Modelos de Regresión Binaria Bayesiana Skew Probit

Omar Chocotea Poca

En memoria a Gualberto Poca,
mi Padre Corazonológico, Amigo y Mentor

Agradecimientos

A Dios, por su amor y protección.

A mi familia, abuelita Manuela, mamita Elena Mery, tíos Gualberto (†) y Eugenio y hermanitos Iveth Luz y Daniel, por darme su amistad, su respeto y cariño, por tener un corazón como una casa de puertas abiertas, por darme palabras de fuerza y de fe, y por sus sonrisas y abrazos efectivos a cada llegada.

A mi tutor Dindo Valdez, por la confianza, por la paciencia infinita, y por la orientación inestimable.

A los miembros del tribunal Raúl Delgado y Juan Carlos Flores, por disponer de sus tiempos valiosos, por sus comentarios y sugerencias.

A mis profesores Rubén Belmonte, Nicolás Chávez, Ramiro Coa, Zenón Condori, Lucy Cuarita, Raúl Delgado, Juan Carlos Flores, Jaime Pinto, María de los Ángeles Ramos, Fernando Rivero, Augusto Solís y Dindo Valdez, por contribuir en mi formación académica.

A mis amigos Mauro Marca, Ángel Pairumani, Saúl Peñaloza y Gerinel Ugarte, por los momentos de desconcentración y aprendizaje.

Resumen

En esta memoria, se presenta y discute detalles de la distribución skew normal y de dos nuevos modelos de regresión binaria con perspectiva Bayesiana, el skew probit CDS y el skew probit BBB aplicables cuando hay probabilidades extremas. La distribución skew normal incluye a la distribución normal y posee resultados exquisitos y fundamentales en la especificación de la estructura jerárquica de los modelos. Las distribuciones a posteriori de los modelos son difíciles de obtener, por lo tanto las buenas aproximaciones son dadas bajo el muestreo de Gibbs. En la aplicación, los modelos son comparados con el probit, el logit, el cloglog, el scobit y el power logit, y el veredicto de los tres criterios de bondad de ajuste, criterio de la información de la devianza, criterio de la información de Akaike y criterio de la información Bayesiana, favorece como mejor modelo al skew probit CDS.

Palabras clave: Modelos de regresión binaria, perspectiva Bayesiana, probabilidades extremas, estructura jerárquica, distribuciones a posteriori, muestreo de Gibbs, criterios de bondad de ajuste.

Índice general

Índice de figuras

Índice de cuadros

Símbolos y Notaciones

Básicos

df	Función doble factorial
fda	Función de distribución acumulada
fdp	Función de densidad de probabilidad
fmp	Función masa de probabilidad
ig	Iterando y generalizando
mlg	Modelo lineal general
pp	Método por partes

Distribuciones de Probabilidad

$\mathcal{B}er$	Distribución Bernoulli
$\mathcal{B}urr - II$	Distribución Burr tipo II
χ^2	Distribución Chi–cuadrado
$\mathcal{I}nv - \chi^2$	Distribución Chi–cuadrado Inversa
$\mathcal{G}$	Distribución Gamma
$\mathcal{GI}$	Distribución Gamma Inversa
$\mathcal{G}u$	Distribución Gumbel
$\mathcal{HN}$	Distribución Half Normal
$\mathcal{L}o$	Distribución Logística
$\mathcal{N}_n$	Distribución Normal
$\mathcal{S}c$	Distribución Scobit
$\mathcal{SN}_n$	Distribución Skew Normal
$\mathcal{T}$	Distribución t de Student
$\mathcal{U}$	Distribución Uniforme

Capítulo 1

Introducción

La distribución normal es una función de densidad de probabilidad (fdp) importante, y es ampliamente utilizada en estadísticas y otras ciencias, tratamientos completos aparecen en Tong (1990), Bryc (1995), Flury (1997), Kotz *et al.* (2000), Rencher y Schaalje (2008), Balakrishnan y Lai (2009) y Ahsanullah *et al.* (2014), entre otros. Los distintos aplicativos estadísticos en su gran mayoría invitan a utilizar o cumplir el supuesto de normalidad, pero ¿porque no ser anormales?, cuando el permiso lo dio Azzalini (1985) al introducir la distribución skew normal estándar, teniendo como caso particular a la normal estándar. No podemos decir que la acogida fue inmediata, pero es un gran inicio.

A lo que nos concierne, empecemos con el ejemplo de la elección presidencial, actualmente la oposición pretende mandar un solo candidato, dando al ciudadano dos opciones, pero lo que los partidos políticos desean conocer previamente son las características que influyen en su decisión. Nótese que, la regresada es una variable binaria. En este tipo de escenarios de investigación se desenvuelven los modelos de regresión binaria.

Bliss (1935) introduce el primer modelo de regresión binaria, el probit, aún es utilizado, sin olvidarnos de sus contendientes históricos, el logit (Berkson, 1944) y el cloglog (Gumbel, 1935), todos disponibles en programas comerciales, pero no nos preguntamos si ¿hay alternativas? y ¿cuál es el mejor?

1.1. Objetivo

Esta memoria tiene el objetivo de presentar y discutir detalles de la distribución skew normal (Azzalini, 1985, 1986, 2005, 2014) y de dos nuevos modelos de regresión binaria con perspectiva Bayesiana, el skew probit CDS (Chen *et al.*, 1999; Chen, 2004; Bazán *et al.*, 2006; Bazán *et al.*, 2010; Farias y Branco, 2011, 2012) y el skew probit BBB (Bazán *et al.*, 2006, 2014; Bazán *et al.*, 2010) aplicables cuando hay probabilidades extremas (Chen *et al.*, 1999; Chen, 2004; Bazán *et al.*, 2006, 2014; Bermúdez *et al.*, 2008; Bazán *et al.*, 2010; Sáez–Castillo *et al.*, 2010; Farias y Branco, 2011, 2012; Pérez–Sánchez *et al.*, 2014).

Para una lectura agradable se recomienda revisar detalladamente los buenos resultados de la distribución normal y familiarizarse con las distribuciones de probabilidad situadas en el apéndice A.

Capítulo 2

La Regresión Binaria

Los modelos de regresión binaria son utilizados para predecir la probabilidad de una respuesta binaria (0/1) en función de diversas variables explicativas. La aplicación se encuentra en un rango más amplio de escenarios de investigación que el análisis discriminante. Para un estudio casi detallado, el capítulo se remite a abordar tres secciones.

2.1. Una Introducción al Modelo Lineal General

El modelo lineal general (mlg) es introducido por Nelder y Wedderburn (1972), es una síntesis de otros modelos y se basa en la clase exponencial.

La siguiente definición y proposición es una extensión de los resultados presentados por Nelder y Wedderburn (1972) y McCullagh y Nelder (1989).

Definición 2.1. Sea y_i una variable aleatoria, $\xi_i \in \Theta_i \subset \mathbb{R}$ un parámetro de interés y $\omega \in \mathbb{R}_+$ un parámetro de molestia. Denotando por $a_i(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$, $b(\cdot) : \mathbb{R} \to \mathbb{R}$ y $c(\cdot;\cdot) : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ a las funciones. Una distribución pertenece a la clase exponencial, si su fmp (función masa de probabilidad) o fdp está dada por

$$f(y_i; \xi_i, \omega) = \exp\left\{ \frac{y_i \xi_i - b(\xi_i)}{a_i(\omega)} + c(y_i; \omega) \right\}. \tag{2.1.1}$$

ξ_i es el parámetro natural y Θ_i es el espacio del parámetro natural.

Proposición 2.2. *La esperanza y la varianza de la distribución perteneciente a la clase exponencial, están dadas por*

$$\mathbb{E}[y_i] = \frac{\partial b(\xi_i)}{\partial \xi_i} \triangleq \mu_i(\xi_i) \qquad \text{y} \qquad \mathbb{V}[y_i] = a_i(\omega) \frac{\partial^2 b(\xi_i)}{\partial \xi_i^2} \triangleq \sigma_i^2(\xi_i). \tag{2.1.2}$$

Demostración. Es inmediata utilizando los dos primeros momentos de la función score, $\mathbb{E}[S(\xi_i)] = 0$ y $\mathbb{E}[S^2(\xi_i)] = -\mathbb{E}[\partial S(\xi_i)/\partial \xi_i]$. □

Ziegler (2011, pp. 21) introduce la siguiente definición.

Definición 2.3. Sea $\mathbf{y} = (y_1, y_2, \ldots, y_n)'$ un vector aleatorio, $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)'$ una matriz de diseño, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_k)'$ un vector de parámetros y $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)'$ un vector aleatorio de errores. Se asume que, los pares $(y_i, \mathbf{x}_i)$ son independientes, las condicionales $y_i|\mathbf{x}_i$ son distribuidas idénticamente para $i = 1, 2, \ldots, n$ y la matriz $\frac{1}{n}\mathbf{X}'\mathbf{X}$ converge a una matriz regular no estocástica $\mathbf{Q}$ cuando $n \to \infty$.

En el mlg, la variable aleatoria y_i se descompone aditivamente en términos de un componente sistemático μ_i y un error ε_i,

$$y_i = \mu_i + \varepsilon_i, \tag{2.1.3}$$

donde ε_i y $\mathbf{x}_i$ son independientes, i.e. $\mathbb{E}[\varepsilon_i|\mathbf{x}_i] = 0$ y $\mathbb{E}[y_i|\mathbf{x}_i] = \mu_i$ (McCullagh y Nelder, 1989; Ziegler, 2011; Kroese y Chan, 2014). La fmp o fdp condicional $f(y_i|\xi_i)$ forma parte de la clase exponencial con parámetro natural ξ_i (McCullagh y Nelder, 1989; Ziegler, 2011; Kroese y Chan, 2014). Además, la esperanza μ_i está relacionada con el predictor lineal $\mathbf{x}_i'\boldsymbol{\beta}$ por un link $g(\cdot) : \mathbb{R} \to \mathbb{R}$ monótono y diferenciable: $g(\mu_i) = \mathbf{x}_i'\boldsymbol{\beta}$ (McCullagh y Nelder, 1989; Ziegler, 2011; Kroese y Chan, 2014). Cuando $\xi_i = \mathbf{x}_i'\boldsymbol{\beta}$, el mlg es llamado mlg con link natural, un caso especial es el modelo de regresión lineal,

$$\begin{aligned} y_i &= \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i \\ \varepsilon_i &\sim \mathcal{N}(0, \sigma^2) \end{aligned} \tag{2.1.4}$$

(McCullagh y Nelder, 1989; Ziegler, 2011; Kroese y Chan, 2014). Tratamientos completos del mlg aparecen en Nelder y Wedderburn (1972), McCullagh y Nelder (1989), Lindsey (1997), Dobson (2002), Yan y Su (2009), Bingham y Fry (2010), Myers *et al.* (2010), Ziegler (2011), Fahrmeir *et al.* (2013) y Kroese y Chan (2014), entre otros.

2.2. La Regresión Binaria Utilizando Variables Latentes

La siguiente definición y proposición es una extensión de los resultados presentados por Albert y Chib (1995), Fahrmeir *et al.* (2013) y Kroese y Chan (2014).

Definición 2.4. Sea $\mathbf{y} = (y_1, y_2, \ldots, y_n)'$ un vector de n variables aleatorias independientes binarias (0/1), $\mathbf{x}_i = (1, x_{i2}, \ldots, x_{ik})'$ un vector de diseño y $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_k)'$ un vector de coeficientes de regresión. El modelo de regresión binaria, está dado por

$$p_i = \Pr(y_i = 1) = F(\mathbf{x}_i'\boldsymbol{\beta}), \tag{2.2.1}$$

donde $F(\cdot)$ denota a la fda, su inverso $F^{-1}(\cdot)$ de acuerdo al mlg es llamado link.

Un link resulta ser simétrico cuando la fda procede de una fdp simétrica (Chen *et al.*, 1999; Bermúdez *et al.*, 2008; Bazán *et al.*, 2010). Por supuesto, un link resulta ser asimétrico cuando la fda procede de una fdp asimétrica (Chen *et al.*, 1999; Bermúdez *et al.*, 2008; Bazán *et al.*, 2010). También, un link asimétrico puede reducirse a un link simétrico (Stukel, 1988; Czado, 1994; Chen *et al.*, 1999; Bazán *et al.*, 2010).

La función de verosimilitud para el modelo de regresión binaria, está dada por

$$\mathcal{L}(\boldsymbol{\beta}|n, \mathbf{y}, \mathbf{X}) = \prod_{i=1}^{n} [F(\mathbf{x}_i'\boldsymbol{\beta})]^{y_i}[1 - F(\mathbf{x}_i'\boldsymbol{\beta})]^{1-y_i}. \tag{2.2.2}$$

Para simplificar la estimación e inferencia se puede utilizar la data aumentada (Bazán *et al.*, 2006, 2014; Bazán *et al.*, 2010; Kroese y Chan, 2014). La idea general detrás de la data aumentada es, incluir variables latentes (o ocultas) en el modelo para simplificar el análisis (Kroese y Chan, 2014). Pormenores importantes de modelos de variables latentes en el análisis de datos categóricos aparecen en Agresti y Kateri (2014).

Proposición 2.5. *El modelo de regresión binaria utilizando variables latentes, está dado por*

$$\begin{aligned} y_i &= \begin{cases} 1 & \text{si} \quad z_i > 0 \\ 0 & \text{si} \quad z_i \leq 0 \end{cases} \\ z_i &= \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i \\ \varepsilon_i &\sim F(\varepsilon_i), \end{aligned} \tag{2.2.3}$$

donde y_i *determina el signo de la variable latente* z_i *y* $F(\cdot)$ *denota a la fda.*

Demostración.

$$\begin{aligned} \Pr(y_i = 1) &= \Pr(z_i > 0) \\ &= \Pr(\varepsilon_i > -\mathbf{x}_i'\boldsymbol{\beta}) \\ &= 1 - F(-\mathbf{x}_i'\boldsymbol{\beta}). \\ \Pr(y_i = 0) &= F(-\mathbf{x}_i'\boldsymbol{\beta}). \\ y_i &\sim \mathcal{B}er(1 - F(-\mathbf{x}_i'\boldsymbol{\beta})). \end{aligned}$$

□

En la proposición anterior el residuo latente es $\varepsilon_i(z_i, \boldsymbol{\beta}) = z_i - \mathbf{x}_i'\boldsymbol{\beta}$ (Farias y Branco, 2012), fue definido por Albert y Chib (1995) para detectar valores extremos o *outliers*. Nótese que, si $F(\cdot)$ es la fda de un fdp simétrica alrededor del origen con soporte en la recta real $y_i \sim \mathcal{B}er(F(\mathbf{x}_i'\boldsymbol{\beta}))$ (Fahrmeir *et al.*, 2013; Kroese y Chan, 2014).

Las variables latentes también evitan trabajar con la verosimilitud tipo Bernoulli (Bazán *et al.*, 2006, 2014; Bazán *et al.*, 2010; Bolfarine y Bazán, 2010). La función de verosimilitud de data completa para el modelo de regresión binaria, está dada por

$$\mathcal{L}(\boldsymbol{\beta}|n, \mathbf{y}, \mathbf{X}, \mathbf{z}) = \prod_{i=1}^{n} f(z_i - \mathbf{x}_i'\boldsymbol{\beta})\{I_{\{1\}}(y_i)I_{\mathbb{R}_+}(z_i) + I_{\{0\}}(y_i)I_{\mathbb{R}_-}(z_i)\}, \tag{2.2.4}$$

donde $f(\cdot)$ denota a la fdp de la fda $F(\cdot)$ e $I_A(\cdot)$ es la función indicadora en el set A.

