



COLEGIO DE POSTGRADUADOS

INSTITUCIÓN DE ENSEÑANZA E INVESTIGACIÓN
EN CIENCIAS AGRÍCOLAS

CAMPUS MONTECILLO

POSTGRADO DE SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA
ESTADÍSTICA

**Algunos Aspectos de Inferencia Estadística en la
Distribución Normal Asimétrica**

Paulino Pérez Rodríguez

T E S I S

PRESENTADA COMO REQUISITO PARCIAL PARA
OBTENER EL GRADO DE:

DOCTOR EN CIENCIAS

MONTECILLO, TEXCOCO, EDO. DE MÉXICO
2008

La presente tesis titulada: **Algunos Aspectos de Inferencia Estadística en la Distribución Normal Asimétrica**, realizada por el alumno: **Paulino Pérez Rodríguez**, bajo la dirección del Consejo Particular indicado ha sido aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

DOCTOR EN CIENCIAS

SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA ESTADÍSTICA

CONSEJO PARTICULAR

CONSEJERO _____
Dr. José A. Villaseñor Alva

ASESOR _____
Dr. Gilberto Rendón Sánchez

ASESOR _____
Dr. Sergio Pérez Elizalde

ASESOR _____
Dr. Gabriel Rodríguez Yam

ASESOR _____
Dr. Barry C. Arnold

Algunos Aspectos de Inferencia Estadística en la Distribución Normal Asimétrica

Paulino Pérez Rodríguez

Colegio de Postgraduados, 2008

En este escrito se tratan algunos aspectos relacionados con inferencia estadística en la distribución normal asimétrica. El trabajo puede ser dividido en dos grandes apartados. El primer apartado aborda el problema de estimación en esta familia de distribuciones desde el punto de vista bayesiano, se muestra además cómo extender la metodología de estimación propuesta a modelos de regresión, con errores con distribución normal asimétrica. Con la finalidad de ilustrar la aplicabilidad de la metodología propuesta se presentan algunos ejemplos que se resuelven empleando el algoritmo de Metropolis, cuyo código se presenta en los apéndices. Se explora también una nueva técnica de Cadenas de Markov Monte Carlo denominada algoritmo “t-walk” que también permite resolver de manera eficiente el problema de estimación.

En el segundo apartado se estudian pruebas de hipótesis. El primer problema que se trata en este apartado es el de pruebas de bondad de ajuste, se hacen algunas propuestas, se estudian sus potencias, tamaños y se comparan con procedimientos de prueba existentes. Otro tema tratado en este apartado es el de una prueba para normalidad dentro de la familia normal asimétrica. También se trata el problema de decidir si el parámetro de localización de la distribución toma un valor particular, se deriva una estadística de prueba que no requiere la estimación de ningún parámetro y se obtiene la distribución asintótica de la estadística de prueba. Finalmente se estudia el tema clásico de comparación de dos poblaciones.

Palabras clave: Algoritmo de Metropolis, algoritmo “t-walk”, pruebas de hipótesis, bootstrap, algoritmo EM, estadística señal-ruido.

Some Aspects of Statistical Inference in The Skew Normal Distribution

Paulino Pérez Rodríguez

Colegio de Postgraduados, 2008

In this work, we discuss some aspects related to statistical inference in the skew normal distribution. The paper can be divided into two great sections. The first section deals with the problem of estimation in this family of distributions from the Bayesian point of view, we also show how to extend the proposed estimation methodology to regression models with errors with skew normal distribution. In order to illustrate the applicability of the proposed methodology, we show some examples which are solved using the Metropolis algorithm, whose code is presented in the annexes. Moreover, we explore a new Markov Monte Carlo Chain technique called t-walk algorithm, which also allows to solve the problem of estimation efficiently.

In the second section we study hypothesis tests. The first problem discussed in this section is that of goodness of fit tests. Some proposals are made, their power and size are studied and then compared with existing test procedures. Another topic discussed in this section is a normality test within the skew normal family. We also discuss the problem of deciding if the location parameter of the distribution takes on a particular value. We take a test statistic that requires no estimation from any parameter, and its asymptotic distribution is obtained. Finally, we study the classic topic of comparison of two populations.

Key words: Metropolis algorithm, “t-walk” algorithm, hypothesis tests, bootstrap, EM algorithm, signal noise-statistic.

AGRADECIMIENTOS

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo económico brindado durante la realización de mis estudios de postgrado.

Al Colegio de Postgraduados, por haberme brindado la oportunidad de seguir mi formación académica en sus aulas.

A los integrantes de mi Consejo Particular:

Dr. José Villaseñor Alva, por su excelente dirección, sin la cual me hubiera sido más difícil la culminación del presente trabajo.

Dr. Sergio Pérez Elizalde, por sus observaciones, consejos y ayuda desinteresada en la realización del trabajo.

Dr. Gabriel Rodríguez Yam por sus comentarios y revisión detallada del trabajo.

A los doctores Gilberto Rendón Sánchez y Barry C. Arnold por sus observaciones y por haber revisado el escrito.

A mis profesores, compañeros de clases y todos aquellos que de alguna u otra manera fueron copartícipes de esta tarea.

DEDICATORIA

A mis **padres:** Paulino Pérez Orona y Socorro Rodríguez Fernández, por todo el cariño y apoyo brindado y darme la oportunidad de seguir estudiando.

A mis **hermanos:** Jaime, Álvaro y Aracely.

A mi **esposa:** Vianey y mis **hijos:** Paola y Samuel.

Índice

1. Introducción	1
2. Objetivos	3
3. Revisión de Literatura	4
3.1. Distribución normal asimétrica	4
3.2. Pruebas de hipótesis	7
3.2.1. Prueba de razón de verosimilitudes generalizada	8
3.2.2. Pruebas de bondad de ajuste	9
3.3. Teorema central del límite multivariado	10
3.4. Los algoritmos EM y EM Generalizado	10
3.5. El algoritmo de Metropolis	12
3.6. El algoritmo “t-walk”	12
4. Métodos alternativos de estimación	13
4.1. Estimación desde el punto de vista clásico	13
4.1.1. Estimación con el método de momentos para la parametrización centrada	14
4.2. Estimación Bayesiana	16

4.2.1.	Ejemplo 1: Datos de Azzalini	20
4.2.2.	Ejemplo 2: Datos de Sartori	26
4.2.3.	Ejemplo 3: Regresión lineal simple	29
5.	Pruebas de bondad de ajuste	34
5.1.	Pruebas propuestas	35
5.1.1.	Kolmogorov-Smirnov	35
5.1.2.	Coefficiente de correlación	38
5.1.3.	Bootstrap paramétrico	43
5.2.	Estudios de simulación	44
5.2.1.	Pruebas con probabilidad de error tipo I máximo	44
5.2.2.	r_n^* con bootstrap paramétrico	48
5.3.	Ejemplos de aplicación	48
6.	Pruebas para normalidad	50
6.1.	Prueba basada en el índice de asimetría	50
6.2.	Prueba de razón de verosimilitudes generalizada	51
6.3.	Potencia y tamaño de las pruebas	52
6.3.1.	Tamaño de las pruebas	52
6.3.2.	Potencia de las pruebas	53
7.	Pruebas de Hipótesis Para la Distribución Normal Asimétrica	55
7.1.	Pruebas para el parámetro de localidad	55
7.1.1.	Estadística Señal-Ruido	55

Índice

7.1.2. Estadística de Cox y Oakes	65
7.1.3. Razón de verosimilitudes	67
7.1.4. Comparación de potencias	68
7.1.5. Ejemplo de aplicación en epidemiología	68
7.2. Comparación de medias de dos poblaciones normales asimétricas . . .	70
7.2.1. Dos poblaciones con parámetros de escala y forma comunes . .	70
7.2.2. Dos poblaciones con parámetros de localidad y forma comunes	76
7.2.3. Dos poblaciones con parámetro de forma común	78
8. Conclusiones	85
Referencias	86
Apéndices	91
Apéndice A: Densidades a posteriori de los parámetros de la distribución normal asimétrica	91
Apéndice B: Matriz de información de Fisher observada para comparación de poblaciones normales asimétricas con diferente parámetro de localidad	98
Apéndice C: Cálculo de estimadores de máxima verosimilitud para mezclas de dos componentes SN con el algoritmo EM Generalizado	104

Índice de tablas

4.1. Datos de estimación en la frontera de Azzalini	20
4.2. Prueba de convergencia de Gelman y Rubin para los datos de Azzalini con las a prioris propuestas.	23
4.3. Estadísticas resumen para cadenas transformadas con las a prioris propuestas para los datos de Azzalini.	23
4.4. Datos de Sartori	26
4.5. Prueba de convergencia de Gelman y Rubin para los datos de Sartori con las a prioris propuestas.	27
4.6. Estadísticas resumen para cadenas transformadas con las a prioris propuestas para los datos de Sartori.	28
4.7. Prueba de convergencia de Gelman y Rubin para el problema de regresión con las a prioris propuestas.	30
4.8. Estadísticas resumen para cadenas transformadas con las a prioris propuestas para el modelo de regresión.	31
5.1. Valores críticos de la estadística D_n	38
5.2. Valores críticos de la estadística r_n	42
5.3. Valores críticos de la estadística r_n^*	43
5.4. Tamaño de la prueba D_n	44
5.5. Tamaño de la prueba r_n	44

Índice de tablas

5.6. Tamaño de la prueba r_n^*	44
5.7. Comparación de potencias de las pruebas estudiadas por Mateu <i>et al.</i> y las propuestas para $n = 50$	46
5.8. Comparación de potencias de las pruebas estudiadas por Mateu <i>et al.</i> y las propuestas para $n = 100$	46
5.9. Comparación de potencias de las pruebas estudiadas por Mateu <i>et al.</i> y las propuestas para $n = 150$	47
5.10. Comparación de potencias de las estadística propuesta por Meintanis y r_n^* para $n = 50$	48
5.11. IQ para hombres blancos.	48
5.12. IQ para hombres de otras razas.	49
6.1. Tamaño de la prueba basada en el índice de asimetría para probar normal vs normal asimétrica	53
6.2. Tamaño de la prueba de razón de verosimilitudes para probar normal vs normal asimétrica	53
7.1. Valores críticos de la estadística R	58
7.2. Comparación de cuantiles de RC empírica, RC asintótica y RC^*	64
7.3. Valores críticos de la estadística de Cox y Oakes	67
7.4. Datos de resfriados de gemelos de Roberts (1966)	69
7.5. Tamaño de la prueba de r. v. para probar si dos poblaciones SN tienen la misma media, parámetros de escala y forma comunes pero desconocidos	74
7.6. Potencia de la prueba de r. v. para mezclas de poblaciones SN	84

Índice de figuras

3.1. Densidad SN para diferentes valores del parámetro γ	5
4.1. Verosimilitud perfil relativa para los datos de Azzalini.	21
4.2. Densidades a posteriori para los datos de Azzalini usando la parametrización centrada.	22
4.3. Diagnóstico de convergencia, con las a prioris propuestas para los datos de Azzalini.	23
4.4. Densidades a posteriori para los datos de Azzalini usando la parametrización directa.	24
4.5. Densidad marginal a posteriori para γ , reimpresión del artículo de Liseo y Loperfido (2006).	24
4.6. Datos de estimación en la frontera y modelos ajustados.	25
4.7. Densidades a posteriori para los datos de Sartori usando la parametrización centrada.	27
4.8. Densidades a posteriori para los datos de Sartori usando la parametrización directa.	28
4.9. Densidades a posteriori con a prioris propuestas para el modelo de regresión usando la parametrización centrada.	31
4.10. Densidades a posteriori con a prioris propuestas para el modelo de regresión usando la parametrización centrada.	32
4.11. Residuales para el modelo de regresión.	32

Índice de figuras

5.1. Distribución empírica de la estadística D_n	37
5.2. Valores críticos para las estadísticas r_n y r_n^* como función de γ para $n = 50$	42
6.1. Potencia de la prueba basada en el índice de asimetría y r. v. para probar normal vs normal asimétrica	54
6.2. Comparación de potencias de la prueba basada en el índice de asimetría y r.v. para probar normal vs normal asimétrica	54
7.1. Distribución empírica de la estadística R	57
7.2. Distribución empírica de la estadística RC y la aproximación normal, $n = 100, 200$	59
7.3. Distribución empírica de la estadística RC y la aproximación normal, $n = 400, 800$	60
7.4. Distribución empírica de la estadística R y otra aproximación normal, $n = 200, 300$	63
7.5. Distribución empírica de la estadística Cox	66
7.6. Potencia de la prueba R , Cox y de razón de verosimilitudes, $n = 100$	68
7.7. Comparación de medias de dos poblaciones normales asimétricas con parámetros de escala y forma comunes, pero desconocidos	70
7.8. Potencia de la prueba de rangos y r. v. para probar si dos poblaciones SN tienen la misma media, parámetros de escala y forma comunes, $\gamma = 0$	75
7.9. Potencia de la prueba de rangos y r. v. para probar si dos poblaciones SN tienen la misma media, parámetros de escala y forma comunes, $\gamma \neq 0$	75
7.10. Comparación de medias de dos poblaciones normales asimétricas con parámetros de localidad y forma comunes, pero desconocidos	76
7.11. Potencia de la prueba de r. v. para probar si dos poblaciones SN tienen la misma media, parámetros de localidad y forma comunes pero desconocidos, $\gamma = 3$	78

7.12. Comparación de dos poblaciones normales asimétricas con parámetro de forma común, pero desconocido.	79
---	----

Capítulo 1

Introducción

Las distribuciones normales asimétricas constituyen una familia de distribuciones de tres parámetros: localidad, escala y forma, la cual contiene a la familia normal cuando el parámetro de forma es 0 y a la distribución media-normal cuando dicho parámetro tiende a infinito. Esta familia de distribuciones tiene algunas de las propiedades de la familia normal [ver por ejemplo [Arnold y Lin \(2004\)](#), [Gupta *et al.* \(2004\)](#) y [Azzalini \(2005\)](#)], lo que la hace atractiva desde el punto de vista de aplicaciones.

La familia apareció de forma independiente varias veces en la literatura estadística [ver [Roberts \(1966\)](#), [O'Hagan y Leonard \(1976\)](#)]; sin embargo, fue [Azzalini \(1985\)](#) quien estudió sus principales propiedades, propuso algunas generalizaciones y le dio el nombre con el cual se le conoce actualmente. Desde entonces y hasta nuestros días sus aplicaciones han sido numerosas, ver por ejemplo [Russell y González \(2002\)](#), [Chen *et al.* \(2004\)](#), [Genton \(2004\)](#), [Azzalini \(2005\)](#), [Liseo y Loperfido \(2006\)](#).

A pesar de ser un modelo flexible, en el sentido de poder modelar datos asimétricos, heredar propiedades de la distribución normal, y ser tratable matemáticamente, presenta algunos problemas desde el punto de vista de inferencia, el primero de los cuales es la estimación de parámetros, la cual no es nada fácil, por ejemplo si se considera el caso estándar (parámetro de localidad 0 y escala 1) y todas las observaciones son positivas (negativas), la función de verosimilitud es monótona creciente (decreciente) y no puede maximizarse, por lo tanto el estimador del parámetro de forma no puede obtenerse. El método de momentos presenta también dificultades. El problema de estimación está aún vigente aunque ya existen algunas propuestas, tanto desde el punto de vista clásico como del Bayesiano [ver [Sartori \(2006\)](#), [Liseo y Loperfido \(2006\)](#)].

Ahora bien, dado un conjunto de datos, ¿cómo saber si pueden ser modelados con la distribución normal asimétrica?, la respuesta es obvia, utilizar una prueba de bondad ajuste, pero aunque parezca sorprendente pocos trabajos se han hecho al respecto [ver [Salvan \(1986\)](#), [Gupta y Chen \(2001\)](#), [Mateu *et al.* \(2007\)](#) y [Meintanis \(2007\)](#)].

1. Introducción

Más aún no existe referencia alguna en la literatura a extensión de pruebas clásicas que se realizan con datos normales (pruebas de medias, comparación de poblaciones, entre otras) al caso de la distribución normal asimétrica.

En el presente escrito se abordan algunos de los problemas mencionados en los párrafos anteriores. En el capítulo 2 se presentan los objetivos del trabajo. En el capítulo 3 se presenta una revisión de literatura de la distribución normal asimétrica, técnicas y resultados necesarios para el desarrollo del trabajo. En el capítulo 4 se aborda el problema de estimación, desde el punto de vista Bayesiano, se muestra como extender el algoritmo empleado al caso de problemas de regresión con errores normales asimétricos utilizando el algoritmo de Metropolis y una nueva técnica de Cadenas de Markov Monte Carlo muy poderosa denominada algoritmo “t-walk”. En el capítulo 5 se estudia el problema de pruebas de bondad de ajuste. En el capítulo 6 se estudia el problema de decidir si una muestra es normal o normal asimétrica. En el capítulo 7 se estudian pruebas de hipótesis en la familia normal asimétrica, las que incluyen pruebas para el parámetro de localización, pruebas sobre las medias considerando varios casos, el último de los cuales conduce a la utilización de mezclas. En el capítulo 8 se presentan las conclusiones del trabajo y finalmente se presenta un apéndice con programas en [R](#), C++.

Capítulo 2

Objetivos

- Presentar un método alternativo de estimación de parámetros que sea de fácil implementación y extensión a otros modelos, por ejemplo el de regresión.
- Desarrollar pruebas de bondad de ajuste para la distribución normal asimétrica.
- Estudiar la potencia de pruebas para saber si un conjunto de datos viene de la distribución normal o de la distribución normal asimétrica.
- Desarrollar pruebas para el parámetro de localización de la distribución.
- Estudiar el problema de comparación de poblaciones normales asimétricas.

Capítulo 3

Revisión de Literatura

En este capítulo se hace una revisión de los aspectos relevantes de la distribución normal asimétrica, así como conceptos y metodología utilizados para el desarrollo del trabajo.

3.1. Distribución normal asimétrica

Definición 3.1 Una v.a. Z es normal asimétrica con parámetro de forma γ si su función de densidad es:

$$f_Z(z; \gamma) = 2\phi(z) \Phi(\gamma z) I_{(-\infty, \infty)}(z), \quad (3.1)$$

donde $\phi(\cdot)$ y $\Phi(\cdot)$ denotan la función de densidad y de distribución de una variable normal estándar, respectivamente y $\gamma \in \mathbb{R}$.

Para denotar que Z tiene la función de densidad (3.1) se escribe $Z \sim SN(\gamma)$.

Algunas propiedades

- 1) Si $\gamma = 0$ en (3.1) entonces $Z \sim N(0, 1)$
- 2) Si $Z \sim SN(\gamma)$, entonces $-Z \sim SN(-\gamma)$
- 3) Si $\gamma \rightarrow \infty$, entonces (3.1) converge a la distribución media-normal, $2\phi(z)I_{(0, \infty)}(z)$
- 4) Si $Z \sim SN(\gamma)$, entonces $Z^2 \sim \chi_1^2$

3.1. Distribución normal asimétrica

5) Si $Z \sim SN(\gamma)$, $W \sim N(0, 1)$ entonces $|Z| \stackrel{d}{=} |W|$, donde $\stackrel{d}{=}$ denota igualdad en distribución, ver [Arnold y Lin \(2004\)](#).

Suponga que $Z \sim SN(\gamma)$ y que

$$Y = \xi + \omega Z, \quad (3.2)$$

con $\xi \in \mathbb{R}$, $\omega \in \mathbb{R}^+$, entonces la función de densidad de Y es:

$$f_Y(y; \xi, \omega, \gamma) = 2 \frac{1}{\omega} \phi \left(\frac{y - \xi}{\omega} \right) \Phi \left[\gamma \left(\frac{y - \xi}{\omega} \right) \right] I_{(-\infty, \infty)}(y), \quad (3.3)$$

y se denota $Y \sim SN_D(\xi, \omega, \gamma)$, donde el subíndice D indica el uso de la “parametrización directa” ([Azzalini y Capitanio, 1999](#)).

En (3.2) ξ es un parámetro de localidad, ω es un parámetro de escala y γ es un parámetro de forma.

En la Figura 3.1 se observa la función de densidad normal asimétrica para diferentes valores del parámetro de forma γ .

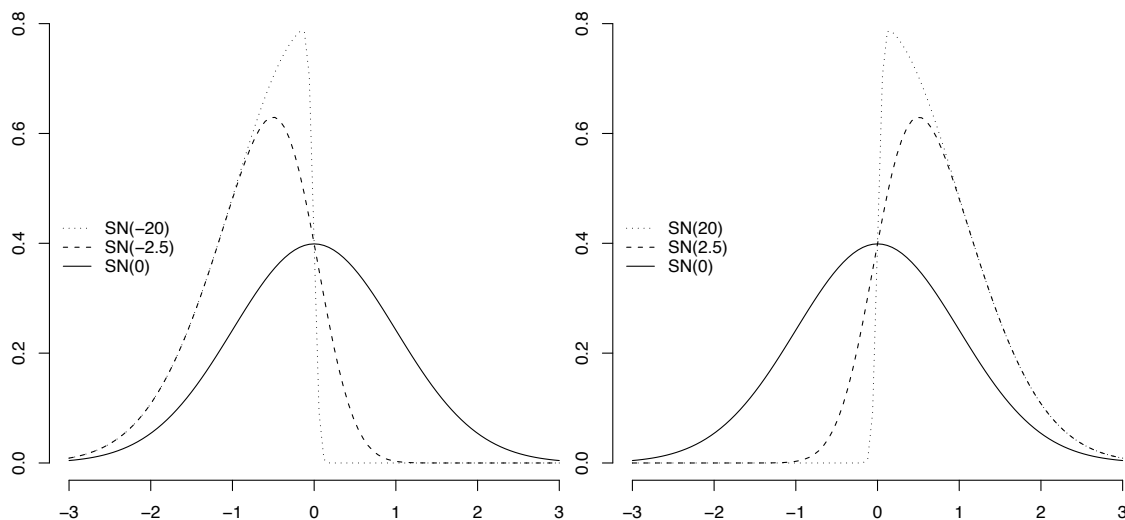


Figura 3.1: Densidad SN para diferentes valores del parámetro γ .

La función generatriz de momentos de (3.1), obtenida por [Azzalini \(1985\)](#), está dada por:

3.1. Distribución normal asimétrica

$$M_Z(t) = 2 \exp(t^2/2) \Phi \left(\frac{\gamma t}{\sqrt{1 + \gamma^2}} \right). \quad (3.4)$$

A partir de la cual se pueden obtener la media y la varianza:

$$\begin{aligned} E(Z) &= \sqrt{\frac{2}{\pi}} \frac{\gamma}{\sqrt{1 + \gamma^2}}, \\ \text{Var}(Z) &= 1 - \frac{2}{\pi} \frac{\gamma^2}{1 + \gamma^2}. \end{aligned}$$

El índice de asimetría de Y está dado por:

$$\gamma_1 = \frac{\kappa_3}{\kappa_2^{3/2}} = \frac{4 - \pi}{2} \frac{(E(Z))^3}{\{1 - (E(Z))^2\}^{3/2}}, \quad (3.5)$$

donde κ_2, κ_3 son los cumulantes de segundo y tercer orden respectivamente. El rango de γ_1 es $(-c_1, c_1)$ donde $c_1 = \sqrt{2}(4 - \pi)/(\pi - 2)^{3/2} \approx 0.99527$.

La función de distribución acumulada de (3.1) está dada por:

$$F_Z(z) = P(Z \leq z) = \int_{-\infty}^z \frac{1}{\pi} \exp(-x^2/2) \int_{-\infty}^{\gamma x} \exp(-t^2/2) dt dx, \quad (3.6)$$

la cual fue tabulada por [Gupta y Chen \(2001\)](#) para diferentes valores del parámetro γ . También es posible calcular el valor de esta función empleando la expresión general dada por [Azzalini \(1985\)](#):

$$F_Z(z) = \phi(z) - 2T(z, \gamma),$$

donde $T(\cdot)$ es la función de [Owen \(1956\)](#), dada por:

$$T(h, a) = (2\pi)^{-1} \int_0^a \frac{\exp\{-\frac{1}{2}h^2(1 + x^2)\}}{1 + x^2} dx, \quad -\infty < h, a < +\infty. \quad (3.7)$$

Existen rutinas para calcular de manera eficiente el valor de la integral en (3.7) como por ejemplo la propuesta por [Young y Minder \(1974\)](#).

3.2. Pruebas de hipótesis

Definición 3.2 Una hipótesis estadística es una conjetura o aseveración acerca de la distribución de una o más variables aleatorias. Si en la hipótesis estadística la distribución se encuentra completamente especificada, entonces es llamada hipótesis simple; de otra manera, es llamada hipótesis compuesta.

Sea X_1, \dots, X_n una muestra aleatoria de $F(x; \boldsymbol{\theta})$ con $\boldsymbol{\theta} \in \Omega \subset \mathbb{R}^m$, $m \geq 1$. Entonces usualmente se tiene interés en probar el juego de hipótesis:

$$H_0 : \boldsymbol{\theta} \in \Omega_0 \quad \text{vs} \quad H_1 : \boldsymbol{\theta} \in \Omega - \Omega_0.$$

Es decir, con base en la muestra se desea decidir si se rechaza o no se rechaza la hipótesis nula H_0 en favor de la hipótesis alterna H_1 .

Definición 3.3 Una prueba de hipótesis es una regla que especifica:

- i) Para qué valores de la muestra no se rechaza H_0 .
- ii) Para qué valores de la muestra se rechaza H_0 en favor de H_1 .

Usualmente la prueba de hipótesis se especifica usando una *estadística de prueba* $W = w(X_1, \dots, X_n)$. Si Ξ denota el conjunto de todos los valores posibles de \mathbf{X} , entonces una prueba de hipótesis es una partición de Ξ en los conjuntos Ξ_A, Ξ_R . Es decir, Ξ_A, Ξ_R son disjuntos y $\Xi = \Xi_A \cup \Xi_R$. Si la realización \mathbf{x} de \mathbf{X} es tal que $\mathbf{x} \in \Xi_A$ no se rechaza H_0 y si $\mathbf{x} \in \Xi_R$ entonces se rechaza H_0 . A Ξ_A y Ξ_R se les llama la región de aceptación y rechazo respectivamente.

Ξ_R define una prueba $\Xi_R = \{\mathbf{x} \in \Xi : w(\mathbf{x}) > c\}$. A cada prueba Ξ_R se le asocia una función indicadora $\phi_{\Xi_R}(\mathbf{x}) : \Xi \rightarrow \{0, 1\}$ de tal modo que:

$$\phi_{\Xi_R}(\mathbf{x}) = \begin{cases} 1 & \text{si } \mathbf{x} \in \Xi_R, \\ 0 & \text{si } \mathbf{x} \in \Xi_A. \end{cases}$$

Cuando se realiza una prueba de hipótesis se puede cometer uno de los errores siguientes:

Error tipo I: Rechazar H_0 cuando de hecho es verdadera.

Error tipo II: No rechazar H_1 cuando de hecho es falsa.

3.2. Pruebas de hipótesis

Desgraciadamente no existe una prueba que minimice ambos tipos de errores a la vez, y lo que usualmente se hace es fijar la probabilidad de cometer el error tipo I y tratar de minimizar la probabilidad de cometer el error tipo II.

Definición 3.4 *La función de potencia de una prueba de hipótesis con región de rechazo Ξ_R es una función de θ dada por $\beta_\phi(\theta) = P_\theta(\mathbf{X} \in \Xi_R)$.*

Una buena prueba tiene una función de potencia cercana a 1 cuando $\theta \in \Omega - \Omega_0$ y cercana a 0 cuando $\theta \in \Omega_0$.

Definición 3.5 *Para $0 \leq \alpha \leq 1$, una prueba con función de potencia $\beta_\phi(\theta)$ es una prueba de tamaño α si*

$$\sup_{\theta \in \Omega_0} \beta_\phi(\theta) = \alpha.$$

Existen varios métodos para construir pruebas de hipótesis (Casella y Berger, 2002) algunos de los cuales toman ventaja de diferentes aspectos del problema de interés, pero existe un método general denominado *Prueba de razón de verosimilitudes generalizada* que casi siempre es aplicable y posee propiedades óptimas en ciertos casos.

3.2.1. Prueba de razón de verosimilitudes generalizada

Sea X_1, \dots, X_n una muestra aleatoria de la densidad $f_X(x; \theta)$, la función de verosimilitud se define como:

$$L(\theta; x_1, \dots, x_n) := L(\theta; \mathbf{x}) = f(\mathbf{x}; \theta) = \prod_{i=1}^n f_{X_i}(x_i; \theta). \quad (3.8)$$

Sea Ω el espacio paramétrico, en seguida se define la prueba de razón de verosimilitudes.

Definición 3.6 *La estadística de razón de verosimilitudes para probar $H_0 : \theta \in \Omega_0$ vs $H_1 : \theta \in \Omega - \Omega_0$ es*

$$\lambda(\mathbf{x}) = \frac{\sup_{\theta \in \Omega_0} L(\theta; \mathbf{x})}{\sup_{\theta \in \Omega} L(\theta; \mathbf{x})}. \quad (3.9)$$

3.2. Pruebas de hipótesis

Una prueba de razón de verosimilitudes tiene una región de rechazo de la forma $\{\mathbf{x} : \lambda(\mathbf{x}) \leq c\}$ donde $c \in [0, 1]$. Usualmente es muy difícil obtener la distribución de $\lambda(\mathbf{X})$. Una alternativa es utilizar su distribución asintótica, de la cual se trata en los teoremas siguientes.

Teorema 3.1 (Espacio paramétrico univariado) Para probar $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$, suponga que se tiene una muestra aleatoria X_1, \dots, X_n de $f_X(x; \theta)$ y que $f_X(x; \theta)$ satisface ciertas condiciones de regularidad (Casella y Berger, 2002, pág. 516), entonces bajo la hipótesis nula, cuando $n \rightarrow \infty$,

$$-2 \log \lambda(\mathbf{X}) \xrightarrow{d} \chi_1^2,$$

donde χ_1^2 es una variable aleatoria χ^2 con 1 grado de libertad.

Se rechaza H_0 al nivel de significancia α si y solo sí $-2 \log \lambda(\mathbf{X}) \geq \chi_{1,1-\alpha}^2$ donde $\chi_{1,1-\alpha}^2$ es el cuantil $1 - \alpha$ de la distribución χ^2 con 1 grado de libertad.

Teorema 3.2 (Espacio paramétrico multivariado) Sea X_1, \dots, X_n una muestra aleatoria de $f_X(x; \boldsymbol{\theta})$ que satisface ciertas condiciones de regularidad y se desea probar $H_0 : \boldsymbol{\theta} \in \Omega_0$ vs $H_1 : \boldsymbol{\theta} \in \Omega - \Omega_0$, entonces bajo la hipótesis nula, cuando $n \rightarrow \infty$,

$$-2 \log \lambda(\mathbf{X}) \xrightarrow{d} \chi_\nu^2,$$

donde χ_ν^2 es una variable aleatoria χ^2 con ν grados de libertad, los cuales se determinan a partir de la diferencia entre el número de parámetros libres bajo H_0 y el número de parámetros libres en todo el espacio paramétrico.

Se rechaza H_0 al nivel de significancia α si y solo sí $-2 \log \lambda(\mathbf{X}) \geq \chi_{\nu,1-\alpha}^2$ donde $\chi_{\nu,1-\alpha}^2$ es el cuantil $1 - \alpha$ de la distribución χ^2 con ν grados de libertad.

Para mayores detalles ver Mood *et al.* (1974), Casella y Berger (2002).

3.2.2. Pruebas de bondad de ajuste

Sean x_1, \dots, x_n observaciones independientes de una variable aleatoria con función de distribución $F(x)$ la cual es desconocida. Supóngase que se desea probar:

$$H_0 : F(x) = F_0(x), \tag{3.10}$$

3.3. Teorema central del límite multivariado

donde $F_0(x)$ es una distribución particular, la cual puede ser continua o discreta. El problema de probar (3.10) es denominado el problema de la “prueba de bondad de ajuste” (Kendall y Stuart, 1973). Es decir, el problema consiste en verificar si un conjunto de datos vienen de una distribución en particular, para lo cual se observa los mismos y posteriormente se decide si H_0 se rechaza o no.

3.3. Teorema central del límite multivariado

Teorema 3.3 (TCL multivariado) Sean $\mathbf{X}_1, \dots, \mathbf{X}_n$ vectores aleatorios independientes e idénticamente distribuidos (i.i.d) con vector de medias $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)'$ y matriz de varianzas y covarianzas $\boldsymbol{\Sigma} = (\sigma_{ij})$, $i, j = 1, \dots, p$, entonces, conforme $n \rightarrow \infty$

$$\sqrt{n}(\bar{\mathbf{X}}_n - \boldsymbol{\mu}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma})$$

donde

$$\begin{aligned}\mathbf{X} &= (X_1, \dots, X_p)', \\ \bar{\mathbf{X}}_n &= (\bar{X}_1, \dots, \bar{X}_p), \\ \bar{X}_i &= \frac{1}{n} \sum_{k=1}^n X_{ik}, \quad i = 1, \dots, p.\end{aligned}$$

Teorema 3.4 (Taylor multivariado) Sea $\{\mathbf{T}_n\}$ una sucesión de vectores aleatorios tal que $\sqrt{n}(\mathbf{T}_n - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma})$ y considere la función $g(\mathbf{T}_n)$ tal que $\dot{g}(\boldsymbol{\theta}) = \partial g(\mathbf{x})/\partial \mathbf{x}|_{\mathbf{x}=\boldsymbol{\mu}}$ es no nula y continua en una vecindad de $\boldsymbol{\mu}$. Entonces:

$$\sqrt{n}(g(\mathbf{T}_n) - g(\boldsymbol{\theta})) \xrightarrow{d} N(0, \gamma^2) \quad \text{donde } \gamma^2 = [\dot{g}(\boldsymbol{\theta})]' \boldsymbol{\Sigma} [\dot{g}(\boldsymbol{\theta})].$$

Una revisión detallada del teorema central del límite multivariado y Taylor multivariado puede verse en Sen y Singer (1993) capítulo 3 y en el libro de Lehmann y Casella (1998).

3.4. Los algoritmos EM y EM Generalizado

El algoritmo Esperanza Maximización (EM) es una técnica iterativa para el cálculo de estimadores de máxima verosimilitud en problemas donde la función de verosimilitud es muy compleja, como por ejemplo en presencia de datos faltantes u ocultos. Dicho

3.4. Los algoritmos EM y EM Generalizado

algoritmo fue introducido por [Dempster, Laird y Rubin \(1977\)](#), consiguiendo formalizar y justificar una idea que habían explorado otros investigadores: Dada una primera estimación de los parámetros, predecir los datos faltantes, re-estimar parámetros y continuar iterando hasta la convergencia.

