, G. A., Sanhueza, A., 2008b. *Generalized Birnbaum-Saunders distribution applied to air pollutant concentration*. Environmetrics, **19**, 235-249.
- [44] Leiva, V., Sanhueza, A., Angulo, J. M., 2009. *A length-biased version of the Birnbaum-Saunders distribution with application in water quality*. Stochastic Environmental Research and Risk Assessment, **23**, 299-307.
- [45] Lin GD, Stoyanov J. 2009, *The logarithmic skew-normal distributions are moment-indeterminate*. Journal of Applied Probability, **46**, 909-916.
- [46] Liseo, B., 1990, *The skew-normal class of densities: inferential aspects from a Bayesian viewpoint*. Statistica, **50**, 59-70.
- [47] Liseo, B. and Loperfido, N., 2003. *A Bayesian interpretation of the multivariate skew-normal distribution*. Statistics Probability, **61(4)**, 395-401.

- [48] Liseo, B. and Loperfido, N., 2006. *A note on reference priors for the scalar skew-normal distribution*. J. Statist. Plann. Infer., **136**, 373-389.
- [49] Marchenko YV, Genton MG., 2010. *Multivariate log-skew-elliptical distributions with applications to precipitation data*. Environmetrics, **21**, 318-340.
- [50] Martínez, E.H., Varela, H., Gómez, H.W., Bolfarine, H., 2008. *A note on the likelihood and moments of the skew-normal distribution*. Statistics and Operations Research Transactions, **32(1)**, 57-66.
- [51] Martínez-Flórez G., Bolfarine H., Gómez H., 2014. *An alpha-power extension for the Birnbaum-Saunders distribution*. A Journal of Theoretical and Applied Statistics, **48(4)**, 896-912.
- [52] Mehrabi, Y. & Matthews, J. N. S., 1995. *Likelihood-based methods for bias reduction in limiting dilution assays*. Biometrics, **51**, 1543-1549.
- [53] Nadarajah, S., Kotz, S. (2003). *Skewed distributions generated by the normal kernel*. Statistics Probability Letters, **65**, 269-277.
- [54] Nadarajah, S., 2008. *A truncated inverted beta distribution with application to air pollution data*. Stochastic Environmental Research and Risk Assessment, **22**, 285-289.
- [55] Pettitt, A. N., Kelly, J. M. & Gao, J. T., 1998. *Bias correction for censored data with exponential lifetimes*. Statistica Sinica, **8**, 941-964.
- [56] Pewsey, A., 2000. *Problems of inference for Azzalini's skew-normal distribution*. Journal of Applied Statistics, **27(7)**, 859-870.

- [57] Quenouille, M. H., 1956. *Notes on bias in estimation*. *Biometrika*, **43**, 353-360.
- [58] Rotnitzky, A., Cox, D. R., Bottai, M., Robins, J., 2000. *Likelihood-based inference with singular information matrix*. *Bernoulli*, **6**, 243-284.
- [59] Sampford, M.R., 1953. *Some Inequalities on Mill's Ratio and Related Functions*. *The Annals of Mathematical Statistics*, **24(1)**, 130-132.
- [60] Sartori, N., 2006. *Bias prevention of maximum likelihood estimates for scalar skew normal and skew t distributions*. *Journal of Statistical Planning and Inference*, **136**, 4259-4275.
- [61] Sever, M., Lajovic, J., Rajer B., 2005. *Robustness of the Fisher's discriminant function to skew-curved normal distribution*. *Metodoloski zvezki*, **2(2)**, 231-242.
- [62] Sen PK, Singer J., 2000. *Lange Sample Methods in Statistics: An Introduction with Applications*. Chapman & Hall/CRC.
- [63] Tanner, M.A., 1996, Tools for statistical inference. In: *Methods for the Exploration of Posterior Distributions and Likelihood Functions*. 3rd ed. New York: Springer-Verlag.
- [64] Vilca, F., Sanhueza, A., Leiva, V., Christakos, G., 2010. *An extended Birnbaum-Saunders model and its application in the study of environmental quality in Santiago, Chile*. *Stochastic Environmental Research and Risk Assessment* (in press, DOI 10.1007/s00477-009-0363-6).
- [65] Wang, J., Boyer, J. and Genton, M.G., 2004. *A skew-symmetric representation of multivariate distributions*. *Statistica Sinica*, **14**, 1259-1270.

- [66] Wolfram Research, Inc., Mathematica, Version 10.0, Champaign, IL (2014).
- [67] Zorn, C., 2005. *A solution to separation in binary response models*. Political Analysis, **13**, 157-170.