2.3. Los Detalles de Algunos Modelos de Regresión Binaria

Analicemos los tres modelos de regresión binaria comúnmente utilizados en la práctica.

- El probit (Bliss, 1935), tiene la fda procedente de la fdp normal estándar ($\mathcal{N}(0, 1)$), en consecuencia el link es simétrico. Utilizando variables latentes: $y_i = 1$ si $z_i \in \mathbb{R}_+$, $y_i = 0$ si $z_i \in \mathbb{R}_-$, $z_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i$ y $\varepsilon_i \sim \mathcal{N}(0, 1)$ (Fahrmeir *et al.*, 2013).
- El logit (Berkson, 1944), tiene la fda procedente de la fdp logística estándar ($\mathcal{L}o(0, 1)$), en consecuencia el link es simétrico. Utilizando variables latentes: $y_i = 1$ si $z_i \in \mathbb{R}_+$, $y_i = 0$ si $z_i \in \mathbb{R}_-$, $z_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i$ y $\varepsilon_i \sim \mathcal{L}o(0, 1)$ (Fahrmeir *et al.*, 2013).
- El cloglog (Gumbel, 1935), tiene la fda procedente de la fdp Gumbel estándar ($\mathcal{G}u(0, 1)$), en consecuencia el link es asimétrico. Utilizando variables latentes: $y_i = 1$ si $z_i \in \mathbb{R}_+$, $y_i = 0$ si $z_i \in \mathbb{R}_-$, $z_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i$ y $\varepsilon_i \sim \mathcal{G}u(0, 1)$ (Fahrmeir *et al.*, 2013).

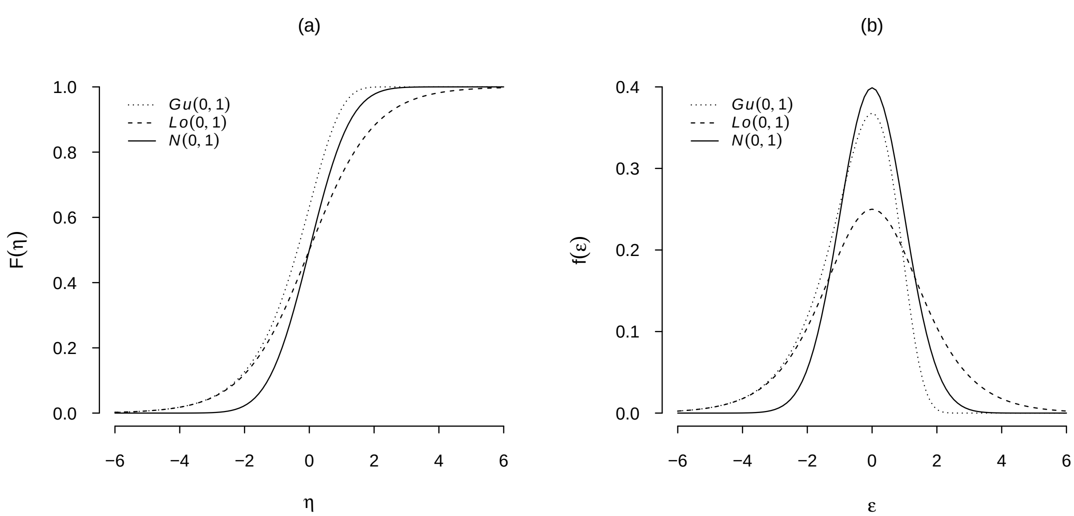

Figura 2.3.1. Plots de las (a) fda's $F(\eta)$ y (b) fdp's $f(\varepsilon)$

Revisemos dos modelos que modifican el predictor lineal del logit y del probit.

- El logit general (Stukel, 1988) es caracterizado por

$$p_i = \Pr(y_i = 1) = F(\mathbf{x}_i'\boldsymbol{\beta}; \lambda_1, \lambda_2) = \frac{\exp(m(\mathbf{x}_i'\boldsymbol{\beta}; \lambda_1, \lambda_2))}{1 + \exp(m(\mathbf{x}_i'\boldsymbol{\beta}; \lambda_1, \lambda_2))}, \tag{2.3.1}$$

donde para $\mathbf{x}_i'\boldsymbol{\beta} \geq 0$

$$m(\mathbf{x}_i'\boldsymbol{\beta}; \lambda_1, \lambda_2) = \begin{cases} \lambda_1^{-1}[\exp(\lambda_1 \mathbf{x}_i'\boldsymbol{\beta}) - 1] & \text{si} \quad \lambda_1 > 0 \\ \mathbf{x}_i'\boldsymbol{\beta} & \text{si} \quad \lambda_1 = 0 \\ -\lambda_1^{-1} \ln(1 - \lambda_1 \mathbf{x}_i'\boldsymbol{\beta}) & \text{si} \quad \lambda_1 < 0 \end{cases}$$

y para $\mathbf{x}_i'\boldsymbol{\beta} \leq 0$

$$m(\mathbf{x}_i'\boldsymbol{\beta}; \lambda_1, \lambda_2) = \begin{cases} -\lambda_2^{-1}[\exp(-\lambda_2 \mathbf{x}_i'\boldsymbol{\beta}) - 1] & \text{si} \quad \lambda_2 > 0 \\ \mathbf{x}_i'\boldsymbol{\beta} & \text{si} \quad \lambda_2 = 0 \\ \lambda_2 \ln(1 + \lambda_2 \mathbf{x}_i'\boldsymbol{\beta}) & \text{si} \quad \lambda_2 < 0. \end{cases}$$

λ_1 y λ_2 son los parámetros de forma. Para λ_1 y λ_2 fijos, el link está dado por

$$g(p_i) = m^{-1}\left[\ln\left(\frac{p_i}{1 - p_i}\right)\right] = \mathbf{x}_i'\boldsymbol{\beta}. \tag{2.3.2}$$

Cuando $\lambda_1 = \lambda_2 = 0$, se obtiene el link logit. Si $\lambda_1 = \lambda_2$, el link es simétrico, caso contrario es asimétrico.

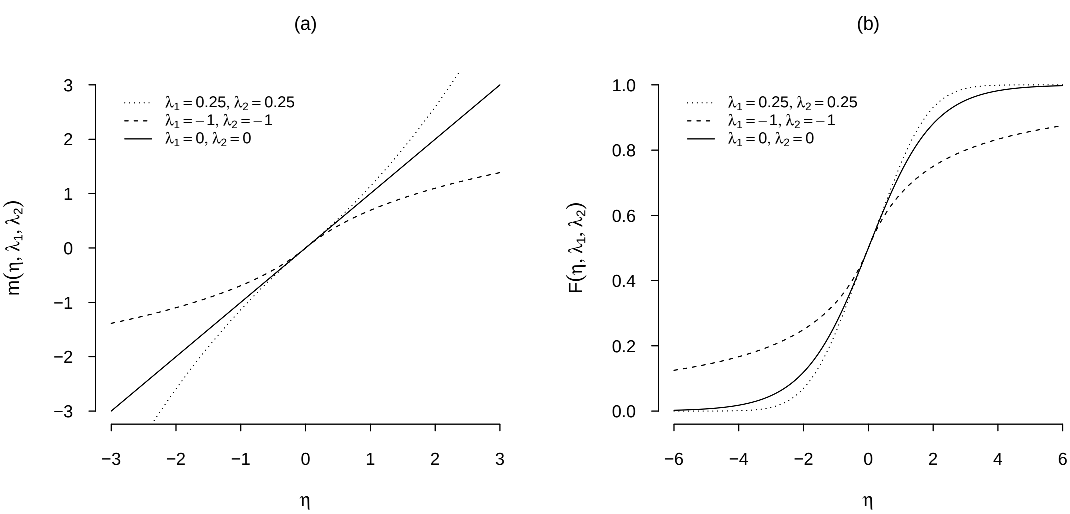

Figura 2.3.2. Plots de la (a) función $m(\eta; \lambda_1, \lambda_2)$ y (b) fda $F(\eta; \lambda_1, \lambda_2)$

- El probit general (Czado, 1994) es caracterizado por

$$p_i = \Pr(y_i = 1) = F(\mathbf{x}_i'\boldsymbol{\beta}; \lambda) = \Phi(m(\mathbf{x}_i'\boldsymbol{\beta}; \lambda)), \qquad (2.3.3)$$

donde $\Phi(\cdot)$ denota la fda de la $\mathcal{N}(0,1)$ y

$$m(\mathbf{x}_i'\boldsymbol{\beta}; \lambda) = \begin{cases} \lambda^{-1}[(\mathbf{x}_i'\boldsymbol{\beta} + 1)^{\lambda} - 1] & \text{si} \quad \mathbf{x}_i'\boldsymbol{\beta} \geq 0 \\ \mathbf{x}_i'\boldsymbol{\beta} & \text{si} \quad \mathbf{x}_i'\boldsymbol{\beta} < 0 \end{cases}$$

o

$$m(\mathbf{x}_i'\boldsymbol{\beta}; \lambda) = \begin{cases} \mathbf{x}_i'\boldsymbol{\beta} & \text{si} \quad \mathbf{x}_i'\boldsymbol{\beta} \geq 0 \\ -\lambda^{-1}[(-\mathbf{x}_i'\boldsymbol{\beta} + 1)^{\lambda} - 1] & \text{si} \quad \mathbf{x}_i'\boldsymbol{\beta} < 0. \end{cases}$$

$\lambda \in \mathbb{R}_+$ es el parámetro de forma. Cuando $\lambda = 0$ se obtiene el link probit. Si $\lambda \neq 0$, el link es asimétrico.

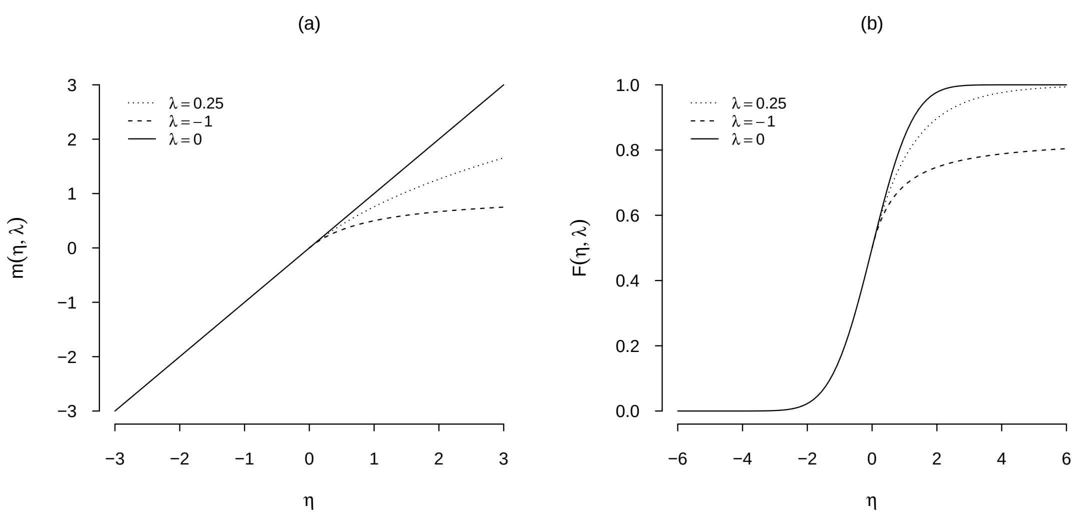

Figura 2.3.3. Plots de la (a) función $m(\eta; \lambda)$ y (b) fda $F(\eta; \lambda)$

Con fines comparativos, agregamos dos modelos de regresión binaria, que también capturan al logit, sus nombres van de acuerdo a Achen (2002) y Bolfarine y Bazán (2010).

- El scobit (Prentice, 1976; Achen, 2002; Bolfarine y Bazán, 2010), es caracterizado por

$$p_i = \Pr(y_i = 1) = F_4(\mathbf{x}_i'\boldsymbol{\beta}; \lambda) = 1 - [1 + \exp(\mathbf{x}_i'\boldsymbol{\beta})]^{-\lambda}, \qquad (2.3.4)$$

donde $\lambda \in \mathbb{R}_+$ es el parámetro de forma, la fda procede de la fdp asimétrica scobit estándar ($\mathcal{S}c(\lambda)$), en consecuencia el link es asimétrico y puede reducirse al link logit si

$\lambda = 1$. Utilizando variables latentes: $y_i = 1$ si $z_i \in \mathbb{R}_+$, $y_i = 0$ si $z_i \in \mathbb{R}_-$, $z_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i$ y $\varepsilon_i \sim \mathcal{S}c(\lambda)$.

- El power logit (Prentice, 1976; Nagler, 1994; Achen, 2002; Bolfarine y Bazán, 2010), es caracterizado por

$$p_i = \Pr(y_i = 1) = F_5(\mathbf{x}_i'\boldsymbol{\beta}; \lambda) = [1 + \exp(-\mathbf{x}_i'\boldsymbol{\beta})]^{-\lambda}, \tag{2.3.5}$$

  donde $\lambda \in \mathbb{R}_+$ es el parámetro de forma, la fda procede de la fdp asimétrica Burr tipo II estándar ($\mathcal{B}urr - II(\lambda)$), en consecuencia el link es asimétrico y puede reducirse al link logit si $\lambda = 1$. Utilizando variables latentes: $y_i = 1$ si $z_i \in \mathbb{R}_+$, $y_i = 0$ si $z_i \in \mathbb{R}_-$, $z_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i$ y $\varepsilon_i \sim \mathcal{B}urr - II(\lambda)$.

Nótese que, en el modelo scobit utilizando variables latentes, la probabilidad para $y_i = 1$, está dada por

$$\begin{aligned}\Pr(y_i = 1) &= \Pr(z_i > 0)\\ &= \Pr(\varepsilon_i > -\boldsymbol{x}_i'\boldsymbol{\beta})\\ &= 1 - F_4(-\mathbf{x}_i'\boldsymbol{\beta}; \lambda)\\ &= F_5(\mathbf{x}_i'\boldsymbol{\beta}; \lambda),\end{aligned} \tag{2.3.6}$$

naturalmente $\Pr(y_i = 0) = 1 - F_5(\mathbf{x}_i'\boldsymbol{\beta}; \lambda)$, en consecuencia $y_i \sim \mathcal{B}er(F_5(\mathbf{x}_i'\boldsymbol{\beta}; \lambda))$ y obviamente en el modelo power logit utilizando variables latentes $y_i \sim \mathcal{B}er(F_4(\mathbf{x}_i'\boldsymbol{\beta}; \lambda))$ (Bolfarine y Bazán, 2010).