Sea \mathbf{y} el vector de datos observados y Ω_Y su correspondiente espacio muestral. Suponga que existe otro vector de datos no observado, digamos \mathbf{x} , perteneciente al espacio muestral Ω_X , y una función ψ , $\psi : \Omega_X \rightarrow \Omega_Y$.

La iteración i -ésima del algoritmo ($i = 1, 2, \dots$) consta de dos etapas:

- 1) **Etapas E** (Esperanza). Calcular $Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(i-1)}) = E(\log f(\mathbf{x}; \boldsymbol{\theta}) | \mathbf{y}, \boldsymbol{\theta}^{(i-1)})$
- 2) **Etapas M** (Maximización). Obtener $\boldsymbol{\theta}^{(i)}$ que maximiza a $Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(i-1)})$ en $\boldsymbol{\theta}$

La regla de parada para el proceso iterativo puede basarse en una distancia entre $\boldsymbol{\theta}^{(i)}$ y $\boldsymbol{\theta}^{(i-1)}$ o bien entre $Q(\boldsymbol{\theta}^{(i-1)}; \boldsymbol{\theta}^{(i-1)})$ y $Q(\boldsymbol{\theta}^{(i)}; \boldsymbol{\theta}^{(i-1)})$.

La maximización de $Q(\boldsymbol{\theta}; \boldsymbol{\theta}^*)$ en $\boldsymbol{\theta}$ puede resultar muy complicada, ya que puede no existir una expresión explícita para el estimador ([Lange, 1999](#)). En tales casos es necesario aplicar algún tipo de algoritmo de optimización para realizar la maximización, sin embargo este paso puede evitarse.

Definición 3.7 (Algoritmo EM Generalizado). *Es una variante del algoritmo EM que consiste en calcular $\boldsymbol{\theta}$ en la etapa M de modo que $Q(\boldsymbol{\theta}; \boldsymbol{\theta}^*) \geq Q(\boldsymbol{\theta}^*; \boldsymbol{\theta}^*)$ sin llegar a comprobar que sea un máximo.*

El algoritmo EM no genera estimadores para la matriz de varianzas y covarianzas de los estimadores, razón por la cual se han hecho modificaciones al mismo con la finalidad de salvar este problema. Una de tales modificaciones es la hecha por [Oakes \(1999\)](#), y según [Robert y Casella \(2004\)](#) dicha modificación es simple y muy útil.

[Oakes \(1999\)](#) mostró que si $L(\mathbf{y}; \boldsymbol{\theta})$ es la log-verosimilitud de la muestra entonces:

$$\frac{\partial^2 L(\mathbf{y}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2} = \left(\frac{\partial^2 Q(\boldsymbol{\theta}'; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}'^2} + \frac{\partial^2 Q(\boldsymbol{\theta}'; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}' \partial \boldsymbol{\theta}} \right)_{\boldsymbol{\theta}' = \boldsymbol{\theta}}. \quad (3.11)$$

La varianza aproximada de $\hat{\boldsymbol{\theta}}$ se calcula con:

$$Var \hat{\boldsymbol{\theta}} \approx \left[\frac{\partial^2 L(\mathbf{y}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2} \right]^{-1}. \quad (3.12)$$

3.5. El algoritmo de Metropolis

Con esto se tiene una expresión para calcular de manera aproximada la varianza de los estimadores, una desventaja es que las derivadas pueden ser difíciles de obtener.

3.5. El algoritmo de Metropolis

El algoritmo de Metropolis es un método de Cadenas de Markov Monte Carlo (MCMC por sus siglas en inglés) que ha sido ampliamente utilizado en Física, restauración de imágenes y recientemente en Estadística (Carlin y Louis, 2000). Dicho algoritmo, propuesto por Metropolis *et al.* (1953), se describe a continuación.

Supóngase que se desea muestrear de una distribución $p(\boldsymbol{\theta})$ completamente conocida (excepto, quizás por una constante de proporcionalidad). El algoritmo de Metropolis permite resolver este problema, como sigue:

- 1) Seleccionar un punto inicial $\boldsymbol{\theta}^0$ en el soporte de $p(\cdot)$
- 2) Para $t = 1, 2, \dots$
 - a) Generar $\boldsymbol{\theta}^*$ al tiempo t , de $J_t(\boldsymbol{\theta}^*, \boldsymbol{\theta}^{(t-1)})$, dicha distribución debe ser simétrica en sus argumentos, es decir, $J_t(\boldsymbol{\theta}_a, \boldsymbol{\theta}_b) = J_t(\boldsymbol{\theta}_b, \boldsymbol{\theta}_a)$ para todo $\boldsymbol{\theta}_a, \boldsymbol{\theta}_b, t$
 - b) Calcular el cociente $r = p(\boldsymbol{\theta}^*)/p(\boldsymbol{\theta}^{t-1})$
 - c) Hacer

$$\boldsymbol{\theta}^t = \begin{cases} \boldsymbol{\theta}^* & \text{con probabilidad } \min(r, 1) \\ \boldsymbol{\theta}^{t-1} & \text{de otro modo} \end{cases}$$

Los libros de Carlin y Louis (2000) y Gelman *et al.* (2004) tienen abundantes ejemplos en los que se utiliza este algoritmo.

3.6. El algoritmo “t-walk”

El algoritmo “t-walk” (Christen y Fox, 2007) es un método de Cadenas de Markov Monte Carlo que permite muestrear de distribuciones continuas con estructuras de correlación arbitrarias, construido como un algoritmo de Metropolis-Hastings en el espacio producto. El algoritmo, que trabaja como una caja negra, requiere una función objetivo, $p(\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \mathbb{R}^m$ que puede ser una distribución a posteriori o cualquier otra distribución de la que se desee muestrear y dos puntos $\boldsymbol{\theta}_0, \boldsymbol{\theta}_1$ tales que $p(\boldsymbol{\theta}_1) > 0$, $p(\boldsymbol{\theta}_2) > 0$, a partir de los cuales se generan movimientos en varios sentidos para generar nuevos candidatos. El algoritmo parece trabajar de manera adecuada, en el

3.6. El algoritmo “t-walk”

escrito de los mencionados autores se encuentran varios ejemplos de su uso, quienes además implementaron el algoritmo en el programa [R](#) y en C++ ([Stroustrup, 1984](#)).

Capítulo 4

Métodos alternativos de estimación

En este capítulo se presentan métodos alternativos de estimación de los tres parámetros de la distribución normal asimétrica, tanto desde el punto de vista clásico como el Bayesiano.

4.1. Estimación desde el punto de vista clásico

Para estimar los parámetros desde el punto de vista clásico existen varias alternativas:

i) Estimadores de momentos

Los estimadores de momentos no dan buenos resultados, en el sitio web de Azzalinni se pueden ver algunos ejemplos. [Pewsey \(2000\)](#) da razones adicionales del porque los estimadores de momentos con la “parametrización directa” no deben utilizarse.

ii) Estimadores de máxima verosimilitud

Aunque la función de verosimilitud puede calcularse sin mayores problemas, al momento de maximizarla se presentan varios problemas; por ejemplo cuando $\xi = 0$, $\omega = 1$, y todas las observaciones son positivas o negativas, entonces la función de verosimilitud es monótona creciente o decreciente por lo que el estimador de máxima verosimilitud de γ no existe ([Azzalini, 1985](#)). Este problema en particular aún está abierto a la investigación, aunque ya existen algunas propuestas, entre las que destaca la de [Sartori \(2006\)](#), quien utiliza una modificación de la función *score* para estimar el parámetro γ en combinación con estimadores de máxima verosimilitud para los otros parámetros. El método propuesto por [Sartori](#) no presenta ninguna mejora para los parámetros de localidad y escala con respecto al método de máxima

4.1. Estimación desde el punto de vista clásico

verosimilitud convencional.

En el caso de parámetros ξ , ω , γ desconocidos el problema puede ser aún más grave. Siempre hay un punto de inflexión en $\gamma = 0$ para la función de verosimilitud, con lo cual el Hessiano se vuelve singular (Russell y González, 2002). Este problema se resuelve empleando una reparametrización de la función de densidad Azzalini y Capitanio (1999).

Desde el punto de vista Bayesiano actualmente solo se dispone del método propuesto por Liseo y Loperfido (2006), y su aportación se describe con mayor detalle en la sección respectiva.

4.1.1. Estimación con el método de momentos para la parametrización centrada

La reparametrización de la que se habló en *ii*) genera un método para obtener estimadores de ξ , ω , γ de forma semiparamétrica. La reparametrización, propuesta por Azzalini y Capitanio (1999), consiste en reparametrizar de (ξ, ω, γ) a (μ, σ, γ_1) y se obtiene al escribir (3.2) de la forma siguiente:

$$Y = \xi + \omega Z = \mu + \sigma Z_0, \quad Z_0 = \{Z - E(Z)\} / \sqrt{1 - (E(Z))^2},$$

donde Z_0 es una variable aleatoria estandarizada con media cero y varianza uno, μ y σ denotan la media y la desviación estándar de Y respectivamente y γ_1 es el índice de asimetría de Y y Z . En esta parametrización $E(Y) = \mu$, $Var(Y) = \sigma^2$, y usualmente se escribe $Y \sim SN_C(\mu, \sigma, \gamma_1)$.

La idea es estimar primero el índice de asimetría utilizando un método no paramétrico. Dadas las observaciones y_1, \dots, y_n , el estimador muestral de γ_1 es:

$$\check{\gamma}_1 = \frac{\sqrt{n} \sum_{i=1}^n (y_i - \bar{y})^3}{\left(\sum_{i=1}^n (y_i - \bar{y})^2 \right)^{3/2}}.$$

Se sabe que $\check{\gamma}_1$ es un estimador sesgado de γ_1 , el estimador de γ_1 está dado por:

$$\tilde{\gamma}_1 = \frac{\sqrt{n(n-1)}}{n-2} \check{\gamma}_1. \quad (4.1)$$

4.1. Estimación desde el punto de vista clásico

Es importante recordar que $\gamma_1 \in (-c_1, c_1)$, por lo tanto si $\tilde{\gamma}_1$ no está contenido en este intervalo hay que hacer un ajuste, es decir si $|\tilde{\gamma}_1| > c_1$ entonces usualmente se toma $\tilde{\gamma}_1 = \text{sgn}(\tilde{\gamma}_1)0.95 c_1$. Así una vez que se estima usando (4.1) entonces se utiliza (3.5) para estimar γ , de donde se obtiene:

$$\tilde{\gamma} = \text{sgn}(\gamma_1) \left(\frac{2}{4 - \pi} \right)^{1/3} |\tilde{\gamma}_1|^{1/3} \left\{ \frac{2}{\pi} + |\tilde{\gamma}_1|^{2/3} \left(\frac{2}{4 - \pi} \right)^{2/3} \left(\frac{2}{\pi} - 1 \right) \right\}^{-1/2}. \quad (4.2)$$

Los estimadores de ξ y ω se calculan a partir de los estimadores de momentos de μ y σ y usando las relaciones:

$$\begin{aligned} \text{Var}(Y) &= \sigma^2 = \omega^2 \text{Var}(Z), \\ E(Y) &= \mu = \xi + \omega E(Z) \end{aligned}$$

De donde se obtiene:

$$\tilde{\omega} = \frac{\tilde{\sigma}}{\sqrt{\widetilde{\text{Var}}(Z)}} \quad (4.3)$$

$$\tilde{\xi} = \tilde{\mu} - \tilde{\omega} \widetilde{E}(Z) \quad (4.4)$$

En la que:

$$\widetilde{E}(Z) = \sqrt{\frac{2}{\pi}} \frac{\tilde{\gamma}}{\sqrt{1 + \tilde{\gamma}^2}}, \quad \widetilde{\text{Var}}(Z) = 1 - \frac{2}{\pi} \frac{\tilde{\gamma}^2}{1 + \tilde{\gamma}^2}$$

Entonces se puede cambiar fácilmente de la parametrización centrada a la directa utilizando la ecuación siguiente:

4.2. Estimación Bayesiana

$$\begin{aligned}
 \gamma &= \tilde{\gamma} = \text{sgn}(\gamma_1) \left(\frac{2}{4 - \pi} \right)^{1/3} |\tilde{\gamma}_1|^{1/3} \left\{ \frac{2}{\pi} + |\tilde{\gamma}_1|^{2/3} \left(\frac{2}{4 - \pi} \right)^{2/3} \left(\frac{2}{\pi} - 1 \right) \right\}^{-1/2}, \\
 \xi &= \mu - \sigma \left(\frac{2}{4 - \pi} \gamma_1 \right)^{1/3}, \\
 \omega &= \sigma \left\{ 1 + \left(\frac{2}{4 - \pi} \gamma_1 \right)^{2/3} \right\}^{1/2}.
 \end{aligned} \tag{4.5}$$

El obtener los estimadores de momentos usando la “parametrización centrada” tiene varias ventajas sobre los estimadores de máxima verosimilitud:

- Son fáciles de calcular.
- Siempre existen.

[Pewsey \(2000\)](#) señala que en muchas situaciones prácticas la inferencia se centra en μ y σ con γ_1 siendo un parámetro de “estorbo”, en estos casos hay poca ganancia al utilizar estimación por máxima verosimilitud. Dicho autor también indica que los estimadores de momentos obtenidos al reparametrizar se desempeñan de manera adecuada para tamaños de muestra pequeños de distribuciones moderadamente asimétricas.

4.2. Estimación Bayesiana

Sea Y_1, \dots, Y_n una muestra aleatoria de $SN_D(\xi, \omega, \gamma)$ y supóngase que se desea obtener estimadores Bayesianos de los parámetros. [Liseo y Loperfido \(2006\)](#) abordan el problema utilizando la distribución a priori no informativa de Jeffreys cuando $\xi = 0$, $\omega = 1$ y γ es desconocido y en el caso general cuando los tres parámetros son desconocidos. Dichos autores centran su inferencia en γ , considerando a ξ, ω parámetros de “estorbo” que son eliminados mediante integración. La distribución a priori de referencia para el caso de los tres parámetros obtenida por estos autores esta dada por:

$$p(\xi, \omega, \gamma) = p(\xi, \omega)p(\gamma) \propto \frac{1}{\omega} g^{\frac{1}{2}}(\gamma),$$

donde:

4.2. Estimación Bayesiana

$$\begin{aligned}
 g(\gamma) &= \frac{i_{11}i_{22}i_{33} + 2i_{12}i_{13}i_{23} - i_{11}i_{23}^2 - i_{33}i_{12}^2 - i_{22}i_{13}^2}{i_{33}i_{22} - i_{23}^2}, \\
 \{i_{jk}\} &= \begin{pmatrix} a_2 & -\gamma a_2/\omega & \left(\frac{b}{(1+\gamma^2)^{3/2}} - \gamma a_1\right)/\omega \\ -\gamma a_2/\omega & (2 + \gamma^2 a_2)/\omega^2 & \left(b\delta\frac{1+2\gamma^2}{1+\gamma^2} + \gamma^2 a_1\right)/\omega^2 \\ \left(\frac{b}{(1+\gamma^2)^{3/2}} - \gamma a_1\right)/\omega & \left(b\delta\frac{1+2\gamma^2}{1+\gamma^2} + \gamma^2 a_1\right)/\omega^2 & (1 + \gamma^2 a_0)/\omega^2 \end{pmatrix}, \\
 b &= \sqrt{\frac{2}{\pi}}, \\
 \delta &= \frac{\gamma}{\sqrt{1 + \gamma^2}}, \\
 a_j &= a_j(\gamma) = \int_{\mathbb{R}} 2z^j \frac{\phi(\gamma z)\phi(z)}{\Phi(\gamma z)} dz, \quad j = 0, 1, 2.
 \end{aligned}$$

Es claro entonces que es complicado trabajar con la distribución a priori de referencia.

De manera similar a [Azzalini \(1985\)](#), [Azzalini y Capitanio \(1999\)](#) y [Pewsey \(2000\)](#) en la estimación en el contexto clásico, aquí se propone utilizar la “parametrización centrada” para obtener estimadores bayesianos para los parámetros de interés.

Aquí se proponen las siguientes distribuciones a priori:

$$\begin{aligned}
 p(\mu) &\propto 1, \\
 p(\sigma) &\propto \frac{1}{\sigma}, \\
 p(\gamma_1) &\propto I_{(-c_1, c_1)}(\gamma_1).
 \end{aligned}$$

Bajo el supuesto de independencia a priori, la distribución a priori conjunta es:

$$p(\mu, \sigma, \gamma_1) \propto \frac{1}{\sigma} I_{(-c_1, c_1)}(\gamma_1). \tag{4.6}$$

La densidad de Y en la “parametrización centrada” puede obtenerse fácilmente utilizando el teorema de cambio de variable. Es decir,

4.2. Estimación Bayesiana

$$\begin{aligned} f_Y(y, \mu, \sigma, \gamma_1) &= f_Z(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| \\ &= 2\phi(g^{-1}(y)) \Phi(\gamma g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| \end{aligned}$$

donde

$$\begin{aligned} g^{-1}(y) &= \mu_Z + \frac{\sqrt{1 - \mu_Z^2}(y - \mu)}{\sigma}, \\ \gamma &= \frac{\mu_Z}{\sqrt{\frac{2}{\pi} - \mu_Z^2}}, \\ \mu_Z &= E(Z) = \text{sgn}(\gamma_1) \sqrt{\frac{(2\gamma_1)^{2/3}}{(4 - \pi)^{2/3} + (2\gamma_1)^{2/3}}}. \end{aligned}$$

Por lo tanto

$$f_y(y, \mu, \sigma, \gamma_1) = 2\phi(g^{-1}(y)) \Phi(\gamma g^{-1}(y)) \frac{\sqrt{1 - \mu_Z^2}}{\sigma} \quad (4.7)$$

Aplicando el teorema de Bayes, de (4.6) y (4.7) se obtiene que la distribución a posteriori de μ , σ y γ_1 es:

$$\begin{aligned} p(\mu, \sigma, \gamma_1 | \mathbf{y}) &\propto \left\{ \prod_{i=1}^n f_{Y_i}(y_i; \mu, \sigma, \gamma_1) \right\} p(\mu, \sigma, \gamma_1) \\ &= \left\{ \prod_{i=1}^n f_{Y_i}(y_i; \mu, \sigma, \gamma_1) \right\} \frac{1}{\sigma} I_{(-c_1, c_1)}(\gamma_1). \end{aligned} \quad (4.8)$$

Ahora el problema es obtener las distribuciones marginales de interés $p(\mu | \mathbf{y})$, $p(\sigma | \mathbf{y})$, $p(\gamma_1 | \mathbf{y})$. El problema no tiene solución analítica, por lo cual hay que utilizar alguna de las técnicas de Cadenas de Markov Monte Carlo:

- 1) Muestreador Gibbs
- 2) Metropolis
- 3) "t-walk"

4.2. Estimación Bayesiana

En el caso del muestreador de Gibbs, es necesario contar con las condicionales completas $p(\mu|\sigma, \gamma_1, \mathbf{y})$, $p(\sigma|\mu, \gamma_1, \mathbf{y})$, $p(\gamma_1|\mu, \sigma, \mathbf{y})$ para implementarlo, pero no se dispone de la forma cerrada de las mismas, por lo tanto no es posible aplicar esta técnica de manera directa. [Gilks y Wild \(1992\)](#) desarrollaron un algoritmo que permite muestrear de cualquier distribución absolutamente continua, con función de densidad log-cóncava, y que ha sido aplicado exitosamente en situaciones como la descrita antes, ver por ejemplo [Dellaportas y Smith \(1993\)](#). Por lo tanto si se quisiera aplicar este algoritmo tendría que mostrarse primero que (4.8) es log-cóncava.

Para aplicar el algoritmo de Metropolis hay que seleccionar primero la distribución generadora de candidatos, en este paso hay que ser muy cuidadosos, ya que una elección inadecuada de dicha distribución puede provocar que el algoritmo de Metropolis tenga un desempeño muy pobre debido a la alta tasa de rechazos.

[Pewsey \(2000\)](#) obtuvo resultados asintóticos para los estimadores de momentos obtenidos al utilizar la “parametrización centrada”. Si $Y \sim SN_C(\mu, \sigma, \gamma_1)$,

$$\begin{aligned}\beta_2 &= 3 + \tau^4(\pi - 3), \\ \beta_3 &= 10\gamma_1 + \tau^5(3\pi^2 - 40\pi + 96)/4, \\ \beta_4 &= 15 \{1 + \tau^4(2\pi - 6)\} - \tau^6(9\pi^2 - 80\pi + 160)/2,\end{aligned}$$

con $\tau = (2/(4 - \pi)\gamma_1)^{1/3}$. Utilizando el método delta ([Pewsey, 2000](#)) se obtiene que:

$$\begin{aligned}\text{Var}(\tilde{\mu}) &= \sigma^2/n, \\ \text{Var}(\tilde{\sigma}) &= \sigma^2(\beta_2 - 1)/4n + O(n^{-3/2}), \\ \text{Var}(\tilde{\gamma}_1) &= \{9 - 6\beta_2 - 3\gamma_1\beta_3 + \beta_4 + \gamma_1^2(35 + 9\beta_2)/4\} /n + O(n^{-3/2}), \\ \text{Cov}(\tilde{\mu}, \tilde{\sigma}) &= \sigma^2\gamma_1/2n + O(n^{-3/2}), \\ \text{Cov}(\tilde{\mu}, \tilde{\gamma}_1) &= \sigma(\beta_2 - 3 - 3\gamma_1^2/2)/n + O(n^{-3/2}), \\ \text{Cov}(\tilde{\sigma}, \tilde{\gamma}_1) &= \sigma \{2\beta_3 - \gamma_1(5 + 3\beta_2)\} /4n + O(n^{-3/2}),\end{aligned}\tag{4.9}$$

donde $\tilde{\mu}$, $\tilde{\sigma}$ y $\tilde{\gamma}_1$ son los estimadores de momentos de μ , σ y γ_1 respectivamente. De (4.9), la distribución asintótica conjunta de $\tilde{\mu}$, $\tilde{\sigma}$ y $\tilde{\gamma}_1$ es normal trivariada. Por lo tanto se pueden aplicar estos resultados para seleccionar la distribución normal multivariada como la generadora de candidatos en el algoritmo de Metropolis.

Entonces para arrancar el algoritmo de Metropolis se usa la distribución normal trivariada $N_3(\boldsymbol{\theta}^{(t-1)}, \tilde{\boldsymbol{\Sigma}})$ donde

4.2. Estimación Bayesiana

$$\tilde{\Sigma} = \begin{pmatrix} \widetilde{Var}(\tilde{\mu}) & \widetilde{Cov}(\tilde{\mu}, \tilde{\sigma}) & \widetilde{Cov}(\tilde{\mu}, \tilde{\gamma}_1) \\ \widetilde{Cov}(\tilde{\mu}, \tilde{\sigma}) & \widetilde{Var}(\tilde{\sigma}) & \widetilde{Cov}(\tilde{\sigma}, \tilde{\gamma}_1) \\ \widetilde{Cov}(\tilde{\mu}, \tilde{\gamma}_1) & \widetilde{Cov}(\tilde{\sigma}, \tilde{\gamma}_1) & \widetilde{Var}(\tilde{\gamma}_1) \end{pmatrix}, \quad (4.10)$$

y $\boldsymbol{\theta} = (\mu, \sigma, \gamma_1)'$ se obtiene de los estimadores de momentos utilizando la “parametrización centrada”. Es importante notar que $\tilde{\Sigma}$ dada en la ecuación (4.10) puede ajustarse con la finalidad de imitar la distribución a posteriori (Carlin y Louis, 2000). En el apéndice A se muestran rutinas en el paquete R para obtener las distribuciones a posteriori de μ , σ y γ_1 . Dichas rutinas están diseñadas para actualizar la matriz $\tilde{\Sigma}$ después de que se han obtenido s muestras Monte Carlo de los parámetros de interés, dicha matriz se calcula en la forma usual:

$$\tilde{\Sigma} = \frac{1}{s} \sum_{j=1}^s (\boldsymbol{\theta}_j - \bar{\boldsymbol{\theta}}) (\boldsymbol{\theta}_j - \bar{\boldsymbol{\theta}})', \quad (4.11)$$

donde j indexa las muestras Monte Carlo.

Con la finalidad de verificar si hay convergencia en el algoritmo de Metropolis se puede utilizar la prueba de convergencia de Gelman y Rubin (1992) implementada en la biblioteca de funciones boa de Smith (2007) o bien la biblioteca coda de Plummer *et al.* (2007), las cuales pueden utilizarse sin mayores problemas en el paquete R.

En la siguiente sección se muestran ejemplos de aplicación de la estimación Bayesiana. La idea es mostrar que dicha técnica es una alternativa viable al método de máxima verosimilitud y que trabaja de manera adecuada en casos en los que el método de máxima verosimilitud falla.

4.2.1. Ejemplo 1: Datos de Azzalini

El siguiente conjunto de datos es una muestra aleatoria de tamaño $n = 50$ de $SN_D(0, 1, 5)$. El estimador de máxima verosimilitud de γ no existe.

Tabla 4.1: Datos de estimación en la frontera de Azzalini.

0.1169	1.7311	0.0144	0.4881	2.3877	-0.0853	0.0522	0.7226	0.8718	3.0415
0.6363	1.3590	1.5958	0.4567	0.9885	1.7001	1.0380	1.0195	0.3528	1.4249
0.3170	1.3276	-0.1032	1.1568	0.0699	1.6802	0.2470	0.4147	1.6882	0.7256
1.3568	1.1091	0.0500	2.2886	1.4985	2.7261	1.8443	-0.0687	0.9441	0.6872
0.5258	0.4743	0.4240	0.7349	0.4428	0.1880	0.4642	0.2786	0.2742	0.5678

4.2. Estimación Bayesiana

En la Figura 4.1 se muestra la verosimilitud perfil relativa para los datos de Azzalini, donde se observa que conforme n crece la verosimilitud en γ también crece.

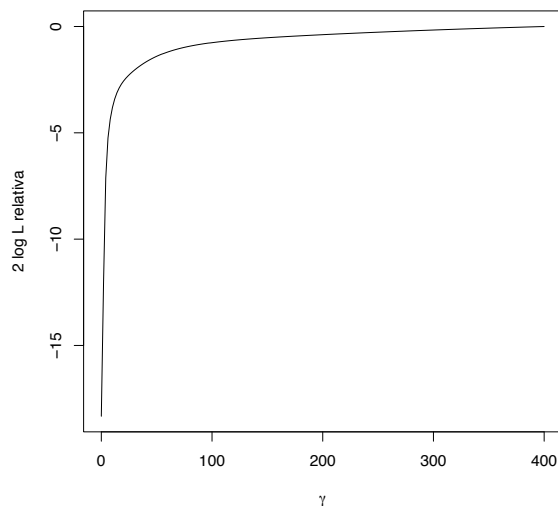


Figura 4.1: Verosimilitud perfil relativa para los datos de Azzalini.

Empleando las rutinas del apéndice A, se estimaron las densidades marginales a posteriori de μ , σ y γ_1 a partir de 3 cadenas de Markov Monte Carlo con 30,000 muestras cada una, desechando las primeras 25,000 que fueron utilizadas como calentamiento. La tasa de aceptación global del algoritmo de Metropolis obtenida fue alrededor del 30 %.

En la Figura 4.2 se muestra la traza y la densidad estimada para las últimas 5,000 muestras Monte Carlo para las tres cadenas, donde se observa que las cadenas se mezclan muy bien.

Para verificar la convergencia se emplea la prueba de Gelman y Rubin. En la Figura 4.3 se observa la manera en la que va cambiando la estadística de prueba conforme el número de iteraciones aumenta. En la Tabla 4.2 se muestra el valor de la estadística de prueba para las últimas 5,000 iteraciones. De la Figura 4.3 y la Tabla 4.2 se concluye que no hay problemas con convergencia y por lo tanto las muestras Monte Carlo obtenidas con el algoritmo de Metropolis pueden utilizarse para hacer inferencia.

En la Figura 4.4 se muestra la densidad a posteriori de ξ , ω y γ obtenidas al aplicar (4.5) a las cadenas generadas previamente.

En la Tabla 4.3 se presenta un resumen para las cadenas de Markov Monte Carlo transformadas, a partir del cual se pueden obtener estimadores puntuales para ξ , ω y γ . Si se usa la moda a posteriori entonces $\check{\xi} = -0.0391(0.1315)$, $\check{\omega} = 1.1625(0.1461)$, $\check{\gamma} = 3.7584(9.8517)$.

Es importante notar que los estimadores puntuales de los parámetros de localidad

4.2. Estimación Bayesiana

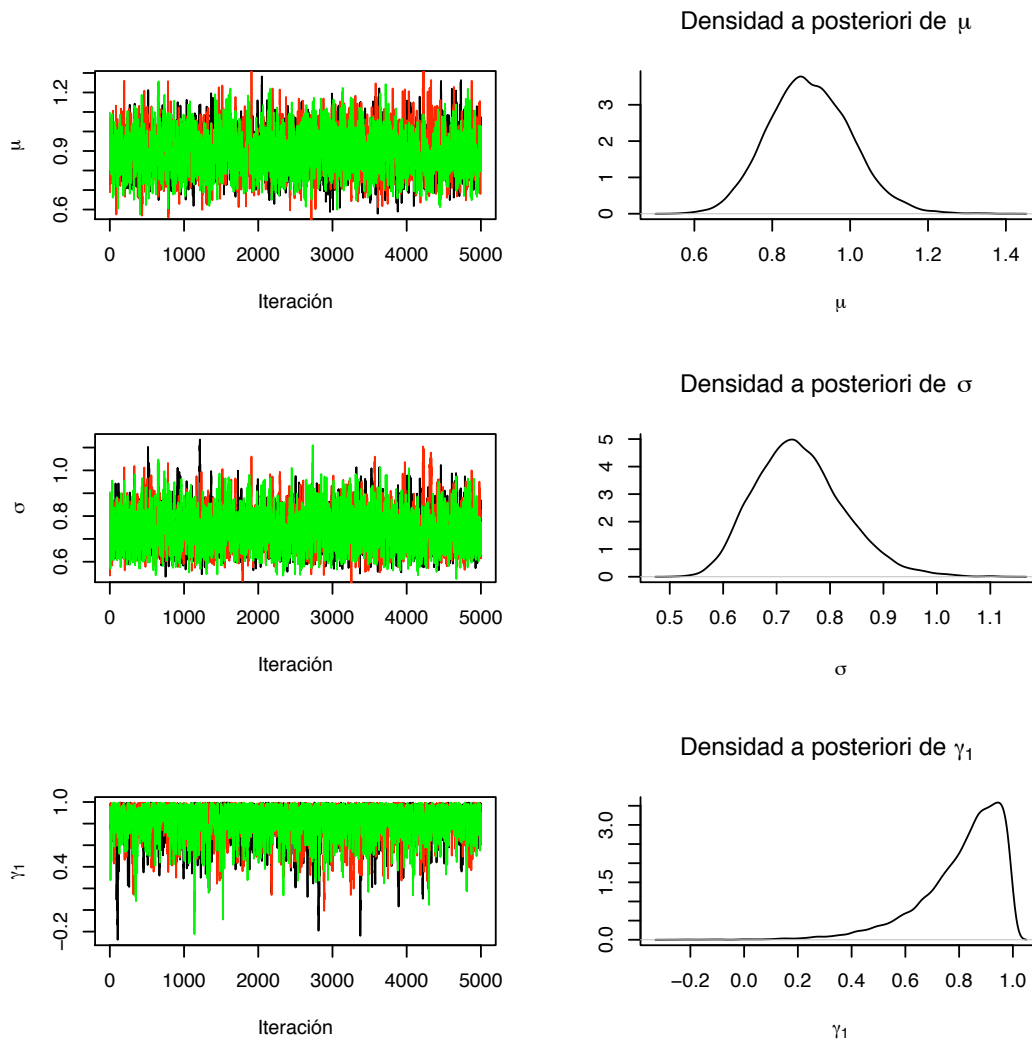


Figura 4.2: Densidades a posteriori para los datos de Azzalini usando la parametrización centrada.

y escala son muy parecidos a los valores verdaderos usados por Azzalini en la simulación, y además el estimador puntual obtenido de γ es finito. Los estimadores puntuales son parecidos a los obtenidos con el método propuesto por Sartori (2006), $\hat{\xi} = -0.0689(0.0860)$, $\hat{\omega} = 1.1978(0.1378)$, $\hat{\gamma} = 9.1474(5.7747)$.

En la Figura 4.5 se muestra la densidad a posteriori de γ , obtenida por Liseo y Loperfido (2006), la cual no tiene media finita, pero los autores mencionados sugieren utilizar la mediana a posteriori como estimador puntual, en este caso 2.14 parece un valor adecuado.

En la Figura 4.6 se muestran los datos de estimación en la frontera de Azzalini en el eje horizontal junto con el estimador no paramétrico y los modelos ajustados obtenidos utilizando el método de estimación de Sartori, Liseo y el método propuesto.

4.2. Estimación Bayesiana

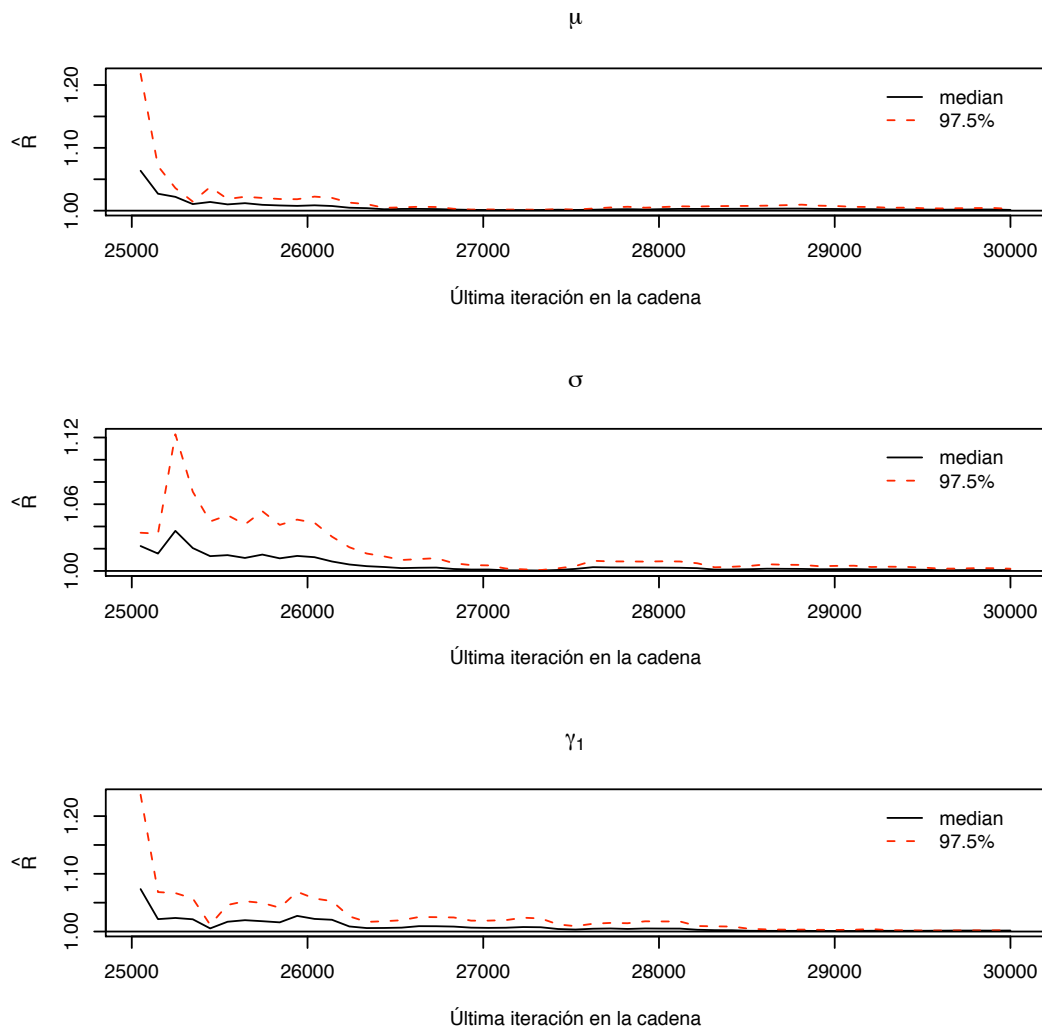


Figura 4.3: Diagnóstico de convergencia, con las a priori propuestas para los datos de Azzalini.

Tabla 4.2: Prueba de convergencia de Gelman y Rubin para los datos de Azzalini con las a priori propuestas.

	Estimación puntual de R	Cuantil 97.5 %
μ	1.00	1.00
σ	1.00	1.00
γ_1	1.00	1.00

Tabla 4.3: Estadísticas resumen para cadenas transformadas con las a priori propuestas para los datos de Azzalini.

	Media	Desviación estándar	Cuantil 2.5 %	Mediana	Moda	Cuantil 97.5 %
ξ	-0.0112	0.1315	-0.2141	-0.02164	-0.0391	0.2427
ω	1.1701	0.1461	0.9146	1.1600	1.1625	1.4852
γ	6.6639	9.8517	1.7459	4.8446	3.7584	21.7363

4.2. Estimación Bayesiana

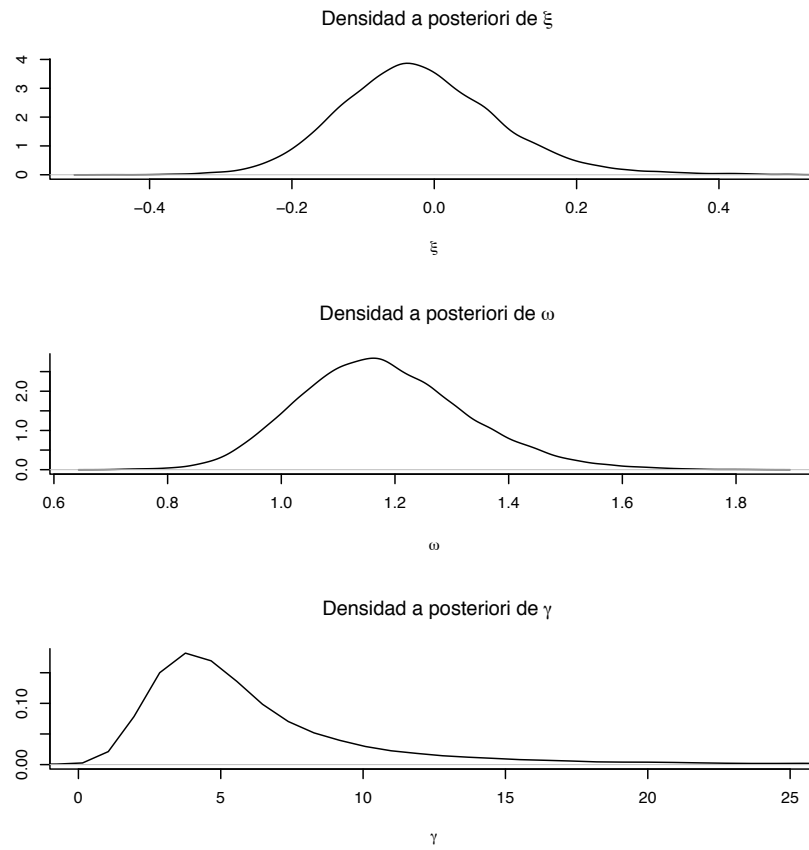


Figura 4.4: Densidades a posteriori para los datos de Azzalini usando la parametrización directa.

4.2. Estimación Bayesiana

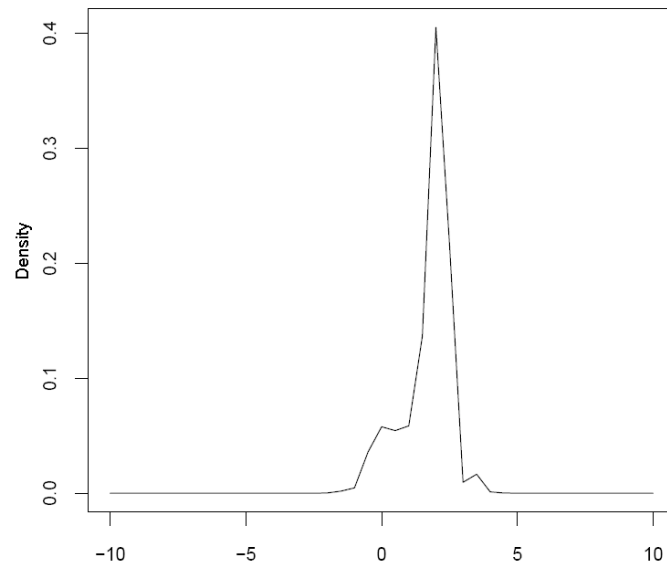


Figura 4.5: Densidad marginal a posteriori para γ , reimpressa del artículo de [Liseo y Loperfido \(2006\)](#).

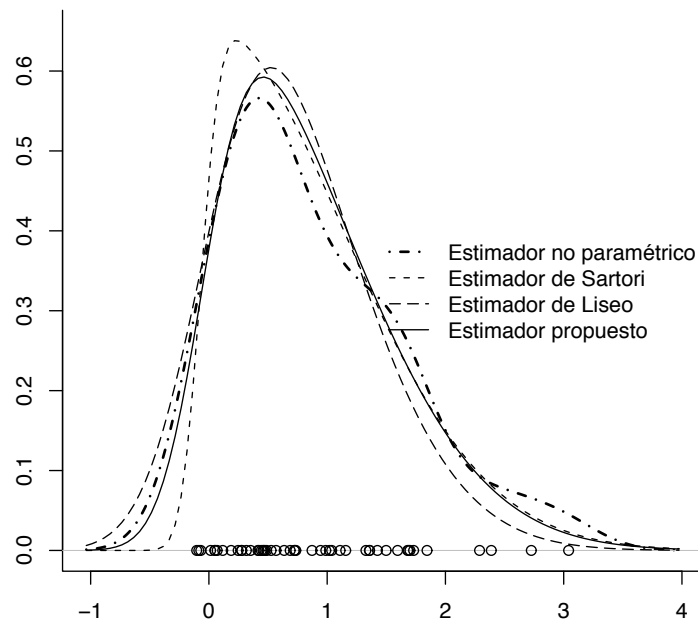


Figura 4.6: Datos de estimación en la frontera y modelos ajustados.

4.2. Estimación Bayesiana

4.2.2. Ejemplo 2: Datos de Sartori

Los datos de la Tabla 4.4 son una muestra aleatoria de tamaño 20 de $SN_D(0, 1, 10)$ que aparecen publicados en el artículo de Sartori (2006). En este caso el estimador de máxima verosimilitud convencional y el obtenido por el método de Sartori (2006) para γ son ambos finitos.

Tabla 4.4: Datos de Sartori.

0.195	0.847	0.726	-0.139	1.788	0.570	2.069	0.452	0.868	1.199
0.894	0.887	1.258	0.918	0.496	0.183	0.119	1.207	0.446	2.579

Empleando las rutinas del apéndice A, se estimaron las densidades marginales a posteriori de μ , σ y γ_1 utilizando 3 cadenas Monte Carlo con 30,000 muestras cada una desechando las primeras 25,000 que fueron utilizadas solo como calentamiento. La tasa de aceptación global del algoritmo de Metropolis obtenida fue alrededor del 35 %.

En la Figura 4.7 se muestran la traza y la densidad a posteriori estimada para las últimas 5,000 muestras Monte Carlo para las tres cadenas, donde se observa que las cadenas se mezclan muy bien. Para verificar la convergencia se emplea la prueba de Gelman y Rubin.

En la Tabla 4.5 se muestra el valor de la estadística de prueba para las últimas 5,000 iteraciones, de donde se concluye que no hay problemas con la convergencia y por lo tanto las muestras Monte Carlo obtenidas con el algoritmo de Metropolis pueden utilizarse para hacer inferencias.

En la Figura 4.8 se muestra la densidad a posteriori de μ , ω y γ obtenidas al aplicar (4.5) a las cadenas generadas previamente.

En la Tabla 4.6 se presenta un resumen de las Cadenas de Markov Monte Carlo transformadas, a partir del cual se pueden obtener estimadores puntuales para μ , ω y γ , usando una de las modas a posteriori, entonces, $\tilde{\xi} = 0.1488(0.4693)$, $\tilde{\omega} = 0.9466(0.2218)$ y $\tilde{\gamma} = 1.8349(4.1089)$.

Hay que notar que los estimadores puntuales de los parámetros de localidad y escala y forma son muy parecidos a los obtenidos por el método de Sartori $\hat{\xi} = 0.1612(0.2376)$, $\hat{\omega} = 0.9592(0.2341)$ y $\hat{\gamma} = 2.7727(2.2300)$.

4.2. Estimación Bayesiana

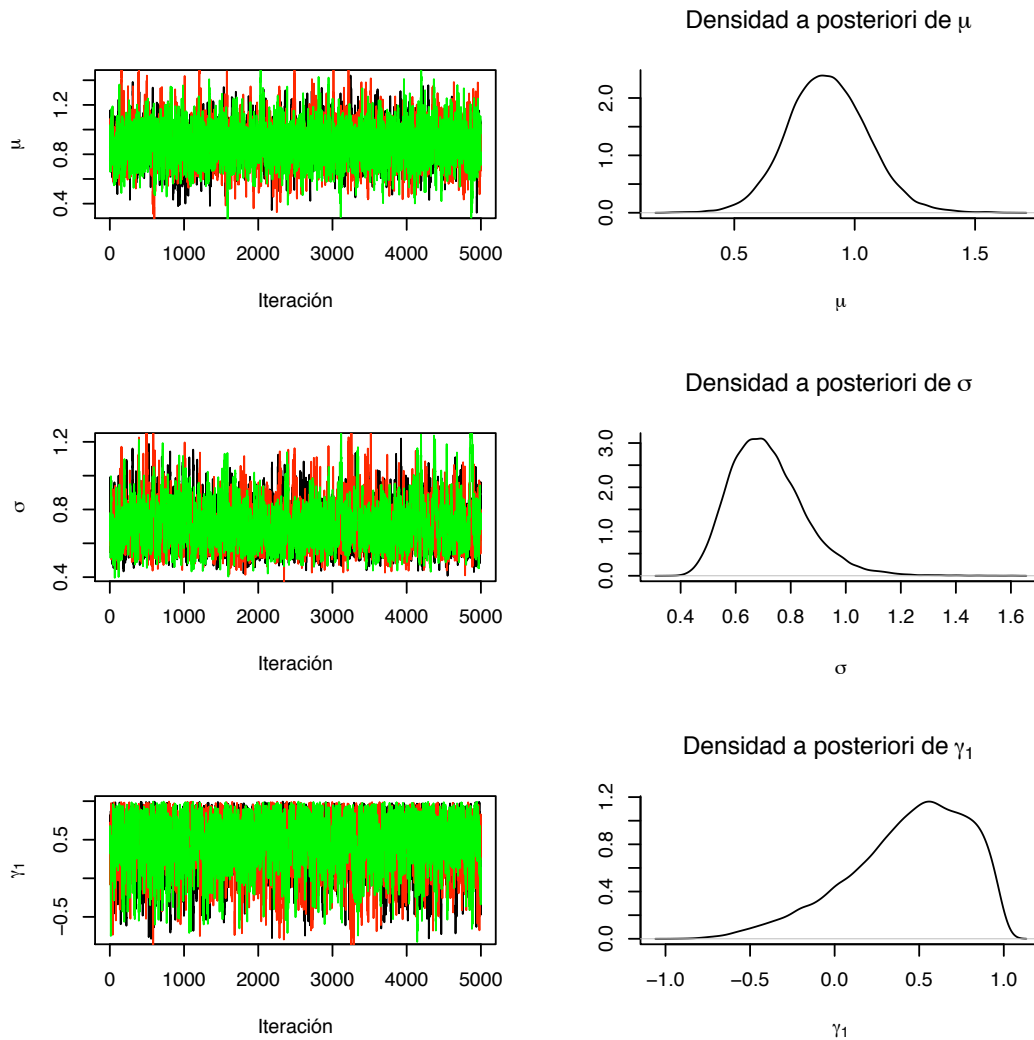


Figura 4.7: Densidades a posteriori para los datos de Sartori usando la parametrización centrada.

Tabla 4.5: Prueba de convergencia de [Gelman y Rubin](#) para los datos de Sartori con las a priori propuestas.

	Estimación puntual de R	Cuantil 97.5 %
μ	1.00	1.00
σ	1.00	1.00
γ_1	1.00	1.00

4.2. Estimación Bayesiana

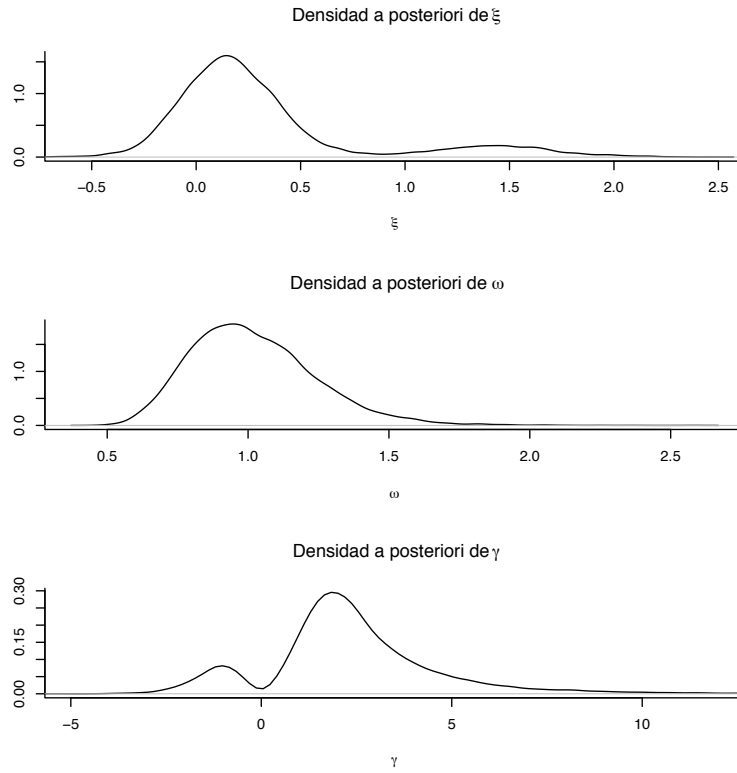


Figura 4.8: Densidades a posteriori para los datos de Sartori usando la parametrización directa.

Tabla 4.6: Estadísticas resumen para cadenas transformadas con las a priori propuestas para los datos de Sartori.

	Media	Desviación estándar	Cuantil 2.5 %	Mediana	Moda*	Cuantil 97.5 %
ξ	0.3095	0.4693	-0.2570	0.1900	0.1488	1.627
ω	1.0143	0.2218	0.6575	0.9914	0.9466	1.522
γ	2.5564	4.1089	-1.6591	2.1617	1.8349	8.888

4.2.3. Ejemplo 3: Regresión lineal simple

Azzalini y Capitanio (1999) estudian el modelo de regresión donde los errores siguen una distribución normal asimétrica. Rusell y González (2002) retoman el problema considerando el caso univariado, donde se tiene una variable dependiente Y , una variable explicatoria W no aleatoria y los errores independientes e idénticamente distribuidos $SN_D(0, \omega, \gamma)$. La forma del modelo es:

$$Y_i = \beta_0 + \beta_1 W_i + \epsilon_i, \quad \epsilon_i \text{ i.i.d. } SN_D(0, \omega, \gamma), \quad i = 1, \dots, n \quad (4.12)$$

de donde se sigue que Y_i i.i.d. $SN_D(\beta_0 + \beta_1 W_i, \omega, \gamma)$.

La estimación de parámetros es realizada utilizando el método de máxima verosimilitud. El problema es equivalente a tener Y_i i.i.d. $SN_C(\beta_0^* + \beta_1^* W_i, \sigma, \gamma_1)$. Una vez hecha la estimación usando la parametrización centrada se pueden recuperar los parámetros en la parametrización directa usando (4.5) para ω y γ . Además:

$$\beta_0 = \beta_0^* - \sigma \frac{\mu_Z}{\sqrt{1 - \mu_Z^2}}, \quad (4.13)$$

para el caso de β_0 y $\beta_1 = \beta_1^*$ (Azzalini y Capitanio, 1999).

Desde el punto de vista Bayesiano, hay que especificar distribuciones a priori para β_0^* , β_1^* , σ y γ_1 , para lo cual se proponen las siguientes:

$$\begin{aligned} p(\beta_0^*, \beta_1^*, \sigma) &\propto \frac{1}{\sigma}, \\ p(\gamma_1) &\propto I_{(-c_1, c_1)}(\gamma_1). \end{aligned}$$

Bajo el supuesto de independencia la distribución a priori conjunta para β_0^* , β_1^* , σ y γ_1 está dada por:

$$p(\beta_0^*, \beta_1^*, \sigma, \gamma_1) \propto \frac{1}{\sigma} I_{(-c_1, c_1)}(\gamma_1).$$

Aplicando el Teorema de Bayes, la distribución a posteriori conjunta de β_0^* , β_1^* , σ y γ_1 es:

4.2. Estimación Bayesiana

$$\begin{aligned}
 p(\beta_0^*, \beta_1^*, \sigma, \gamma_1 | \mathbf{y}) &\propto \left\{ \prod_{i=1}^n f_{Y_i}(y_i; \beta_0^*, \beta_1^*, \sigma, \gamma_1) \right\} p(\beta_0^*, \beta_1^*, \sigma, \gamma_1) \\
 &= \left\{ \prod_{i=1}^n f_{Y_i}(y_i; \beta_0^*, \beta_1^*, \sigma, \gamma_1) \right\} \frac{1}{\sigma} I_{(-c_1, c_1)}(\gamma_1). \quad (4.14)
 \end{aligned}$$

Empleando el algoritmo de Metropolis se obtienen muestras Monte Carlo de (4.14) de manera totalmente similar a los ejemplos previos. Para ilustrar el método y ver cómo funciona, se consideran datos de mediciones biomédicas de atletas del Instituto Australiano del Deporte, estudiados por [Azzalini y Capitanio \(1999\)](#), y que aparecen en el libro de [Cook y Weisberg \(1994\)](#). Se toma como variable dependiente a bmi (índice de masa corporal) y como variable explicatoria a lbm (masa magra del cuerpo), se obtienen estimadores Bayesianos de los parámetros para el modelo (4.12). En la Figura 4.9 se muestra la densidad a posteriori estimada de β_0^* , β_1^* , σ y γ_1 , generadas a partir de tres cadenas con 20,000 muestras Monte Carlo cada una, desechando las primeras 15,000. La tasa de aceptación global del algoritmo fue de alrededor del 30 %.

De la Tabla 4.7 se concluye que no hay problemas de convergencia, y por lo tanto las muestras Monte Carlo obtenidas pueden utilizarse para hacer inferencias.

En la Figura 4.10 se muestran la densidades a posteriori de β_0 , β_1 , ω y γ obtenidas al aplicar (4.5) a las cadenas generadas previamente.

En la Tabla 4.8 se muestra un resumen de la cadena de Markov Monte Carlo transformadas, a partir de la cual se pueden obtener estimadores puntuales para β_0 , β_1 , ω y γ . Si se utiliza la moda a posteriori, se obtiene $\check{\beta}_0 = 11.3755(0.5888)$, $\check{\beta}_1 = 0.1438(0.0087)$, $\check{\omega} = 3.0657(0.2138)$ y $\check{\gamma} = 3.5490(1.0042)$. Los estimadores de máxima verosimilitud sí se pueden calcular en este caso, obteniéndose $\hat{\beta}_0 = 11.2781(0.5811)$, $\hat{\beta}_1 = 0.1432(0.0085)$, $\hat{\omega} = 3.1139(0.2083)$, $\hat{\gamma} = 4.1525(1.0503)$, note que los estimadores obtenidos utilizando el algoritmo de Metropolis y los de máxima verosimilitud son muy parecidos.

Tabla 4.7: Prueba de convergencia de [Gelman y Rubin](#) para el problema de regresión con las a priors propuestas.

	Estimación puntual de R	Cuantil 97.5 %
β_0^*	1.00	1.00
β_1^*	1.00	1.00
σ	1.00	1.01
γ_1	1.00	1.00

En la Figura 4.11 se muestra una gráfica de residuales para el modelo de regresión

4.2. Estimación Bayesiana

Tabla 4.8: Estadísticas resumen para cadenas transformadas con las a priori propuestas para el modelo de regresión.

	Media	Desviación estándar	Cuantil 2.5 %	Mediana	Moda	Cuantil 97.5 %
β_0	11.2727	0.5888	10.1024	11.2772	11.3755	12.4031
β_1	0.1437	0.0087	0.1267	0.1436	0.1438	0.1612
ω	3.1099	0.2138	2.7076	3.1046	3.0657	3.5425
γ	3.9601	1.0042	2.3687	3.8267	3.5490	6.3220

con lo cual se puede verificar de manera informal el supuesto distribucional de los errores.

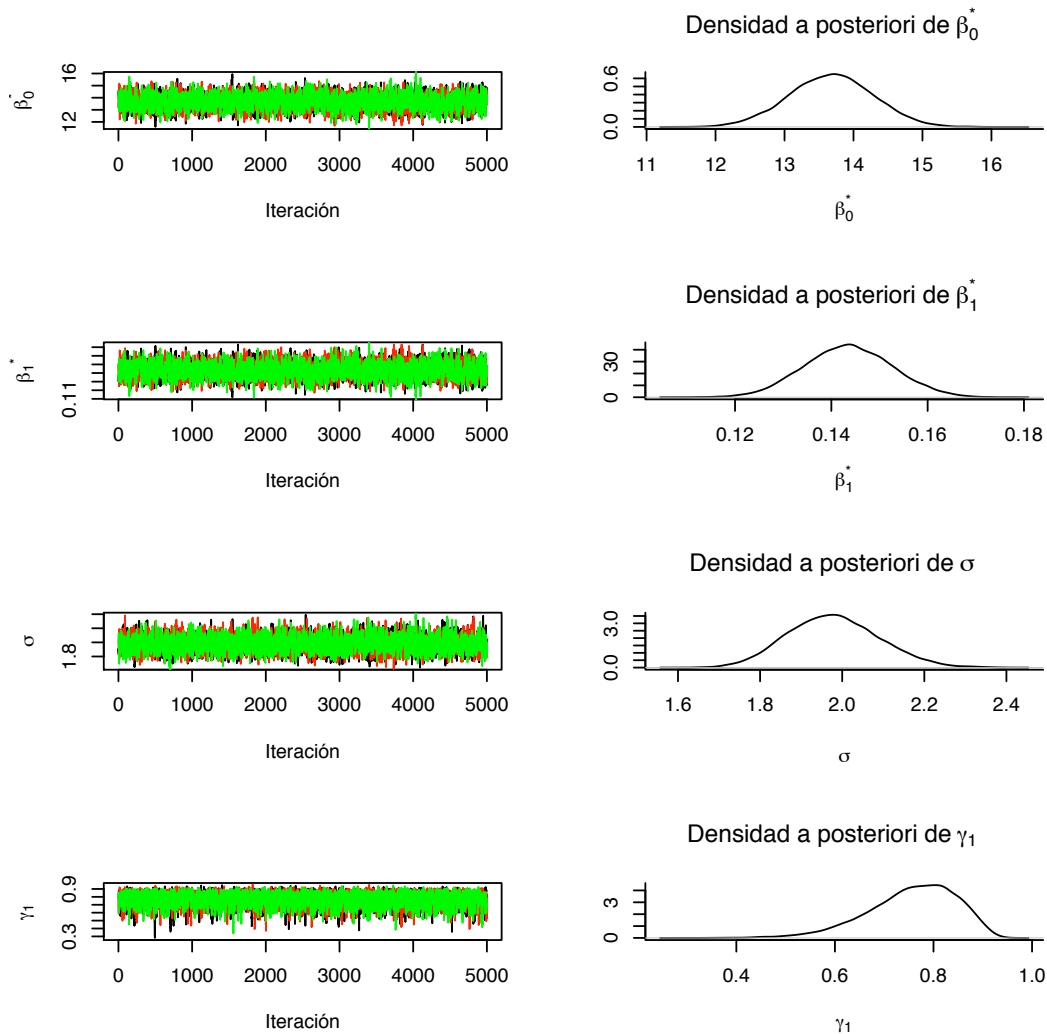


Figura 4.9: Densidades a posteriori con a priori propuestas para el modelo de regresión usando la parametrización centrada.

Las distribuciones a posteriori también se pueden obtener utilizando el algoritmo “t-walk”. En el apéndice A se muestran rutinas necesarias para obtener la distribuciones

4.2. Estimación Bayesiana

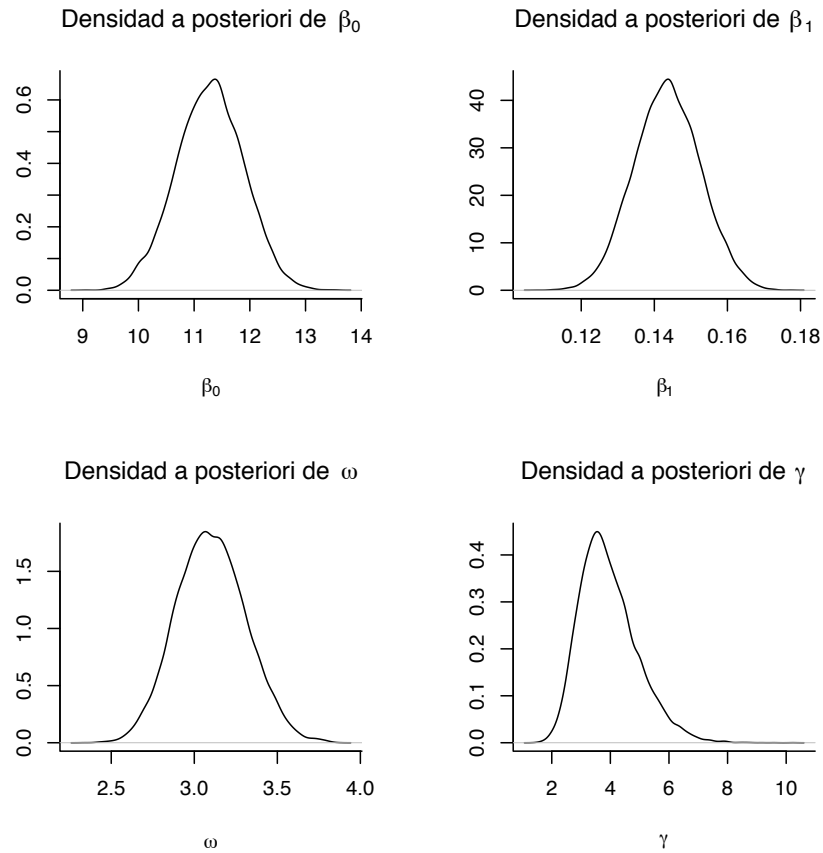


Figura 4.10: Densidades a posteriori con a priori propuestas para el modelo de regresión usando la parametrización centrada.

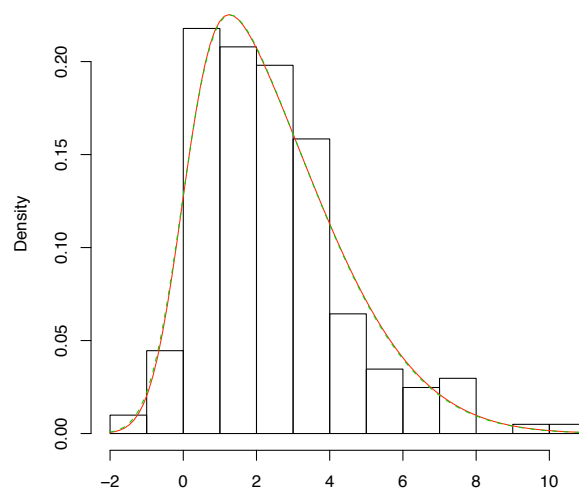


Figura 4.11: Residuales para el modelo de regresión.

a posteriori $p(\mu|\mathbf{y})$, $p(\sigma|\mathbf{y})$ y $p(\gamma_1|\mathbf{y})$. Los resultados obtenidos son similares a los del algoritmo de Metropolis. La tasa de aceptación es de alrededor del 20% para los

4.2. Estimación Bayesiana

ejemplos presentados.

Es muy importante notar también que al ser la distribución a priori propia para γ_1 esto asegura que la distribución a posteriori, si existe, deberá ser propia . Esto es interesante desde el punto de vista práctico ya que la distribución a posteriori puede emplearse en el contexto de pruebas de hipótesis para probar $H_0 : \gamma_1 = 0$ vs $H_1 : \gamma_1 \neq 0$ u otras empleando los factores de Bayes.

Capítulo 5

Pruebas de bondad de ajuste

Las propiedades estadísticas de la distribución normal asimétrica han sido estudiadas ampliamente. El problema de probar si una muestra aleatoria Y_1, \dots, Y_n viene de la distribución normal asimétrica está poco estudiado. [Gupta y Chen \(2001\)](#) abordan este problema considerando el caso de la distribución en su forma estándar ($\xi = 0$, $\omega = 1$) con el parámetro $\gamma = \gamma_0$ conocido. Si $f(y)$ es la función de densidad de las Y_i 's, [Gupta y Chen \(2001\)](#) prueban:

$$\begin{aligned} H_0 : f_0(y; 0, 1, \gamma_0) &= 2\phi(y) \Phi(\gamma_0 y) I_{(-\infty, \infty)}(y) \\ &\text{vs} \\ H_1 : f(y; \cdot) &\neq f_0(y; 0, 1, \gamma_0). \end{aligned}$$

Para probar estas hipótesis utilizan dos pruebas bien conocidas: la χ^2 de Pearson y la de Kolmogorov-Smirnov.

[Mateu et al. \(2007\)](#) consideran el problema de la prueba de bondad de ajuste con los parámetros ξ , ω desconocidos, estudian varias pruebas y presentan Tablas de valores críticos para varios valores del parámetro γ . [Meintanis \(2007\)](#) aborda también el problema utilizando como estadística de prueba la función generatriz de momentos empírica.

En este trabajo se proponen tres pruebas de bondad de ajuste para la distribución normal asimétrica que son independientes del parámetro γ utilizando las propiedades de esta distribución.

Sea \mathcal{F}_{SN} la clase de densidades normales asimétricas con parámetros de localización, escala y forma desconocidos. La hipótesis de interés es:

5.1. Pruebas propuestas

$$H_0 : f_Y \in \mathcal{F}_{SN} \text{ vs } H_1 : f_Y \notin \mathcal{F}_{SN} \quad (5.1)$$

5.1. Pruebas propuestas

5.1.1. Kolmogorov-Smirnov

Distribución χ_1^2 , estimadores de máxima verosimilitud

De la ecuación (3.2) se despeja Z y se obtiene:

$$\frac{Y - \xi}{\omega} = Z \sim SN(\gamma). \quad (5.2)$$

Utilizando la Propiedad 4) de la distribución normal asimétrica se obtiene que:

$$\left(\frac{Y - \xi}{\omega} \right)^2 = Z^2 \sim \chi_1^2. \quad (5.3)$$

Es claro que se ha eliminado el parámetro γ del problema, pero aún falta conocer el valor de los parámetros ξ y ω . Si dichos parámetros se sustituyen por estimadores consistentes, digamos $\hat{\xi}$ y $\hat{\omega}$ entonces para tamaños de muestra grandes:

$$W := \left(\frac{Y - \hat{\xi}}{\hat{\omega}} \right)^2 \sim \chi_1^2. \quad (5.4)$$

Dadas la muestra aleatoria Y_1, \dots, Y_n el procedimiento de prueba consiste en calcular $W_i, i = 1, \dots, n$ utilizando la ecuación (5.4), luego verificar si las W_i 's vienen de una distribución χ_1^2 con alguno de los procedimientos de prueba ya conocidos.

Para poder construir la prueba se requieren buenos estimadores de los parámetros ξ y ω .

En el caso de la prueba de bondad de ajuste que se plantea solo interesan los estimadores de los parámetros ξ y ω siendo en este caso γ un parámetro de “estorbo”. Para eliminarlo podría utilizarse la verosimilitud integrada (Berger *et al.*, 1999). En el caso de la distribución normal asimétrica la verosimilitud integrada uniforme está dada

5.1. Pruebas propuestas

por:

$$\begin{aligned}
 L^U(\xi, \omega) &= \int L(\xi, \omega, \gamma) d\gamma \\
 &= \int \prod_{i=1}^n f_{Y_i}(y_i; \xi, \omega, \gamma) d\gamma \\
 &= \left(\frac{2}{\omega}\right)^n \prod_{i=1}^n \phi\left(\frac{y_i - \xi}{\omega}\right) \int \prod_{i=1}^n \Phi\left[\gamma\left(\frac{y_i - \xi}{\omega}\right)\right] d\gamma. \quad (5.5)
 \end{aligned}$$

De (5.5) se observa que la integral diverge, por lo tanto $L^U(\xi, \omega)$ no puede maximizarse para obtener los estimadores de los parámetros de interés, por lo tanto para estimar los parámetros se propone utilizar los estimadores de Sartori (2006) o bien los de máxima verosimilitud. En este caso como solo se requiere estimadores de parámetros de localidad y escala se puede utilizar cualquiera de los dos métodos de estimación.

Finalmente para probar si las W'_i s vienen de una distribución χ_1^2 se emplea la prueba de Kolmogorov-Smirnov, sin embargo como se están estimando parámetros entonces no se pueden utilizar los valores críticos de las Tablas estándar de esta prueba.