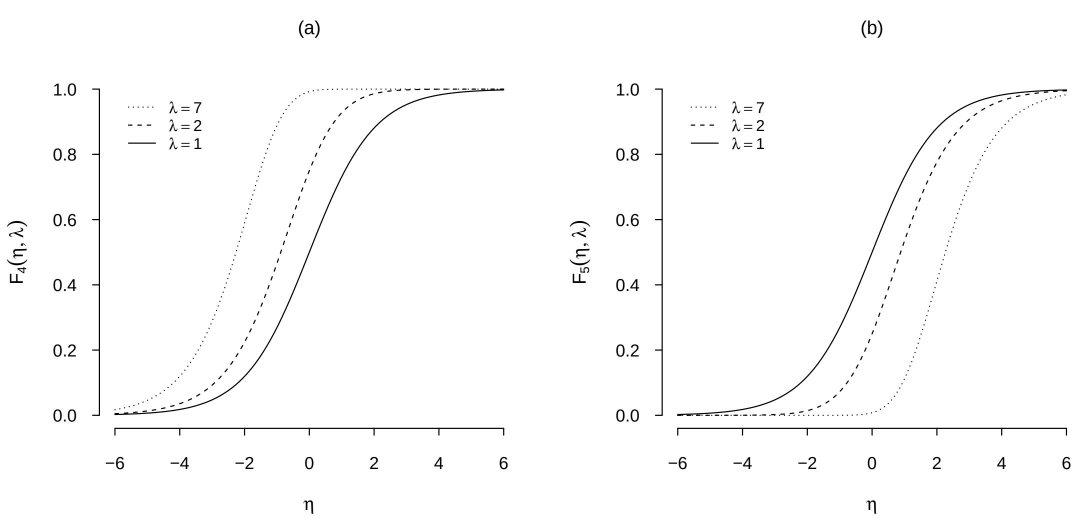

Figura 2.3.4. Plots de las fda's (a) $F_4(\eta; \lambda)$ y (b) $F_5(\eta; \lambda)$

Capítulo 3

La Distribución Skew Normal

El desarrollo de clases paramétricas y el estudio de sus propiedades siempre ha sido un tema persistente en la literatura estadística, aunque por supuesto no constantemente, ni con la misma intensidad. En los años recientes, el interés por la clase de distribuciones skew normal y relacionadas ha crecido enormemente, como la teoría ha avanzado, el desafío de los datos ha crecido y las herramientas computacionales han hecho un progreso sustancial. En este capítulo se intenta proveer una visión general e introductoria de la literatura dedicada a la clase skew normal en el contexto univariado.

3.1. El Modelo Estándar

Azzalini (1985, 1986, 2005, 2014) introduce la siguiente proposición, central para nuestro desarrollo.

Proposición 3.1. *Denotando por $f_0(\cdot)$ a la fdp en $\mathbb{R}^n$, por $G_0(\cdot)$ a la fda continua en $\mathbb{R}$, y por $h(\cdot)$ a la función de valor real en $\mathbb{R}^n$, tal que $f_0(-\mathbf{x}) = f_0(\mathbf{x})$, $G_0(-y) = 1 - G_0(y)$ y $h(-\mathbf{x}) = -h(\mathbf{x})$ para $\mathbf{x} \in \mathbb{R}^n$, $y \in \mathbb{R}$. Entonces*

$$f_n(\mathbf{x}) = 2f_0(\mathbf{x})G_0\{h(\mathbf{x})\} \tag{3.1.1}$$

es una fdp en $\mathbb{R}^n$.

Demostración. Ver Apéndice B. □

Definición 3.2. Una variable aleatoria x tiene una distribución skew normal estándar con parámetro de forma $\lambda \in \mathbb{R}$ y representaremos por $x \sim \mathcal{SN}(\lambda)$, si su fdp está dada por

$$\phi(x; \lambda) = 2\phi(x)\Phi(\lambda x), \tag{3.1.2}$$

donde $\phi(\cdot)$ y $\Phi(\cdot)$ denotan a la fdp y a la fda de la $\mathcal{N}(0, 1)$, respectivamente.

De acuerdo a la proposición 3.1, la $\mathcal{SN}(\lambda)$ es propia (Azzalini, 1985, 1986, 2005, 2014).

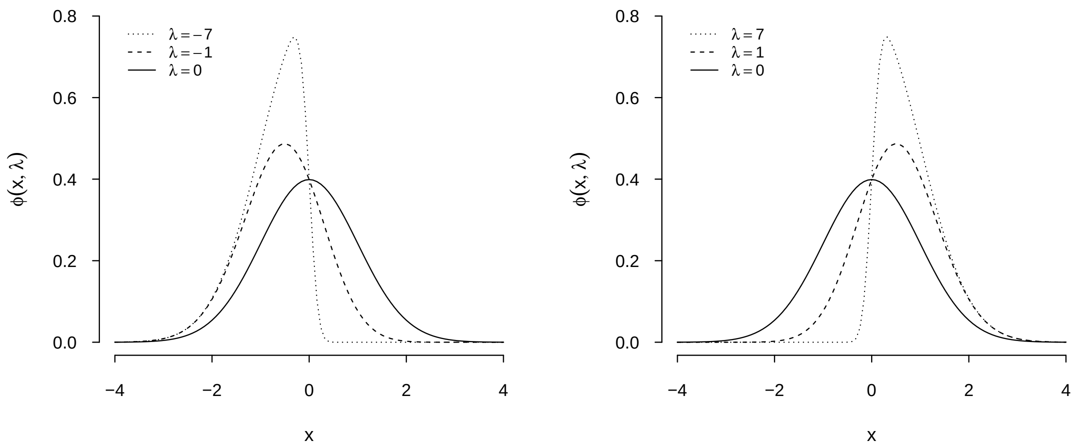

Figura 3.1.1. Plots de la fdp $\phi(x;\lambda)$

La $\mathcal{SN}(\lambda)$ tiene propiedades atractivas convenientes: (a) cuando $\lambda = 0$, $x \sim \mathcal{N}(0,1)$; (b) $|x| \sim \mathcal{HN}(0,1)$; (c) si $\lambda \to \infty$, converge a la $\mathcal{HN}(0,1)$; (d) $x^2 \sim \chi_1^2$; (e) $-x \sim \mathcal{SN}(-\lambda)$; (f) es fuertemente unimodal i.e. el $\ln\phi(x;\lambda)$ es una función cóncava de x; y (g) si $\epsilon \sim \mathcal{N}(0,1)$ es independiente de x, entonces $\frac{a\epsilon+bx}{\sqrt{a^2+b^2}} \sim \mathcal{SN}\Big(\frac{b\lambda}{\sqrt{a^2(1+\lambda^2)+b^2}}\Big)$ (Azzalini, 1985, 1986, 2005, 2014; Dalla–Valle, 2004; Gupta *et al.*, 2004; Bagui y Bagui, 2006; Mameli y Musio, 2013; Ahsanullah *et al.*, 2014; Sarısoy *et al.*, 2014).

Vidal *et al.* (2006) expresan la distancia media $L_1(\cdot)$ entre la $\mathcal{SN}(\lambda)$ y la $\mathcal{N}(0,1)$,

$$L_1(\lambda) = \frac{1}{2}\int_{-\infty}^{\infty} |\phi(x;\lambda) - \phi(x)|dx = \frac{1}{2} - \frac{1}{\pi}\arccos\left(\frac{|\lambda|}{\sqrt{1+\lambda^2}}\right). \tag{3.1.3}$$

$L_1(\cdot) : \mathbb{R} \to [0, 0.5]$ (ver Apéndice B), $L_1(\lambda) = 0$ indica que $\phi(x;0) = \phi(x)$ y $L_1(\lambda) = 0.5$ indica discrepancia (Weiss, 1996; Vidal *et al.*, 2006).

3.1.1. La representación estocástica

Henze (1986) introduce la siguiente proposición, da una representación estocástica a la $\mathcal{SN}(\lambda)$ en términos de la $\mathcal{N}(0,1)$ y la $\mathcal{HN}(0,1)$, i.e. revela la estructura de la $\mathcal{SN}(\lambda)$.

Proposición 3.3. *Si $\epsilon \sim \mathcal{N}(0,1)$ es independiente de $\tau \sim \mathcal{HN}(0,1)$, entonces*

$$x = \sqrt{1-\delta^2}\epsilon + \delta\tau \sim \mathcal{SN}(\lambda), \tag{3.1.4}$$

donde $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}} \in [-1,1]$.

Demostración.

$$\begin{aligned}
\Pr(\mathrm{x} \leq x) &= \Pr\left(\epsilon \leq \frac{x-\delta\tau}{\sqrt{1-\delta^2}}\right) \\
&= \mathbb{E}\left[\Pr\left(\epsilon \leq \frac{x-\delta\tau}{\sqrt{1-\delta^2}}\middle|\, \tau\right)\right] \\
&= \int_0^\infty \Phi\left(\frac{x-\delta\tau}{\sqrt{1-\delta^2}}\right)\psi(\tau)d\tau \ldots
\end{aligned}$$

□

Proposición 3.4. *Si* $x|\tau \sim \mathcal{N}(\delta\tau, 1-\delta^2)$ *y* $\tau \sim \mathcal{HN}(0,1)$*, entonces*

$$x \sim \mathcal{SN}(\lambda), \tag{3.1.5}$$

donde $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}} \in [-1,1]$*.*

Demostración.

$$\begin{aligned}
\frac{\partial \Pr(\mathrm{x} \leq x)}{\partial x} &= \int_0^\infty \phi(x|\tau;\delta\tau,1-\delta^2)\psi(\tau)d\tau \\
&= 2\phi(x)\int_0^\infty \phi(\tau;\delta x, 1-\delta^2)d\tau \\
&= 2\phi(x)\Pr\left(\frac{\tau-\delta x}{\sqrt{1-\delta^2}} > -\lambda x\right) \\
&= 2\phi(x)\Phi(\lambda x).
\end{aligned}$$

□

Nótese que, la proposición anterior es complementaria a la proposición 3.3.

3.1.2. Los momentos

La subsección cubre cuatro alternativas para poder determinar los momentos de la $\mathcal{SN}(\lambda)$.

Lema 3.5. *Si* $\epsilon \sim \mathcal{N}(0,1)$ *y* λ_0*,* $\lambda_1 \in \mathbb{R}$*, entonces*

$$\mathbb{E}[\Phi(\lambda_0\epsilon + \lambda_1)] = \Phi\left(\frac{\lambda_1}{\sqrt{1+\lambda_0^2}}\right), \tag{3.1.6}$$

donde $\Phi(\cdot)$ *denota a la fda de la* $\mathcal{N}(0,1)$*.*

Demostración. Ver Apéndice B o Zacks (1981, pp. 53–54). □

Azzalini (1985, 1986, 2005, 2014) introduce la siguiente proposición.

Proposición 3.6. *La función generatriz de momentos de la $\mathcal{SN}(\lambda)$, está dada por*

$$\mathbb{M}_x(t) = 2\exp\left(\frac{1}{2}t^2\right)\Phi(\delta t), \tag{3.1.7}$$

donde $\Phi(\cdot)$ *denota a la fda de la* $\mathcal{N}(0,1)$ *y* $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}} \in [-1,1]$.

Demostración.

$$\begin{aligned}
\mathbb{M}_x(t) &= 2\int_{-\infty}^{\infty} \exp(tx)\phi(x)\Phi(\lambda x)dx \\
&= 2\exp\left(\frac{1}{2}t^2\right)\int_{-\infty}^{\infty} \Phi(\lambda x)\phi(x-t)dx \\
&= 2\exp\left(\frac{1}{2}t^2\right)\mathbb{E}[\Phi(\lambda\epsilon + \lambda t)] \\
&= 2\exp\left(\frac{1}{2}t^2\right)\Phi(\delta t).
\end{aligned}$$

□

Los momentos de orden par de la $\mathcal{SN}(\lambda)$ son iguales a los momentos de orden par de la $\mathcal{N}(0,1)$ (Azzalini, 1985, 1986, 2005, 2014; Henze, 1986; Dalla–Valle, 2004; Arellano–Valle *et al.*, 2005; Bagui y Bagui, 2006; Martínez *et al.*, 2008; Ahsanullah *et al.*, 2014).

Henze (1986) introduce la siguiente proposición.

Proposición 3.7. *Los momentos de orden impar de la $\mathcal{SN}(\lambda)$, están dados por*

$$\mathbb{E}[x^{2n+1}] = \sqrt{\frac{2}{\pi}}\frac{\lambda}{(1+\lambda^2)^{(2n+1)/2}}\frac{(2n+1)!}{2^n}\sum_{k=0}^{n}\frac{k!(2\lambda)^{2k}}{(2k+1)!(n-k)!}, \quad n = 0,1,\ldots \tag{3.1.8}$$

Demostración.

$$\begin{aligned}
\mathbb{E}[x^{2n+1}] &= \mathbb{E}[(\sqrt{1-\delta^2}\epsilon + \delta\tau)^{2n+1}] \\
&= \left(\frac{\lambda}{\sqrt{1+\lambda^2}}\right)^{2n+1}\sum_{k=0}^{2n+1}\binom{2n+1}{k}\lambda^{-k}\mathbb{E}[\epsilon^k]\mathbb{E}[\tau^{2n+1-k}] \\
&\overset{\text{ig}}{=} \left(\frac{\lambda}{\sqrt{1+\lambda^2}}\right)^{2n+1}\sum_{k=0}^{n}\binom{2n+1}{2k}\lambda^{-2k}\mathbb{E}[\epsilon^{2k}]\mathbb{E}[\tau^{2(n-k)+1}] \\
&= \sqrt{\frac{2}{\pi}}\frac{\lambda}{(1+\lambda^2)^{(2n+1)/2}}\frac{(2n+1)!}{2^n}\sum_{k=0}^{n}\frac{(n-k)!(2\lambda)^{2(n-k)}}{(2(n-k)+1)!k!} \\
&\overset{\text{ig}}{=} \sqrt{\frac{2}{\pi}}\frac{\lambda}{(1+\lambda^2)^{(2n+1)/2}}\frac{(2n+1)!}{2^n}\sum_{k=0}^{n}\frac{k!(2\lambda)^{2k}}{(2k+1)!(n-k)!}.
\end{aligned}$$

□

Martínez *et al.* (2008) introducen las siguientes proposiciones.

Proposición 3.8. *Los momentos de orden impar de la $\mathcal{SN}(\lambda)$, también están dados por*

$$\mathbb{E}[x^{2n+1}] = 2n\mathbb{E}[x^{2n-1}] + \sqrt{\frac{2}{\pi}}\frac{\lambda}{(1+\lambda^2)^{(2n+1)/2}}\frac{(2n)!}{2^n n!}, \quad n = 1, 2, \ldots \tag{3.1.9}$$

Demostración.

$$\begin{aligned}
\mathbb{E}[x^{2n-1}] &= 2\int_{-\infty}^{\infty} x^{2n-1}\phi(x)\Phi(\lambda x)dx \\
&\overset{\text{pp}}{=} \frac{x^{2n}}{n}\phi(x)\Phi(\lambda x)\Big|_{-\infty}^{\infty} + \frac{\mathbb{E}[x^{2n+1}]}{2n} - \frac{\lambda}{n}\int_{-\infty}^{\infty} x^{2n}\phi(x)\phi(\lambda x)dx \\
&= \frac{\mathbb{E}[x^{2n+1}]}{2n} - \frac{1}{n}\frac{\lambda}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x^{2n}\phi(\sqrt{1+\lambda^2}x)dx \\
&= \frac{\mathbb{E}[x^{2n+1}]}{2n} - \frac{1}{2n}\sqrt{\frac{2}{\pi}}\frac{\lambda}{(1+\lambda^2)^{(2n+1)/2}}\mathbb{E}[\epsilon^{2n}]. \\
\mathbb{E}[x^{2n+1}] &= 2n\mathbb{E}[x^{2n-1}] + \sqrt{\frac{2}{\pi}}\frac{\lambda}{(1+\lambda^2)^{(2n+1)/2}}\frac{(2n)!}{2^n n!}.
\end{aligned}$$

□

Hasta aquí las tres alternativas son exigentes, i.e. los cálculos para obtener los momentos son extensos, porque para determinar el momento de interés se requiere: en la proposición 3.6, la derivada con su respectiva evaluación en $t = 0$; en la proposición 3.7, las funciones factorial y sumatoria; y en la proposición anterior, la función factorial y la recursividad. En la siguiente proposición, Martínez *et al.* (2008) ubican la función doble factorial en la proposición 3.8 y luego de manipulaciones algebraicas identifican un triángulo tipo Pascal que simplifica en gran medida los cálculos para obtener los coeficientes de los momentos de orden impar de la $\mathcal{SN}(\lambda)$.