Stephens y D'Agostino (1986) señalan que la distribución muestral de estadísticas basadas en la función de distribución empírica (como la de Kolmogorov-Smirnov) cuando se tienen parámetros de localidad y escala desconocidos no depende del valor verdadero de dichos parámetros, siempre y cuando sean estimados usando estimadores de localidad y escala equivariantes. En el caso de la distribución normal asimétrica se tiene un parámetro adicional de forma, pero como la prueba estará basada en las W'_i s cuya distribución es independiente de dicho parámetro, se espera que la distribución de la estadística de prueba sea independiente de todos los parámetros. Para obtener los valores críticos para la prueba se emplea el procedimiento descrito a continuación:

- 1) Fijar n , $\xi = 0$, $\omega = 1$ y γ .
- 2) Simular una muestra tamaño n de $SN_D(\xi, \omega, \gamma)$.
- 3) Calcular los estimadores de máxima verosimilitud de los parámetros ξ y ω usando el procedimiento propuesto por Sartori.
- 4) Calcular w_i , $i = 1, \dots, n$ utilizando la ecuación (5.4).
- 5) Ordenar las w_i en forma ascendente.
- 6) Calcular $w_i^* = F^*(w_i)$, $i = 1, \dots, n$ donde $F^*(w_i)$ es la función de distribución de $W \sim \chi_1^2$.

5.1. Pruebas propuestas

7) Calcular

$$\begin{aligned} D_n^+ &:= \max_{1 \leq i \leq n} \left\{ \frac{i}{n} - w_i^* \right\}, \\ D_n^- &:= \max_{1 \leq i \leq n} \left\{ w_i^* - \frac{i-1}{n} \right\}, \\ D_n &:= \sqrt{n} \max_{1 \leq i \leq n} \{ D_n^+, D_n^- \}. \end{aligned}$$

8) Repetir B veces los pasos 2 a 7.

Al terminar el proceso de simulación se tendrán B realizaciones de D_n .

Ya se ha mencionado antes que se espera que la distribución de la estadística de prueba sea independiente del parámetro γ , sin embargo cuando se obtiene la densidad empírica de dicha estadística para diferentes valores de este parámetro se puede observar que existe una dependencia del mismo.

En la Figura 5.1 se observa la densidad empírica de la estadística de prueba D_n obtenida mediante simulación con $B = 10,000$ muestras de tamaño $n = 150$ y $\gamma = \{0, 10\}$.

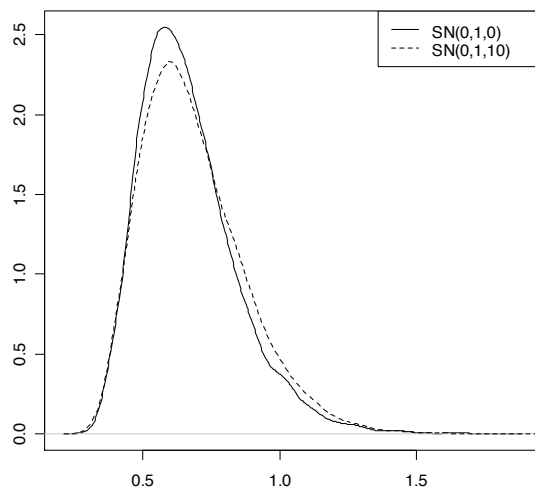


Figura 5.1: Distribución empírica de la estadística D_n para $n = 150$ obtenida mediante simulación con $B = 10,000$.

Se rechaza la hipótesis nula planteada en (5.1) al nivel de significancia α si $D_n \geq C_n(\alpha)$, donde $C_n(\alpha)$ debe ser tal que:

$$\alpha = \max_{\gamma} P(\text{Rechazar } H_0 | H_0) = \max_{\gamma} P(D_n \geq C_n(\alpha)). \quad (5.6)$$

5.1. Pruebas propuestas

La distribución de D_n no depende del signo de γ , por lo tanto la constante crítica en (5.6) es tal que:

$$\alpha = \max_{\gamma \geq 0} P(\text{Rechazar } H_0 | H_0) = \max_{\gamma \geq 0} P(D_n \geq C_n(\alpha)).$$

De los estudios de simulación realizados se observó que es necesario tener un γ grande para que la prueba sea de tamaño α , con fines prácticos puede tomarse cualquier $\gamma \geq 20$ ya que cuando esto sucede, las distribuciones normales asimétricas para ξ y ω fijos son prácticamente indistinguibles entre ellas, o bien generar muestras directamente de la distribución media-normal.

Los valores de la constante crítica se determinan con los cuantiles $(1 - \alpha)100$ de la distribución empírica de D_n . En la Tabla 5.1 se muestran los valores críticos de la estadística D_n obtenida mediante simulación, usando el procedimiento descrito en los pasos 1-8 con $B = 10,000$ muestras de los tamaños n indicados.

Tabla 5.1: Valores críticos para la estadística D_n , obtenida mediante Simulación Monte Carlo $B = 10,000$ muestras de tamaño n .

n	α			
	0.10	0.05	0.025	0.01
20	0.9212	1.0167	1.1133	1.2387
50	0.9395	1.0364	1.1398	1.2688
100	0.9697	1.0704	1.1610	1.2750
150	0.9654	1.0718	1.1669	1.3107

5.1.2. Coeficiente de correlación

De la ecuación (3.2):

$$Y - \xi = \omega Z.$$

De la Propiedad 4) de la distribución normal asimétrica se tiene que:

$$X := (Y - \xi)^2 = \omega^2 Z^2 \sim \Gamma(1/2, 2\omega^2). \quad (5.7)$$

Es claro que se ha eliminado el parámetro γ del problema, pero aún falta conocer el valor de los parámetros ξ y ω .

Para ξ fijo y conocido, digamos $\xi = \xi_0$, X tiene distribución de escala,

5.1. Pruebas propuestas

$$P(X \leq x) = F_1\left(\frac{x}{2\omega^2}\right),$$

donde F_1 es la función de distribución de una variable aleatoria $\Gamma(1/2, 1)$. Así, dada la muestra y_1, \dots, y_n se calcula x_1, \dots, x_n utilizando la ecuación (5.7). Luego la función de probabilidades empírica dada por (5.8) es un estimador consistente de $P(X \leq x)$,

$$F_n(x) = \begin{cases} 0, & x < x_{(1)}, \\ i/n, & x_{(i)} \leq x < x_{(i+1)}, \\ 1, & x_{(n)} < x \end{cases} \quad (5.8)$$

donde $x_{(1)} \leq \dots \leq x_{(n)}$ son los valores ordenados de las x'_i s, entonces:

$$F_1\left(\frac{x}{2\omega^2}\right) \approx F_n(x),$$

por lo tanto:

$$F_1^{-1}\left(F_1\left(\frac{x}{2\omega^2}\right)\right) \approx F_1^{-1}(F_n(x)),$$

es decir:

$$u := F_1^{-1}(F_n(x)) \approx \frac{x}{2\omega^2}. \quad (5.9)$$

De la ecuación (5.9) bajo la hipótesis nula planteada en (5.1) se espera una fuerte relación lineal entre las x'_i s y las u'_i s. Si en la ecuación (5.7) se sustituye ξ por un estimador consistente del mismo, digamos $\hat{\xi}$, también se espera que esta relación lineal todavía exista, por lo tanto se propone probar la existencia de esta relación lineal utilizando algún procedimiento de prueba adecuado.

Otro procedimiento de prueba puede ser desarrollado como sigue. Para γ fijo, digamos $\gamma = \gamma_0$, Y tiene una distribución de localidad y escala

$$P(Y \leq y) = F_{\gamma_0}\left(\frac{y - \xi}{\omega}\right),$$

donde F_{γ_0} es la función de distribución de $Z \sim SN(\gamma_0)$ dada en (3.6).

5.1. Pruebas propuestas

Dada la muestra y_1, \dots, y_n , la función de distribución empírica $F_n(y)$ es un estimador consistente de $P(Y \leq y)$, entonces:

$$F_{\gamma_0} \left(\frac{y - \xi}{\omega} \right) \approx F_n(y).$$

Por lo tanto

$$F_{\gamma_0}^{-1} \left(F_{\gamma_0} \left(\frac{y - \xi}{\omega} \right) \right) \approx F_{\gamma_0}^{-1} (F_n(y)).$$

Es decir:

$$v := F_{\gamma_0}^{-1} (F_n(y)) \approx \frac{y - \xi}{\omega}. \quad (5.10)$$

De la ecuación (5.10) la hipótesis nula planteada en (5.1) se espera una fuerte relación lineal entre las y'_i s y las v'_i s. Si γ_0 en la ecuación (5.10) se sustituye por un estimador consistente, $\hat{\gamma}_0$, se espera que dicha relación se conserve.

Prueba basada en las x'_i s y u'_i s

Para probar si existe una fuerte relación lineal entre las x'_i s y las u'_i s se propone utilizar el coeficiente de correlación, dado por:

$$r_n = \text{Corr}(X, U) = \frac{\sum_{i=1}^n (X_i - \bar{X})(U_i - \bar{U})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (U_i - \bar{U})^2}}. \quad (5.11)$$

La hipótesis nula (5.1) se rechaza al nivel de significancia α si $r_n \leq C_n(\alpha)$, donde $C_n(\alpha)$ debe ser tal que:

$$\alpha = \max_{\gamma} P(\text{Rechazar } H_0 | H_0) = \max_{\gamma} P(r_n \leq C_n(\alpha)). \quad (5.12)$$

Bajo la hipótesis nula, la distribución de r_n para cada valor fijo de γ puede obtenerse utilizando simulación de Monte Carlo, ya que r_n es invariante a transformaciones de

5.1. Pruebas propuestas

escala y la distribución de X en la ecuación (5.7) no depende de ξ . Por lo tanto se utiliza el siguiente procedimiento.

- 1) Fijar n , $\xi = 0$, $\omega = 1$ y γ .
- 2) Simular una muestra de tamaño n de $SN_D(\xi, \omega, \gamma)$.
- 3) Calcular el estimador de máxima verosimilitud del parámetro ξ usando el método propuesto por Sartori.
- 4) Calcular x_i , $i = 1, \dots, n$ usando la ecuación (5.7).
- 5) Ordenar las x_i 's en forma ascendente.
- 6) Calcular $u_i = F_1^{-1}(F_n(x_i))$, $i = 1, \dots, n$, donde F_1^{-1} es la función de cuantiles de la distribución $\Gamma(1/2, 1)$.
- 7) Con las x_i , u_i generados en los pasos 5 y 6 calcular r_n en (5.11).
- 8) Repetir B veces los pasos 2 a 7.

Al terminar el proceso de simulación se tendrán B realizaciones de r_n . El valor de las constantes críticas $C_n(\alpha)$ se obtienen con los cuantiles 100α de la distribución empírica de r_n .

En la Figura 5.2 se presentan gráficas de $C_n(\alpha)$ como una función de γ para $\alpha = \{0.01, 0.025, 0.05\}$ y $n = 50$ que muestra claramente que la distribución de r_n bajo H_0 depende del valor verdadero del parámetro desconocido γ .

Note que es suficiente considerar $Y \sim SN_D(\xi, \omega, \gamma)$ con $\gamma \geq 0$ ya que $-Y \sim SN_D(-\xi, \omega, -\gamma)$. Por lo tanto la distribución de r_n bajo H_0 no depende de el signo de γ , y las constantes críticas $C_n(\alpha)$ en (5.12) deben ser tal que:

$$\alpha = \max_{\gamma \geq 0} P(\text{Rechazar } H_0 | H_0) = \max_{\gamma \geq 0} P(r_n \leq C_n(\alpha)). \quad (5.13)$$

De la Figura 5.2 y de (5.13) los valores de la constantes críticas se determinan con los cuantiles 100α de la distribución empírica de r_n obtenida mediante simulación con $\gamma \geq 20$. Con fines prácticos se puede utilizar $\gamma = 20$ o bien generar muestras directamente de la distribución media-normal.

En la Tabla 5.2 se presentan los valores críticos de la estadística r_n obtenidos mediante simulación, utilizando el procedimiento descrito en los pasos 1-8 con $B = 5,000$ muestras de Monte Carlo de los tamaños n indicados.

5.1. Pruebas propuestas

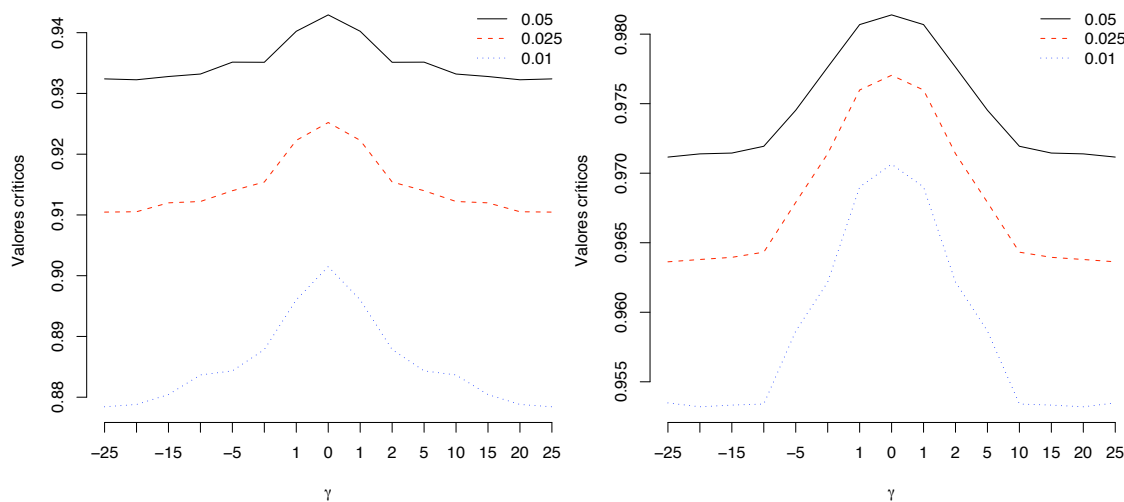


Figura 5.2: Valores críticos como función de γ para $n = 50$, $B = 5,000$ para las estadísticas r_n (izquierda) y r_n^* (derecha)

Tabla 5.2: Valores críticos de la estadística r_n obtenida mediante simulación de Monte Carlo $B = 5,000$ muestras de tamaño n .

n	α			
	0.01	0.025	0.05	0.1
10	0.8512	0.8682	0.8851	0.9078
20	0.8514	0.8810	0.9042	0.9279
30	0.8536	0.8902	0.9147	0.9387
50	0.8745	0.9087	0.9312	0.9512
100	0.9124	0.9340	0.9537	0.9678
150	0.9283	0.9493	0.9618	0.9732
200	0.9360	0.9537	0.9668	0.9768
300	0.9498	0.9666	0.9761	0.9827
500	0.9638	0.9754	0.9825	0.9881

Prueba basada en las y'_i s y v'_i s

Para probar si hay una fuerte relación lineal entre las y'_i s y las v'_i s se utiliza nuevamente el estimador de correlación muestral, la única diferencia es que las v'_i s en la ecuación (5.10) se calculan utilizando la ecuación (4.2). La estadística resultante se denominará r_n^* . El proceso de simulación para obtener r_n^* bajo la hipótesis nula H_0 es análogo al caso de r_n , debido a que r_n^* es invariante a transformaciones de localidad y escala, pero nuevamente, su distribución depende del valor verdadero del parámetro desconocido γ (ver Figura 5.2).

En la Tabla 5.3 se muestran los valores críticos de la estadística r_n^* obtenida mediante

5.1. Pruebas propuestas

simulación Monte Carlo con $B = 5,000$ de los tamaños n indicados.

Tabla 5.3: Valores críticos de la estadística r_n^* obtenida mediante simulación de Monte Carlo $B = 5,000$ muestras de tamaño n .

n	α			
	0.01	0.025	0.05	0.1
10	0.8783	0.9028	0.9217	0.9388
20	0.9092	0.9298	0.9440	0.9577
30	0.9295	0.9443	0.9576	0.9678
50	0.9534	0.9631	0.9704	0.9772
100	0.9724	0.9782	0.9826	0.9862
150	0.9798	0.9844	0.9872	0.9897
200	0.9840	0.9878	0.9898	0.9917
300	0.9893	0.9912	0.9924	0.9937
500	0.9928	0.9939	0.9948	0.9958

5.1.3. Bootstrap paramétrico

En esta sección se aborda el problema de probar (5.1) utilizando bootstrap paramétrico utilizando la estadística r_n^* . El procedimiento para obtener los valores críticos de la estadística de prueba es similar al descrito en el artículo de Meintanis (2007) y se resume en seguida.

- 1) Dadas las observaciones y_1, \dots, y_n calcular el estimador de momentos de γ , digamos $\tilde{\gamma}$
- 2) Calcular r_n^* usando $\tilde{\gamma}$
 - a) Generar una muestra bootstrap de tamaño n de $SN(\tilde{\gamma})$
 - b) Dada la muestra bootstrap generada previamente, calcular el estimador de momentos de γ , digamos, $\tilde{\gamma}^*$
 - c) Calcular el valor de la estadística de prueba, \tilde{r}_n^* usando $\tilde{\gamma}^*$ y la muestra bootstrap
- 3) Repetir los pasos a), b) y c) 1,000 veces para obtener \tilde{r}_{nj}^* , $j = 1, \dots, 1000$
- 4) Obtener $\tilde{C}_n(0.05)$ como $\tilde{r}_{n(50)}^*$, donde $\tilde{r}_{n(j)}^*$, $j = 1, \dots, 1000$ denota los valores ordenados de \tilde{r}_n^* .

5.2. Estudios de simulación

5.2.1. Pruebas con probabilidad de error tipo I máximo

En las Tablas 5.4, 5.5 y 5.6 se presentan los resultados de estimaciones del tamaño de las pruebas D_n , r_n y r_n^* , respectivamente obtenidas mediante simulación Monte Carlo. Los tamaños de muestra seleccionados fueron 50, 100 y 150, el valor del parámetro $\gamma = \{0, 1, 2, 3, 5, 10, 20\}$, $\xi = 0$ y $\omega = 1$. De los resultados mostrados en estas Tablas se observa que los tamaños estimados casi siempre están por debajo del nivel de significancia α , aún con $\gamma = 20$. Es evidente también que la prueba r_n^* es muy conservadora para $\gamma \leq 10$, por lo tanto se espera que esta prueba tenga mayor potencia para distribuciones sesgadas que para distribuciones simétricas.

Tabla 5.4: Tamaño de la prueba utilizando la estadística D_n obtenida con simulación Monte Carlo con $B = 5,000$ muestras del tamaño n con $\alpha = 0.05$.

n	γ							
	0	1	2	3	5	7	10	20
50	0.0269	0.0224	0.0158	0.0162	0.0206	0.0296	0.0372	0.0426
100	0.0228	0.0188	0.0080	0.1072	0.0166	0.0248	0.0336	0.0448
150	0.0276	0.0270	0.0070	0.0070	0.0170	0.0214	0.0342	0.0396

Tabla 5.5: Tamaño de la prueba utilizando la estadística r_n obtenida con simulación Monte Carlo con $B = 5,000$ muestras del tamaño n con $\alpha = 0.05$.

n	γ							
	0	1	2	3	5	7	10	20
50	0.0298	0.0370	0.0462	0.0508	0.0510	0.0513	0.0472	0.0498
100	0.0420	0.0484	0.0510	0.0504	0.0498	0.0482	0.0484	0.0528
150	0.0322	0.0448	0.0435	0.0414	0.0498	0.0448	0.0483	0.0512

Tabla 5.6: Tamaño de la prueba utilizando la estadística r_n^* obtenida con simulación Monte Carlo con $B = 5,000$ muestras del tamaño n con $\alpha = 0.05$.

n	γ							
	0	1	2	3	5	7	10	20
50	0.0092	0.0112	0.0226	0.0262	0.0380	0.0409	0.0392	0.0438
100	0.0061	0.0122	0.0226	0.0315	0.0343	0.0400	0.0430	0.0482
150	0.0061	0.0084	0.0215	0.0363	0.0328	0.0323	0.0360	0.0431

Para analizar el comportamiento de las pruebas propuestas se consideran algunas alternativas diferentes a la normal asimétrica. Las distribuciones seleccionadas con este fin son: Logística, t de Student, Cauchy, Exponencial, Ji Cuadrada, Gumbel, Lognormal y algunas estables [ver Nolan (1999)].

5.2. Estudios de simulación

Las potencias estimadas se comparan con los resultados obtenidos utilizando el procedimiento propuesto por [Mateu *et al.* \(2007\)](#), para 5 pruebas basadas en la función de distribución de probabilidades empíricas: Cramér-von Mises, W^2 , U^2 de Watson, A^2 de Anderson-Darling, D de Kolmogorov-Smirnov y Kuiper V . Los resultados se muestran en las Tablas [5.7](#), [5.8](#), [5.9](#) en las que se observa que las pruebas r_n y r_n^* presentan potencias más altas para varias de las alternativas consideradas. De estas Tablas se observa también que en general la prueba r_n es más potente que la prueba r_n^* . Las pruebas propuestas poseen una ventaja adicional sobre las estudiadas por [Mateu *et al.*](#), al maximizar sobre γ se elimina la dependencia de este parámetro, con lo cual no es necesario interpolar en las Tablas de los mencionados autores para obtener las constantes críticas.

Tabla 5.7: Comparación de potencias de las pruebas estudiadas por *Mateu et al.* y las propuestas para $n = 50$.

		Alternativa	A^2	W^2	U^2	D	V	D_n	r_n^*	r_n	
Distrib. simétricas	Cola ligera	Logística estándar	0.1058	0.0849	0.1108	0.0672	0.1062	0.0284	0.1212	0.1607	
		t(12)	0.0564	0.0464	0.0665	0.0419	0.0664	0.0256	0.0709	0.1156	
		t(4)	0.3293	0.2862	0.3228	0.2286	0.2933	0.0874	0.3963	0.3896	
		pesada	0.9946	0.9942	0.9950	0.9904	0.9936	0.9606	0.9942	0.9710	
	Cola ligera	Cauchy estándar	0.7412	0.7184	0.5231	0.6667	0.5156	0.7292	0.5148	0.3310	
		Exp. estándar	0.0999	0.1064	0.0828	0.1065	0.0744	0.1356	0.1893	0.1704	
		Ji Cuadrada(4)	0.9967	0.9920	0.9666	0.9885	0.9656	0.9740	0.8856	0.5430	
	Distrib. asimétricas	Cola pesada	Weibull(0.75,1)	0.0565	0.0514	0.0661	0.0543	0.0637	0.0394	0.1397	0.1600
			Gumbel estándar	0.1326	0.1387	0.1280	0.1335	0.1145	0.1480	0.3064	0.2672
		Log-Normal(0,0.5)	0.5895	0.5435	0.5729	0.4859	0.5389	0.3342	0.6613	0.6606	
Sta(1.8,0.25,1,0;0)		0.2773	0.2398	0.2648	0.2147	0.2511	0.1296	0.3603	0.4036		
Sta(1.8,0.50,1,0;0)	0.2704	0.2344	0.2577	0.2092	0.2492	0.1340	0.3588	0.4070			

Tabla 5.8: Comparación de potencias de las pruebas estudiadas por *Mateu et al.* y las propuestas para $n = 100$.

		Alternativa	A^2	W^2	U^2	D	V	D_n	r_n^*	r_n	
Distrib. simétricas	Cola ligera	Logística estándar	0.2220	0.1902	0.1990	0.1382	0.1773	0.0316	0.1962	0.2570	
		t(12)	0.1165	0.0953	0.1032	0.0747	0.0931	0.0262	0.0958	0.1772	
		t(4)	0.6369	0.5776	0.5859	0.4762	0.5293	0.1474	0.6207	0.6064	
		pesada	1.0000	1.0000	1.0000	1.0000	1.0000	0.9991	1.0000	1.0000	
	Cola ligera	Cauchy estándar	0.9521	0.9416	0.8084	0.9037	0.8077	0.9222	0.7627	0.4780	
		Exp. estándar	0.1377	0.1393	0.1302	0.1307	0.1121	0.1350	0.2996	0.2562	
		Ji Cuadrada(4)	1.0000	1.0000	1.0000	1.0000	1.0000	0.9994	0.9919	0.7453	
	Distrib. asimétricas	Cola pesada	Weibull(0.75,1)	0.1106	0.0947	0.0999	0.0872	0.0944	0.0376	0.2212	0.2332
			Gumbel estándar	0.2647	0.2525	0.2374	0.2272	0.2005	0.1954	0.4530	0.4094
		Log-Normal(0,0.5)	0.8713	0.8416	0.8448	0.7875	0.8144	0.5540	0.8900	0.8782	
Sta(1.8,0.25,1,0;0)		0.5108	0.4533	0.4616	0.3968	0.4302	0.2178	0.5874	0.6240		
Sta(1.8,0.50,1,0;0)	0.4890	0.4329	0.4397	0.3780	0.4117	0.2214	0.5754	0.6238			

Tabla 5.9: Comparación de potencias de las pruebas estudiadas por [Mateu et al.](#) y las propuestas para $n = 150$.

		Alternativa	A^2	W^2	U^2	D	V	D_n	r_n^*	r_n
Distrib. simétricas	Cola ligera	Logística estándar	0.3227	0.2804	0.2836	0.2102	0.2472	0.0402	0.2450	0.3016
		$t(12)$	0.1506	0.1254	0.1277	0.0960	0.1173	0.0288	0.1256	0.2022
	Cola pesada	$t(4)$	0.8022	0.7628	0.7644	0.6616	0.7118	0.2134	0.7632	0.7130
		Cauchy estándar	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	Cola ligera	Exp. estándar	0.9924	0.9899	0.9279	0.9751	0.9312	0.9778	0.8836	0.5230
		Ji Cuadrada(4)	0.1762	0.1702	0.1699	0.1487	0.1523	0.1632	0.3576	0.2890
Distrib. asimétricas	Cola pesada	Weibull(0.75,1)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.8588
		Gumbel estándar	0.1518	0.1298	0.1304	0.1176	0.1211	0.0512	0.2618	0.2748
	Cola pesada	Log-Normal(0,0.5)	0.4011	0.3737	0.3497	0.3328	0.2986	0.2826	0.5731	0.4874
		Sta(1.6,0.25,1,0;0)	0.9571	0.9430	0.9434	0.9126	0.9263	0.7312	0.9632	0.9470
		Sta(1.8,0.25,1,0;0)	0.6648	0.6022	0.6059	0.5389	0.5693	0.3028	0.7245	0.7703
		Sta(1.8,0.50,1,0;0)	0.6374	0.5765	0.5775	0.5097	0.5394	0.3092	0.7103	0.7432

5.3. Ejemplos de aplicación

5.2.2. r_n^* con bootstrap paramétrico

Con la finalidad de estudiar el desempeño de la prueba r_n^* , se compara con la prueba $\tilde{T}_{n,a}$ propuesta por Meintanis (2007). Se realizó un estudio de simulación Monte Carlo para las pruebas bootstrap con $B = 100$ muestras de tamaño $n = 50$ y las mismas distribuciones alternativas estudiadas por el autor mencionado: normal asimétrica $SN(\gamma)$, t asimétrica $ST(\gamma, \nu)$, g de Tukey $TU(g)$ y Laplace asimétrica $AL(\phi)$. Los resultados se presentan en la Tabla 5.10.

Tabla 5.10: Comparación de potencias de las estadísticas $\tilde{T}_{n,a}$ (izquierda), r_n^* (derecha) para las distribuciones $SN(\gamma)$, $ST(\gamma, \nu)$, $TU(g)$ y $AL(\phi)$ con $n = 50$, $\alpha = 0.05$, $B = 100$, obtenidas utilizando bootstrap paramétrico

$\gamma = 0.25$		$\gamma = 0.50$		$\gamma = 0.75$		$\gamma = 1.0$		$\gamma = 2.0$	
6	4	6	3	5	3	4	5	4	5
$\nu = 2$		$\nu = 3$		$\nu = 5$		$\nu = 10$		$\nu = 15$	
48	87	33	65	11	30	7	15	6	11
$g = 0.25$		$g = 0.50$		$g = 1.0$		$g = 1.5$		$g = 2.0$	
3	11	25	31	85	91	98	99	100	100
$\phi = 1.0$		$\phi = 1.25$		$\phi = 1.55$		$\phi = 1.75$		$\phi = 2.0$	
9	56	20	45	27	46	34	43	42	42

Los resultados mostrados en la Tabla 5.10 muestran que la versión bootstrap de la prueba r_n^* domina la prueba $\tilde{T}_{n,a}$ en términos de potencia. También se puede observar que la prueba r_n^* tiene un buen control sobre el nivel de significancia.

5.3. Ejemplos de aplicación

Los procedimientos de prueba son aplicados a los datos de las Tablas 5.11 y 5.12 de Gupta y Brown (2001). Estos datos que también fueron analizados por Meintanis (2007), dan los IQ de 87 hombres blancos y 52 hombres de otras razas, reportados por una compañía de seguros en 1971.

Tabla 5.11: IQ para hombres blancos.

124	106	108	112	113	122	100	108	108	94	102	120	101	118	113
117	100	106	111	107	112	120	102	135	125	98	121	117	124	114
103	122	122	113	113	104	103	113	120	106	132	106	112	118	113
112	112	121	112	85	117	109	104	129	140	106	115	109	122	108
119	121	113	107	122	103	97	116	114	131	94	112	108	118	112
	116	113	111	122	112	136	116	108	112	108	116	103		

5.3. Ejemplos de aplicación

Tabla 5.12: IQ para hombres de otras razas.

91	102	100	117	122	115	97	109	108	104	108	118	103	123	123
103	106	102	118	100	103	107	108	107	97	95	119	102	108	103
	102	112	99	116	114	102	111	104	122	103	111	101	91	
			99	121	97	109	106	102	104	107	95			

Suponga que se desea probar el juego de hipótesis planteado en (5.1) para los IQ utilizando las pruebas r_n y r_n^* .

Prueba r_n para los datos de la Tabla 5.11

Los estimadores de máxima verosimilitud son: $\hat{\xi} = 105.89$, $\hat{\omega} = 11.80$, $n = 87$ y $r_n = 0.9856$, de la Tabla 5.2 se tiene que para $\alpha = 0.05$ la constante crítica es 0.9478, entonces como $0.9478 < 0.9856$ no se rechaza la hipótesis nula.

Prueba r_n^ para los datos de la Tabla 5.12*

En este caso $n = 52$. Los estimadores de momentos son $\tilde{\xi} = 98.63$, $\tilde{\omega} = 11.55$, $r_n^* = 0.9709$. De la Tabla 5.3 se tiene que para $\alpha = 0.05$ la constante crítica es 0.9708, entonces como $0.9708 < 0.9709$ no se rechaza la hipótesis nula.

Capítulo 6

Pruebas para normalidad

En este capítulo se estudia la prueba de razón de verosimilitudes generalizada para poder decidir si una muestra es de la distribución normal o bien de la distribución normal asimétrica. Cuando el parámetro de forma toma el valor 0 en (3.3) el modelo se reduce al modelo normal usual. El procedimiento de prueba se aplica una vez que se ha hecho una prueba de bondad de ajuste para verificar si los datos vienen de la distribución normal asimétrica y se cree que es posible trabajar con un modelo más pequeño. Dada la muestra aleatoria Y_1, \dots, Y_n de $SN_D(\xi, \omega, \gamma)$, el juego de hipótesis de interés es:

$$H_0 : \gamma = 0 \text{ vs } H_1 : \gamma \neq 0. \quad (6.1)$$

6.1. Prueba basada en el índice de asimetría

Salvan (1986) estudió familias de densidades de la forma $g_\theta(x; a, b)$ con a y b parámetros de localidad y escala respectivamente, mientras que θ es un parámetro de forma. Su interés se centró en probar hipótesis del tipo

$$H_0 : \theta = \theta_0.$$

Usualmente los parámetros a y b son desconocidos, por lo tanto se tiene una hipótesis compuesta, pero debido a que las estadísticas de prueba resultantes son invariantes a transformaciones de localidad y escala entonces la hipótesis compuesta se reduce a una hipótesis simple.

6.2. Prueba de razón de verosimilitudes generalizada

Para probar (6.1), [Salvan \(1986\)](#) mostró que la prueba local uniformemente más potente está basada en el índice de asimetría. La regla de decisión es:

Rechazar H_0 si $T < C_1(n, \frac{\alpha}{2})$ o $T > C_2(n, 1 - \frac{\alpha}{2})$ donde $C_1(n, \frac{\alpha}{2})$ y $C_2(n, 1 - \frac{\alpha}{2})$ son los cuantiles $(\alpha/2)100$ y $(1 - \alpha/2)100$ de la distribución de T bajo la hipótesis nula, y T está dada por:

$$T = \frac{\sum(Y_i - \bar{Y})^3}{\{\sum(Y_i - \bar{Y})^2\}^{\frac{3}{2}}}. \quad (6.2)$$

Bajo la hipótesis nula, cuando n es grande $\sqrt{n}T \approx N(0, 6/n)$, por lo tanto se pueden obtener las constantes críticas utilizando la distribución normal.

Hay que notar que la estadística de prueba es esencialmente una función del estimador de momentos de γ_1 dado por (4.1) obtenido al reparametrizar bajo H_0 de (ξ, ω, γ) a (μ, σ, γ_1) tal y como sugieren [Azzalini y Capitanio \(1999\)](#). La prueba resultante es muy intuitiva, ya que si la hipótesis nula es cierta entonces el valor de la estadística de prueba debe estar cercano a 0.

6.2. Prueba de razón de verosimilitudes generalizada

Para resolver el problema de probar (6.1) se puede emplear la prueba de razón de verosimilitudes generalizada, para lo cual es necesario calcular los estimadores bajo la hipótesis nula y en todo el espacio paramétrico.

La función de verosimilitud está dada por:

$$L(\boldsymbol{\theta}; \mathbf{y}) = \prod_{i=1}^n f_{Y_i}(y_i; \xi, \omega, \gamma) = \prod_{i=1}^n \frac{2}{\omega} \phi\left(\frac{y_i - \xi}{\omega}\right) \Phi\left[\gamma\left(\frac{y_i - \xi}{\omega}\right)\right]. \quad (6.3)$$

Estimadores de los parámetros bajo la hipótesis nula

Bajo la hipótesis nula Y_1, \dots, Y_n es una muestra aleatoria de la distribución $N(\xi, \omega^2)$, entonces los estimadores de máxima verosimilitud de los parámetros se pueden obtener fácilmente de (6.3), y están dados por:

6.3. Potencia y tamaño de las pruebas

$$\hat{\xi} = \frac{1}{n} \sum_{i=1}^n y_i = \bar{y},$$
$$\hat{\omega} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}.$$

El numerador de $\lambda(\mathbf{y})$ en (3.9) es $L(\hat{\xi}, \hat{\omega}, 0; \mathbf{y})$, y está dado por:

$$L(\hat{\xi}, \hat{\omega}, 0; \mathbf{y}) = \left(\frac{n}{2\pi \sum_{i=1}^n (y_i - \bar{y})^2} \right)^{n/2} e^{-n/2}.$$

Estimadores de los parámetros en todo el espacio paramétrico

Los estimadores de ξ , ω y γ se obtienen utilizando el procedimiento propuesto por Sartori (2006). Sean $\tilde{\xi}$, $\tilde{\omega}$ y $\tilde{\gamma}$ los estimadores de máxima verosimilitud ξ , ω y γ , entonces el denominador de $\lambda(\mathbf{y})$ en (3.9) es $L(\tilde{\omega}, \tilde{\gamma}, \tilde{\xi}; \mathbf{y})$.