Proposición 3.9. *Los momentos de orden impar de la $\mathcal{SN}(\lambda)$, también están dados por*

$$\mathbb{E}[x^{2n-1}] = \sqrt{\frac{2}{\pi}}\frac{1}{(1+\lambda^2)^{(2n-1)/2}}\sum_{k=1}^{n} a_n(k)\lambda^{2k-1}, \quad n = 1, 2, \ldots \tag{3.1.10}$$

donde

$$\begin{aligned}
&a_1(1) = 1; \\
&a_n(1) = (2n-1)a_{n-1}(1);\ n \geq 2; \\
&a_n(k) = (2n-2)[a_{n-1}(k) + a_{n-1}(k-1)];\ n \geq 2,\ 1 < k < n; \\
&a_n(n) = (2n-2)a_{n-1}(n-1).
\end{aligned}$$

Demostración.

$$\begin{aligned}
\mathbb{E}[x^{2n+1}] &= 2n\mathbb{E}[x^{2n-1}] + \sqrt{\frac{2}{\pi}}\frac{\lambda}{(1+\lambda^2)^{(2n+1)/2}}\frac{(2n)!}{2^n(n)!} \\
&\overset{\text{df}}{=} 2n\mathbb{E}[x^{2n-1}] + \sqrt{\frac{2}{\pi}}\frac{\lambda}{(1+\lambda^2)^{(2n+1)/2}}(2n-1)!! \\
&\overset{\text{ig}}{=} \sqrt{\frac{2}{\pi}}\sum_{k=0}^{n}\frac{(2n)!!(2k-1)!!}{(2k)!!}\frac{\lambda}{(1+\lambda^2)^{(2k+1)/2}} \\
&= \sqrt{\frac{2}{\pi}}\frac{\lambda}{(1+\lambda^2)^{(2n+1)/2}}\sum_{k=0}^{n}\frac{(2n)!!(2k-1)!!}{(2k)!!}(1+\lambda^2)^{n-k} \\
&= \sqrt{\frac{2}{\pi}}\frac{\lambda}{(1+\lambda^2)^{(2n+1)/2}}\sum_{k=0}^{n}\sum_{j=0}^{n-k}\frac{(2n)!!(2k-1)!!}{(2k)!!}\binom{n-k}{j}\lambda^{2j}.
\end{aligned}$$

Iterando y centrando atención en los coeficientes de λ^{2j+1}, formamos un triángulo tipo Pascal, esto prueba la proposición. $\square$

Ilustremos la proposición anterior con $n = 4$, los coeficientes se obtienen realizando

$$\begin{gathered}
a_1(1) = 1 \\
a_2(1) = 3a_1(1) \quad a_2(2) = 2a_1(1) \\
a_3(1) = 5a_2(1) \quad a_3(2) = 4[a_2(1) + a_2(2)] \quad a_3(3) = 4a_2(2) \\
a_4(1) = 7a_3(1) \quad a_4(2) = 6[a_3(1) + a_3(2)] \quad a_4(3) = 6[a_3(2) + a_3(3)] \quad a_4(4) = 6a_3(3),
\end{gathered}$$

disponiéndolos en el triángulo tipo Pascal,

$$\begin{array}{ccccccc}
 & & & 1 & & & \\
 & & 3 & & 2 & & \\
 & 15 & & 20 & & 8 & \\
105 & & 210 & & 168 & & 48
\end{array}$$

tendríamos inmediatamente los cuatro primeros momentos de orden impar, por ejemplo

$$\mathbb{E}[x^7] = \sqrt{\frac{2}{\pi}}\frac{\lambda}{(1+\lambda^2)^{7/2}}(105 + 210\lambda^2 + 168\lambda^4 + 48\lambda^6).$$

Proposición 3.10. *La esperanza y la varianza de la $\mathcal{SN}(\lambda)$, están dadas por*

$$\mu_x = \sqrt{\frac{2}{\pi}}\delta \qquad y \qquad \sigma_x^2 = 1 - \frac{2}{\pi}\delta^2, \tag{3.1.11}$$

donde $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}} \in [-1, 1]$.

Demostración. Es inmediata utilizando (3.1.7) o combinando (A.0.1) con (3.1.8) o (3.1.9) o (3.1.10). □

Proposición 3.11. *Los coeficientes de asimetría y curtosis de la* $\mathcal{SN}(\lambda)$*, están dados por*

$$\gamma = \frac{4-\pi}{2}\operatorname{sgn}(\lambda)\left[\frac{\mu_x}{\sigma_x}\right]^3 \qquad y \qquad \kappa = 2(\pi-3)\left[\frac{\mu_x}{\sigma_x}\right]^4, \tag{3.1.12}$$

donde $\operatorname{sgn}(\cdot)$ *denota a la función signo.*

Demostración. Es inmediata utilizando (3.1.7) o combinando (A.0.1) con (3.1.8) o (3.1.9) o (3.1.10). □

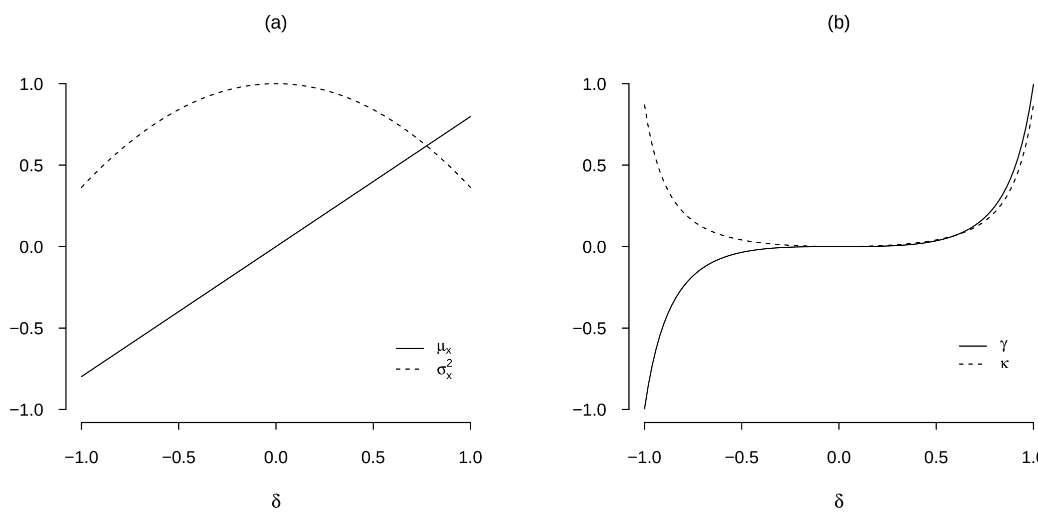

Figura 3.1.2. Plots de (a) μ_x, σ_x^2, (b) γ y κ para $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}} \in [-1, 1]$.

No dejan de ser interesantes los intervalos de $\mu_x \in [-0.797, 0.797]$, $\sigma_x^2 \in [0.363, 1]$, $\gamma \in [-0.995, 0.995]$ y $\kappa \in [0, 0.869]$ (Azzalini, 1985, 1986, 2005, 2014; Dalla–Valle, 2004; Bazán *et al.*, 2006).

3.1.3. La función de distribución acumulada

La subsección cubre tres alternativas para poder determinar la fda de la $\mathcal{SN}(\lambda)$.

Azzalini (1985, 1986, 2005, 2014) introduce la siguiente proposición, utiliza la función de Owen (1956)

$$\mathcal{O}(x;\lambda) = \int_x^\infty \int_0^{\lambda s} \phi(s)\phi(t)dtds, \quad x, \lambda \in \mathbb{R}_+$$

(ver Apéndice B) para obtener una expresión amable con la fda de la $\mathcal{SN}(\lambda)$.

Proposición 3.12. *La fda de la $\mathcal{SN}(\lambda)$, está dada por*

$$\Phi(x;\lambda) = \Phi(x) - 2\mathcal{O}(x;\lambda), \quad x,\lambda \in \mathbb{R}_+ \tag{3.1.13}$$

donde $\Phi(\cdot)$ y $\mathcal{O}(\cdot)$ denotan a la fda de la $\mathcal{N}(0,1)$ y a la función de Owen, respectivamente.

Demostración.

$$\Phi(x;\lambda) = 2\left[\int_{-\infty}^{x}\int_{-\infty}^{0}\phi(s)\phi(t)dtds - \mathcal{O}(x;\lambda)\right] = \Phi(x) - 2\mathcal{O}(x;\lambda).$$

□

También y sin olvidarnos, la fda de la $\mathcal{SN}(\lambda)$ tiene propiedades atractivas convenientes: (a) $\Phi(x;-\lambda) = 1-\Phi(-x;\lambda)$; (b) $\Phi(x;1) = \Phi^2(x)$; (c) $\sup_x|\Phi(x;\lambda)-\Phi(x)| = \frac{1}{\pi}\arctan|\lambda|$; y (d) $\Phi(0;\lambda) = \frac{1}{2} - \frac{1}{\pi}\arctan\lambda$ (Azzalini, 1985, 1986, 2014; Chiogna, 1998; Azzalini y Chiogna, 2004; Dalla–Valle, 2004; Mameli y Musio, 2013).

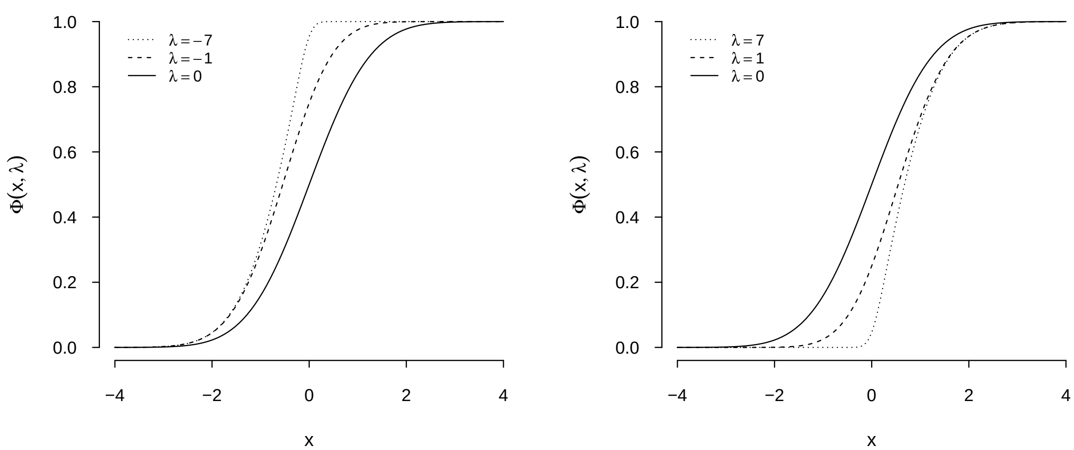

Figura 3.1.3. Plots de la fda $\Phi(x;\lambda)$

Bazán *et al.* (2006, 2014) y Bazán *et al.* (2010) introducen las siguientes proposiciones.

Proposición 3.13. *La fda de la $\mathcal{SN}(\lambda)$, también está dada por*

$$\Phi(x;\lambda) = \int_{0}^{\infty}\Phi(x\sqrt{1+\lambda^2} - \lambda\tau)\psi(\tau)d\tau, \tag{3.1.14}$$

donde $\Phi(\cdot)$ y $\psi(\cdot)$ denotan a la fda y a la fdp de la $\mathcal{N}(0,1)$ y la $\mathcal{HN}(0,1)$, respectivamente.

Demostración. Es inmediata utilizando (3.1.4) o (3.1.5) □

Proposición 3.14. *La fda de la $\mathcal{SN}(\lambda)$, también está dada por*

$$\Phi(x;\lambda) = 2\Phi_2\left(\begin{pmatrix} x \\ 0 \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & -\delta \\ -\delta & 1 \end{pmatrix}\right) \tag{3.1.15}$$

donde $\Phi_2(\cdot;\mathbf{0},\mathbf{\Sigma})$ denota a la fda de la $\mathcal{N}_2(\mathbf{0},\mathbf{\Sigma})$ y $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}} \in [-1,1]$.

Demostración.

$$\begin{aligned}
\Phi(x;\lambda) &= 2\int_{-\infty}^{x}\int_{-\infty}^{\lambda s_1} \phi(s_1)\phi(t)dtds_1 \\
&= 2\int_{-\infty}^{x}\int_{-\infty}^{0} \phi(s_1)\sqrt{1+\lambda^2}\phi(\lambda s_1 + \sqrt{1+\lambda^2}s_2)ds_2ds_1 \\
&= 2\int_{-\infty}^{x}\int_{-\infty}^{0} \left(\frac{1}{\sqrt{2\pi}}\right)^2 \sqrt{1+\lambda^2}\exp\left\{-\frac{1}{2}[s_1^2 + (\lambda s_1 + \sqrt{1+\lambda^2}s_2)^2]\right\} ds_2ds_1 \\
&= 2\int_{-\infty}^{x}\int_{-\infty}^{0} \left(\frac{1}{\sqrt{2\pi}}\right)^2 \frac{1}{\sqrt{1-\delta^2}}\exp\left\{-\frac{1}{2}\frac{1}{1-\delta^2}[s_1^2 - 2(-\delta)s_1s_2 + s_2^2]\right\} ds_2ds_1 \\
&= 2\int_{-\infty}^{x}\int_{-\infty}^{0} \phi_2\left(\begin{pmatrix} s_1 \\ s_2 \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & -\delta \\ -\delta & 1 \end{pmatrix}\right) ds_2ds_1 \\
&= 2\Phi_2\left(\begin{pmatrix} x \\ 0 \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & -\delta \\ -\delta & 1 \end{pmatrix}\right).
\end{aligned}$$

□

Para el cálculo de la probabilidad en las proposiciones 3.12 y 3.14, se recurre a métodos numéricos computacionales disponibles en los programas RStudio y R (Gupta y Chen, 2001; Bagui y Bagui, 2006; Genz y Bretz, 2009; Azzalini, 2014). En el capítulo 4, se da detalles de la procedencia de la proposición 3.13.

3.2. El Modelo

En esta sección, veremos que, el desarrollo del modelo y todas las proposiciones son una extensión del modelo estándar y sus proposiciones.

Definición 3.15. Una variable aleatoria y tiene una distribución skew normal con parámetros de localización $\xi \in \mathbb{R}$, escala $\omega \in \mathbb{R}_+$ y forma $\lambda \in \mathbb{R}$ y representaremos por $y \sim \mathcal{SN}(\xi,\omega^2,\lambda)$, si su fdp está dada por

$$\phi(y;\xi,\omega^2,\lambda) = \frac{2}{\omega}\phi\left(\frac{y-\xi}{\omega}\right)\Phi\left(\lambda\frac{y-\xi}{\omega}\right), \tag{3.2.1}$$

donde $\phi(\cdot)$ y $\Phi(\cdot)$ denotan a la fdp y a la fda de la $\mathcal{N}(0,1)$, respectivamente.

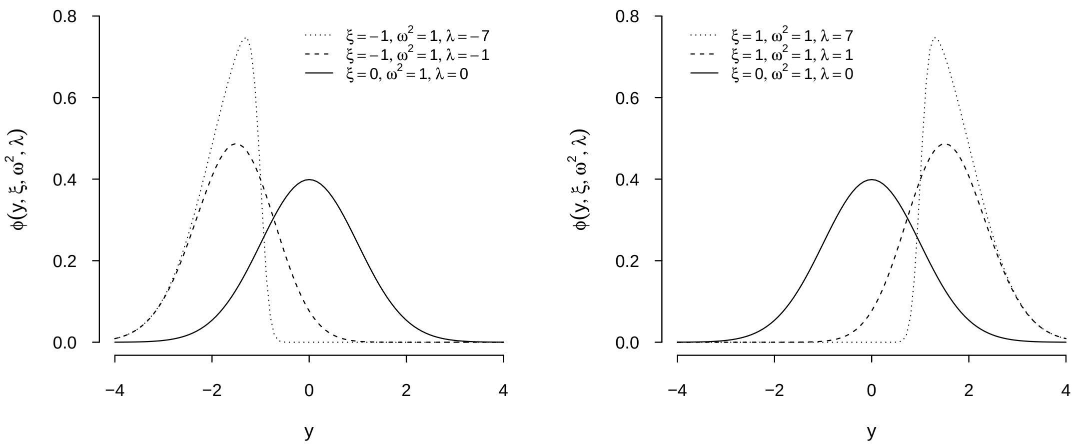

Figura 3.2.1. Plots de la fdp $\phi(y; \xi, \omega^2, \lambda)$

Proposición 3.16. *Si* $x \sim \mathcal{SN}(\lambda)$, *entonces*

$$y = \xi + \omega x \sim \mathcal{SN}(\xi, \omega^2, \lambda). \tag{3.2.2}$$

Demostración. Es inmediata utilizando la transformación. □

Corolario 3.17. *Si* $y \sim \mathcal{SN}(\xi, \omega^2, \lambda)$ *y* $x \sim \mathcal{SN}(\lambda)$, *entonces*

$$\frac{y-\xi}{\omega} \sim \mathcal{SN}(\lambda) \qquad \text{y} \qquad -\xi - \omega x \sim \mathcal{SN}(-\xi, \omega^2, -\lambda). \tag{3.2.3}$$

3.2.1. La representación estocástica

Proposición 3.18. *Si* $\epsilon \sim \mathcal{N}(0,1)$ *es independiente de* $\tau \sim \mathcal{HN}(0,1)$, *entonces*

$$y = \xi + \omega(\sqrt{1-\delta^2}\epsilon + \delta\tau) \sim \mathcal{SN}(\xi, \omega^2, \lambda), \tag{3.2.4}$$

donde $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}} \in [-1, 1]$.

Demostración. Es inmediata utilizando (3.1.4) y (3.2.2). □

Proposición 3.19. *Si* $y|\tau \sim \mathcal{N}(\xi + \omega\delta\tau, \omega^2(1-\delta^2))$ *y* $\tau \sim \mathcal{HN}(0,1)$, *entonces*

$$y \sim \mathcal{SN}(\xi, \omega^2, \lambda), \tag{3.2.5}$$

donde $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}} \in [-1, 1]$.

Demostración. Es inmediata utilizando (3.1.5) y (3.2.2). □

3.2.2. Los momentos

Proposición 3.20. *La función generatriz de momentos de la $\mathcal{SN}(\xi, \omega^2, \lambda)$, está dada por*

$$\mathbb{M}_y(t) = 2\exp\left(\xi t + \frac{1}{2}\omega^2 t^2\right)\Phi(\omega\delta t), \tag{3.2.6}$$

donde $\Phi(\cdot)$ denota a la fda de la $\mathcal{N}(0,1)$ y $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}} \in [-1, 1]$.

Demostración. Es inmediata utilizando (3.1.7) y (3.2.2). □

Proposición 3.21. *Los momentos respecto al origen de la $\mathcal{SN}(\xi, \omega^2, \lambda)$, están dados por*

$$\mathbb{E}[y^n] = \sum_{k=0}^{n} \binom{n}{k} \xi^k \omega^{n-k} \mathbb{E}[x^{n-k}], \quad n = 1, 2, \ldots \tag{3.2.7}$$

donde $y = \xi + \omega x$ y $x \sim \mathcal{SN}(\lambda)$.

Demostración. Es inmediata utilizando la expansión binomial. □

Proposición 3.22. *La esperanza y la varianza de la $\mathcal{SN}(\xi, \omega^2, \lambda)$, están dadas por*

$$\mu_y = \xi + \omega\mu_x \qquad \text{y} \qquad \sigma_y^2 = \omega^2(1 - \mu_x^2). \tag{3.2.8}$$

Demostración. Es inmediata utilizando (3.2.6) o (3.2.7). □

Los coeficientes de asimetría y curtosis de la $\mathcal{SN}(\xi, \omega^2, \lambda)$ son iguales a los coeficientes de asimetría y curtosis de la $\mathcal{SN}(\lambda)$ debido a que,

$$\frac{y - \mu_y}{\sigma_y} = \frac{(\xi + \omega x) - \mu_{\xi+\omega x}}{\sigma_{\xi+\omega x}} = \frac{x - \mu_x}{\sigma_x}$$

(Azzalini, 1985, 1986, 2005, 2014; Dalla–Valle, 2004; Arellano–Valle *et al.*, 2005; Ahsanullah *et al.*, 2014).

3.2.3. La función de distribución acumulada

Proposición 3.23. *La fda de la $\mathcal{SN}(\xi, \omega^2, \lambda)$, está dada por*

$$\Phi(y; \xi, \omega^2, \lambda) = \Phi\left(\frac{y - \xi}{\omega}\right) - 2\mathcal{O}\left(\frac{y - \xi}{\omega}, \lambda\right), \quad y - \xi, \lambda \in \mathbb{R}_+ \tag{3.2.9}$$

donde $\Phi(\cdot)$ y $\mathcal{O}(\cdot)$ denotan a la fda de la $\mathcal{N}(0,1)$ y a la función de Owen, respectivamente.

Demostración. Es inmediata utilizando (3.1.13) y (3.2.2). □

Proposición 3.24. *La fda de la $\mathcal{SN}(\xi, \omega^2, \lambda)$, también está dada por*

$$\Phi(y; \xi, \omega^2, \lambda) = \int_0^{\infty} \Phi\left(\frac{y - \xi}{\omega}\sqrt{1 + \lambda^2} - \lambda\tau\right)\psi(\tau)d\tau, \tag{3.2.10}$$

donde $\Phi(\cdot)$ y $\psi(\cdot)$ denotan a la fda y a la fdp de la $\mathcal{N}(0, 1)$ y la $\mathcal{HN}(0, 1)$, respectivamente.

Demostración. Es inmediata utilizando (3.1.14) y (3.2.2). □

Proposición 3.25. *La fda de la $\mathcal{SN}(\xi, \omega^2, \lambda)$, también está dada por*

$$\Phi(y; \xi, \omega^2, \lambda) = 2\Phi_2\left(\begin{pmatrix} y \\ 0 \end{pmatrix}; \begin{pmatrix} \xi \\ 0 \end{pmatrix}, \begin{pmatrix} \omega^2 & -\delta\omega \\ -\delta\omega & 1 \end{pmatrix}\right), \tag{3.2.11}$$

donde $\Phi_2(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denota a la fda de la $\mathcal{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ y $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}} \in [-1, 1]$.

Demostración. Es inmediata utilizando (3.1.15) y (3.2.2). □

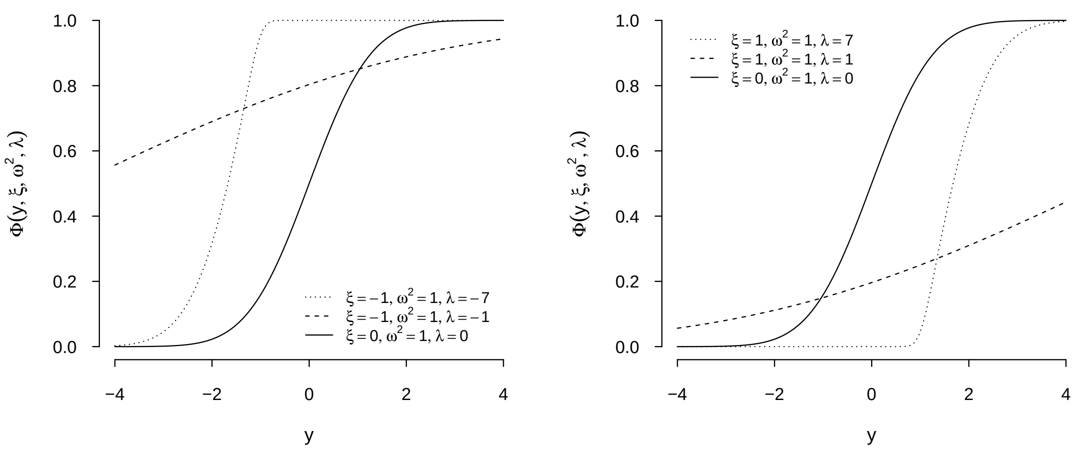

Figura 3.2.2. Plots de la fda $\Phi(y; \xi, \omega^2, \lambda)$

3.3. Los Modelos Muestrales

Un detalle también importante, es el estudio de las distribuciones muestrales de $\bar{x}$, s^2 y $\sqrt{n}\bar{x}/s$.

Gupta y Chen (2003) introducen las siguientes proposiciones.

Nótese que, cuando $\mathbf{x} = (x_1, x_2, \ldots, x_n)' \sim \mathcal{SN}_n(\mathbf{0}, \mathbf{I}_n, \boldsymbol{\lambda})$ no implica que $x_1, x_2, \ldots, x_n$ sean independientes.

Proposición 3.26. *Si* $\mathbf{x} = (x_1, x_2, \ldots, x_n)' \sim \mathcal{SN}_n(\mathbf{0}, \mathbf{I}_n, \boldsymbol{\lambda})$*, entonces*

- *la esperanza muestral* $\sqrt{n}\bar{x} \sim \mathcal{SN}(\lambda)$*,*
- *la varianza muestral* $(n-1)s^2 \sim \chi^2_{n-1}$ *y*
- $\bar{x}$ *y* s^2 *son independientes.*

Demostración. Consideremos la matriz de Helmert, definida por

$$\mathbf{A} = \mathbf{D}^{-1} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & -1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & -(n-1) \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

donde $\mathbf{D}^2$ es una matriz diagonal con elementos diagonales $d_{ii}^2 = n, 1\times 2, 2\times 3, \ldots, (n-1)n$ (Lancaster, 1965). $\mathbf{A}$ es ortogonal, i.e. $\mathbf{A}'\mathbf{A} = \mathbf{I}_n$, $\mathbf{A}\mathbf{A}' = \mathbf{I}_n$ y $|\mathbf{A}| = 1$. También, consideremos la transformación $\mathbf{v} = \mathbf{A}\mathbf{x}$, donde $\mathbf{v} = (v_1, v_2, \ldots, v_n)'$ y $v_1 = \sqrt{n}\bar{x}$. Realicemos particiones: (a) en $\mathbf{v}$, con $\mathbf{w} = (v_2, v_3, \ldots, v_n)'$ se tiene que $\mathbf{v} = (v_1, \mathbf{w}')'$; y (b) en $\mathbf{A}$, con $\mathbf{a}_i = (a_{i1}, a_{i2}, \ldots, a_{in})'$ se tiene que $\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n)'$. Observemos que, $(n-1)s^2 = \mathbf{x}'\mathbf{x} - (\sqrt{n}\bar{x})^2 = \mathbf{w}'\mathbf{w} = \mathbf{x}'(\mathbf{a}_2, \mathbf{a}_3, \ldots, \mathbf{a}_n)(\mathbf{a}_2, \mathbf{a}_3, \ldots, \mathbf{a}_n)'\mathbf{x}$.

Por las proposiciones 5 y 7 de Azzalini y Capitanio (1999)

$$\sqrt{n}\bar{x} \sim \mathcal{SN}(\lambda) \qquad \text{y} \qquad (n-1)s^2 \sim \chi^2_{n-1}$$

respectivamente. Después de una transformación ortogonal $\bar{x}$ y s^2 son independientes. $\square$

Antes de continuar, demos la definición de la distribución skew–t estándar univariada.

Definición 3.27. Una variable aleatoria u tiene una distribución skew–t estándar con parámetros de forma $\lambda \in \mathbb{R}$ y grados de libertad $\nu \in \mathbb{R}_+$ y representaremos por $u \sim \mathcal{ST}(\lambda, \nu)$, si su fdp está dada por

$$t(u; \lambda, \nu) = 2t(u; \nu) T\left(\lambda u \sqrt{\frac{\nu+1}{\nu+u^2}}; \nu+1\right), \tag{3.3.1}$$

donde $t(\cdot; \varrho)$ y $T(\cdot; \varrho)$ denotan a la fdp y a la fda de la $\mathcal{T}(0, 1, \varrho)$, respectivamente.

Tratamientos completos de la distribución skew–t aparecen en Kim (2002), Azzalini y Capitanio (2003), Gupta (2003) y Gupta *et al.* (2013).

Proposición 3.28. *Sea* $\mathbf{x} = (x_1, x_2, \ldots, x_n)'$ *una muestra aleatoria simple procedente de la población* $\mathcal{SN}(\lambda)$*. Si* $\mathbf{x} = (x_1, x_2, \ldots, x_n)' \sim \mathcal{SN}_n(\mathbf{0}, \mathbf{I}_n, \boldsymbol{\lambda})$*, entonces*

$$\frac{\sqrt{n}\bar{x}}{s} \sim \mathcal{ST}(\lambda, n-1). \tag{3.3.2}$$

Demostración. La fdp conjunta de $(u, v) = (\sqrt{n}\bar{x}/s, (n-1)s^2)$ tiene la forma

$$f_2(u, v) = \frac{2}{(2\pi\nu)^{1/2} 2^{\nu/2}\Gamma(\nu/2)} v^{(\nu-1)/2} \exp\left[-\frac{1}{2}v\left(1+\frac{u^2}{\nu}\right)\right]\Phi\left(\lambda u\sqrt{\frac{v}{\nu}}\right),$$

donde $\nu = n-1$. Por el lema 1 de Azzalini y Capitanio (2003) o Gupta (2003) o Gupta *et al.* (2013, pp. 299)

$$\frac{\sqrt{n}\bar{x}}{s} \sim \mathcal{ST}(\lambda, n-1).$$

$\square$

Chen *et al.* (2004) introducen la siguiente proposición.