La distribución de la estadística de razón de verosimilitudes es difícil de obtener en esta situación, pero se puede aplicar el Teorema 3.1 y por lo tanto cuando n es grande se rechaza H_0 en (6.1) al nivel de significancia α si y solo si:

$$-2\{\log L(\hat{\xi}, \hat{\omega}, 0; \mathbf{y}) - \log L(\tilde{\xi}, \tilde{\omega}, \tilde{\gamma}; \mathbf{y})\} \geq \chi_{1,1-\alpha}^2.$$

6.3. Potencia y tamaño de las pruebas

En esta sección se presenta un estudio de simulación Monte Carlo utilizado para evaluar el tamaño y potencia de la prueba basada en el coeficiente de asimetría y la prueba de razón de verosimilitudes descrita en la sección previa. Todos los cálculos fueron hechos utilizando el paquete R y la biblioteca de funciones `sn` de Azzalini.

6.3.1. Tamaño de las pruebas

De la Tabla 6.1 se observa que la prueba basada en el índice de asimetría funciona de manera adecuada aún y cuando se utilizó la distribución asintótica de la misma para obtener el valor de las constantes críticas. Por lo tanto se puede aplicar con muestras de tamaño moderado sin ningún problema. De la Tabla 6.2 se observa que

6.3. Potencia y tamaño de las pruebas

Tabla 6.1: Tamaño estimado de la prueba basada en el índice de asimetría, obtenida mediante simulación Monte Carlo con $B = 1,000$ muestras de tamaño n , $\xi = 0, \omega = 1$.

n	α		
	0.025	0.05	0.1
30	0.017	0.034	0.074
50	0.023	0.042	0.079
75	0.023	0.040	0.090
100	0.022	0.043	0.091
150	0.024	0.047	0.093
200	0.024	0.046	0.094

Tabla 6.2: Tamaño estimado de la prueba de razón de verosimilitudes, obtenida mediante simulación Monte Carlo con $B = 1,000$ muestras de tamaño n , $\xi = 0, \omega = 1$.

n	α		
	0.025	0.05	0.1
30	0.085	0.114	0.191
50	0.039	0.067	0.132
75	0.028	0.064	0.114
100	0.024	0.049	0.105
150	0.026	0.051	0.107
200	0.021	0.052	0.101

es necesario tener muestras de tamaño $n > 100$ para que la prueba de razón de verosimilitudes funcione de manera adecuada. Para muestras pequeñas no es recomendable aplicar la prueba de razón de verosimilitudes, ya que no respeta el tamaño de prueba especificado.

6.3.2. Potencia de las pruebas

En la Figura 6.1 se muestra la potencia estimada de la prueba basada en el índice de asimetría y la de razón de verosimilitudes, de donde se observa que ambas pruebas son consistentes, ya que a medida que aumenta el tamaño de la muestra la potencia también se incrementa. En la Figura 6.2 se compara la potencia de las pruebas para $n = 150$. En este caso las potencias son prácticamente iguales, por lo tanto para muestras grandes podemos utilizar cualquiera de las dos pruebas, pero por facilidad de cálculo obviamente es más sencillo de realizar la prueba basada en el índice de asimetría.

6.3. Potencia y tamaño de las pruebas

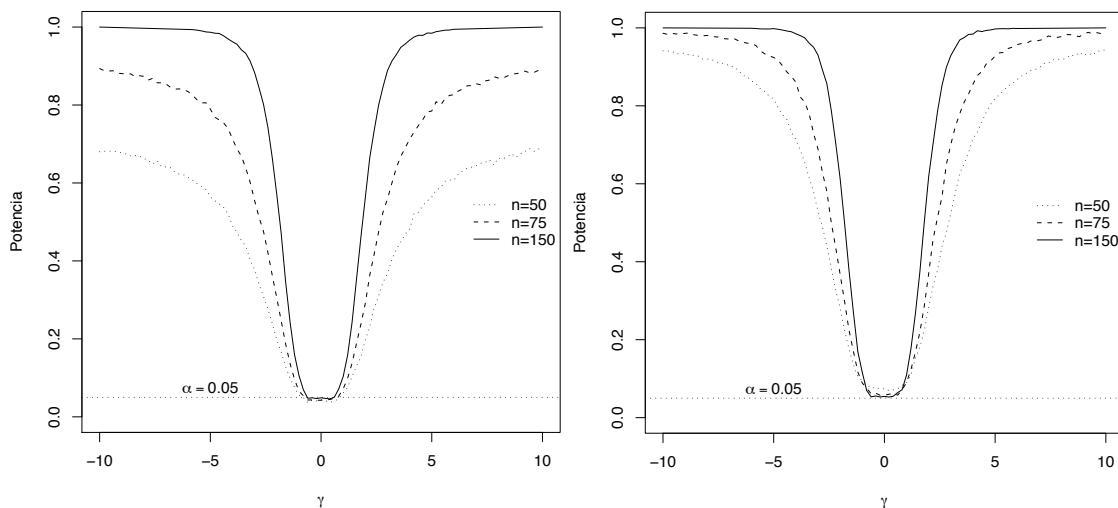


Figura 6.1: Potencia de la prueba en el índice de asimetría (izquierda) y razón de verosimilitudes (derecha) para probar normal vs normal asimétrica, para varios tamaños de muestra, $B = 5,000$, $\xi = 0$ y $\omega = 1$.

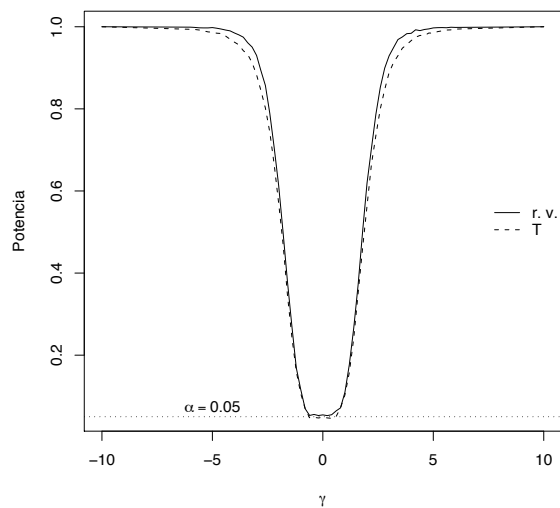


Figura 6.2: Comparación de potencias de la prueba basada en el índice de asimetría y razón de verosimilitudes para probar normal vs normal asimétrica, $B = 5,000$, $\xi = 0$, $\omega = 1$ y $n = 150$.

Capítulo 7

Pruebas de Hipótesis Para la Distribución Normal Asimétrica

7.1. Pruebas para el parámetro de localidad

Sea Y_1, \dots, Y_n una m. a. de $SN_D(\xi, \omega, \gamma)$ y suponga que se desea probar el juego de hipótesis:

$$H_0 : \xi = \xi_0, \omega > 0, \gamma \in \mathbb{R} \text{ vs } H_1 : \xi \neq \xi_0, \omega > 0, \gamma \in \mathbb{R}, \quad (7.1)$$

donde ξ_0 es un valor conocido.

Es decir, se está interesado en probar si el parámetro de localidad toma un valor en particular. Este juego de hipótesis puede ser probado empleando la prueba de razón de verosimilitudes generalizada; sin embargo, es posible construir pruebas en las que no es necesario estimar ningún parámetro. El procedimiento para construir dichas pruebas se presenta en seguida.

7.1.1. Estadística Señal-Ruido

Sean las variables aleatorias $Z_i = \left(\frac{Y_i - \xi_0}{\omega}\right)^2$, $i = 1, \dots, n$, entonces $Z_i \sim \chi_1^2$, y considere la estadística:

$$R = \frac{\bar{Z}}{\sqrt{S^2}}, \quad (7.2)$$

7.1. Pruebas para el parámetro de localidad

donde $\bar{Z} = n^{-1} \sum_{i=1}^n Z_i$, $S^2 = n^{-1} \sum_{i=1}^n (Z_i - \bar{Z})^2$. Entonces bajo H_0 la distribución de R es independiente de los parámetros desconocidos ω y γ .

Demostración

$$\begin{aligned} R &= \frac{\frac{1}{n} \sum_{i=1}^n Z_i}{\sqrt{\frac{1}{n} \sum_{i=1}^n \left(Z_i - \frac{1}{n} \sum_{i=1}^n Z_i \right)^2}} \\ &= \frac{\frac{1}{n} \sum_{i=1}^n (Y_i - \xi_0)^2}{\sqrt{\frac{1}{n} \sum_{i=1}^n \left((Y_i - \xi_0)^2 - \frac{1}{n} \sum_{i=1}^n (Y_i - \xi_0)^2 \right)^2}}. \end{aligned}$$

Sea $W_i = (Y_i - \xi_0)^2$, entonces la ecuación (7.2) es equivalente a:

$$R = \frac{\bar{W}}{\sqrt{\frac{1}{n} \sum_{i=1}^n (W_i - \bar{W})^2}}, \quad (7.3)$$

Bajo la hipótesis nula la distribución de R en (7.3) es libre de parámetros, no así del tamaño de la muestra. Note que valores grandes y pequeños de R apoyan la hipótesis alternativa.

La regla de decisión es:

Rechazar H_0 si $R < C_1(n, \frac{\alpha}{2})$ o $R > C_2(n, 1 - \frac{\alpha}{2})$, donde $C_1(n, \frac{\alpha}{2})$ y $C_2(n, 1 - \frac{\alpha}{2})$ son los cuantiles $(\alpha/2)100$ y $(1 - \alpha/2)100$ de la distribución de R bajo la hipótesis nula.

Distribución de la estadística de prueba

La distribución empírica de R puede obtenerse utilizando simulación de Monte Carlo con base en el siguiente algoritmo:

- 1) Fijar n , $\xi = \xi_0$, ω y γ .
- 2) Generar una muestra de tamaño n de la distribución $SN_D(\xi, \omega, \gamma)$.

7.1. Pruebas para el parámetro de localidad

- 3) Calcular R .
- 4) Repetir los pasos 2-3 B veces.

Al final de las iteraciones se tendrán B realizaciones de R a partir de las cuales se obtiene la distribución empírica de R y de donde se pueden obtener $C_1(n, \alpha/2)$ y $C_2(n, 1 - \alpha/2)$ para n y α fijos.

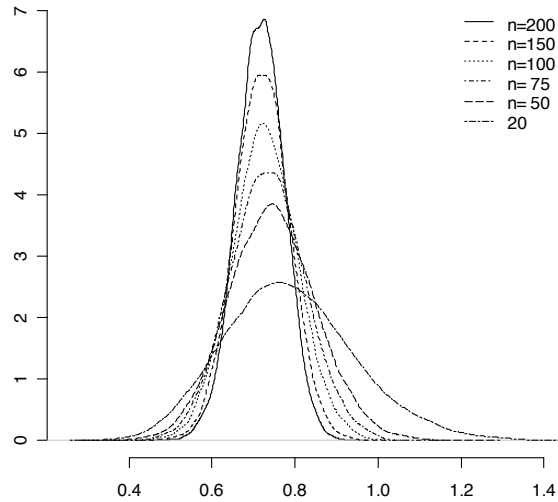


Figura 7.1: Distribución empírica de la estadística R para varios tamaños de muestra obtenida mediante simulación Monte Carlo con $B = 20,000$ réplicas de tamaño n .

De la Figura 7.1 se observa que la distribución de R depende fuertemente del tamaño de muestra.

En la Tabla 7.1 se muestran los valores críticos $C_1(n, \alpha/2)$ y $C_2(n, 1 - \alpha/2)$ para varios tamaños de muestra n y $\alpha = \{0.01, 0.025, 0.05, 0.1\}$, obtenidos a partir de la distribución empírica de R .

La distribución asintótica de la estadística de R puede obtenerse empleando el TCL y el teorema de Taylor, ambos en su versión multivariada. En efecto, si $\{X_n\}$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas con media μ y varianza σ^2 , y además $EX_1^4 < \infty$ y si se define $\mathbf{T}_n = [\bar{X}_n - \mu, S_n^2 - \sigma^2]'$ donde $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$ y $S^2 = n^{-1} \sum_{i=1}^n (X_i - \bar{X})^2$ entonces $\sqrt{n}\mathbf{T}_n \xrightarrow{d} N(\mathbf{0}, \Sigma)$, donde

$$\Sigma = \begin{pmatrix} \sigma^2 & \mu_3 \\ \mu_3 & \mu_4 - \sigma^4 \end{pmatrix}$$

Los detalles de la demostración pueden verse en [Sen y Singer \(1993\)](#). Sea $\boldsymbol{\theta} = (\mu, \sigma^2)'$

7.1. Pruebas para el parámetro de localidad

Tabla 7.1: Valores críticos de la estadística R obtenida mediante Simulación Monte Carlo con $B = 20,000$ réplicas de tamaño n .

n	Nivel de significancia α							
	0.01		0.025		0.05		0.10	
	C_1	C_2	C_1	C_2	C_1	C_2	C_1	C_2
10	0.4257	1.6384	0.4612	1.4828	0.4965	1.3756	0.5382	1.2579
20	0.4353	1.2711	0.4737	1.1903	0.5062	1.1281	0.5462	1.0647
30	0.4489	1.1495	0.4839	1.0885	0.5163	1.0437	0.5546	0.9941
50	0.4765	1.0368	0.5099	0.9972	0.5398	0.9611	0.5709	0.9236
60	0.4903	1.0102	0.5248	0.9712	0.5487	0.9401	0.5807	0.9058
70	0.4969	0.9849	0.5260	0.9478	0.5526	0.9178	0.5832	0.8886
80	0.5027	0.9611	0.5362	0.9302	0.5627	0.9032	0.5903	0.8753
90	0.5116	0.9452	0.5425	0.9151	0.5676	0.8900	0.5962	0.8654
100	0.5188	0.9350	0.5442	0.9071	0.5680	0.8827	0.5960	0.8581
120	0.5311	0.9161	0.5601	0.8901	0.5817	0.8667	0.6057	0.8423
140	0.5431	0.8983	0.5668	0.8737	0.5868	0.8536	0.6096	0.8314
160	0.5466	0.8831	0.5710	0.8614	0.5926	0.8444	0.6145	0.8250
180	0.5545	0.8732	0.5789	0.8550	0.5976	0.8377	0.6193	0.8190
200	0.5631	0.8653	0.5838	0.8469	0.6018	0.8322	0.6219	0.8124
250	0.5752	0.8467	0.5958	0.8300	0.6108	0.8162	0.6282	0.8011
300	0.5853	0.8364	0.6039	0.8220	0.6194	0.8077	0.6350	0.7924

y $g(x, y) = x/\sqrt{y}$, es decir, $g(\bar{X}_n, S_n^2) = \bar{X}_n/\sqrt{S_n^2} = \nu_n$ y $g(\boldsymbol{\theta}) = \mu/\sigma = \nu$. Además

$$\begin{aligned}\frac{\partial g(x, y)}{\partial x} \Big|_{(x,y)=\boldsymbol{\theta}} &= \frac{1}{\sqrt{y}} \Big|_{(x,y)=\boldsymbol{\theta}} = \frac{1}{\sigma}, \\ \frac{\partial g(x, y)}{\partial y} \Big|_{(x,y)=\boldsymbol{\theta}} &= -\frac{x}{2y^{\frac{3}{2}}} \Big|_{(x,y)=\boldsymbol{\theta}} = -\frac{\mu}{2\sigma^3}.\end{aligned}$$

Por lo tanto, al aplicar el teorema de Taylor multivariado se tiene:

$$\sqrt{n}(\nu_n - \nu) \xrightarrow{d} N(0, \beta^2) \text{ donde } \beta^2 = \frac{4\sigma^6 - 4\mu\sigma^2\mu_3 + \mu^2(-\sigma^4 + \mu_4)}{4\sigma^6}.$$

En el caso de la estadística R , bajo la hipótesis nula Z_1, \dots, Z_n , son variables aleatorias independientes e idénticamente distribuidas χ_1^2 . Entonces $E\{Z_1\} = \mu = 1$, $Var(Z_1) = \sigma^2 = 2$, $E(Z - \mu)^3 = \mu_3 = 8$, $E(Z - \mu)^4 = \mu_4 = 60$, entonces $\beta^2 = 3/4$, por lo tanto cuando $n \rightarrow \infty$,

$$RC := \sqrt{n} \left(R - \frac{1}{\sqrt{2}} \right) \xrightarrow{d} N \left(0, \frac{3}{4} \right). \quad (7.4)$$

7.1. Pruebas para el parámetro de localidad

En las Figuras 7.2 y 7.3 se muestra la distribución empírica de RC y su distribución asintótica para varios tamaños de muestra. La distribución empírica de la estadística de RC fue obtenida utilizando simulación Monte Carlo siguiendo un procedimiento análogo al empleado para obtener la distribución empírica de la estadística R bajo H_0 . En general, se puede observar que la aproximación normal es bastante mala (sobre todo en la cola izquierda) aún con muestras de tamaño $n = 400$.

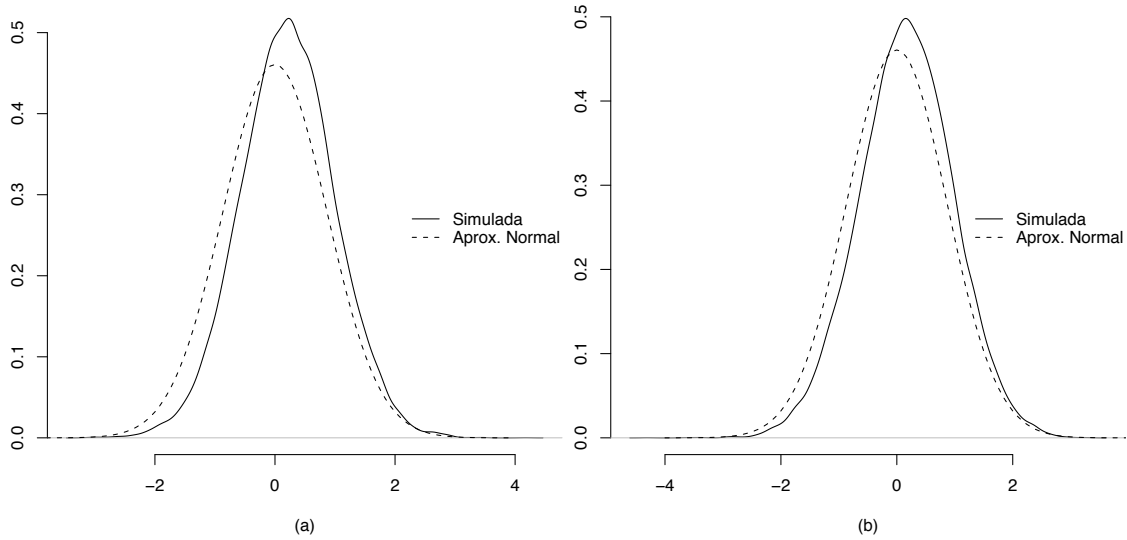


Figura 7.2: Distribución empírica de la estadística RC y la aproximación normal, $B = 20,000$ réplicas de tamaño n , (a) $n = 100$, (b) $n = 200$.

Otra alternativa para aproximar los valores críticos $C_1(n, \alpha/2)$ y $C_2(n, 1 - \alpha/2)$ para la estadística R , para n finito, es suponer que su distribución es aproximadamente normal y calcular valores apropiados para su media y su varianza.

Mood *et al.* (1974) proporcionan expresiones para aproximar la media y la varianza de funciones de dos variables aleatorias:

$$\begin{aligned}
 E\{g(X, Y)\} &\approx g(\mu_X, \mu_Y) + \frac{1}{2}Var\{X\} \frac{\partial^2}{\partial x^2} g(x, y) \Big|_{\mu_X, \mu_Y} \\
 &\quad + \frac{1}{2}Var\{Y\} \frac{\partial^2}{\partial y^2} g(x, y) \Big|_{\mu_X, \mu_Y} \\
 &\quad + Cov\{X, Y\} \frac{\partial^2}{\partial x \partial y} g(x, y) \Big|_{\mu_X, \mu_Y} .
 \end{aligned} \tag{7.5}$$

7.1. Pruebas para el parámetro de localidad

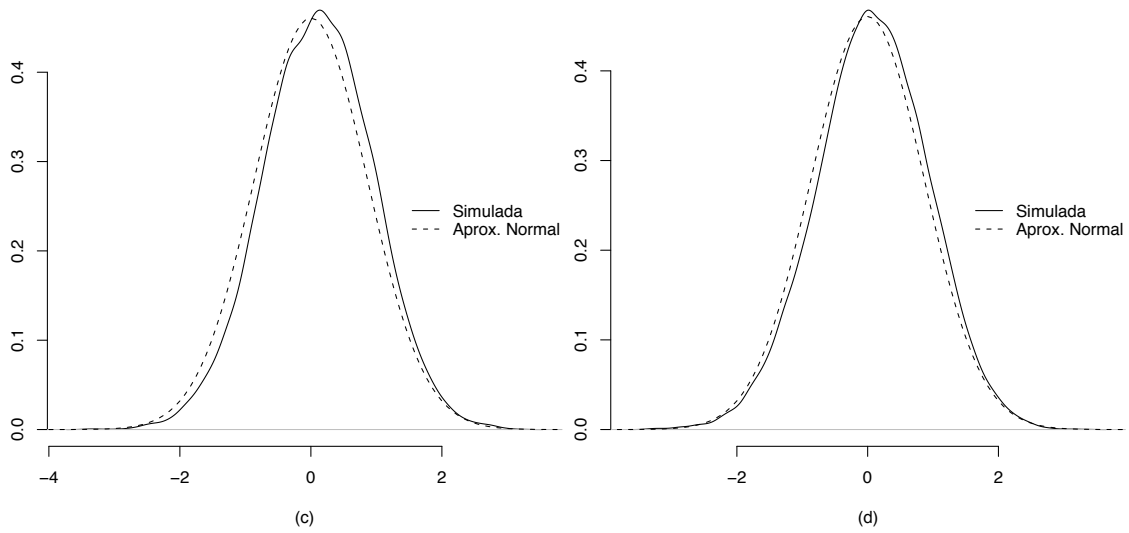


Figura 7.3: Distribución empírica de la estadística RC y la aproximación normal, $B = 20,000$ réplicas de tamaño n , (c) $n = 400$, (d) $n = 800$.

$$\begin{aligned}
 Var\{g(X, Y)\} &\approx 2Cov\{X, Y\} \left\{ \frac{\partial}{\partial x} g(x, y) \Big|_{\mu_X, \mu_Y} \frac{\partial}{\partial y} g(x, y) \Big|_{\mu_X, \mu_Y} \right\} \\
 &\quad + Var\{X\} \left\{ \frac{\partial}{\partial x} g(x, y) \Big|_{\mu_X, \mu_Y} \right\}^2 \\
 &\quad + Var\{Y\} \left\{ \frac{\partial}{\partial y} g(x, y) \Big|_{\mu_X, \mu_Y} \right\}^2, \tag{7.6}
 \end{aligned}$$

donde:

$$E\{X\} = \mu_x, \quad E\{Y\} = \mu_Y.$$

En este caso $X := \bar{Z}$, $Y := S^2$ y $g(X, Y) = \frac{X}{\sqrt{Y}}$.

En seguida se calcula $E\{X\}$, $E\{Y\}$, $Var\{X\}$, $Var\{Y\}$, $Cov\{X, Y\}$

$$\begin{aligned}
 E\{\bar{Z}\} &= 1 \\
 Var\{\bar{Z}\} &= \frac{2}{n}
 \end{aligned}$$

7.1. Pruebas para el parámetro de localidad

$$E\{\bar{Z}^2\} = E\left\{\left(\frac{1}{n}\sum_{i=1}^n Z_i\right)^2\right\} = \frac{1}{n^2}E\left\{\left(\sum_{i=1}^n Z_i\right)^2\right\}$$

Sea $W := \sum_{i=1}^n Z_i \sim \chi_n^2$, entonces $E\{W^2\} = 2n + n^2$, por lo tanto

$$E\{\bar{Z}^2\} = \frac{2n + n^2}{n^2} = 1 + \frac{2}{n}.$$

Además:

$$E\{S^2\} = E\left\{\frac{1}{n}\sum_{i=1}^n (Z_i - \bar{Z})^2\right\} = \frac{1}{n}E\left\{\sum_{i=1}^n Z_i^2 - n\bar{Z}^2\right\} = 2(n-1)/(n)$$

Ahora si $\theta_1 = E\{Z_i\}$ y $\theta_j = E\{(Z_i - \theta_1)^j\}$, $j = 2, 3, 4$, entonces

$$Var\{S^2\} = \frac{(n-1)^2}{n^3} \left(\theta_4 - \frac{n-3}{n-1}\theta_2^2 \right).$$

Es decir,

$$\begin{aligned} Var\{S^2\} &= \frac{(n-1)^2}{n^3} \left(E\{(Z_i - 1)^4\} - \frac{n-3}{n-1} (E\{(Z_i - 1)^2\})^2 \right) \\ &= \frac{(n-1)^2}{n^3} \left(60 - 4\frac{n-3}{n-1} \right) \end{aligned}$$

También,

$$Cov\{\bar{Z}, S^2\} = E\{\bar{Z}S^2\} - E\{\bar{Z}\}E\{S^2\} = E\{\bar{Z}S^2\} - E\{S^2\}$$

Sea $W_i = Z_i - 1$, $i = 1, \dots, n$ entonces

7.1. Pruebas para el parámetro de localidad

$$\begin{aligned}
 E\{W_i\} &= 0, \\
 \bar{Z} &= \frac{1}{n} \sum_{i=1}^n Z_i = \frac{1}{n} \sum_{i=1}^n (W_i + 1) = \bar{W} + 1, \\
 S_*^2 &= \frac{1}{n} \sum_{i=1}^n (W_i - \bar{W})^2 = S^2.
 \end{aligned}$$

Finalmente

$$Cov\{\bar{Z}, S^2\} = E\{(\bar{W} + 1)S_*^2\} - E\{S^2\} = E\{\bar{W}S_*^2\} + E\{S_*^2\} - E\{S^2\} = E\{\bar{W}S_*^2\}$$

La siguiente identidad es de gran utilidad ([Casella y Berger, 2002](#)) para calcular $E\{\bar{W}S_*^2\}$:

$$S_*^2 = \frac{1}{n} \sum_{i=1}^n (W_i - \bar{W})^2 = \frac{1}{2n^2} \sum_{i=1}^n \sum_{j=1}^n (W_i - W_j)^2.$$

Entonces

$$\begin{aligned}
 E\{\bar{W}S_*^2\} &= E\left\{\frac{1}{n} \sum_{k=1}^n W_k \frac{1}{2n^2} \sum_{i=1}^n \sum_{j=1}^n (W_i - W_j)^2\right\} \\
 &= \frac{1}{2n^3} E\left\{\sum_{k=1}^n W_k \sum_{i=1}^n \sum_{j=1}^n (W_i - W_j)^2\right\}.
 \end{aligned}$$

Cuando $i = j$ el término $W_i - W_j = 0$, por lo tanto la doble sumatoria tiene $n(n-1)$ términos distintos de cero. Para cada k la esperanza es distinta de cero para $2(n-1)$ términos $k = i$ o $k = j$, por lo tanto:

$$\begin{aligned}
 E\{\bar{W}S_*^2\} &= \frac{2n(n-1)}{2n^3} E\{W_i(W_i - W_j)^2\} = \frac{n-1}{n^2} E\{W_i^3\} \\
 &= \frac{n-1}{n^2} E\{(Z_i - 1)^3\} = \frac{8(n-1)}{n^2}.
 \end{aligned}$$

7.1. Pruebas para el parámetro de localidad

Sustituyendo los resultados previos en las ecuaciones (7.5) y (7.6) se obtiene:

$$E \left\{ \frac{\bar{Z}}{S} \right\} \approx \frac{1}{\sqrt{2}} + \frac{3}{32\sqrt{2}} \frac{(n-1)^2}{n^3} \left(60 - 4 \frac{n-3}{n-1} \right) - \frac{\sqrt{2}(n-1)}{n^2} := \mu_n. \quad (7.7)$$

$$Var \left\{ \frac{\bar{Z}}{S} \right\} \approx \frac{1}{n} + \frac{(n-1)^2}{32n^3} \left(60 - 4 \frac{n-3}{n-1} \right) - \frac{2(n-1)}{n^2} := \sigma_n^2. \quad (7.8)$$

Entonces podemos aproximar la distribución de R con la distribución $N(\mu_n, \sigma_n^2)$, con μ_n y σ_n^2 dados en las ecuaciones (7.7) y (7.8). En la Figura 7.4 se muestra la distribución empírica de R y la distribución normal con la media y la varianza obtenida usando las ecuaciones (7.5) y (7.6). La distribución empírica de la estadística R fue obtenida utilizando simulación Monte Carlo.

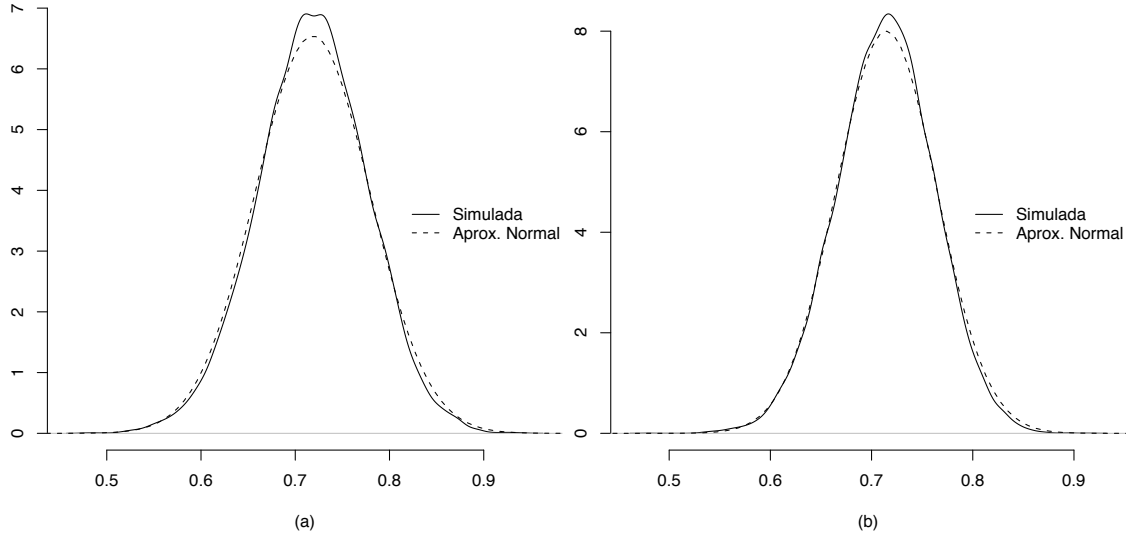


Figura 7.4: Distribución empírica de la estadística R y otra aproximación normal, $B = 20,000$ réplicas de tamaño n , (a) $n = 200$, (b) $n = 300$.

Es importante notar que esta aproximación normal coincide con la obtenida anteriormente en la ecuación (7.4), en efecto,

$$RC := \sqrt{n} \left(R - \frac{1}{\sqrt{2}} \right)$$

7.1. Pruebas para el parámetro de localidad

$$\begin{aligned}
 \mu_n^* &:= E(RC) \approx E\sqrt{n} \left(R - \frac{1}{\sqrt{2}} \right) \\
 &= \sqrt{n} \left[\frac{3}{32\sqrt{2}} \frac{(n-1)^2}{n^3} \left(60 - 4\frac{n-3}{n-1} \right) - \frac{\sqrt{2}(n-1)}{n^2} \right] \\
 \sigma_n^{2*} &:= Var(RC) \approx Var \left[\sqrt{n} \left(R - \frac{1}{\sqrt{2}} \right) \right] \\
 &= n \left[\frac{1}{n} + \frac{(n-1)^2}{32n^3} \left(60 - 4\frac{n-3}{n-1} \right) - \frac{2(n-1)}{n^2} \right]
 \end{aligned}$$

al tomar el límite:

$$\begin{aligned}
 \lim_{n \rightarrow \infty} \mu_n^* &= 0, \\
 \lim_{n \rightarrow \infty} \sigma_n^{2*} &= \frac{3}{4}.
 \end{aligned}$$

En la Tabla 7.2 se comparan los cuantiles de $\alpha/2$ y $1 - \alpha/2$ para $\alpha = 0.05$ de RC empírica, RC asintótica de la ecuación (7.4) y RC aproximada por una normal, con media μ_n^* y varianza σ_n^{2*} , la cual se denota por RC^* para varios tamaños de muestra. De esta Tabla se observa que cuando el tamaño de muestra es pequeño la distribución asintótica de RC no aproxima bien la cola izquierda de la distribución, algo que ya se esperaba (veáse Figura 7.2). La distribución de RC^* parece no ajustar bien la cola derecha de la distribución. Con fines prácticos puede usarse la distribución aproximada de R con n finito.

Tabla 7.2: Comparación de cuantiles de RC empírica, RC asintótica y RC^* .

n	RC empírica		RC asintótica		RC^*	
	$q_{\alpha/2}$	$q_{1-\alpha/2}$	$q_{\alpha/2}$	$q_{1-\alpha/2}$	$q_{\alpha/2}$	$q_{1-\alpha/2}$
50	-1.1993	1.7983	-1.6973	1.6973	-1.3598	1.9792
100	-1.3490	1.7667	-1.6973	1.6973	-1.4589	1.9077
150	-1.4395	1.7354	-1.6973	1.6973	-1.5033	1.8726
200	-1.4698	1.7374	-1.6973	1.6973	-1.5297	1.8509
300	-1.5400	1.7051	-1.6973	1.6973	-1.5610	1.8243
500	-1.5929	1.7099	-1.6973	1.6973	-1.5922	1.7968
1000	-1.6396	1.7057	-1.6973	1.6973	-1.6234	1.7684

7.1. Pruebas para el parámetro de localidad

7.1.2. Estadística de Cox y Oakes

Sea X_1, \dots, X_n una muestra aleatoria de una distribución F y defínanse las estadísticas:

$$\begin{aligned}
 V_{kk} &= \left[n + \frac{1}{\bar{X}} \sum_{i=1}^n X_i \left[\log \frac{1}{\bar{X}} X_i \right]^2 - \frac{\sum_{i=1}^n X_i \left[\log \frac{1}{\bar{X}} X_i \right]^2}{n\bar{X}^2} \right]^{-1}, \\
 U_{k0} &= n + \sum_{i=1}^n \log X_i - \frac{\sum_{i=1}^n X_i \log X_i}{\bar{X}}, \\
 Cox &= U_{ko}(V_{kk})^{1/2}.
 \end{aligned} \tag{7.9}$$

Si se desea probar el juego de hipótesis siguiente:

$$H_0 : F \text{ es exponencial vs } H_1 : F \text{ no es exponencial,} \tag{7.10}$$

bajo la hipótesis nula y cuando el tamaño de muestra es grande, la estadística de [Cox y Oakes \(1984\)](#) tiene distribución aproximadamente normal estándar y por lo tanto se rechaza H_0 al nivel de significancia α si $Cox < Z_{\alpha/2}$ o $Cox > Z_{1-\alpha/2}$ donde Z_α es el cuantil α de la distribución normal estándar.