Proposición 3.29. *Sea $\mathbf{x} = (x_1, x_2, \ldots, x_n)'$ una muestra aleatoria simple procedente de la población $\mathcal{SN}(\lambda)$, entonces, la fdp de $\bar{x}$ está dada por*

$$f(\bar{x}) = 2^n\sqrt{n}\phi(\sqrt{n}\bar{x})\Phi_n\left(\lambda\bar{x}\left[\frac{1}{1+\lambda^2}\mathbf{I}_n + \frac{1}{n(1+\lambda^2)}\mathbf{1}\mathbf{1}'\right]\mathbf{1}\right), \tag{3.3.3}$$

donde $\phi(\cdot)$ y $\Phi_n(\cdot; \mathbf{0}, \mathbf{\Sigma})$ denotan a la fdp y a la fda de la $\mathcal{N}(0,1)$ y la $\mathcal{N}_n(\mathbf{0}, \mathbf{\Sigma})$, respectivamente.

Demostración.

$$\begin{aligned} f_n(\mathbf{x}) &= 2^n\prod_{i=1}^{n}\phi(x_i)\Phi(\lambda x_i) \\ &= 2^n\phi_n(\mathbf{x}; \mathbf{0}, \mathbf{I}_n)\prod_{i=1}^{n}\int_{-\infty}^{\lambda x_i}\phi(s_i)ds_i. \end{aligned}$$

Con las consideraciones ubicadas en la demostración de la proposición 3.26, realicemos particiones: (c) en $\mathbf{a}_i$, con $\mathbf{b}_i = (a_{i2}, a_{i3}, \ldots, a_{in})'$ se tiene que $\mathbf{a}_i = (a_{i1}, \mathbf{b}_i')'$; y (d) en $\mathbf{A}$, con $\mathbf{c} = (a_{11}, a_{21}, \ldots, a_{n1})'$ y $\mathbf{B} = (\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n)'$ se tiene que $\mathbf{A} = (\mathbf{c}, \mathbf{B})$. Observemos que, $\mathbf{v}'\mathbf{v} = \mathbf{x}'\mathbf{x} = v_1^2 + \mathbf{w}'\mathbf{w}$, $\mathbf{a}_i'\mathbf{v} = a_{i1}v_1 + \mathbf{b}_i'\mathbf{w}$ y $\mathbf{B}'\mathbf{B} = \mathbf{I}_{n-1}$. Entonces

$$\begin{aligned} f_n(\mathbf{v}) &= 2^n\phi_n(\mathbf{v}; \mathbf{0}, \mathbf{I}_n)\prod_{i=1}^{n}\int_{-\infty}^{\lambda\mathbf{a}_i'\mathbf{v}}\phi(t_i)dt_i \\ &= 2^n\phi(v_1)\left(\frac{1}{\sqrt{2\pi}}\right)^{n-1}\exp\left(-\frac{1}{2}\mathbf{w}'\mathbf{w}\right)\prod_{i=1}^{n}\int_{-\infty}^{\lambda(a_{i1}v_1+\mathbf{b}_i'\mathbf{w})}\phi(t_i)dt_i \\ &= 2^n\phi(v_1)\left(\frac{1}{\sqrt{2\pi}}\right)^{n-1}\exp\left(-\frac{1}{2}\mathbf{w}'\mathbf{w}\right) \\ &\quad\times\int_{-\infty}^{\lambda(a_{i1}v_1+\mathbf{b}_1'\mathbf{w})}\int_{-\infty}^{\lambda(a_{i1}v_1+\mathbf{b}_2'\mathbf{w})}\cdots\int_{-\infty}^{\lambda(a_{i1}v_1+\mathbf{b}_n'\mathbf{w})}\left(\frac{1}{\sqrt{2\pi}}\right)^{n}\exp\left(-\frac{1}{2}\mathbf{t}'\mathbf{t}\right)d\mathbf{t}. \end{aligned}$$

Para comprimir el producto de probabilidades consideremos el recinto

$$R^* = (-\infty, \lambda a_{11} v_1) \times (-\infty, \lambda a_{21} v_1) \times \cdots \times (-\infty, \lambda a_{n1} v_1).$$

Entonces

$$\begin{aligned}
f_n(\mathbf{v}) &= 2^n \phi(v_1) \left(\frac{1}{\sqrt{2\pi}}\right)^{-1} \int_{R^*} \exp\left\{-\frac{1}{2}[\mathbf{w}'\mathbf{w} + (\mathbf{t} - \lambda \mathbf{B}\mathbf{w})'(\mathbf{t} - \lambda \mathbf{B}\mathbf{w})]\right\} d\mathbf{t} \\
&= 2^n \phi(v_1) \left(\frac{1}{\sqrt{2\pi}}\right)^{-1} \int_{R^*} \exp\left\{-\frac{1}{2}[\mathbf{t}'\mathbf{t} + (2\mathbf{t}'\lambda \mathbf{B}\mathbf{w} + \lambda^2 \mathbf{w}'\mathbf{B}'\mathbf{B}\mathbf{w} + \mathbf{w}'\mathbf{w})]\right\} d\mathbf{t} \\
&= 2^n \phi(v_1) \left(\frac{1}{\sqrt{2\pi}}\right)^{-1} \int_{R^*} \exp\left\{-\frac{1}{2}[\mathbf{t}'\boldsymbol{\Omega}^{-1}\mathbf{t} + (\mathbf{w} - \boldsymbol{\vartheta})'\boldsymbol{\Upsilon}^{-1}(\mathbf{w} - \boldsymbol{\vartheta})]\right\} d\mathbf{t} \\
&= 2^n \phi(v_1) \int_{R^*} |\boldsymbol{\Omega}|^{1/2} \phi_n(\mathbf{t}; \mathbf{0}, \boldsymbol{\Omega}) |\boldsymbol{\Upsilon}|^{1/2} \phi_{n-1}(\mathbf{w}; \boldsymbol{\vartheta}, \boldsymbol{\Upsilon}) d\mathbf{t} \\
&= 2^n \phi(v_1) \int_{R^*} \phi_n(\mathbf{t}; \mathbf{0}, \boldsymbol{\Omega}) \phi_{n-1}(\mathbf{w}; \boldsymbol{\vartheta}, \boldsymbol{\Upsilon}) d\mathbf{t},
\end{aligned}$$

donde $\boldsymbol{\Omega}^{-1} = \mathbf{I}_n - \lambda^2 \mathbf{B}(\mathbf{I}_{n-1} + \lambda^2 \mathbf{B}'\mathbf{B})^{-1}\mathbf{B}' = \mathbf{I}_n - \frac{\lambda^2}{1+\lambda^2}\mathbf{B}\mathbf{B}'$, $\boldsymbol{\vartheta} = (\mathbf{I}_{n-1} + \lambda^2 \mathbf{B}'\mathbf{B})\lambda \mathbf{B}'\mathbf{t} = \lambda(1+\lambda^2)\mathbf{B}'\mathbf{t}$, $\boldsymbol{\Upsilon}^{-1} = \mathbf{I}_{n-1} + \lambda^2 \mathbf{B}'\mathbf{B} = (1+\lambda^2)\mathbf{I}_{n-1}$, $|\boldsymbol{\Omega}| = (1+\lambda^2)^{n-1}$ y $|\boldsymbol{\Upsilon}| = (1+\lambda^2)^{-(n-1)}$. La variable aleatoria de interés es y_1, su fdp tiene la forma

$$\begin{aligned}
f_1(v_1) &= \int_{\mathbb{R}^{n-1}} \left[2^n \phi(v_1) \int_{R^*} \phi_n(\mathbf{t}; \mathbf{0}, \boldsymbol{\Omega}) \phi_{n-1}(\mathbf{w}; \boldsymbol{\vartheta}, \boldsymbol{\Upsilon}) d\mathbf{t}\right] d\mathbf{w} \\
&= 2^n \phi(v_1) \int_{R^*} \phi_n(\mathbf{t}; \mathbf{0}, \boldsymbol{\Omega}) \left[\int_{\mathbb{R}^{n-1}} \phi_{n-1}(\mathbf{w}; \boldsymbol{\vartheta}, \boldsymbol{\Upsilon}) d\mathbf{w}\right] d\mathbf{t} \\
&= 2^n \phi(v_1) \int_{R^*} \phi_n(\mathbf{t}; \mathbf{0}, \boldsymbol{\Omega}) d\mathbf{t} \\
&= 2^n \phi(v_1) \Phi_n(\boldsymbol{\Omega}^{1/2} \boldsymbol{\Lambda}),
\end{aligned}$$

donde $\boldsymbol{\Lambda} = v_1 \frac{\lambda}{\sqrt{n}} \mathbf{1}$. Realizando la transformación $\bar{x} = v_1/\sqrt{n}$ se prueba la proposición. $\square$

3.4. Análisis Bayesiano

En el modelo estadístico Bayesiano, la fmp o fdp del vector de parámetros es llamada previa, la función de verosimilitud va coherente con el principio de verosimilitud (Berger y Wolpert, 1988; Pericchi, 1998), y la esencia es la fmp o fdp a posteriori, por el teorema de Bayes, es proporcional al producto de la verosimilitud y la previa. Tratamientos completos de la inferencia Bayesiana aparecen en Bernardo y Smith (2000), Press (2003), Robert (2007), Carlin y Louis (2009), Savchuk y Tsokos (2011), Wakefield (2013), Held y Bové (2014), entre otros. Un artículo dedicado a la historia de la estadística Bayesiana aparece en Leonard (2014).

3.4.1. Especificación de las distribuciones previas

Bayes y Branco (2007), siguen el principio de Bayes (1763) y Laplace (1812) y establecen que, la previa no informativa (i.e. que no contiene información sobre el parámetro), propia y natural para el parámetro δ del modelo $\mathcal{SN}(\lambda)$ es la fdp uniforme.

Proposición 3.30. *Si* $\delta \sim \mathcal{U}(-1,1)$, *entonces*

$$\lambda \sim \mathcal{T}(0, 0.5, 2). \tag{3.4.1}$$

donde $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}} \in [-1,1]$ *y* $\lambda = \frac{\delta}{\sqrt{1-\delta^2}} \in \mathbb{R}$.

Demostración. Es inmediata utilizando la transformación. □

En la proposición anterior, la previa del parámetro λ del modelo $\mathcal{SN}(\lambda)$ es la fdp t de Student (Bayes y Branco, 2007) y es informativa (Pericchi, 1998). La alternativa indudable es la previa de Jeffreys (1946) ($\pi(\lambda) \propto \sqrt{I(\lambda)}$, $I(\cdot)$ es la información esperada de Fisher (1930)) desarrollada por Liseo (2004) (Liseo y Loperfido, 2006; Bayes y Branco 2007; Cabras y Castellanos, 2009; Branco *et al.*, 2012; Rubio y Liseo, 2014).

Proposición 3.31. *La previa de Jeffreys para el parámetro* λ *del modelo* $\mathcal{SN}(\lambda)$, *está dada por*

$$\pi(\lambda) \propto \sqrt{\int_0^\infty x^2 \phi(x) \frac{\phi^2(\lambda x)}{\Phi(\lambda x)\left[1-\Phi(\lambda x)\right]} dx}, \tag{3.4.2}$$

donde $\phi(\cdot)$ *y* $\Phi(\cdot)$ *denotan a la fdp y a la fda de la* $\mathcal{N}(0,1)$, *respectivamente.*

Demostración.

$$\begin{aligned}
I(\lambda) &\propto \int_{-\infty}^{\infty} \left\{ \frac{\partial \log \phi(x;\lambda)}{\partial \lambda} \right\}^2 \phi(x)\Phi(\lambda x) dx \\
&= \int_0^\infty x^2 \phi(x) \frac{\phi^2(\lambda x)}{\Phi(\lambda x)} dx + \int_{-\infty}^0 x^2 \phi(x) \frac{\phi^2(\lambda x)}{\Phi(\lambda x)} dx \\
&= \int_0^\infty x^2 \phi(x) \frac{\phi^2(\lambda x)}{\Phi(\lambda x)} dx + \int_0^\infty x^2 \phi(x) \frac{\phi^2(\lambda x)}{\left[1-\Phi(\lambda x)\right]} dx \\
&= \int_0^\infty x^2 \phi(x) \frac{\phi^2(\lambda x)}{\Phi(\lambda x)\left[1-\Phi(\lambda x)\right]} dx.
\end{aligned}$$

□

El modelo $\mathcal{SN}(\lambda)$ satisface la condición de Rubio y Liseo (2014). En consecuencia, la previa de Jeffreys: (a) es simétrica respecto al origen; (b) decreciente en $|\lambda|$; (c) integrable; y (d) tiene colas de orden $\mathcal{C}(|\lambda|^{-3/2})$ (Liseo, 2004; Liseo y Loperfido, 2006; Bayes y Branco, 2007; Cabras y Castellanos, 2009; Branco *et al.*, 2012; Rubio y Liseo, 2014). Aun así, es difícil trabajar con la previa de Jeffreys.

Bayes y Branco (2007) utilizan la aproximación de Chaibub–Neto y Branco (2003)

$$\frac{\phi(x)}{\sqrt{\Phi(x)\,[1-\Phi(x)]}} \approx 2\phi(2x/\pi) \tag{3.4.3}$$

para obtener una expresión amable con la previa de Jeffreys (Branco *et al.*, 2012; Azzalini 2014; Rubio y Liseo, 2014).

Proposición 3.32. *La previa de Jeffreys para el parámetro λ del modelo $\mathcal{SN}(\lambda)$, está dada aproximadamente por*

$$\pi(\lambda) \approx t(\lambda; 0, 0.25\pi^2, 0.5), \tag{3.4.4}$$

donde $t(\cdot; 0, \sigma^2, \varrho)$ denota a la fdp de la $\mathcal{T}(0, \sigma^2, \varrho)$.

Demostración.

$$\begin{aligned} I(\lambda) &\propto \int_0^\infty x^2\phi(x)\frac{\phi^2(\lambda x)}{\Phi(\lambda x)\,[1-\Phi(\lambda x)]}dx \\ &\approx 4\int_0^\infty x^2\phi(x)\phi^2(2\lambda x/\pi)dx \\ &\propto \int_0^\infty x^2\phi(x; 0, [1+2\lambda^2/(\pi^2/4)]^{-1})dx \\ &= \left(1+\frac{2\lambda^2}{\pi^2/4}\right)^{-3/2}. \end{aligned}$$

□

Convenientemente, las previas (3.4.1) y (3.4.4) se adaptan a la distribución t de Student

$$\lambda \sim \mathcal{T}(0, \sigma_\lambda^2, \nu_\lambda). \tag{3.4.5}$$

Pero, en la previa (3.4.4) se contempla un problema, el grado de libertad $\nu_\lambda = 0.5$. Para resolver, supongamos un vector de parámetros $(\lambda, \rho)'$, donde el parámetro de interés es λ y el parámetro de molestia es ρ (Basu, 1977; Pericchi, 1998). De manera natural y coherente se introduce la siguiente proposición.