Bajo la hipótesis nula (7.1), dada la muestra aleatoria Y_1, \dots, Y_n , sea

$$W_i = (Y_i - \xi_0)^2 \sim \Gamma\left(\frac{1}{2}, 2\omega^2\right), \quad i = 1, \dots, n.$$

Entonces para cualesquiera parejas $W_i, W_j, i < j, j = 1, \dots, n, X_k = \frac{W_i + W_j}{2} \sim \exp(\lambda)$, $k = 1, \dots, n(n-1)/2, \lambda = \omega^2$, es decir, bajo la hipótesis nula planteada en (7.1) las X'_k s deben ser exponencial. Por lo tanto el juego de hipótesis (7.1) es equivalente a (7.10); sin embargo, en estas circunstancias la distribución de la estadística de Cox y Oakes bajo la hipótesis nula ya no será normal estándar, ya que las X'_k s no son independientes, pero podemos obtener la distribución empírica de la distribución de Cox bajo esta situación, empleando el procedimiento que se describe a continuación:

- 1) Fijar $n, \xi = \xi_0, \omega$ y γ .
- 2) Generar una muestra de tamaño n de la distribución $SN_D(\xi, \omega, \gamma)$.

7.1. Pruebas para el parámetro de localidad

- 3) Generar las X'_k s.
- 4) Calcular Cox usando la ecuación (7.9) y las X'_k s del paso anterior.
- 5) Repetir los pasos 2-4 B veces.

Al final del proceso se tendrán B realizaciones de la estadística de Cox y Oakes.

En la Tabla 7.3 se muestran los valores críticos $C_1(n, \alpha/2)$ y $C_2(n, 1 - \alpha/2)$ para varios tamaños de muestra n y $\alpha = \{0.01, 0.025, 0.05, 0.1\}$, obtenidos a partir de la distribución empírica de Cox .

Se Rechaza la hipótesis planteada en (7.1) al nivel de significancia α si $Cox < C_1(n, \frac{\alpha}{2})$ o $Cox > C_2(n, 1 - \frac{\alpha}{2})$, donde $C_1(n, \frac{\alpha}{2})$ y $C_2(n, 1 - \frac{\alpha}{2})$ son los cuantiles $(\alpha/2)100$ y $(1 - \alpha/2)100$ de la distribución de Cox y Oakes bajo la hipótesis nula.

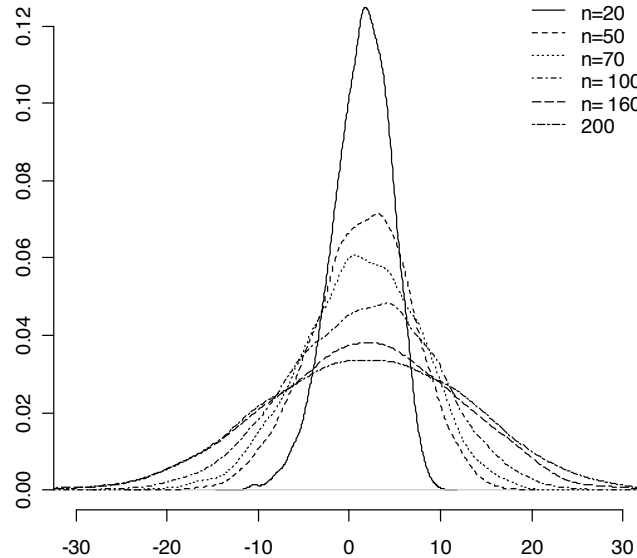


Figura 7.5: Distribución empírica de la estadística Cox para varios tamaños de muestra obtenida mediante simulación Monte Carlo con $B = 1,000$ réplicas de tamaño n .

El juego de hipótesis (7.1) puede probarse usando una prueba de razón de verosimilitudes generalizada, para lo cual hay que estimar los parámetros ω y γ bajo la hipótesis nula y ξ , ω y γ en todo el espacio de parámetros y entonces aplicar los resultados bien conocidos de la distribución asintótica de la estadística de prueba.

7.1. Pruebas para el parámetro de localidad

Tabla 7.3: Valores críticos de la estadística de Cox y Oakes obtenida mediante simulación Monte Carlo con $B = 1,000$ réplicas de tamaño n .

n	Nivel de significancia α							
	0.01		0.025		0.05		0.10	
	C_1	C_2	C_1	C_2	C_1	C_2	C_1	C_2
20	-8.2741	8.1281	-6.8835	7.5214	-5.7502	6.8719	-4.4018	6.2216
30	-11.2864	10.6461	-9.1577	9.5813	-7.5826	8.6365	-6.0052	7.6995
50	-13.4511	14.2103	-11.6506	12.7560	-10.0014	11.5903	-8.1068	10.0390
60	-15.8078	15.5077	-13.1658	13.7763	-11.3936	12.4386	-9.0449	10.7547
70	-16.8367	16.8292	-14.1441	14.9854	-11.9054	13.2793	-9.7844	11.4094
80	-18.5840	17.9767	-15.4448	16.0235	-12.9841	14.5285	-10.4368	12.5344
90	-19.5596	18.8758	-16.5080	16.7524	-14.1359	15.0329	-11.5833	12.9976
100	-20.3390	20.5061	-17.8202	18.4115	-14.9484	16.3708	-12.2494	14.3035
120	-23.5874	24.3895	-20.0977	21.1916	-15.9613	18.2007	-13.0257	15.0859
140	-27.2734	25.0833	-21.5421	20.3905	-20.2733	17.9353	-15.2953	15.8792
160	-27.6644	25.8641	-22.5455	21.7342	-19.7300	19.9780	-16.3596	17.7039
180	-28.0178	26.2460	-24.9755	23.3665	-20.3452	22.0994	-16.8164	19.9522
200	-27.8258	27.3817	-22.2494	24.9795	-19.4636	22.1342	-16.3792	19.6331
250	-32.0626	33.5787	-26.7885	28.6118	-22.8847	25.5988	-18.6430	22.2674
300	-32.0059	37.1124	-29.5188	33.2238	-25.0772	28.9696	-20.2097	24.3302

7.1.3. Razón de verosimilitudes

Para implementar la prueba de razón de verosimilitudes hay que calcular estimadores de ξ , ω y γ bajo la hipótesis nula y bajo la alternativa con la función de verosimilitud dada por la ecuación (6.3). *Estimación de parámetros bajo la hipótesis nula*

Bajo la hipótesis nula, la ecuación (6.3) se puede escribir como:

$$\prod_{i=1}^n f_{Y_i}(y_i; \xi_0, \omega, \gamma) = \prod_{i=1}^n \frac{2}{\omega} \phi\left(\frac{y_i - \xi_0}{\omega}\right) \Phi\left[\gamma\left(\frac{y_i - \xi_0}{\omega}\right)\right]. \quad (7.11)$$

Los estimadores de σ y γ se obtienen al maximizar de manera numérica (7.11) usando rutinas diseñadas para tal fin, como por ejemplo **optim** del paquete R.

Estimación de parámetros en todo el espacio de parámetros

El cálculo de los estimadores de máxima verosimilitud se realiza utilizando el procedimiento de Sartori (2006).

Se rechaza H_0 si y solo sí $-2 \log \lambda(\mathbf{y}) \geq \chi_{1,1-\alpha}^2$

7.1. Pruebas para el parámetro de localidad

7.1.4. Comparación de potencias

En la Figura 7.6 se muestran las potencias de la prueba R , la de Cox y la de razón de verosimilitudes (R. V.) para varios tamaños de muestra y valores del parámetro obtenida mediante simulación Monte Carlo con $B = 10,000$ réplicas de los tamaños indicados.

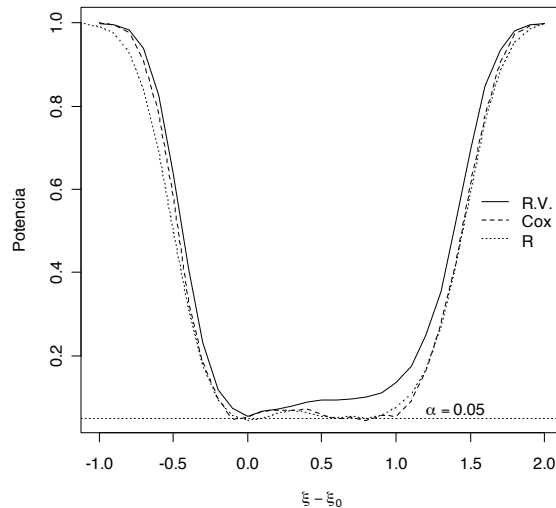


Figura 7.6: Potencia de la prueba R , Cox y de razón de verosimilitudes para muestras de tamaño $n = 100$ de $SN_D(0, 1, 1)$ obtenidas usando simulación con $B = 10,000$.

Cabe destacar que aunque la prueba de razón de verosimilitudes generalizada es ligeramente más potente que la prueba R , esta última presenta algunas ventajas:

- Es simple.
- No se requiere calcular ningún parámetro, lo cual es muy importante ya que en algunas ocasiones los estimadores máximo verosímiles pueden ser difíciles de calcular.
- Los valores críticos se obtienen de la distribución normal estándar para cualquier n .

7.1.5. Ejemplo de aplicación en epidemiología

Roberts (1966) desarrolló un modelo para estudiar la distribución del tiempo en el que un gemelo se resfría por primera vez en su vida. Sean X y Y las edades a las que cada uno de los gemelos se resfría por primera vez. Suponiendo que $X, Y \sim N(\mu, \sigma^2)$

7.1. Pruebas para el parámetro de localidad

y $Cor(X, Y) = \rho$, dicho autor logró mostrar que si $Z = \min\{X, Y\}$ y $W = (Z - \mu)/\sigma$ entonces:

$$f_W(w) = \frac{2}{\sqrt{2\pi}} \Phi \left(-w \sqrt{\frac{1-\rho}{1+\rho}} \right) \exp \left\{ -\frac{w^2}{2} \right\} I_{(-\infty, \infty)}(w). \quad (7.12)$$

Roberts (1966) señala que debido a que los gemelos conviven demasiado, entonces cuando se enferma uno de ellos es casi seguro que el otro también lo hará casi enseñada, por lo tanto solo se podrá observar Z e indica que el modelo puede ser útil para estudiar tiempos de ocurrencia de otras enfermedades en gemelos como leucemia.

De (7.12) se concluye que $W \sim SN_D(0, 1, \gamma)$ con $\gamma = -\sqrt{\frac{1-\rho}{1+\rho}}$ como ha sido observado por Gupta y Chen (2001).

El artículo de Roberts (1966) se centra en inferencias sobre el parámetro ρ asumiendo que μ y σ son conocidos, aunque también señala que en ocasiones se puede estar interesados en probar hipótesis como $H_0 : \mu = \mu_0$, esta hipótesis puede probarse utilizando (7.3) sin necesidad de suponer que σ es conocido.

Roberts reporta el tiempo en el que el primero de cada uno de 10 pares de gemelos se enferma por primera vez en su vida (el tiempo se mide desde el momento de nacimiento, en meses).

Tabla 7.4: Tiempos al primer resfriado de gemelos.

i	1	2	3	4	5	6	7	8	9	10
Z_i	13.88	7.79	17.63	12.32	9.59	15.83	13.10	11.54	10.70	14.72

y supone que $\mu = 14$ y $\sigma = 3$. A continuación se prueba $\mu = 14$ utilizando la prueba R .

De la ecuación (7.3) se obtiene que $R = 0.7959$. Para $\alpha = 0.05$, de la Tabla 7.1, se obtienen las constantes críticas $C_1(10, 0.025) = 0.4965$ y $C_2(10, 0.975) = 1.3756$. Como $0.4965 < 0.7959 < 1.3756$ no hay evidencia para pensar que la suposición hecha por Roberts sea incorrecta.

7.2. Comparación de medias de dos poblaciones normales asimétricas

7.2.1. Dos poblaciones con parámetros de escala y forma comunes

Sean X_1, \dots, X_n una m. a. de $SN_D(\xi_1, \omega, \gamma)$ y Y_1, \dots, Y_m una m. a. de $SN_D(\xi_2, \omega, \gamma)$. Se supone además que las muestras son independientes entre ellas. Se desea probar el juego de hipótesis:

$$H_0 : \xi_1 = \xi_2, \omega > 0, \gamma \in \mathbb{R} \text{ vs } H_1 : \xi_1 \neq \xi_2, \omega > 0, \gamma \in \mathbb{R}. \quad (7.13)$$

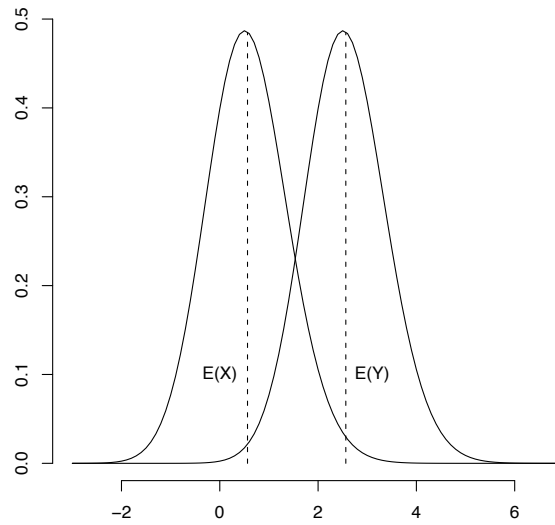


Figura 7.7: Comparación de medias de dos poblaciones normales asimétricas con parámetros de escala y forma comunes, pero desconocidos.

Note que el juego de hipótesis planteado es equivalente a preguntar si las medias de las dos poblaciones son iguales o son diferentes. Además:

$$\begin{aligned} \Omega &= \{\boldsymbol{\theta} = (\xi_1, \xi_2, \omega, \gamma)' \in \mathbb{R}^4 : \xi_1, \xi_2, \gamma \in \mathbb{R}, \omega > 0\} \\ \Omega_0 &= \{\boldsymbol{\theta} = (\xi, \omega, \gamma)' \in \mathbb{R}^3 : \xi, \gamma \in \mathbb{R}, \omega > 0\}. \end{aligned}$$

Para resolver el problema se propone utilizar la prueba de razón de verosimilitudes generalizada, para lo cual es necesario calcular estimadores de λ en la ecuación (3.9), en este caso se puede escribir de la siguiente manera:

7.2. Comparación de medias de dos poblaciones normales asimétricas

$$\lambda(\mathbf{x}, \mathbf{y}) = \frac{\sup_{\boldsymbol{\theta} \in \Omega_0} L(\boldsymbol{\theta}; \mathbf{x}, \mathbf{y})}{\sup_{\boldsymbol{\theta} \in \Omega} L(\boldsymbol{\theta}; \mathbf{x}, \mathbf{y})}. \quad (7.14)$$

Estimación de parámetros bajo la hipótesis nula

Para estimar los parámetros bajo la hipótesis nula se utiliza el método de máxima verosimilitud modificado propuesto por Sartori (2006). Bajo H_0 todas las observaciones vienen de una misma población, es decir, $\mathbf{Z} := (X_1, \dots, X_n, Y_1, \dots, Y_m)'$ es una muestra aleatoria de tamaño $n + m$ de la distribución $SN_D(\xi, \omega, \gamma)$. Sean $\hat{\xi}$, $\hat{\omega}$ y $\hat{\gamma}$ los estimadores de dichos parámetros, entonces el numerador de (7.14) es $L(\hat{\xi}, \hat{\omega}, \hat{\gamma}; \mathbf{x}, \mathbf{y})$

Estimación de parámetros en todo el espacio de parámetros

La función de verosimilitud conjunta para las muestras es:

$$\begin{aligned} L(\xi_1, \xi_2, \omega, \gamma; \mathbf{x}, \mathbf{y}) &= \prod_{i=1}^n f_{X_i}(x_i; \xi_1, \sigma, \gamma) \prod_{j=1}^m f_{Y_j}(y_j; \xi_2, \sigma, \gamma) \\ &\propto \omega^{-(n+m)} \exp \left\{ -\frac{1}{2\omega^2} \left[\sum_{i=1}^n (x_i - \xi_1)^2 + \sum_{j=1}^m (y_j - \xi_2)^2 \right] \right\} \\ &\quad \prod_{i=1}^n \Phi \left[\gamma \left(\frac{x_i - \xi_1}{\omega} \right) \right] \prod_{j=1}^m \Phi \left[\gamma \left(\frac{y_j - \xi_2}{\omega} \right) \right]. \end{aligned}$$

Sean $l(\xi_1, \xi_2, \omega, \gamma; \mathbf{x}, \mathbf{y}) := \log L(\xi_1, \xi_2, \omega, \gamma; \mathbf{x}, \mathbf{y})$ y $\Lambda(x) = \log \Phi(x)$, entonces:

$$\begin{aligned} l(\xi_1, \xi_2, \omega, \gamma; \mathbf{x}, \mathbf{y}) &\propto -(n+m) \log \omega - \frac{1}{2\omega^2} \left[\sum_{i=1}^n (x_i - \xi_1)^2 + \sum_{i=1}^m (y_i - \xi_2)^2 \right] \\ &\quad + \sum_{i=1}^n \Lambda \left[\gamma \left(\frac{x_i - \xi_1}{\omega} \right) \right] + \sum_{j=1}^m \Lambda \left[\gamma \left(\frac{y_j - \xi_2}{\omega} \right) \right]. \end{aligned}$$

Las derivadas parciales con respecto a cada uno de los parámetros de interés se presentan a continuación.

7.2. Comparación de medias de dos poblaciones normales asimétricas

$$\frac{\partial l(\xi_1, \xi_2, \omega, \gamma; \mathbf{x}, \mathbf{y})}{\partial \xi_1} = \frac{1}{\omega^2} \sum_{i=1}^n (x_i - \xi_1) - \frac{\gamma}{\omega} \sum_{i=1}^n \Lambda' \left[\gamma \left(\frac{x_i - \xi_1}{\omega} \right) \right]. \quad (7.15)$$

$$\frac{\partial l(\xi_1, \xi_2, \omega, \gamma; \mathbf{x}, \mathbf{y})}{\partial \xi_2} = \frac{1}{\omega^2} \sum_{j=1}^m (y_j - \xi_2) - \frac{\gamma}{\omega} \sum_{j=1}^m \Lambda' \left[\gamma \left(\frac{y_j - \xi_2}{\omega} \right) \right]. \quad (7.16)$$

$$\begin{aligned} \frac{\partial l(\xi_1, \xi_2, \omega, \gamma; \mathbf{x}, \mathbf{y})}{\partial \omega} &= -\frac{\gamma}{\omega^2} \left\{ \sum_{i=1}^n (x_i - \xi_1) \Lambda' \left[\gamma \left(\frac{x_i - \xi_1}{\omega} \right) \right] \right\} \\ &\quad - \frac{\gamma}{\omega^2} \left\{ \sum_{j=1}^m (y_j - \xi_2) \Lambda' \left[\gamma \left(\frac{y_j - \xi_2}{\omega} \right) \right] \right\} \\ &\quad - \frac{(n+m)}{\omega} + \frac{1}{\omega^3} \left[\sum_{i=1}^n (x_i - \xi_1)^2 + \sum_{j=1}^m (y_j - \xi_2)^2 \right]. \end{aligned} \quad (7.17)$$

$$\begin{aligned} \frac{\partial l_1(\xi_1, \xi_2, \omega, \gamma; \mathbf{x}, \mathbf{y})}{\partial \gamma} &= \frac{1}{\omega} \left\{ \sum_{i=1}^n (x_i - \xi_1) \Lambda' \left[\gamma \left(\frac{x_i - \xi_1}{\omega} \right) \right] \right\} \\ &\quad + \frac{1}{\omega} \left\{ \sum_{j=1}^m (y_j - \xi_2) \Lambda' \left[\gamma \left(\frac{y_j - \xi_2}{\omega} \right) \right] \right\}. \end{aligned} \quad (7.18)$$

Es claro que es necesario igualar y resolver las ecuaciones (7.15), (7.16), (7.17) y (7.18) de manera numérica o bien maximizar de forma directa $l(\xi_1, \xi_2, \omega, \gamma; \mathbf{x}, \mathbf{y})$ utilizando algún algoritmo que no requiera derivadas; sin embargo en cualquiera de los casos hay que tener cuidado con la solución obtenida por el método numérico, ya que puede darnos máximos o mínimos locales e incluso puntos de inflexión.

Cero trivial

Sea $\hat{\xi}_1 = \bar{x}$, $\hat{\xi}_2 = \bar{y}$, $\hat{\omega}^2 = \frac{1}{n+m} \left[\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{j=1}^m (y_j - \bar{y})^2 \right]$ y $\hat{\gamma} = 0$. Las ecuaciones (7.15), (7.16), (7.17) y (7.18) evaluadas en $\hat{\theta}_0 = (\hat{\xi}_1, \hat{\xi}_2, \hat{\omega}, 0)'$ son 0 (ver apéndice B), pero no se puede concluir nada acerca de la naturaleza de esta solución, lo cual se muestra al obtener la matriz de información de Fisher observada.

7.2. Comparación de medias de dos poblaciones normales asimétricas

Matriz de información de Fisher observada

Sea $\boldsymbol{\theta} = (\xi_1, \xi_2, \omega, \gamma)'$, la matriz de información de Fisher observada $IO(\boldsymbol{\theta})$ está dada por:

$$IO(\boldsymbol{\theta}) = - \left[\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right], \quad i, j = 1, \dots, 4. \quad (7.19)$$

La Matriz de Información de Fisher evaluada en $\hat{\boldsymbol{\theta}}_0$ está dada por:

$$IO(\hat{\boldsymbol{\theta}}_0) = - \begin{bmatrix} -\frac{n}{\hat{\omega}^2} & 0 & 0 & -\frac{n}{\hat{\omega}} \sqrt{\frac{2}{\pi}} \\ 0 & -\frac{m}{\hat{\omega}^2} & 0 & -\frac{m}{\hat{\omega}} \sqrt{\frac{2}{\pi}} \\ 0 & 0 & -\frac{2(n+m)}{\hat{\omega}^2} & 0 \\ -\frac{n}{\hat{\omega}} \sqrt{\frac{2}{\pi}} & -\frac{m}{\hat{\omega}} \sqrt{\frac{2}{\pi}} & 0 & -\frac{2(n+m)}{\pi} \end{bmatrix}, \quad (7.20)$$

como se muestra en el apéndice B.

Como el determinante de IO es 0 no se puede concluir nada acerca del punto $\hat{\boldsymbol{\theta}}_0 = (\hat{\xi}_1, \hat{\xi}_2, \hat{\omega}, 0)$, ver [Apostol \(1967\)](#). Por lo tanto se debe ser cuidadoso al momento de realizar la optimización de la función $l(\xi_1, \xi_2, \omega, \gamma; \mathbf{x}, \mathbf{y})$, ya que algoritmos de optimización tipo Newton o bien otros que utilicen derivadas podrían verse afectados por este hecho.

Sean $\tilde{\xi}_1, \tilde{\xi}_2, \tilde{\omega}$ y $\tilde{\gamma}$ los estimadores de máxima verosimilitud de ξ_1, ξ_2, ω y γ respectivamente, entonces el denominador de (7.14) es $L(\tilde{\xi}_1, \tilde{\xi}_2, \tilde{\omega}, \tilde{\gamma}; \mathbf{x}, \mathbf{y})$. Aplicando el Teorema 3.2 y para n y m son grandes se rechaza H_0 en (7.13) al nivel de significancia α si y solo si:

$$-2\{\log L(\hat{\xi}_1, \hat{\xi}_2, \hat{\omega}, \hat{\gamma}; \mathbf{x}, \mathbf{y}) - \log L(\tilde{\xi}, \tilde{\omega}, \tilde{\gamma}; \mathbf{x}, \mathbf{y})\} \geq \chi_{1,1-\alpha}^2.$$

Estudio de tamaño y potencia de la prueba

Tamaño

En la Tabla 7.5 se muestran los tamaños de la prueba de razón de verosimilitudes obtenidos mediante un experimento de simulación Monte Carlo con $B = 1,000$ muestras de los tamaños $n = m$ indicados y $\xi_1 = \xi_2 = 0$, $\omega = 1$ y γ como se muestra en la Tabla. Todos los cálculos fueron realizados utilizando $\alpha = 0.05$.

7.2. Comparación de medias de dos poblaciones normales asimétricas

Tabla 7.5: Tamaño de la prueba de razón de verosimilitudes para probar si dos poblaciones SN tienen la misma media, parámetros de escala y forma comunes pero desconocidos, $B = 1,000$, $n = m$, $\alpha = 0.05$.

n	γ					
	0	1	3	5	7	10
50	0.052	0.058	0.052	0.065	0.056	0.055
100	0.052	0.050	0.045	0.053	0.062	0.061
150	0.047	0.051	0.041	0.063	0.044	0.048
200	0.046	0.052	0.043	0.058	0.057	0.055

Potencia

Con la finalidad de estudiar el desempeño de la potencia de la prueba de razón de verosimilitudes descrita anteriormente, se compara con la prueba no paramétrica de Wilcoxon-Mann-Whitney como se describe en [Conover y Iman \(1981\)](#). La estadística de prueba para probar el juego de hipótesis planteado es:

$$t_R = \frac{T}{\left[\frac{N-1}{N-2} - \frac{1}{N-2}T^2\right]^{1/2}}, \quad (7.21)$$

donde:

$$T = \frac{S - n(N+1)/2}{\left[\frac{nm}{N(N-1)} \sum_{i=1}^N R_i^2 - \frac{nm(N+1)^2}{4(N-1)}\right]^{1/2}},$$

$$N = n + m,$$

$$S = \sum_{i=1}^n R_i.$$

R_i es el rango de las X 's obtenido considerando el conjunto completo de observaciones $\mathbf{Z} := (X_1, \dots, X_n, Y_1, \dots, Y_m)'$.

Se rechaza la hipótesis nula al nivel de significancia α si $t_R > t_{1-\alpha/2, n+m-2}$ o si $t_R < -t_{1-\alpha/2, n+m-2}$.

En las Figuras 7.8 y 7.9 se observa la potencia de las dos pruebas para $\omega = 1$, $\gamma = \{0, 3, 5\}$ y $\Delta\xi = \xi_1 - \xi_2 \in [-1, 1]$ obtenidas mediante simulación Monte Carlo con $B = 1,000$ muestras de tamaño $n = m = 50$ y $\alpha = 0.05$. En dichas Figuras se observa que cuando $\gamma = 0$, las potencias de las dos pruebas son muy similares, y es

7.2. Comparación de medias de dos poblaciones normales asimétricas

prácticamente imposible decidir cual de las dos pruebas tiene mejor potencia. Cuando $\gamma = \{3, 5\}$ resulta claro que la prueba de razón de verosimilitudes generalizada tiene mayor potencia que la prueba basada en rangos.

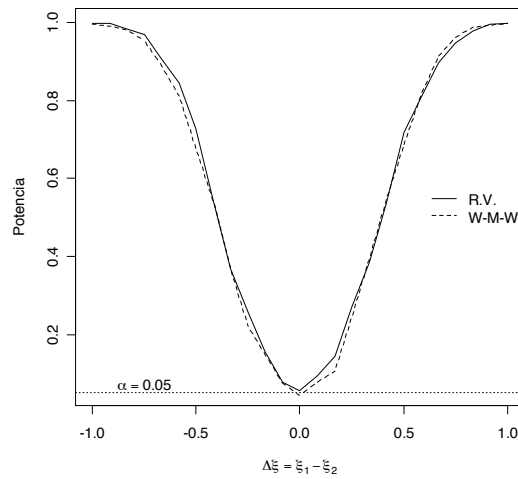


Figura 7.8: Potencia de la prueba de rangos y razón de verosimilitudes (r. v.) para probar si dos poblaciones SN tienen la misma media, parámetros de escala y forma comunes, $\gamma = 0$.

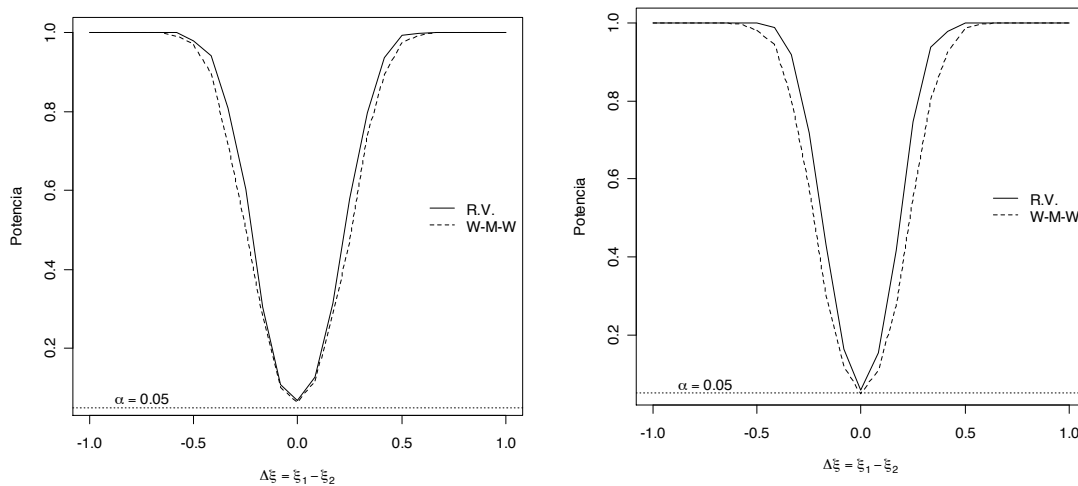


Figura 7.9: Potencia de la prueba de rangos y razón de verosimilitudes (r. v.) para probar si dos poblaciones SN tienen la misma media, parámetros de escala y forma comunes, $\gamma = 3$ (izquierda), $\gamma = 5$ (derecha).

7.2. Comparación de medias de dos poblaciones normales asimétricas

7.2.2. Dos poblaciones con parámetros de localidad y forma comunes

Sean X_1, \dots, X_n una m. a. de $SN_D(\xi, \omega_1, \gamma)$ y Y_1, \dots, Y_m una m. a. de $SN_D(\xi, \omega_2, \gamma)$. Se supone además que las muestras son independientes entre ellas. Se desea probar el juego de hipótesis:

$$H_0 : \omega_1 = \omega_2, \xi, \gamma \in \mathbb{R} \text{ vs } H_1 : \omega_1 \neq \omega_2, \xi, \gamma \in \mathbb{R}. \quad (7.22)$$

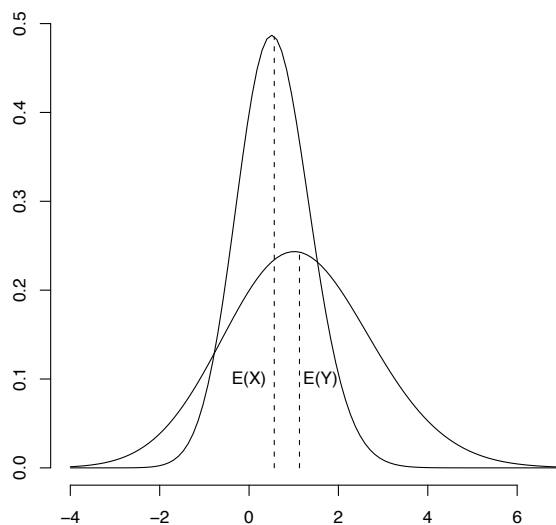


Figura 7.10: Comparación de medias de dos poblaciones normales asimétricas con parámetros de localidad y forma comunes, pero desconocidos.

Note que el juego de hipótesis planteado es equivalente a preguntar si las medias de las dos poblaciones son iguales o diferentes. Además:

$$\begin{aligned} \Omega &= \{\boldsymbol{\theta} = (\xi, \omega_1, \omega_2, \gamma)' \in \mathbb{R}^4 : \xi, \gamma \in \mathbb{R}, \omega_1, \omega_2 > 0\}, \\ \Omega_0 &= \{\boldsymbol{\theta} = (\xi, \omega, \gamma)' \in \mathbb{R}^3 : \xi, \gamma \in \mathbb{R}, \omega > 0\}. \end{aligned}$$

Nuevamente se recurre a la prueba de razón de verosimilitudes para este juego de hipótesis de interés. Para aplicar la prueba será necesario calcular los estimadores de los parámetros bajo la hipótesis nula (el espacio restringido Ω_0) y en todo el espacio de parámetros Ω .

7.2. Comparación de medias de dos poblaciones normales asimétricas

Estimación de parámetros bajo la hipótesis nula

Al igual que en el caso anterior Bajo H_0 todas las observaciones vienen de una misma población, es decir, $\mathbf{Z} := (X_1, \dots, X_n, Y_1, \dots, Y_m)'$ es una muestra aleatoria de tamaño $n+m$ de la distribución $SN_D(\xi, \omega, \gamma)$. Los estimadores de los parámetros se obtienen utilizando el método propuesto por Sartori (2006). Sean $\hat{\xi}$, $\hat{\omega}$, $\hat{\gamma}$ los estimadores de dichos parámetros, entonces el numerador de (7.14) es $L(\hat{\xi}, \hat{\omega}, \hat{\gamma}; \mathbf{x}, \mathbf{y})$

Estimación de parámetros para todo el espacio de parámetros

La función de verosimilitud para las muestras es:

$$\begin{aligned} L(\xi, \omega_1, \omega_2, \gamma; \mathbf{x}, \mathbf{y}) &= \prod_{i=1}^n f_{X_i}(x_i; \xi, \omega_1, \gamma) \prod_{j=1}^m f_{Y_j}(y_j; \xi, \omega_2, \gamma) \\ &\propto \omega_1^{-n} \omega_2^{-m} \exp \left\{ -\frac{1}{2\omega_1^2} \sum_{i=1}^n (x_i - \xi)^2 - \frac{1}{2\omega_2^2} \sum_{j=1}^m (y_j - \xi)^2 \right\} \\ &\quad \prod_{i=1}^n \Phi \left[\gamma \left(\frac{x_i - \xi}{\omega_1} \right) \right] \prod_{j=1}^m \Phi \left[\gamma \left(\frac{y_j - \xi}{\omega_2} \right) \right]. \end{aligned}$$

Sea $l(\xi, \omega_1, \omega_2, \gamma; \mathbf{x}, \mathbf{y}) := \log L_1(\xi, \omega_1, \omega_2, \gamma; \mathbf{x}, \mathbf{y})$ entonces:

$$\begin{aligned} l(\xi, \omega_1, \omega_2, \gamma; \mathbf{x}, \mathbf{y}) &\propto -n \log \omega_1 - m \log \omega_2 - \frac{1}{2\omega_1^2} \sum_{i=1}^n (x_i - \xi)^2 - \frac{1}{2\omega_2^2} \sum_{j=1}^m (y_j - \xi)^2 \\ &\quad + \sum_{i=1}^n \log \Phi \left[\gamma \left(\frac{x_i - \xi}{\omega_1} \right) \right] + \sum_{j=1}^m \log \Phi \left[\gamma \left(\frac{y_j - \xi}{\omega_2} \right) \right]. \end{aligned}$$

La maximización de $l(\xi, \omega_1, \omega_2, \gamma; \mathbf{x}, \mathbf{y})$ se puede realizar utilizando métodos numéricos, como por ejemplo las implementadas en **optim** del paquete R.

Sean $\tilde{\xi}$, $\tilde{\omega}_1$, $\tilde{\omega}_2$ y $\tilde{\gamma}$ los estimadores de dichos parámetros, entonces el denominador de (7.14) es $L(\tilde{\xi}, \tilde{\omega}_1, \tilde{\omega}_2, \tilde{\gamma}; \mathbf{x}, \mathbf{y})$.

Aplicando el Teorema 3.2, para n, m grandes se rechaza H_0 en (7.22) al nivel de significancia α si y solo si:

7.2. Comparación de medias de dos poblaciones normales asimétricas

$$-2\{\log L(\hat{\xi}, \hat{\omega}, \hat{\gamma}; \mathbf{x}, \mathbf{y}) - \log L(\tilde{\xi}, \tilde{\omega}_1, \tilde{\omega}_2, \tilde{\gamma}; \mathbf{x}, \mathbf{y})\} \geq \chi_{1,1-\alpha}^2.$$

Potencia de la prueba

En la Figura 7.11 se observa la potencia de la prueba de razón de verosimilitudes obtenida mediante simulación Monte Carlo con $B = 1,000$ muestras de los tamaños indicados, $\xi_1 = \xi_2 = 0$, $\rho = \omega_2/\omega_1 \in [1, 2]$ y $\gamma = \{0, 3, 5\}$.

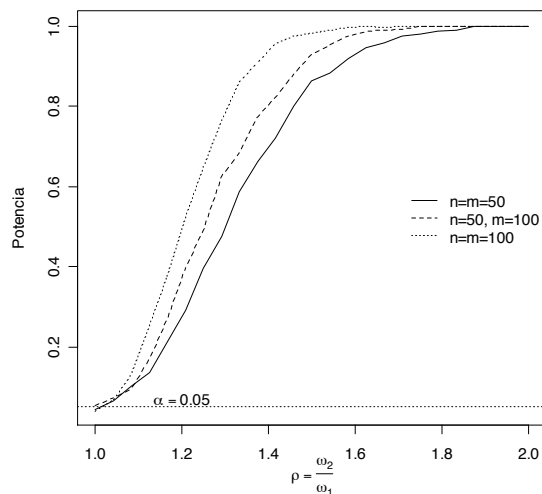


Figura 7.11: Potencia de la prueba de r. v. para probar si dos poblaciones SN tienen la misma media, parámetros de localización y forma comunes pero desconocidos, $\gamma = 3$.

7.2.3. Dos poblaciones con parámetro de forma común

Sean X_1, \dots, X_n una m. a. de $SN_D(\xi_1, \omega_1, \gamma)$ y Y_1, \dots, Y_m una m. a. de $SN_D(\xi_2, \omega_2, \gamma)$. Se desea probar si las observaciones vienen de una distribución normal asimétrica o de una mezcla de dos distribuciones normales asimétricas. Si los datos son de una mezcla, es de interés estimar los parámetros y la proporción de los componentes individuales que forman la misma.

En esta sección se consideran mezclas de dos componentes con el mismo parámetro de forma γ pero que tienen diferente media.

La función de densidad de la mezcla está dada por:

7.2. Comparación de medias de dos poblaciones normales asimétricas

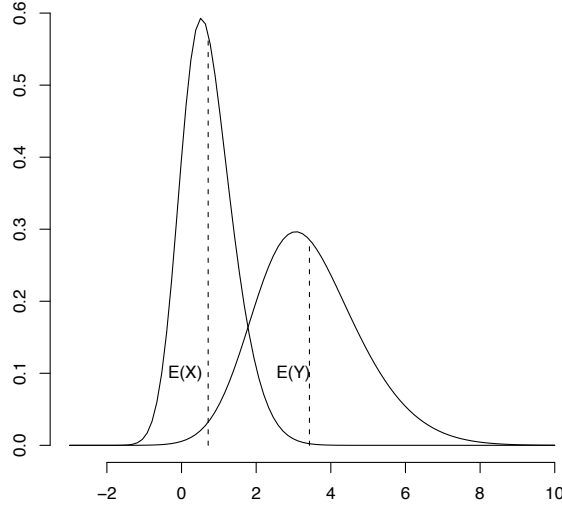


Figura 7.12: Comparación de dos poblaciones normales asimétricas con parámetro de forma común, pero desconocido.

$$f_W(w; \pi, \xi_1, \omega_1, \xi_2, \omega_2, \gamma) = \pi f_{Z_1}(w; \xi_1, \omega_1, \gamma) + (1 - \pi) f_{Z_2}(w; \xi_2, \omega_2, \gamma), \quad (7.23)$$

donde $Z_1 \sim SN_D(\xi_1, \omega_1, \gamma)$, $Z_2 \sim SN_D(\xi_2, \omega_2, \gamma)$ y $\pi \in [0, 1]$ es la proporción del primer componente de la mezcla.

Dadas las muestras X_1, \dots, X_n y Y_1, \dots, Y_m se forma un nuevo vector \mathbf{Z} conteniendo las dos muestras, es decir, $\mathbf{Z} = (X_1, \dots, X_n, Y_1, \dots, Y_m)'$. Bajo el supuesto que \mathbf{Z} tiene densidad (7.23) se plantea el siguiente juego de hipótesis:

$$\begin{aligned} H_0 : \xi_1 = \xi_2, \omega_1 = \omega_2 > 0, \quad \pi \in [0, 1], \gamma \in \mathbb{R}, \\ \text{vs} \\ H_1 : \xi_1 \neq \xi_2 \text{ ó } \omega_1 \neq \omega_2, \quad \pi \in [0, 1], \gamma \in \mathbb{R}. \end{aligned} \quad (7.24)$$

Este juego de hipótesis permitirá decidir si las muestras vienen de una sola población normal asimétrica o de una mezcla de dos normales asimétricas diferentes. Para probar este juego de hipótesis se emplea una prueba de razón de verosimilitudes generalizada para lo cual será necesario calcular los estimadores de todos los parámetros de interés bajo la hipótesis nula y en todo el espacio de parámetros.

7.2. Comparación de medias de dos poblaciones normales asimétricas

Estimadores para todo el espacio de parámetros

En este caso es necesario estimar 6 parámetros, a saber, el vector $\boldsymbol{\theta} = (\pi, \xi_1, \omega_1, \xi_2, \omega_2, \gamma)'$, la función de densidad de Z es dado por una mezcla. Para estimar los parámetros se utiliza el algoritmo EM Generalizado, ya que como se verá más adelante se presentan algunas dificultades al realizar el paso de maximización en el algoritmo iterativo.

Se parte del hecho de que se tiene una sola muestra Z_1, \dots, Z_{n+m} de una mezcla de dos normales asimétricas. Se define $\delta_i = 1$ si la observación i -ésima viene del primer componente de la mezcla y $\delta_i = 0$ en caso contrario. Entonces la verosimilitud para los datos completos W_1, \dots, W_{n+m} con $W_i = (Z_i, \delta_i)$ está dada por:

$$f(w_1, \dots, w_{n+m}; \boldsymbol{\theta}) = \prod_{i=1}^{n+m} \pi^{\delta_i} (1 - \pi)^{1 - \delta_i} (f_{Z_1}(z_i; \xi_1, \omega_1, \gamma))^{\delta_i} (f_{Z_2}(z_i; \xi_2, \omega_2, \gamma))^{1 - \delta_i}. \quad (7.25)$$

Etapa E

Al tomar el logaritmo de (7.25) se obtiene:

$$\begin{aligned} l(\boldsymbol{\theta}) &= \log \left(\prod_{i=1}^{n+m} \pi^{\delta_i} (1 - \pi)^{1 - \delta_i} (f_{Z_1}(z_i; \xi_1, \omega_1, \gamma))^{\delta_i} (f_{Z_2}(z_i; \xi_2, \omega_2, \gamma))^{1 - \delta_i} \right) \\ &= \sum_{i=1}^{n+m} \delta_i \log \pi + (1 - \delta_i) \log(1 - \pi) + \delta_i \log \frac{2}{\omega_1} \phi \left(\frac{z_i - \xi_1}{\omega_1} \right) \Phi \left(\gamma \frac{z_i - \xi_1}{\omega_1} \right) \\ &\quad + (1 - \delta_i) \log \frac{2}{\omega_2} \phi \left(\frac{z_i - \xi_2}{\omega_2} \right) \Phi \left(\gamma \frac{z_i - \xi_2}{\omega_2} \right), \end{aligned}$$

es decir,

$$\begin{aligned} l(\boldsymbol{\theta}) &\propto \sum_{i=1}^{n+m} \delta_i \log \pi + (1 - \delta_i) \log(1 - \pi) \\ &\quad - \sum_{i=1}^{n+m} \delta_i \left(\log \omega_1 + \frac{1}{2} \left(\frac{z_i - \xi_1}{\omega_1} \right)^2 - \log \Phi \left(\gamma \frac{z_i - \xi_1}{\omega_1} \right) \right) \\ &\quad - \sum_{i=1}^{n+m} (1 - \delta_i) \left(\log \omega_2 + \frac{1}{2} \left(\frac{z_i - \xi_2}{\omega_2} \right)^2 - \log \Phi \left(\gamma \frac{z_i - \xi_2}{\omega_2} \right) \right). \end{aligned} \quad (7.26)$$

7.2. Comparación de medias de dos poblaciones normales asimétricas

Como el operador esperanza es lineal, la función

$$Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(j-1)}) = E \left\{ \log f(w_1, \dots, w_{n+m}; \boldsymbol{\theta}) | z_1, \dots, z_{n+m}, \boldsymbol{\theta}^{(j-1)} \right\},$$

es la log-verosimilitud de los datos completos, pero se sustituye δ_i por su valor esperado, es decir,

$$\begin{aligned} p_i &= E\{\delta_i | z_1, \dots, z_{n+m}, \boldsymbol{\theta}^{(j-1)}\} \\ &= 0P(\delta_i = 0 | z_1, \dots, z_{n+m}, \boldsymbol{\theta}^{(j-1)}) + 1P(\delta_i = 1 | z_1, \dots, z_{n+m}, \boldsymbol{\theta}^{(j-1)}) \\ &= P(\delta_i = 1 | z_1, \dots, z_{n+m}, \boldsymbol{\theta}^{(j-1)}). \end{aligned}$$

Luego por el teorema de Bayes:

$$p_i = \frac{P(\delta_i = 1)P(z_i, \boldsymbol{\theta}^{(j-1)} | \delta_i = 1)}{P(\delta_i = 0)P(z_i, \boldsymbol{\theta}^{(j-1)} | \delta_i = 0) + P(\delta_i = 1)P(z_i, \boldsymbol{\theta}^{(j-1)} | \delta_i = 1)}, \quad (7.27)$$

donde

$$\begin{aligned} P(\delta_i = 1)P(z_i, \boldsymbol{\theta}^{(j-1)} | \delta_i = 1) &= \frac{\pi^{(j-1)}}{\omega_1^{(j-1)}} \phi \left(\frac{z_i - \xi_1^{(j-1)}}{\omega_1^{(j-1)}} \right) \Phi \left(\gamma^{(j-1)} \frac{z_i - \xi_1^{(j-1)}}{\omega_1^{(j-1)}} \right), \\ P(\delta_i = 0)P(z_i, \boldsymbol{\theta}^{(j-1)} | \delta_i = 0) &= \frac{(1 - \pi)^{(j-1)}}{\omega_2^{(j-1)}} \phi \left(\frac{z_i - \xi_2^{(j-1)}}{\omega_2^{(j-1)}} \right) \Phi \left(\gamma^{(j-1)} \frac{z_i - \xi_2^{(j-1)}}{\omega_2^{(j-1)}} \right). \end{aligned}$$

Etapa M

Hay que maximizar

7.2. Comparación de medias de dos poblaciones normales asimétricas

$$\begin{aligned}
 Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(j-1)}) &\propto - \sum_{i=1}^{n+m} p_i \left(\log \omega_1 + \frac{1}{2} \left(\frac{z_i - \xi_1}{\omega_1} \right)^2 - \log \Phi \left(\gamma \frac{z_i - \xi_1}{\omega_1} \right) \right) \\
 &\quad - \sum_{i=1}^{n+m} (1 - p_i) \left(\log \omega_2 + \frac{1}{2} \left(\frac{z_i - \xi_2}{\omega_2} \right)^2 - \log \Phi \left(\gamma \frac{z_i - \xi_2}{\omega_2} \right) \right) \\
 &\quad + \sum_{i=1}^{n+m} p_i \log \pi + (1 - p_i) \log(1 - \pi). \tag{7.28}
 \end{aligned}$$

La etapa de maximización no se puede hacer de manera analítica, por lo que se propone recurrir al algoritmo EM Generalizado.

En el EM Generalizado, nada asegura que al seleccionar $\boldsymbol{\theta}$ tal que $Q(\boldsymbol{\theta}; \boldsymbol{\theta}^*) \geq Q(\boldsymbol{\theta}^*; \boldsymbol{\theta}^*)$, π tome valores solo entre 0 y 1. El problema anterior puede evitarse reduciendo la dimensionalidad de la función a maximizar. Si se considera que $\xi_1, \omega_1, \xi_2, \omega_2, \gamma$ son fijos, entonces $Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(j-1)})$ es solo función de π . Ya que

$$\frac{\partial Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(j-1)})}{\partial \pi} = \sum_{i=1}^{n+m} \frac{p_i}{\pi} - \frac{1 - p_i}{1 - \pi} = 0,$$

se obtiene que el valor de π que maximiza $Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(j-1)})$ es:

$$\hat{\pi} = \frac{1}{n + m} \sum_{i=1}^{n+m} p_i.$$

Una vez que conocemos el valor del parámetro π se buscan $\xi_1, \omega_1, \xi_2, \omega_2$ y γ tal que $Q(\boldsymbol{\theta}; \boldsymbol{\theta}^*) \geq Q(\boldsymbol{\theta}^*; \boldsymbol{\theta}^*)$ para lo cual se puede utilizar algún método basado en gradientes.

Sea $H(\boldsymbol{\theta}; \boldsymbol{\theta}^{(j-1)}) = -Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(j-1)})$ entonces el problema de buscar $\boldsymbol{\theta}$ tal que $Q(\boldsymbol{\theta}; \boldsymbol{\theta}^*) \geq Q(\boldsymbol{\theta}^*; \boldsymbol{\theta}^*)$ es equivalente a buscar $\boldsymbol{\theta}$ de tal modo que $H(\boldsymbol{\theta}; \boldsymbol{\theta}^*) \leq H(\boldsymbol{\theta}^*; \boldsymbol{\theta}^*)$, para resolver este problema se puede utilizar el método de descenso más rápido (ver [Burden y Faires, 1996](#)). Para el problema de interés adquiere la forma siguiente:

- 1) Dar un valor inicial a $\boldsymbol{\theta}^{\text{ini}}$.
- 2) Calcular $\mathbf{g} = -\nabla Q(\boldsymbol{\theta}^{\text{ini}}; \boldsymbol{\theta}^*)$.
- 3) Calcular $\boldsymbol{\theta} = \boldsymbol{\theta}^{\text{ini}} + \beta \mathbf{g}$.

7.2. Comparación de medias de dos poblaciones normales asimétricas

Con esta estrategia se asegura que $H(\boldsymbol{\theta}; \boldsymbol{\theta}^*) \leq H(\boldsymbol{\theta}^*; \boldsymbol{\theta}^*)$ para β suficientemente pequeño. Existen varios métodos para seleccionar β como el método de búsqueda lineal o cúbica, o bien dejar un valor fijo.

Las etapas de Esperanza-Maximización se realizan de manera alternada hasta que se alcanza la convergencia, la regla de parada seleccionada para tal fin está basada en la diferencia entre $H(\boldsymbol{\theta}^{(i-1)}; \boldsymbol{\theta}^{(i-1)})$ y $H(\boldsymbol{\theta}^{(i)}; \boldsymbol{\theta}^{(i-1)})$. En el presente trabajo se escribió un programa para R que junto con la biblioteca de funciones `sn` de [Azzalini](#) sirve para calcular los parámetros de una mezcla de dos poblaciones normales asimétricas. El listado del programa puede verse en el apéndice C.

Estimadores para todo el espacio de parámetros

El cálculo de los estimadores bajo la hipótesis nula se realiza utilizando el método de [Sartori \(2006\)](#).

Una vez que se tienen los estimadores bajo la hipótesis nula y en todo el espacio de parámetros se realiza una prueba de razón de verosimilitudes para probar la hipótesis de interés, se rechaza la hipótesis nula al nivel de significancia α si:

$$-2 \log \lambda(\mathbf{Z}) \geq \chi_{2,1-\alpha}^2.$$

Potencia de la prueba

Con la finalidad de estudiar el comportamiento de la prueba de razón de verosimilitudes, se consideran algunas alternativas. En la Tabla 7.6 se presentan los resultados obtenidos realizados mediante simulación Monte Carlo con $B = 1,000$ réplicas de los tamaños indicados con $\alpha = 0.05$. En dicha Tabla se observa que a medida que la distancia entre los parámetros de localidad se incrementa la potencia también se incrementa, como era de esperarse.

Los resultados mostrados en la Tabla 7.6 son congruentes con los obtenidos por [Mendell et al. \(1991\)](#) quienes estudiaron el problema de mezclas de dos componentes en el caso de la distribución normal.

7.2. Comparación de medias de dos poblaciones normales asimétricas

Tabla 7.6: Potencia de la prueba de razón de verosimilitudes para mezclas de poblaciones SN, obtenida mediante simulación Monte Carlo con $B = 1,000$, $\gamma = 1$, $\omega_1 = \omega_2 = 1$.

Tamaño de muestra(n)	Diferencia entre parámetros de loc.		
	$\xi_2 - \xi_1 = 2$	$\xi_2 - \xi_1 = 3$	$\xi_2 - \xi_1 = 4$
	$\pi = 0.5$		
50	0.210	0.633	0.781
100	0.265	0.921	0.952
150	0.371	0.981	0.988
	$\pi = 0.6$		
50	0.258	0.657	0.912
100	0.403	0.739	0.984
150	0.439	0.988	0.988
	$\pi = 0.7$		
50	0.240	0.711	0.937
100	0.394	0.977	0.995
150	0.484	0.997	0.998
	$\pi = 0.8$		
50	0.162	0.399	0.772
100	0.239	0.716	0.985
150	0.281	0.894	0.986
	$\pi = 0.9$		
50	0.102	0.162	0.614
100	0.119	0.267	0.923
150	0.119	0.313	0.970

Capítulo 8

Conclusiones

De los resultados teóricos y de los estudios de simulación realizados en el presente trabajo se puede concluir lo siguiente.

En lo que respecta al problema de estimación de parámetros, las distribuciones a priori propuestas parecen funcionar bien. El estimar los parámetros utilizando la metodología propuesta presenta varias ventajas sobre los métodos de estimación convencionales:

- Proporciona buenos estimadores en casos en los que el método de máxima verosimilitud falla.
- No requiere de maximizar ninguna función.
- Fácil implementación.
- El algoritmo de Metropolis puede extenderse fácilmente al problema de estimación en modelos de regresión con errores normales asimétricos.
- La distribución a posteriori de γ_1 puede utilizarse para decidir si el modelo puede reducirse a una familia más pequeña, es decir la normal y también para analizar violaciones al supuesto de normalidad.

En relación al problema de pruebas de bondad de ajuste,

- Tomando $\gamma \geq 20$ para simular la distribución de las estadísticas D_n , r_n y r_n^* , se asegura que las pruebas sean del tamaño especificado.
- De las Tablas 5.4, 5.5 y 5.6 se observa que los tamaños estimados casi siempre están por debajo del nivel de significancia α , aún con $\gamma = 20$. Es evidente también que la prueba r_n^* es muy conservadora para $\gamma \leq 10$

8. Conclusiones

- Las potencias de las pruebas D_n y D_n^* son muy bajas por lo que no se recomienda su uso.
- Las potencias de las pruebas r_n y r_n^* resultaron ser la más altas para algunas de las alternativas consideradas.
- La prueba r_n^* tiene varias ventajas sobre sus competidores: Es fácil de calcular y el estimador en el que está basada siempre se puede calcular, no es necesario interpolar en Tablas largas para encontrar las constantes críticas.

Finalmente,

- En lo que respecta a Pruebas para el parámetro de localidad la prueba R puede verse como una generalización de la prueba de t clásica realizada cuando se desea probar si la media toma un valor en particular y los datos son normales. La distribución asintótica de la estadística R es útil para obtener los valores críticos necesarios en la prueba, los cuales se pueden obtener directamente de las Tablas de la distribución normal estándar.
- La potencia de las pruebas R y de Cox-Oakes es ligeramente inferior a la de prueba de razón de verosimilitudes generalizada.
- La prueba R presenta ciertas ventajas sobre sus competidores: es simple, no requiere estimar ningún parámetro y los valores críticos se obtienen directamente de la distribución normal estándar.
- En lo que respecta a comparación de medias de dos poblaciones con parámetros de escala y forma comunes pero desconocidos, la prueba de razón de verosimilitudes generalizada resultó ser más potente que la prueba basada en rangos, la diferencia en potencias crece a medida que el parámetro γ crece.
- El algoritmo Esperanza Maximización Generalizado es una herramienta efectiva para la obtención de los estimadores de máxima verosimilitud requeridos para la solución del problema de comparación de poblaciones normales asimétricas con parámetro de forma común. En la Tabla 7.6, se observa que a medida que la distancia entre parámetros de localidad se hace más grande la potencia se incrementa, como era de esperarse.

Referencias

- Apostol, T. M. (1967). *Calculus, Volume 2*. Wiley, New York, segunda edición.
- Arnold, B. C. y Lin, G. D. (2004). Characterizations of the Skew-Normal and Generalized Chi Distributions. *Sankhyā*, 66, 593–606.
- Azzalini, A. (1985). A Class of Distributions which Includes the Normal Ones. *Scandinavian Journal of Statistics*, 12, 171–178.
- Azzalini, A. (2005). The Skew Normal Distribution and Related Multivariate Families. *Scandinavian Journal of Statistics*, 32, 159–188.
- Azzalini, A. (2007). *R package sn: The Skew-normal and Skew-t Distributions (version 0.4-4)*. Università di Padova, Italia.
- Azzalini, A. y Capitanio, A. (1999). Statistical Applications of the Multivariate Skew Normal Distribution. *Journal of the Royal Statistical Society*, 61, 579–602.
- Berger, J., Liseo, B. y Wolpert, R. (1999). Integrated Likelihood Methods for Eliminating Nuisance Parameters. *Statistical Science*, 14, 1–28.
- Burden, R. L. y Faires, J. D. (1996). *Análisis Numérico*. Grupo Editorial Iberoamérica, México, D.F.
- Carlin, B. P. y Louis, T. A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman - Hall/CRC, New York, segunda edición.
- Casella, G. y Berger, R. L. (2002). *Statistical Inference*. DUXBURY, Thomson Learning, segunda edición.
- Chen, J. T., Gupta, A. K. y Nguyen, T. T. (2004). The Density of the Skew Normal Sample Mean and Its Applications. *Journal of Statistical Computation and Simulation*, 74, 487–494.
- Christen, J. A. y Fox, C. (2007). A general purpose scale-independent mcmc algorithm. Inf. Téc. I-07-16, Centro de Investigaciones en Matemáticas, A. C.
- Conover, W. J. y Iman, R. (1981). Rank Transformations as a Bridge Between Parametric and Nonparametric Statistics. *American Statistician*, 35, 124–129.
- Cook, R. D. y Weisberg, S. (1994). *An Introduction to Regression Graphics*. Wiley, New York.

Referencias

- Cox, D. y Oakes, D. (1984). *Analysis of Survival Data*. Chapman Hall.
- Dellaportas, P. y Smith, A. F. M. (1993). Bayesian Inference for Generalized Linear and Proportional Hazards Models via Gibbs Sampling. *Applied Statistics*, 42, 443–459.
- Dempster, A. P., Laird, N. y Rubin, D. (1977). Maximum Likelihood for Incomplete Data Via The EM Algorithm. *Journal of the Royal Statistical Society*, 39, 1–38.
- Gelman, A., Carlin, J. B., Stern, H. S. y Rubin, D. B. (2004). *Bayesian Data Analysis*. Chapman - Hall/CRC, New York, segunda edición.
- Gelman, A. y Rubin, D. B. (1992). Inference From Iterative Simulatiom Using Multiple Sequences. *Statistical Science*, 7, 457–511.
- Genton, M. G., ed. (2004). *Skew-elliptical Distributions and Their Applications: A Journey Beyond Normality*. Chapman - Hall/CRC.
- Gilks, W. R. y Wild, P. (1992). Adaptive Rejection Sampling for Gibbs Sampling. *Applied Statistics*, 41, 337–348.
- Gupta, A. K. y Chen, J. T. (2001). Goodness-Of-Fit Tests for The Skew Normal Distribution. *Communications in Statistics-Simulation and Computation*, 30, 907–930.
- Gupta, A. K., Nguyen, T. T. y Sanqui, J. A. T. (2004). Characterizations of The Skew-Normal Distribution. *Annals of The Institute of Statistical Mathematics*, 56, 351–360.
- Gupta, R. C. y Brown, N. (2001). Reliability Studies Of The Skew Normal Distribution And Its Application To a Strength-Stress Model. *Communications in Statistics-Theory and Methods*, 30, 2427–2445.
- Kendall, M. y Stuart, A. (1973). *The Advanced Theory of Statistics. Volume 2*. Butler and Tanner Ltd.
- Lange, K. (1999). *Numerical Analysis for Statisticians*. Springer-Verlag, New York, segunda edición.
- Lehmann, E. L. y Casella, G. (1998). *Theory of Point Estimation*. Springer-Verlag, New York.
- Liseo, B. y Loperfido, N. (2006). A Note on Reference Priors for the Scalar Skew-Normal Distribution. *Journal of Statistical Planning and Inference*, 136, 373–389.
- Mateu, G., Puig, P. y Pewsey, A. (2007). Goodness-of-Fit Tests for the Skew-Normal Distribution When the Parameters Are Estimated From the Data. *Communications in Statistics-Theory and Methods*, 36, 1735–1755.
- Meintanis, S. G. (2007). A Kolmogorov-Smirnov Type Test for Skew Normal Distributions Based on the Empirical Moment Generating Function. *Journal of Statistical Planning and Inference*, 137, 2681–2688.
- Mendell, N. R., Thode, H. C. y Finch, S. J. (1991). The Likelihood Ratio Test for the Two-Component Normal Mixture Problem: Power and Sample Size Analysis. *Biometrics*, 47, 1143–1148.

Referencias

- Metropolis, N., Rosenbluth, A. W., Teller, M. y Teller, E. (1953). Equations of State Calculations by Fast Computing Machines. *Journal of Chemical Physics*, 21, 1087–1092.
- Mood, A., Graybill, A. y Boes, D. (1974). *Introduction to the Theory of Statistics*. McGraw Hill, tercera edición.
- Nolan (1999). *Stable Distributions*. University Washington DC.
- Oakes, D. (1999). Direct Calculation of the Information Matrix via the EM Algorithm. *Journal of the Royal Statistical Society. Series B*, 61, 479–482.
- O’Hagan, A. y Leonard, T. (1976). Bayes Estimation Subject to Uncertainty About Parameters Constraints. *Biometrika*, 63, 201–203.
- Owen, D. B. (1956). Tables For Computing Bivariate Normal Probabilities. *Annals of Mathematical Statistics*, 27, 1075–1090.
- Pewsey, A. (2000). Problems of Inference for Azzalini’s Skew-Normal Distribution. *Journal of Applied Statistics*, 27, 859–870.
- Plummer, M., Best, N., Cowles, K. y Vines, K. (2007). *coda: Output Analysis and Diagnostics for MCMC*. R package version 0.12-1.
- R Development Core Team (2007). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Robert, C. P. y Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer, New-York, segunda edición.
- Roberts, C. (1966). A Correlation Model Useful in the Study of Twins. *Journal of the American Statistical Association*, 61, 1184–1190.
- Rusell, M. G. y González, G. (2002). Análisis de regresión lineal con errores distribuidos normal sesgados. Inf. Téc. I-02-30, Centro de Investigaciones en Matemáticas, A. C.
- Salvan, A. (1986). Locally Most Powerfull Invariant Test for Normality (In Italian). *Atti della XXXIII Riunione Scientifica della Società Italiana di Statistica*, 3, 173–179.
- 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.
- Sen, P. K. y Singer, J. M. (1993). *Large Sample Methods in Statistics: An Introduction with Applications*. Chapman Hall, New York.
- Smith, B. J. (2007). *boa: Bayesian Output Analysis Program (BOA) for MCMC*. R package version 1.1.6-1.
- Stephens, M. A. y D’Agostino, R. B. (1986). *Tests Based on EDF Statistics*. Marcel Dekker, New York.

Referencias

Stroustrup, B. (1984). The UNIX System: Data Abstraction in C. *AT&T Bell Laboratories Technical Journal*, 63, 8, 1701–1732.

Young, J. C. y Minder, C. E. (1974). An Integral Useful in Calculating Non-Central t and Bivariate Normal Probabilities. *Applied Statistics*, 23, 455–457.

Apéndices

Apéndice A: Densidades a posteriori de los parámetros de la distribución normal asimétrica usando la parametrización centrada

Con el algoritmo de Metropolis

Las rutinas que se presentan en esta subsección funcionan en el paquete R-2.6.0 ó superior, junto con la biblioteca sn de Azzalini, y pueden obtenerse libremente desde la página web del proyecto R, <http://www.r-project.org>

```
library(sn)
library(coda)

#Evalua la media de una variable SN(lambda), dado
#gamma.1
muz=function(gamma.1)
{
  sign(gamma.1)*sqrt(((2*gamma.1)^2)^(1/3)/((4-pi)^(2/3)+((2*gamma.1)^2)^(1/3)))
}

#Evalua la función de densidad de una variable normal asimétrica
#en su parametrización centrada (mu, sigma, gamma.1)
dsn.centrada=function(y,mu,sigma, gamma.1)
{
  mu.z=muz(gamma.1)
  lambda=mu.z/(sqrt(2/pi-mu.z^2))
  z=mu.z+sqrt(1-mu.z^2)*(y-mu)/sigma
  2*dnorm(z)*pnorm(lambda*z)*1/sigma*sqrt(1-mu.z^2)
}
```

Apéndices

```
#Función para calcular la moda de datos continuos
#z es un vector de datos
moda=function(z)
{
  dens=density(z)
  dens$x[dens$y == max(dens$y)] # gives the mode
}

#Función para calcular los estimadores iniciales de loc, escala y forma
#en la parametrización centrada
semillas=function(datos)
{
  n=length(datos)
  X =as.matrix(rep(1,n))
  qrX=qr(X)
  s=sqrt(sum(qr.resid(qrX, datos)^2)/n)
  gamma1=sum(qr.resid(qrX, datos)^3)/(n*s^3)
  if(abs(gamma1) > 0.99527) gamma1=sign(gamma1)*0.95
  c(qr.coef(qrX,datos),s,gamma1)
}

#Función para calcular la matriz de varianzas y covarianzas inicial
#para el algoritmo de Metropolis utilizando la distribución asintótica
#de los estimadores de momentos obtenida por Pewsey(2000),
#Problems of inference for Azzalini's skew normal distribution
Sigma.tilde=function(datos)
{
  n=length(datos)
  theta0=semillas(datos)
  mu=theta0[1]
  sigma=theta0[2]
  gamma.1=theta0[3]
  Sigma=matrix(0,nrow=3,ncol=3)
  c=(2/(4-pi))^(1/3)
  tau=c*sign(gamma.1)*abs(gamma.1)^(1/3)
  beta.2=3+2*tau^4*(pi-3)
  beta.3=10*gamma.1+tau^5*(3*pi^2-40*pi+96)/4
  beta.4=15*(1+tau^4*(2*pi-6))-tau^6*(9*pi^2-80*pi+160)/2
  Sigma[1,1]=sigma^2/n
  Sigma[2,2]=sigma^2*(beta.2-1)/(4*n)
  Sigma[3,3]=(9-6*beta.2-3*gamma.1*beta.3+beta.4+gamma.1^2*(35+9*beta.2))/4/n
  Sigma[1,2]=Sigma[2,1]=sigma^2*gamma.1/(2*n)
  Sigma[1,3]=Sigma[3,1]=sigma*(beta.2-3-3*gamma.1^2/2)/n
  Sigma[2,3]=Sigma[3,2]=sigma*(2*beta.3-gamma.1*(5+3*beta.2))/(4*n)
  Sigma
}
```

Apéndices

```
#Verosimilitud * a priori para la verosimilitud centrada
#p(loc,escala,forma)
#theta=(mu,sigma,gamma.1)
p=function(theta,y)
{
  prod(dsn.centrada(y,theta[1],theta[2],theta[3]))*1/theta[2]
}

#-----
#Empieza el algoritmo de Metropolis
#-----

#Genera una muestra de la distribución sn
#Aquí podemos sustituir el vector de datos por otros
#data(frontier)
#y=frontier
y=c(0.195,0.847,0.726,-0.139,1.788,0.570,2.069,0.452,0.868,
    1.199,0.894,0.887,1.258,0.918,0.469,0.183,0.119,1.207,0.446,2.579)

#Punto inicial para arranque del algoritmo de Metropolis
ini=semillas(y)
theta0=c(ini[1],ini[2],ini[3])
Sigmatilde=Sigma.tilde(y)
limite=sqrt(2)*(4-pi)/(pi-2)^(3/2)

m=30000
s=100                #Después de s iteraciones actualiza Sigmatilde
burnin=25000
i=0
j=0
posterior<-matrix(0,nrow=m,ncol=3)
while(i<m)
{
  j=j+1
  theta_estrella=rmnorm(1,theta0,Sigmatilde)
  if(!(abs(theta_estrella[3])>limite) & theta_estrella[2]>0)
  {
    r=p(theta_estrella,y=y)/p(theta0,y=y)
    if(r>1)
    {
      i=i+1
      posterior[i,]=theta_estrella
      theta0=t(theta_estrella)
    }
  }
  else
  {
    #Aceptar el candidato con probabilidad r
  }
}
```

```
        if(runif(1)<r)
        {
            i=i+1
            posterior[i,]=theta_estrella
            theta0=t(theta_estrella)
        }
    }
    if(i==s)
    {
        Sigmatilde=cov(cbind(posterior[1:s,1],posterior[1:s,2], posterior[1:s,3]))
    }
}

#Tasa de aceptación
tasa=m/j*100
tasa

#Grafica densidades a posteriori
par(mfrow=c(3,1))
plot(density(posterior[burnin:m,1]),
     main=expression(paste("Densidad a posteriori de ",mu)),
     xlab=expression(xi))
plot(density(posterior[burnin:m,2]),
     main=expression(paste("Densidad a posteriori de ",sigma)),
     xlab=expression(log(omega)))
plot(density(posterior[burnin:m,3]),
     main=expression(paste("Densidad a posteriori de ",gamma[1])),
     xlab=expression(gamma))

mu=moda(posterior[burnin:m,1])
sigma=moda(posterior[burnin:m,2])
gamma.1=moda(posterior[burnin:m,3])

cp.to.dp(c(mu,sigma,gamma.1))
```

Con el algoritmo “t-walk”

Las rutinas que se presentan en esta subsección corresponden a la implementación en C++ de la función objetivo $-\log p(\mu, \sigma, \gamma_1 | \mathbf{y})$ utilizando las distribuciones a priori propuestas. Dichas rutinas funcionan junto con la biblioteca de funciones matemáticas del paquete **R** y las proporcionadas por [Christen y Fox \(2007\)](#) en las cuales se implementa el algoritmo “t-walkz que pueden descargarse desde <http://cimat.mx/~jac>. Las pruebas de las rutinas fueron realizadas en **debian GNU/Linux 4.0**.