Proposición 3.33. *Si $\lambda|\rho \sim \mathcal{N}(0, \sigma_\lambda^2/\rho)$ y $\rho \sim \mathcal{G}(0.5\nu_\lambda, 0.5\nu_\lambda)$, entonces*

$$\lambda \sim \mathcal{T}(0, \sigma_\lambda^2, \nu_\lambda). \tag{3.4.6}$$

Demostración.

$$\pi(\lambda) = \int_0^\infty \pi(\lambda|\rho)\pi(\rho)d\rho = t(\lambda; 0, \sigma_\lambda^2, \nu_\lambda).$$

□

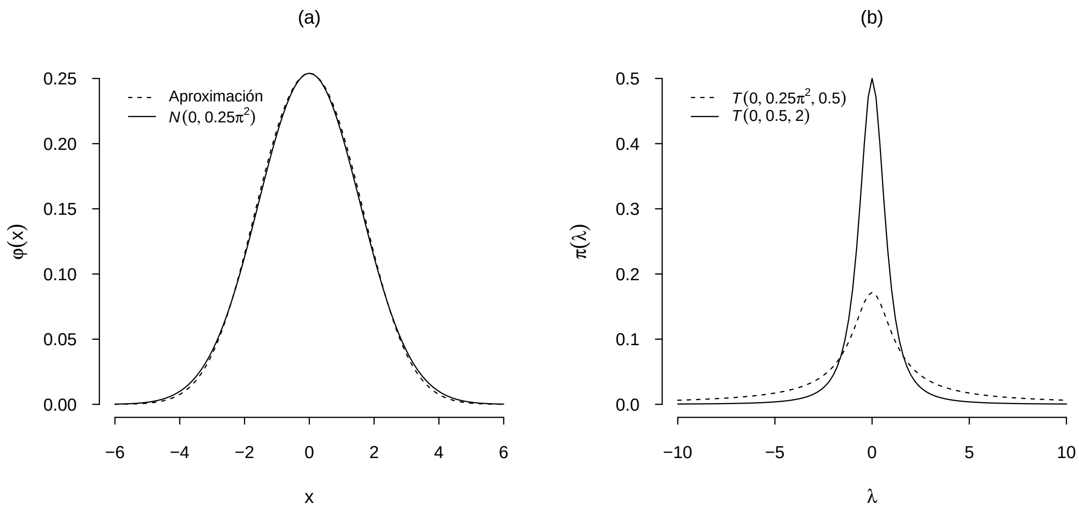

Figura 3.4.1. Plots de la (a) aproximación de Chaibub–Neto y Branco (2003), fdp $\phi(x;0,0.25\pi^2)$ y (b) fdp $\pi(\lambda)$.

Liseo (2004), Liseo y Loperfido (2006), Bayes y Branco (2007), Branco *et al.* (2012) y Rubio y Liseo (2014) introducen la siguiente proposición, consideran independencia entre las previas, i.e.

$$\pi(\xi,\omega^2,\lambda)=\pi(\xi)\pi(\omega^2)\pi(\lambda).$$

Proposición 3.34. *La previa independiente de Jeffreys para el vector de parámetros* (ξ,ω^2,λ) *del modelo* $\mathcal{SN}(\xi,\omega^2,\lambda)$*, está dada por*

$$\pi(\xi,\omega^2,\lambda)\propto\frac{1}{\omega^2}\pi(\lambda) \tag{3.4.7}$$

donde $\pi(\lambda)$ *es la previa específica en (3.4.4).*

Demostración. Las entradas diagonales de la matriz de información de Fisher son

$$\begin{aligned}
I_{\xi,\xi} &\propto \frac{1}{\omega^2}\int_{-\infty}^{\infty}\left\{\frac{1}{\phi(s)}\left[\frac{\partial\phi(s)}{\partial s}\right]+\lambda\frac{\phi(\lambda s)}{\Phi(\lambda s)}\right\}^2\phi(s)\Phi(\lambda s)ds\\
I_{\omega,\omega} &\propto \frac{1}{\omega^2}\int_{-\infty}^{\infty}\left\{1+s\frac{1}{\phi(s)}\left[\frac{\partial\phi(s)}{\partial s}\right]+\lambda t\frac{\phi(\lambda s)}{\Phi(\lambda s)}\right\}^2\phi(s)\Phi(\lambda s)ds\\
I_{\lambda,\lambda} &\propto \int_{0}^{\infty}s^2\phi(s)\frac{\phi^2(\lambda s)}{\Phi(\lambda s)\left[1-\Phi(\lambda s)\right]}ds.
\end{aligned}$$

Nótese que, $I_{\xi,\xi}$ no depende de ξ; $I_{\omega,\omega}$ depende de ω a través de ω^{-2}; e $I_{\lambda,\lambda}$ coincide con $I(\lambda)$ utilizado en la demostración de (3.4.2). $\square$

3.4.2. Especificación jerárquica

Consideremos una muestra aleatoria simple $\mathbf{y} = (y_1, y_2, \ldots, y_n)'$ procedente de la población $\mathcal{SN}(\xi, \omega^2, \lambda)$. Para obtener la distribución a posteriori, también consideremos la proposición 3.3,

$$y_i = \xi + \omega(\sqrt{1-\delta^2}\epsilon_i + \delta\tau_i) \sim \mathcal{SN}(\xi, \omega^2, \lambda). \tag{3.4.8}$$

Un fragmento de la representación jerárquica va de acuerdo proposición 3.4,

$$\begin{aligned} y_i|\tau_i, \xi, \omega^2, \lambda &\sim \mathcal{N}(\xi + \omega\delta\tau_i, \omega^2(1-\delta^2)) \\ \tau_i &\sim \mathcal{HN}(0,1). \end{aligned} \tag{3.4.9}$$

Bayes y Branco (2007) facilitan el trabajo computacional realizando la reparametrización $\alpha = \omega\delta$ y $\beta = \omega\sqrt{1-\delta^2}$. En consecuencia

$$\begin{aligned} y_i|\tau_i, \xi, \alpha, \beta &\sim \mathcal{N}(\xi + \alpha\tau_i, \beta^2) \\ \tau_i &\sim \mathcal{HN}(0,1). \end{aligned} \tag{3.4.10}$$

De la proposición 3.34 obtenemos la nueva previa

$$\pi(\xi, \alpha, \beta, \rho) \propto \frac{\rho^{(\nu_\lambda+1)/2}}{\beta^2} \exp\left\{-\frac{1}{2}\rho\left[\nu_\lambda + \left(\frac{\alpha}{\beta\sigma_\lambda}\right)^2\right]\right\} \tag{3.4.11}$$

(Bayes y Branco, 2007). Las distribuciones condicionales a posteriori completas, están dadas por

$$\begin{aligned} \tau_i|\mathbf{y}, \xi, \alpha, \beta, \rho &\sim \mathcal{N}\left(\frac{(y_i-\xi)\alpha}{\alpha^2+\beta^2}, \frac{\beta^2}{\alpha^2+\beta^2}\right) \\ \xi|\mathbf{y}, \boldsymbol{\tau}, \alpha, \beta, \rho &\sim \mathcal{N}\left(\bar{y} - \alpha\bar{\tau}, \frac{\beta^2}{n}\right) \\ \alpha|\mathbf{y}, \boldsymbol{\tau}, \xi, \beta, \rho &\sim \mathcal{N}\left(\frac{(\mathbf{y}-\xi\mathbf{1})'\boldsymbol{\tau}}{\rho\sigma_\lambda^{-2} + \boldsymbol{\tau}'\boldsymbol{\tau}}, \frac{\beta^2}{\rho\sigma_\lambda^{-2} + \boldsymbol{\tau}'\boldsymbol{\tau}}\right) \\ \frac{1}{\beta^2}|\mathbf{y}, \boldsymbol{\tau}, \xi, \alpha, \rho &\sim \mathcal{G}\left(\frac{n+1}{2}, \frac{1}{2}\left[\rho\left(\frac{\alpha}{\sigma_\lambda}\right)^2 + (\mathbf{y}-\xi\mathbf{1}-\alpha\boldsymbol{\tau})'(\mathbf{y}-\xi\mathbf{1}-\alpha\boldsymbol{\tau})\right]\right) \\ \rho|\mathbf{y}, \boldsymbol{\tau}, \xi, \alpha, \beta &\sim \mathcal{G}\left(\frac{\nu_\lambda+1}{2}, \frac{1}{2}\left[\nu_\lambda + \left(\frac{\alpha}{\beta\sigma_\lambda}\right)^2\right]\right) \end{aligned} \tag{3.4.12}$$

(Bayes y Branco, 2007).

Capítulo 4

La Regresión Binaria Bayesiana Skew Probit

Chen (2004) realizo un estudio de simulación para investigar la importancia de la elección del link en la predicción de una respuesta binaria (0/1). Consideró dos simulaciones: (s1) los datos son generados de acuerdo al modelo probit; y (s2) los datos son generados de acuerdo al modelo cloglog. En ambas situaciones ajusto a los modelos probit, logit y cloglog. Observo que: (o1) cuando el link verdadero es el probit, casi no hay diferencias entre el modelo probit y el logit; mientras el modelo cloglog es inadecuado; y (o2) cuando el link verdadero es el cloglog, los modelos con link simétrico fueron claramente inadecuados. Concluyo que, la elección del link es muy importante, y si no es muy bien especificado puede proveer predicciones pobres (Farias y Branco, 2011, 2012). Previendo esto, el capítulo aborda a los modelos skew probit CDS y BBB.

4.1. La Regresión Skew de Chen, Dey y Shao

Chen (2004) y Farias y Branco (2011, 2012) introducen la siguiente definición, central para nuestro desarrollo.

Definición 4.1. Denotando por $F(\cdot)$ a la fda de una fdp simétrica alrededor del origen con soporte en la recta real y por $G(\cdot)$ a la fda de una fdp asimétrica en $[0, \infty\rangle$. Sea

$$\mathcal{F}_{\mathcal{A}} = \left\{ F_\lambda(\eta) = \int_{[0,\infty\rangle} F(\eta - \lambda\tau) dG(\tau), \quad \lambda \in \mathbb{R} \right\} \tag{4.1.1}$$

una clase de fda's.

Lo más importante es que, el modelo definido en (4.1.1) tiene propiedades atractivas convenientes: (a) cuando $\lambda = 0$ o $G(\cdot)$ es una distribución degenerada, se reduce a un modelo con link simétrico; (b) la asimetría puede ser determinada por λ y $G(\cdot)$; y (c) la cola pesada o ligera puede ser obtenida según $F(\cdot)$ (Chen *et al.*, 1999; Chen, 2004; Bermúdez *et al.*, 2008; Farias y Branco, 2011, 2012).

Proposición 4.2. *Si $F_\lambda(\cdot)$ y $F_{-\lambda}(\cdot)$ son las fda's pertenecientes a la clase $\mathcal{F}_\mathcal{A}$, entonces*

$$F_\lambda(\eta) = 1 - F_{-\lambda}(-\eta). \tag{4.1.2}$$

Demostración. Es inmediata utilizando (4.1.1). □

El modelo de regresión binaria de Chen *et al.* (1999), es caracterizado por

$$p_i = \Pr(y_i = 1) = F_\lambda(\mathbf{x}_i'\boldsymbol{\beta}) = \int_0^\infty F(\mathbf{x}_i'\boldsymbol{\beta} - \lambda\tau_i)dG(\tau_i), \tag{4.1.3}$$

donde $F(\cdot)$ y $G(\cdot)$ son denotados en (4.1.1).

La función de verosimilitud para el modelo de regresión binaria de Chen *et al.* (1999), está dada por

$$\mathcal{L}(\boldsymbol{\beta}, \lambda | n, \mathbf{y}, \mathbf{X}) = \prod_{i=1}^n [F_\lambda(\mathbf{x}_i'\boldsymbol{\beta})]^{y_i}[F_{-\lambda}(-\mathbf{x}_i'\boldsymbol{\beta})]^{1-y_i} \tag{4.1.4}$$

(Bazán *et al.*, 2006, 2014; Bazán *et al.*, 2010).

Proposición 4.3. *El modelo de regresión binaria de Chen et al. (1999), utilizando variables latentes, está dado por*

$$\begin{aligned} y_i &= \begin{cases} 1 & \text{si} \quad z_i > 0 \\ 0 & \text{si} \quad z_i \leq 0 \end{cases} \\ z_i &= \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i \\ \varepsilon_i &= \epsilon_i - \lambda\tau_i \\ \epsilon_i &\sim F(\epsilon_i) \\ \tau_i &\sim G(\tau_i), \end{aligned} \tag{4.1.5}$$

donde $F(\cdot)$ y $G(\cdot)$ son denotados en (4.1.1) y ϵ_i es independiente de τ_i.

Demostración.

$$\begin{aligned} \Pr(y_i = 1) &= \Pr(z_i > 0) \\ &= \Pr(\varepsilon_i > -\mathbf{x}_i'\boldsymbol{\beta}) \\ &= 1 - F_{-\lambda}(-\mathbf{x}_i'\boldsymbol{\beta}) \\ &= F_\lambda(\mathbf{x}_i'\boldsymbol{\beta}). \\ \Pr(y_i = 0) &= 1 - F_\lambda(\boldsymbol{x}_i'\boldsymbol{\beta}). \\ y_i &\sim \mathcal{B}er(F_\lambda(\mathbf{x}_i'\boldsymbol{\beta})). \end{aligned}$$

□

Las funciones de verosimilitud de data completa y verosimilitud de data completa marginal para el modelo de regresión binaria de Chen *et al.* (1999), están dadas por

$$\mathcal{L}(\boldsymbol{\beta}, \lambda | n, \mathbf{y}, \mathbf{X}, \boldsymbol{\tau}) = \prod_{i=1}^{n} [F(\mathbf{x}_i'\boldsymbol{\beta} - \lambda\tau_i)]^{y_i} [1 - F(\mathbf{x}_i'\boldsymbol{\beta} - \lambda\tau_i)]^{1-y_i} g(\tau_i) \tag{4.1.6}$$

y

$$\begin{aligned} \mathcal{L}(\boldsymbol{\beta}, \lambda | n, \mathbf{y}, \mathbf{X}) &= \int_{\mathbb{R}_+^n} \mathcal{L}(\boldsymbol{\beta}, \lambda | n, \mathbf{y}, \mathbf{X}, \boldsymbol{\tau}) d\boldsymbol{\tau} \\ &= \prod_{i=1}^{n} \int_0^{\infty} [F(\mathbf{x}_i'\boldsymbol{\beta} - \lambda\tau_i)]^{y_i} [1 - F(\mathbf{x}_i'\boldsymbol{\beta} - \lambda\tau_i)]^{1-y_i} dG(\tau_i), \end{aligned} \tag{4.1.7}$$

respectivamente (Bazán *et al.*, 2010).