Apéndices

```
#include <iostream>
#include <stdlib.h>
#include <string>
#include <stdio.h>
#include "kernel.h"
#include "twalk.h"

#define MATHLIB_STANDALONE
#include <Rmath.h>

using namespace std;

class sn:public obj_fcn{
protected:
    double *datos;
    int ndatos;

public:
    sn(char *);
    ~sn();
    virtual int insupport(double *);
    virtual double eval(double *);
    double muz(double );
    double ginv(double , double , double , double );
};

//El constructor de la clase
// nombre_archivo indica el nombre de archivo de datos
sn::sn(char *nombre_archivo):obj_fcn(3)
{
    FILE * fptr;
    //Abre archivo de datos
    //El primer renglón corresponde al número de datos
    //y los siguientes a los datos
    fptr = fopen(nombre_archivo,"r");
    fscanf(fptr, "%d",&ndatos);
    printf("%d\n",ndatos);
    datos=vector(ndatos);
    for(int i=0; i<ndatos;i++)
    {
        fscanf(fptr,"%lf",&(datos[i]));
    }
    fclose(fptr);
}

//El destructor de la clase
sn::~sn()
```

Apéndices

```
{
    free(datos);
}
//El punto propuesto está dentro del soporte de la densidad?
int sn::insupport(double *x)
{
    if(fabs(x[2])<0.99527 && x[1]>0)
        return 1;
    else
        return 0;
}

//Evalua la media de una variable SN(lambda), dado gamma_1
double sn::muz(double gamma_1)
{
    int signo=-1;
    if(gamma_1>0) signo=1;
    return(signo*sqrt(pow(pow(2*gamma_1,2.0),1.0/3.0)/
        (pow(pow(4.0-M_PI,2.0),1.0/3.0)+pow(pow(2*gamma_1,2.0),1.0/3.0))));
}

double sn::ginv(double dato, double mu, double sigma, double mu_z)
{
    return(mu_z+sqrt(1.0-pow(mu_z,2.0))*(dato-mu)/sigma);
}

//-log(verosimilitud*priori)
double sn::eval(double *x)
{
    double mu=x[0];
    double sigma=x[1];
    double gamma_1=x[2];
    double mu_z=muz(gamma_1);
    double gamma=mu_z/sqrt(2.0/M_PI-pow(mu_z,2.0));
    double suma1=0;
    double suma2=0;

    for(int i=0; i<ndatos; i++)
    {
        suma1+=dnorm(ginv(datos[i],mu,sigma,mu_z),0,1,1);
        suma2+=pnorm(gamma*ginv(datos[i],mu,sigma,mu_z),0,1,1,1);
    }
    return(-1.0*(ndatos*log(2)+suma1+suma2+ndatos*
        log(sqrt(1.0-pow(mu_z,2.0))/sigma)-log(sigma)));
}

int main(void) {
```

```
double *x; // x, xp: Dos puntos iniciales
double *xp;
int d=3;
x=vector(d);
xp=vector(d);
x[0]=0.9;x[1]=0.7; x[2]=0.85;
xp[0]=0.85;xp[1]=0.75;xp[2]=0.9;

Seed(0); //Inicializa generador de números aleatorios

    sn misn("data/frontier.dat"); // Creación del objeto misn
    twalk mitwalk(misn,x,xp,3,x); // Creación del objeto simulador
    int acc = mitwalk.simulation(10000,"sn.output");// Simulación MCMC
    cout << endl<< "sn example-> acc="
    <<acc << endl; // Imprime la tasa de aceptación
return 1;
}
```


Apéndice B: Matriz de información de Fisher observada para comparación de poblaciones normales asimétricas con diferente parámetro de localidad

Sea $l := l(\xi_1, \xi_2, \omega, \gamma; \mathbf{x}, \mathbf{y})$.

Evaluación de las ecuaciones (7.15), (7.16), (7.17) y (7.18) en $\hat{\theta}_0$

Si $\Lambda'(x) = \frac{d}{dx} \log \Phi(x) = \frac{\phi(x)}{\Phi(x)}$, por lo tanto $\Lambda'(0) = \frac{\phi(0)}{\Phi(0)} = \sqrt{\frac{2}{\pi}}$

$$\frac{\partial l}{\partial \xi_1} \Big|_{\hat{\theta}_0} = \frac{1}{\hat{\omega}^2} \sum_{i=1}^n (x_i - \hat{\xi}_1) - \frac{0}{\hat{\omega}} \sum_{i=1}^n \Lambda' \left[0 \left(\frac{x_i - \hat{\xi}_1}{\omega} \right) \right] = \frac{1}{\hat{\omega}^2} \sum_{i=1}^n (x_i - \hat{\xi}_1) = 0$$

Dado que $\hat{\omega}^2 > 0$ entonces $\sum_{i=1}^n (x_i - \hat{\xi}_1) = 0$, pero $\hat{\xi}_1 = \bar{x}$, entonces esta ecuación de verosimilitud es satisfecha por $\hat{\theta}_0$.

$$\frac{\partial l}{\partial \xi_2} \Big|_{\hat{\theta}_0} = \frac{1}{\hat{\omega}^2} \sum_{j=1}^m (y_j - \hat{\xi}_2) - \frac{0}{\hat{\omega}} \sum_{j=1}^m \Lambda' \left[0 \left(\frac{y_j - \hat{\xi}_2}{\omega} \right) \right] = \frac{1}{\hat{\omega}^2} \sum_{j=1}^m (y_j - \hat{\xi}_2) = 0$$

Dado que $\hat{\omega}^2 > 0$ entonces $\sum_{j=1}^m (y_j - \hat{\xi}_2) = 0$, pero $\hat{\xi}_2 = \bar{y}$, entonces esta ecuación de verosimilitud es satisfecha por $\hat{\theta}_0$.

$$\begin{aligned} \frac{\partial l}{\partial \omega} \Big|_{\hat{\theta}_0} &= -\frac{0}{\hat{\omega}^2} \left\{ \sum_{i=1}^n (x_i - \hat{\xi}_1) \Lambda' \left[0 \left(\frac{x_i - \hat{\xi}_1}{\hat{\omega}} \right) \right] + \sum_{j=1}^m (y_j - \hat{\xi}_2) \Lambda' \left[0 \left(\frac{y_j - \hat{\xi}_2}{\hat{\omega}} \right) \right] \right\} \\ &\quad + \frac{1}{\hat{\omega}^3} \left[\sum_{i=1}^n (x_i - \hat{\xi}_1)^2 + \sum_{j=1}^m (y_j - \hat{\xi}_2)^2 \right] - \frac{(n+m)}{\hat{\omega}} \\ &= -\frac{(n+m)}{\hat{\omega}} + \frac{1}{\hat{\omega}^3} \left[\sum_{i=1}^n (x_i - \hat{\xi}_1)^2 + \sum_{j=1}^m (y_j - \hat{\xi}_2)^2 \right] \\ &= -\frac{(n+m)}{\hat{\omega}} + \frac{(n+m)}{\hat{\omega}^3(n+m)} \left[\sum_{i=1}^n (x_i - \hat{\xi}_1)^2 + \sum_{j=1}^m (y_j - \hat{\xi}_2)^2 \right] \\ &= -\frac{n+m}{\hat{\omega}} + \frac{n+m}{\hat{\omega}^3} \hat{\omega}^2 = -\frac{n+m}{\hat{\omega}} + \frac{n+m}{\hat{\omega}} = 0 \end{aligned}$$

$$\begin{aligned}\frac{\partial l}{\partial \gamma} \Big|_{\hat{\theta}_0} &= \frac{1}{\hat{\omega}} \left\{ \sum_{i=1}^n (x_i - \hat{\xi}_1) \Lambda' \left[0 \left(\frac{x_i - \hat{\xi}_1}{\hat{\omega}} \right) \right] + \sum_{j=1}^m (y_j - \hat{\xi}_2) \Lambda' \left[0 \left(\frac{y_j - \hat{\xi}_2}{\hat{\omega}} \right) \right] \right\} \\ &= \frac{1}{\hat{\omega}} \sqrt{\frac{2}{\pi}} \left\{ \sum_{i=1}^n (x_i - \bar{x}) + \sum_{j=1}^m (y_j - \bar{y}) \right\} = 0\end{aligned}$$

Dado que $\hat{\omega}^2 > 0$ entonces $\sum_{i=1}^n (x_i - \bar{x}) + \sum_{j=1}^m (y_j - \bar{y}) = 0$.

Obtención de la matriz de información de Fisher observada

- 1) La derivada parcial de la log-verosimilitud con respecto de ξ_1 ya fue obtenida antes, y está dada por la ecuación (7.15), a partir de la cual se obtienen las derivadas cruzadas $\frac{\partial^2 l}{\partial \xi_1^2}$, $\frac{\partial^2 l}{\partial \xi_1 \partial \xi_2}$, $\frac{\partial^2 l}{\partial \xi_1 \partial \omega}$, $\frac{\partial^2 l}{\partial \xi_1 \partial \gamma}$.

$$\begin{aligned}\frac{\partial^2 l}{\partial \xi_1^2} &= \frac{\partial}{\partial \xi_1} \left\{ \frac{1}{\omega^2} \sum_{i=1}^n (x_i - \xi_1) - \frac{\gamma}{\omega} \sum_{i=1}^n \Lambda' \left[\gamma \left(\frac{x_i - \xi_1}{\omega} \right) \right] \right\} \\ &= -\frac{n}{\omega^2} + \frac{\gamma^2}{\omega^2} \sum_{i=1}^n \Lambda'' \left[\gamma \left(\frac{x_i - \xi_1}{\omega} \right) \right]\end{aligned}$$

$$\frac{\partial^2 l}{\partial \xi_1 \partial \xi_2} = \frac{\partial}{\partial \xi_2} \left\{ \frac{1}{\omega^2} \sum_{i=1}^n (x_i - \xi_1) - \frac{\gamma}{\omega} \sum_{i=1}^n \Lambda' \left[\gamma \left(\frac{x_i - \xi_1}{\omega} \right) \right] \right\} = 0$$

$$\begin{aligned}\frac{\partial^2 l}{\partial \xi_1 \partial \omega} &= \frac{\partial}{\partial \omega} \left\{ \frac{1}{\omega^2} \sum_{i=1}^n (x_i - \xi_1) - \frac{\gamma}{\omega} \sum_{i=1}^n \Lambda' \left[\gamma \left(\frac{x_i - \xi_1}{\omega} \right) \right] \right\} \\ &= -\frac{2}{\omega^3} \sum_{i=1}^n (x_i - \xi_1) + \frac{\gamma}{\omega^2} \sum_{i=1}^n \Lambda' \left[\gamma \left(\frac{x_i - \xi_1}{\omega} \right) \right] \\ &\quad + \frac{\gamma^2}{\omega^3} \sum_{i=1}^n (x_i - \xi_1) \Lambda'' \left[\gamma \left(\frac{x_i - \xi_1}{\omega} \right) \right]\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 l}{\partial \xi_1 \partial \gamma} &= \frac{\partial}{\partial \gamma} \left\{ \frac{1}{\omega^2} \sum_{i=1}^n (x_i - \xi_1) - \frac{\gamma}{\omega} \sum_{i=1}^n \Lambda' \left[\gamma \left(\frac{x_i - \xi_1}{\omega} \right) \right] \right\} \\ &= -\frac{1}{\omega} \sum_{i=1}^n \Lambda' \left[\gamma \left(\frac{x_i - \xi_1}{\omega} \right) \right] - \frac{\gamma}{\omega^2} \sum_{i=1}^n (x_i - \xi_1) \Lambda'' \left[\gamma \left(\frac{x_i - \xi_1}{\omega} \right) \right]\end{aligned}$$

- 2) La derivada parcial de la log-verosimilitud con respecto de ξ_2 y está dada por (7.16), a partir de esta expresión se obtienen las derivadas cruzadas $\frac{\partial^2 l}{\partial \xi_2^2}$, $\frac{\partial^2 l}{\partial \xi_2 \partial \omega}$, $\frac{\partial^2 l}{\partial \xi_2 \partial \gamma}$.

$$\begin{aligned} \frac{\partial^2 l}{\partial \xi_2^2} &= \frac{\partial}{\partial \xi_2} \left\{ \frac{1}{\omega^2} \sum_{j=1}^m (y_j - \xi_2) - \frac{\gamma}{\omega} \sum_{j=1}^m \Lambda' \left[\gamma \left(\frac{y_j - \xi_2}{\omega} \right) \right] \right\} \\ &= -\frac{m}{\omega^2} + \frac{\gamma^2}{\omega^2} \sum_{j=1}^m \Lambda'' \left[\gamma \left(\frac{y_j - \xi_2}{\omega} \right) \right] \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 l}{\partial \xi_2 \partial \omega} &= \frac{\partial}{\partial \omega} \left\{ \frac{1}{\omega^2} \sum_{j=1}^m (y_j - \xi_2) - \frac{\gamma}{\omega} \sum_{j=1}^m \Lambda' \left[\gamma \left(\frac{y_j - \xi_2}{\omega} \right) \right] \right\} \\ &= -\frac{2}{\omega^3} \sum_{j=1}^m (y_j - \xi_2) + \frac{\gamma}{\omega^2} \sum_{j=1}^m \Lambda' \left[\gamma \left(\frac{y_j - \xi_2}{\omega} \right) \right] \\ &\quad + \frac{\gamma^2}{\omega^3} \sum_{j=1}^m (y_j - \xi_2) \Lambda'' \left[\gamma \left(\frac{y_j - \xi_2}{\omega} \right) \right] \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 l}{\partial \xi_2 \partial \gamma} &= \frac{\partial}{\partial \gamma} \left\{ \frac{1}{\omega^2} \sum_{j=1}^m (y_j - \xi_2) - \frac{\gamma}{\omega} \sum_{j=1}^m \Lambda' \left[\gamma \left(\frac{y_j - \xi_2}{\omega} \right) \right] \right\} \\ &= -\frac{1}{\omega} \sum_{j=1}^m \Lambda' \left[\gamma \left(\frac{y_j - \xi_2}{\omega} \right) \right] - \frac{\gamma}{\omega^2} \sum_{j=1}^m (y_j - \xi_2) \Lambda'' \left[\gamma \left(\frac{y_j - \xi_2}{\omega} \right) \right] \end{aligned}$$

- 3) La derivada parcial de la log-verosimilitud con respecto de ω está dada por la ecuación (7.16), a partir de la cual se obtienen las derivadas cruzadas $\frac{\partial^2 l}{\partial \omega^2}$, $\frac{\partial^2 l}{\partial \omega \partial \gamma}$

$$\begin{aligned} \frac{\partial^2 l}{\partial \omega^2} &= \frac{2\gamma}{\omega^3} \left\{ \sum_{i=1}^n (x_i - \xi_1) \Lambda' \left[\gamma \left(\frac{x_i - \xi_1}{\omega} \right) \right] + \sum_{j=1}^m (y_j - \xi_2) \Lambda' \left[\gamma \left(\frac{y_j - \xi_2}{\omega} \right) \right] \right\} \\ &\quad + \frac{\gamma^2}{\omega^4} \left\{ \sum_{i=1}^n (x_i - \xi_1)^2 \Lambda'' \left[\gamma \left(\frac{x_i - \xi_1}{\omega} \right) \right] + \sum_{j=1}^m (y_j - \xi_2)^2 \Lambda'' \left[\gamma \left(\frac{y_j - \xi_2}{\omega} \right) \right] \right\} \\ &\quad - \frac{3}{\omega^4} \left[\sum_{i=1}^n (x_i - \xi_1)^2 + \sum_{j=1}^m (y_j - \xi_2)^2 \right] + \frac{(n+m)}{\omega^2} \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 l}{\partial \omega \partial \gamma} &= -\frac{1}{\omega^2} \left\{ \sum_{i=1}^n (x_i - \xi_1) \Lambda' \left[\gamma \left(\frac{x_i - \xi_1}{\omega} \right) \right] + \sum_{j=1}^m (y_j - \xi_2) \Lambda' \left[\gamma \left(\frac{y_j - \xi_2}{\omega} \right) \right] \right\} \\ &\quad - \frac{\gamma}{\omega^3} \left\{ \sum_{i=1}^n (x_i - \xi_1)^2 \Lambda'' \left[\gamma \left(\frac{x_i - \xi_1}{\omega} \right) \right] + \sum_{j=1}^m (y_j - \xi_2)^2 \Lambda'' \left[\gamma \left(\frac{y_j - \xi_2}{\omega} \right) \right] \right\} \end{aligned}$$

4) Finalmente

$$\frac{\partial^2 l}{\partial^2 \gamma} = \frac{1}{\omega^2} \left\{ \sum_{i=1}^n (x_i - \xi_1)^2 \Lambda'' \left[\gamma \left(\frac{x_i - \xi_1}{\omega} \right) \right] + \sum_{j=1}^m (y_j - \xi_2)^2 \Lambda'' \left[\gamma \left(\frac{y_j - \xi_2}{\omega} \right) \right] \right\}$$

Solo resta evaluar las derivadas parciales en el punto $\hat{\theta}_0 = (\hat{\xi}_1, \hat{\xi}_2, \hat{\omega}, 0)$. Es conveniente recordar que $\Lambda'(0) = \sqrt{\frac{2}{\pi}}$, $\Lambda''(0) = -\frac{2}{\pi}$.

$$\frac{\partial^2 l}{\partial \xi_1^2} \Big|_{\hat{\theta}_0} = -\frac{n}{\hat{\omega}^2} + \frac{0^2}{\hat{\omega}^2} \sum_{i=1}^n \Lambda'' \left[0 \left(\frac{x_i - \hat{\xi}_1}{\hat{\omega}} \right) \right] = -\frac{n}{\hat{\omega}^2}$$

$$\begin{aligned} \frac{\partial^2 l}{\partial \xi_1 \partial \omega} \Big|_{\hat{\theta}_0} &= -\frac{2}{\hat{\omega}^3} \sum_{i=1}^n (x_i - \hat{\xi}_1) + \frac{0}{\hat{\omega}^2} \sum_{i=1}^n \Lambda' \left[0 \left(\frac{x_i - \hat{\xi}_1}{\hat{\omega}} \right) \right] \\ &\quad + \frac{0^2}{\hat{\omega}^3} \sum_{i=1}^n (x_i - \hat{\xi}_1) \Lambda'' \left[0 \left(\frac{x_i - \hat{\xi}_1}{\hat{\omega}} \right) \right] \\ &= -\frac{2}{\hat{\omega}^3} \sum_{i=1}^n (x_i - \bar{x}) = 0 \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 l}{\partial \xi_1 \partial \gamma} \Big|_{\hat{\theta}_0} &= -\frac{1}{\hat{\omega}} \sum_{i=1}^n \Lambda' \left[0 \left(\frac{x_i - \hat{\xi}_1}{\hat{\omega}} \right) \right] - \frac{0}{\hat{\omega}^2} \sum_{i=1}^n (x_i - \hat{\xi}_1) \Lambda'' \left[0 \left(\frac{x_i - \hat{\xi}_1}{\hat{\omega}} \right) \right] \\ &= -\frac{n}{\hat{\omega}} \sqrt{\frac{2}{\pi}} \end{aligned}$$

$$\frac{\partial^2 l}{\partial \xi_2^2} \Big|_{\hat{\theta}_0} = -\frac{m}{\hat{\omega}^2} + \frac{0^2}{\hat{\omega}^2} \sum_{j=1}^m \Lambda'' \left[0 \left(\frac{y_j - \hat{\xi}_2}{\hat{\omega}} \right) \right] = -\frac{m}{\hat{\omega}^2}$$

$$\begin{aligned}
 \frac{\partial^2 l}{\partial \xi_2 \partial \omega} \Big|_{\hat{\theta}_0} &= -\frac{2}{\hat{\omega}^3} \sum_{j=1}^m (y_j - \hat{\xi}_2) + \frac{0}{\hat{\omega}^2} \sum_{j=1}^m \Lambda' \left[0 \left(\frac{x_j - \hat{\xi}_2}{\hat{\omega}} \right) \right] \\
 &\quad + \frac{0^2}{\hat{\omega}^3} \sum_{j=1}^m (y_j - \hat{\xi}_2) \Lambda'' \left[0 \left(\frac{y_j - \hat{\xi}_2}{\hat{\omega}} \right) \right] \\
 &= -\frac{2}{\hat{\omega}^3} \sum_{j=1}^m (y_j - \hat{\xi}_2) = -\frac{2}{\hat{\omega}^3} \sum_{j=1}^m (y_j - \bar{y}) = 0
 \end{aligned}$$

$$\frac{\partial^2 l}{\partial \xi_2 \partial \gamma} \Big|_{\hat{\theta}_0} = -\frac{m}{\hat{\omega}} \sqrt{\frac{2}{\pi}}$$

$$\begin{aligned}
 \frac{\partial^2 l}{\partial \omega^2} \Big|_{\hat{\theta}_0} &= \frac{(n+m)}{\hat{\omega}^2} + \frac{2(0)}{\hat{\omega}^3} \left\{ \sum_{i=1}^n (x_i - \hat{\xi}_1) \Lambda' \left[0 \left(\frac{x_i - \hat{\xi}_1}{\hat{\omega}} \right) \right] + \sum_{j=1}^m (y_j - \hat{\xi}_2) \Lambda' \left[0 \left(\frac{y_j - \hat{\xi}_2}{\hat{\omega}} \right) \right] \right\} \\
 &\quad + \frac{0^2}{\hat{\omega}^4} \left\{ \sum_{i=1}^n (x_i - \hat{\xi}_1)^2 \Lambda'' \left[0 \left(\frac{x_i - \hat{\xi}_1}{\hat{\omega}} \right) \right] + \sum_{j=1}^m (y_j - \hat{\xi}_2)^2 \Lambda'' \left[0 \left(\frac{y_j - \hat{\xi}_2}{\hat{\omega}} \right) \right] \right\} \\
 &\quad - \frac{3}{\hat{\omega}^4} \left[\sum_{i=1}^n (x_i - \xi_1)^2 + \sum_{j=1}^m (y_j - \xi_2)^2 \right] \\
 &= \frac{(n+m)}{\hat{\omega}^2} - \frac{3(n+m)\hat{\omega}^2}{\hat{\omega}^4} = -\frac{2(n+m)}{\hat{\omega}^2}
 \end{aligned}$$

$$\begin{aligned}
 \frac{\partial^2 l}{\partial \omega \partial \gamma} \Big|_{\hat{\theta}_0} &= -\frac{1}{\hat{\omega}^2} \left\{ \sum_{i=1}^n (x_i - \hat{\xi}_1) \Lambda' \left[0 \left(\frac{x_i - \hat{\xi}_1}{\hat{\omega}} \right) \right] + \sum_{j=1}^m (y_j - \hat{\xi}_2) \Lambda' \left[0 \left(\frac{y_j - \hat{\xi}_2}{\hat{\omega}} \right) \right] \right\} \\
 &\quad - \frac{0}{\hat{\omega}^3} \left\{ \sum_{i=1}^n (x_i - \hat{\xi}_1)^2 \Lambda'' \left[0 \left(\frac{x_i - \hat{\xi}_1}{\hat{\omega}} \right) \right] + \sum_{j=1}^m (y_j - \hat{\xi}_2)^2 \Lambda'' \left[0 \left(\frac{y_j - \hat{\xi}_2}{\hat{\omega}} \right) \right] \right\} \\
 &= -\frac{1}{\hat{\omega}^2} \left\{ \sum_{i=1}^n (x_i - \bar{x}) \sqrt{\frac{2}{\pi}} + \sum_{j=1}^m (y_j - \bar{y}) \sqrt{\frac{2}{\pi}} \right\} = 0
 \end{aligned}$$

$$\begin{aligned}
 \frac{\partial^2 l}{\partial^2 \gamma} \Big|_{\hat{\theta}_0} &= \frac{1}{\hat{\omega}^2} \left\{ \sum_{i=1}^n (x_i - \hat{\xi}_1)^2 \Lambda'' \left[0 \left(\frac{x_i - \hat{\xi}_1}{\hat{\omega}} \right) \right] + \sum_{j=1}^m (y_j - \hat{\xi}_2)^2 \Lambda'' \left[0 \left(\frac{y_j - \hat{\xi}_2}{\hat{\omega}} \right) \right] \right\} \\
 &= -\frac{2}{\pi} \frac{1}{\hat{\omega}^2} \left\{ \sum_{i=1}^n (x_i - \hat{\xi}_1)^2 + \sum_{j=1}^m (y_j - \hat{\xi}_2)^2 \right\} \\
 &= -\frac{2}{\pi} \frac{(n+m)}{\hat{\omega}^2(n+m)} \left\{ \sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{j=1}^m (y_j - \bar{y})^2 \right\} \\
 &= -\frac{2(n+m)}{\pi}
 \end{aligned}$$

La matriz de información de Fisher evaluada en $\hat{\theta}_0$ está dada por:

$$IO(\hat{\theta}_0) = - \begin{pmatrix} -\frac{n}{\hat{\omega}^2} & 0 & 0 & -\frac{n}{\hat{\omega}} \sqrt{\frac{2}{\pi}} \\ 0 & -\frac{m}{\hat{\omega}^2} & 0 & -\frac{m}{\hat{\omega}} \sqrt{\frac{2}{\pi}} \\ 0 & 0 & -\frac{2(n+m)}{\hat{\omega}^2} & 0 \\ -\frac{n}{\hat{\omega}} \sqrt{\frac{2}{\pi}} & -\frac{m}{\hat{\omega}} \sqrt{\frac{2}{\pi}} & 0 & -\frac{2(n+m)}{\pi} \end{pmatrix}$$

Apéndice C: Cálculo de estimadores de máxima verosimilitud para mezclas de dos componentes SN con el algoritmo EM Generalizado

Las rutinas que se presentan en esta sección funcionan en el paquete [R-2.6.0](#) ó superior, junto con la biblioteca `sn` de Azzalini, y pueden obtenerse libremente desde la página web del proyecto R, <http://www.r-project.org>

```
library(sn)
#-----
#La función error
Erf =function(x) 2 * pnorm(x * sqrt(2)) - 1

#-----
#La función error complemento
Erfc=function(x) 1-Erf(x)

#-----
#Renombra algunas funciones y constantes de R para poder utilizar las
#salidas de la función CForm[] de Mathematica sin tener que reescribir las
#fórmulas
Power=function(base, exponente) base^exponente
E=exp(1)
Pi= 3.141593

#-----
#Genera números aleatorios de la mezcla de dos normales asimétricas con
#parámetro de forma común
#p es la probabilidad de que la observación venga del primer componente
#de la mezcla
rsn2mix=function(n=200,p=0.5,loc1=0,escala1=1,loc2=3,escala2=1,forma=1)
{
  cual=rbinom(n,1,p)
  mezcla=rep(0,n)
  for(i in 1:n)
  {
    if(cual[i]==1) mezcla[i]=rsn(1,loc1,escala1,forma)
    else mezcla[i]=rsn(1,loc2,escala2,forma)
  }
  mezcla
}

#-----
#Paso E
#y es el vector de datos (la mezcla)
```

Apéndices

```
#thetaj=(pj,loc1j,escala1j,loc2j,escala2j,formaj)
#pj es la probabilidad de que las observaciones vengan del primer
#componente de la mezcla
#regresa un vector con los cálculos de la esperanza condicional
E.deltai=function(y,thetaj)
{
  pj=thetaj[1]
  loc1j=thetaj[2]
  escala1j=thetaj[3]
  loc2j=thetaj[4]
  escala2j=thetaj[5]
  formaj=thetaj[6]
  num=pj*dsn(y,loc1j,escala1j,formaj)
  den=num+(1-pj)*dsn(y,loc2j,escala2j,formaj)
  num/den
}

#-----
#Paso M
#Esta es la función a maximizar en el paso M
#thetaj=c(pj,loc1j,escala1j,loc2j,escala2j,formaj)
#theta=c(loc1,escala1,loc2,escala2,forma)
#note que he cambiado el signo, para que se convierta en un problema de
#minimización. El problema original es de dimensión 6, pero el estimador
#de p se puede calcular de manera analítica, por lo tanto se puede reducir
#la dimensión del problema
Qj=function(theta,pi,y)
{
  loc1=theta[1]
  escala1=theta[2]
  loc2=theta[3]
  escala2=theta[4]
  forma=theta[5]
  p=mean(pi)      #Es el emv de p, si se puede obtener de manera analítica
  s1=sum(pi*log(p)+(1-pi)*log(1-p))
  s2=-sum(pi*(log(escala1)+0.5*((y-loc1)/escala1)^2
  -log(pnorm(forma*(y-loc1)/escala1))))
  s3=-sum((1-pi)*(log(escala2)+0.5*((y-loc2)/escala2)^2
  -log(pnorm(forma*(y-loc2)/escala2))))
  -(s1+s2+s3)
}

#-----
#Calcula el gradiente de la función de -log-verosimilitud de los datos
#completos
grad=function(theta,pi,y)
{
```


Apéndices

```
loc1=theta[1]
escala1=theta[2]
loc2=theta[3]
escala2=theta[4]
forma=theta[5]

#Calcula los componentes del vector gradiente como si fuera un problema
#de dimensión 5

#Derivada con respecto de loc1
dloc1=sum(pi*(-((-loc1 + y)/Power(escala1,2)) +(forma*sqrt(2/Pi))/
  (Power(E,(Power(forma,2)*Power(-loc1 + y,2))/
    (2.*Power(escala1,2))) *escala1*
    (1 + Erf((forma*(-loc1 + y))/(sqrt(2)*escala1))))),na.rm=TRUE)
descala1=sum( pi*(1/escala1 - Power(-loc1 + y,2)/Power(escala1,3)
  + (forma*sqrt(2/Pi)*(-loc1 + y))/(Power(E,(Power(forma,2)
    *Power(-loc1 + y,2))/(2.*Power(escala1,2))) *Power(escala1,2)
    *(1 + Erf((forma*(-loc1 + y))/(sqrt(2)*escala1))))),na.rm=TRUE)
dloc2=sum((1 - pi)*(-((-loc2 + y)/Power(escala2,2)) +
  (forma*sqrt(2/Pi))/
  (Power(E,(Power(forma,2)*Power(-loc2 + y,2))/
    (2.*Power(escala2,2))) *escala2*
    (1 + Erf((forma*(-loc2 + y))/(sqrt(2)*escala2))))),na.rm=TRUE)
descala2=sum((1 - pi)*(1/escala2 - Power(-loc2 + y,2)/Power(escala2,3)
  + (forma*sqrt(2/Pi)*(-loc2 + y))
  / (Power(E,(Power(forma,2)*Power(-loc2 + y,2))
  / (2.*Power(escala2,2))) *Power(escala2,2)
  * (1 + Erf((forma*(-loc2 + y))/(sqrt(2)*escala2))))),na.rm=TRUE)
dforma=sum((( -1 + pi)*sqrt(2/Pi)*(-loc2 + y)
  / (Power(E,(Power(forma,2)*Power(-loc2 + y,2))
  / (2.*Power(escala2,2))) *escala2
  *(1 + Erf((forma*(-loc2 + y))/(sqrt(2)*escala2))))
  +(pi*sqrt(2/Pi)*(-loc1 + y))
  / (Power(E,(Power(forma,2)*Power(-loc1 + y,2))
  / (2.*Power(escala1,2))) *escala1
  *(-2 + Erfc((forma*(-loc1 + y))/(sqrt(2)*escala1))))),na.rm=TRUE)
c(dloc1,descala1,dloc2,descala2,dforma)
}
#-----

sn2mix.em=function(y)
{
  #Semillas para algoritmo EM
  pr=0.5
  s1=s2=sd(y)
  m1=quantile(y,probs=0.25)
  m2=quantile(y,probs=0.75)
```

```
forma=1
thetaini=c(pr,m1,s1,m2,s2,forma)

#Parámetro para método de descenso más rápido
paso=0.001

#Empieza algoritmo EM
error=1
i=1
pi=E.deltai(y,thetaini)
while (error>0.01 & i<1000)
{
  theta=thetaini[-1]
  L=Qj(theta,pi,y)
  thetanuevo=c(mean(pi),theta-paso*grad(theta,pi,y))
  pnuevo=E.deltai(y,thetanuevo)
  Lnuevo=Qj(thetanuevo[-1],pnuevo,y)
  error=abs(L-Lnuevo)
  if(is.nan(error)) {error=1;i=1000}
  thetaini=thetanuevo
  pi=pnuevo
  i=i+1
}
list(L=L,theta=thetaini,error=error,iter=i)
}

#Ejemplo de llamadas

#500 números aleatorios de una mezcla de dos componentes SN, el
#primer componente tiene parámetros (0,1,1), el segundo componente
#tiene parámetros (3,1,1), proporción de la mezcla = 0.5

a=rsn2mix(500,p=0.5,loc1=0,escala1=1,loc2=3,escala2=1,forma=1)

#estima parámetros

sn2mix.em(a)
```