4.2. La Regresión Skew Probit CDS

En la definición 4.1 y la proposición 4.3 se presenta una mezcla de fda's (Bazán *et al.*, 2010). Cuando $F(\cdot) = \Phi(\cdot)$ es la fda de la $\mathcal{N}(0,1)$ y $G(\cdot) = \Psi(\cdot)$ es la fda de la $\mathcal{HN}(0,1)$ resulta el *modelo de regresión binaria skew probit CDS*, CDS en honor a M.–H. Chen, D. K. Dey y Q.–M. Shao (Chen *et al.*, 1999; Bazán *et al.*, 2006; Bazán *et al.*, 2010; Farias y Branco, 2011, 2012). También y sin olvidarnos, el modelo de regresión binaria skew probit CDS utilizando variables latentes, tiene propiedades atractivas convenientes: (a) cuando $\lambda = 0$, se reduce al modelo de regresión binaria probit utilizando variables latentes, (b) de acuerdo a la proposición 3.18, la variable latente

$$z_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i \sim \mathcal{SN}(\mathbf{x}_i'\boldsymbol{\beta}, 1 + \lambda^2, -\lambda), \tag{4.2.1}$$

y (c) de acuerdo a la proposición 3.19, la condicional

$$z_i | \tau_i \sim \mathcal{N}(\mathbf{x}_i'\boldsymbol{\beta} - \lambda\tau_i, 1) \tag{4.2.2}$$

(Chen *et al.*, 1999; Bazán *et al.*, 2010). La probabilidad para $y_i = 1$, está dada por

$$\begin{aligned} \Pr(y_i = 1) &= \Pr(z_i > 0) \\ &= \Pr(\varepsilon_i > -\mathbf{x}_i'\boldsymbol{\beta}) \\ &= 1 - \Phi(-\mathbf{x}_i'\boldsymbol{\beta}; 0, 1 + \lambda^2, -\lambda) \\ &= \Phi(\mathbf{x}_i'\boldsymbol{\beta}; 0, 1 + \lambda^2, \lambda), \end{aligned} \tag{4.2.3}$$

donde $\Phi(\cdot; \mu, \sigma^2, \zeta)$ denota a la fda de la $\mathcal{SN}(\mu, \sigma^2, \zeta)$ (Chen *et al.*, 1999).

4.3. La Regresión Skew Probit

El modelo de regresión binaria skew probit, es caracterizado por

$$p_i = \Pr(y_i = 1) = \Phi(\mathbf{x}_i'\boldsymbol{\beta}; \xi, \omega^2, \lambda), \tag{4.3.1}$$

donde $\Phi(\cdot; \mu, \sigma^2, \zeta)$ denota a la fda de la $\mathcal{SN}(\mu, \sigma^2, \zeta)$ (Bazán *et al.*, 2010). Son tres modelos especiales que se concentran: cuando $(\xi, \omega^2, \lambda) = (0, 1 + \lambda^2, \lambda)$ resulta el skew probit CDS (Chen *et al.*, 1999; Bazán *et al.*, 2006; Bazán *et al.*, 2010; Farias y Branco, 2011, 2012), cuando $(\xi, \omega^2, \lambda) = (0, 1, \lambda)$ resulta el skew probit BBB (Bazán *et al.*, 2006, 2014; Bazán *et al.*, 2010), y cuando $(\xi, \omega^2, \lambda) = (0, 1, 0)$ resulta el probit (Bliss, 1935). De acuerdo a las proposiciones 3.24 y 3.25

$$\begin{aligned}\Phi(\mathbf{x}_i'\boldsymbol{\beta}; \xi, \omega^2, \lambda) &= \int_0^{\infty} \Phi\left(\frac{\mathbf{x}_i'\boldsymbol{\beta} - \xi}{\omega}\sqrt{1+\lambda^2} - \lambda\tau_i\right)\psi(\tau_i)d\tau_i \\ &= 2\Phi_2\left(\begin{pmatrix}\mathbf{x}_i'\boldsymbol{\beta} \\ 0\end{pmatrix}; \begin{pmatrix}\xi \\ 0\end{pmatrix}, \begin{pmatrix}\omega^2 & -\delta\omega \\ -\delta\omega & 1\end{pmatrix}\right)\end{aligned} \tag{4.3.2}$$

(Bazán *et al.*, 2010; Bazán *et al.*, 2014). La función de verosimilitud, está dada por

$$\mathcal{L}(\boldsymbol{\beta}, \xi, \omega^2, \lambda | n, \mathbf{y}, \mathbf{X}) = \prod_{i=1}^{n} [\Phi(\mathbf{x}_i'\boldsymbol{\beta}; \xi, \omega^2, \lambda)]^{y_i} [1 - \Phi(\mathbf{x}_i'\boldsymbol{\beta}; \xi, \omega^2, \lambda)]^{1-y_i} \tag{4.3.3}$$

(Bazán *et al.*, 2010).

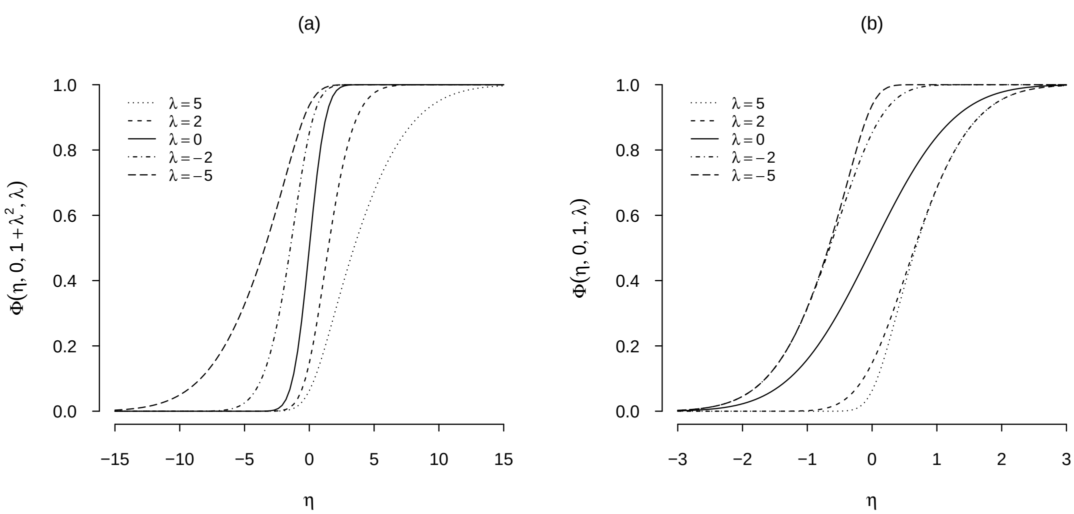

Figura 4.3.1. Curvas de probabilidad del modelo skew probit (a) CDS y (b) BBB

Proposición 4.4. *El modelo de regresión binaria skew probit utilizando variables latentes, está dado por*

$$\begin{aligned}
y_i &= \begin{cases} 1 & \text{si} \quad z_i > 0 \\ 0 & \text{si} \quad z_i \leq 0, \end{cases} \\
z_i &= \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i \\
\varepsilon_i &= -\xi - \omega(\sqrt{1-\delta^2}\epsilon_i + \delta\tau_i) \\
\epsilon_i &\sim \mathcal{N}(0,1) \\
\tau_i &\sim \mathcal{HN}(0,1),
\end{aligned} \tag{4.3.4}$$

donde ϵ_i es independiente de τ_i y $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}} \in [-1,1]$.

Demostración.

$$\begin{aligned}
\Pr(y_i = 1) &= \Pr(z_i > 0) \\
&= \Pr(\varepsilon_i > -\mathbf{x}_i'\boldsymbol{\beta}). \\
\varepsilon_i &\sim \mathcal{SN}(-\xi, \omega^2, -\lambda). \\
\Pr(y_i = 1) &= 1 - \Phi(-\mathbf{x}_i'\boldsymbol{\beta}; -\xi, \omega^2, -\lambda) \\
&= 1 - \Phi((\xi - \mathbf{x}_i'\boldsymbol{\beta})/\omega; 0, 1, -\lambda) \\
&= \Phi((\mathbf{x}_i'\boldsymbol{\beta} - \xi)/\omega; 0, 1, \lambda) \\
&= \Phi(\mathbf{x}_i'\boldsymbol{\beta}; \xi, \omega^2, \lambda). \\
\Pr(y_i = 0) &= 1 - \Phi(\mathbf{x}_i'\boldsymbol{\beta}; \xi, \omega^2, \lambda). \\
y_i &\sim \mathcal{B}er(\Phi(\mathbf{x}_i'\boldsymbol{\beta}; \xi, \omega^2, \lambda)).
\end{aligned}$$

□

Desplegando las estructuras de las proposiciones 3.18 y 3.19 en la estructura de la variable latente de la proposición anterior, obtenemos que,

$$\begin{aligned}
z_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i &\sim \mathcal{SN}(\mathbf{x}_i'\boldsymbol{\beta} - \xi, \omega^2, -\lambda) \\
\varepsilon_i &= -\xi - \omega(\sqrt{1-\delta^2}\epsilon_i + \delta\tau_i) \\
z_i|\tau_i &\sim \mathcal{N}(\mathbf{x}_i'\boldsymbol{\beta} - \xi - \omega\delta\tau_i, (1-\delta^2)\omega^2) \\
\tau_i &\sim \mathcal{HN}(0,1)
\end{aligned} \tag{4.3.5}$$

(Bazán *et al.*, 2010; Bazán *et al.*, 2014). Nótese que, la fdp de z_i tiene la forma

$$\phi(z_i; \mathbf{x}_i'\boldsymbol{\beta} - \xi, \omega^2, -\lambda) = \frac{2}{\omega}\phi\left(\frac{z_i - (\mathbf{x}_i'\boldsymbol{\beta} - \xi)}{\omega}\right)\Phi\left(\lambda\frac{(\mathbf{x}_i'\boldsymbol{\beta} - \xi) - z_i}{\omega}\right), \tag{4.3.6}$$

donde $\phi(\cdot)$ y $\Phi(\cdot)$ denotan a la fdp y a la fda de la $\mathcal{N}(0,1)$, respectivamente (Chen *et al.*, 1999). Con estos resultados, la simulación de z_i puede ser realizada en dos etapas: (e1) generando de acuerdo a τ_i; y (e2) luego generando de acuerdo a $z_i|\tau_i$ (Bazán *et al.*, 2010; Bazán *et al.*, 2014). En consecuencia, la función de verosimilitud de data completa para el modelo de regresión binaria skew probit, está dada por

$$\mathcal{L}(\boldsymbol{\beta},\xi,\omega^2,\lambda|n,\mathbf{y},\mathbf{X},\boldsymbol{\tau},\mathbf{z}) = \prod_{i=1}^{n} \phi(z_i|\tau_i;\mathbf{x}_i'\boldsymbol{\beta}-\xi-\omega\delta\tau_i,(1-\delta^2)\omega^2)\psi(\tau_i)$$
$$\times \{I_{\{1\}}(y_i)I_{\mathbb{R}_+}(z_i|\tau_i)+I_{\{0\}}(y_i)I_{\mathbb{R}_-}(z_i|\tau_i)\}, \qquad (4.3.7)$$

donde $\phi(\cdot;\mu,\sigma^2)$ y $\psi(\cdot)$ denotan a las fdp's de la $\mathcal{N}(\mu,\sigma^2)$ y la $\mathcal{HN}(0,1)$, respectivamente, $\delta=\frac{\lambda}{\sqrt{1+\lambda^2}}\in[-1,1]$ e $I_A(\cdot)$ es la función indicadora en el set A (Bazán *et al.*, 2010; Bazán *et al.*, 2014).

4.3.1. Análisis Bayesiano

4.3.1.1. Especificación de las distribuciones previas

Bazán *et al.* (2010) consideran independencia entre las previas de $(\boldsymbol{\beta},\xi,\omega^2,\lambda)$, i.e.

$$\pi(\boldsymbol{\beta},\xi,\omega^2,\lambda)=\pi_1(\boldsymbol{\beta})\pi_{21}(\xi)\pi_{22}(\omega^2)\pi_{23}(\lambda), \qquad (4.3.8)$$

y sugieren que, $\beta_j\sim\mathcal{N}(\mu_{\beta_j},\sigma^2_{\beta_j})$, $\xi\sim\mathcal{N}(\mu_\xi,\sigma^2_\xi)$, $\omega^2\sim\mathcal{I}nv-\chi^2_{\nu_\omega}(\sigma^2_\omega)$ y $\lambda\sim\mathcal{T}(0,0.5,2)$. De acuerdo a la proposición 3.30, $\delta=\frac{\lambda}{\sqrt{1+\lambda^2}}\sim\mathcal{U}(-1,1)$ (Bazán *et al.*, 2006, 2014; Bazán *et al.*, 2010).

4.3.1.2. Especificación jerárquica

La especificación de la estructura jerárquica de la función de verosimilitud de data completa para el modelo de regresión binaria skew probit, está dada por

$$\begin{aligned} z_i|\tau_i,y_i,\mathbf{x}_i,\boldsymbol{\beta},\xi,\omega^2,\delta &\sim \mathcal{N}(\mathbf{x}_i'\boldsymbol{\beta}-\xi-\omega\delta\tau_i,(1-\delta^2)\omega^2)I(y_i,z_i|\tau_i)\\ \tau_i &\sim \mathcal{HN}(0,1)\\ \beta_j &\sim \mathcal{N}(\mu_{\beta_j},\sigma^2_{\beta_j})\\ \xi &\sim \mathcal{N}(\mu_\xi,\sigma^2_\xi)\\ \omega^2 &\sim Inv-\chi^2_{\nu_\omega}(\sigma^2_\omega)\\ \delta &\sim \mathcal{U}(-1,1). \end{aligned} \qquad (4.3.9)$$

donde $I(y_i,z_i|\tau_i)=\{I_{\{1\}}(y_i)I_{\mathbb{R}_+}(z_i|\tau_i)+I_{\{0\}}(y_i)I_{\mathbb{R}_-}(z_i|\tau_i)\}$ e $I_A(\cdot)$ es la función indicadora en el set A (Bazán *et al.*, 2010; Bazán *et al.*, 2014). Para considerar los modelos skew probit

CDS (Chen *et al.*, 1999; Bazán *et al.*, 2006; Bazán *et al.*, 2010; Farias y Branco, 2011, 2012) y BBB (Bazán *et al.*, 2006, 2014; Bazán *et al.*, 2010), se omite la cuarta y quinta línea, y si además $\delta = 0$ obtenemos el modelo probit (Bliss, 1935; Bazán *et al.*, 2010; Bazán *et al.*, 2014).

Capítulo 5

Los Criterios para Comparar Modelos Bayesianos

Cuando se tiene varios modelos alternativos, surge la pregunta inmediata ¿cómo los comparo y selecciono al mejor? Lo primero que no se debe olvidar es que, no siempre existe un mejor modelo, y se puede tener un conjunto de modelos que ostenten un desempeño similar.

Los tres criterios principales de bondad de ajuste para comparar modelos Bayesianos son, el Criterio de Información de la Devianza (*Deviance Information Criterion*: DIC), el Criterio de Información de Akaike (*Akaike Information Criterion*: AIC) y el Criterio de Información Bayesiana (*Bayesian Information Criterion*: BIC) y están dados por

$$\begin{aligned} \mathrm{DIC} &= 2\mathbb{E}[\mathbb{D}(\boldsymbol{\beta},\boldsymbol{\theta})] - \mathbb{D}[\mathbb{E}(\boldsymbol{\beta}),\mathbb{E}(\boldsymbol{\theta})], \\ \mathrm{AIC} &= \mathbb{E}[\mathbb{D}(\boldsymbol{\beta},\boldsymbol{\theta})] + 2k^* \end{aligned} \tag{5.0.1}$$

y

$$\mathrm{BIC} = \mathbb{E}[\mathbb{D}(\boldsymbol{\beta},\boldsymbol{\theta})] + k^* \log n, \tag{5.0.2}$$

donde $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_k)'$ es el vector de coeficientes de regresión, $\boldsymbol{\theta} = (\xi, \omega^2, \lambda)'$ es el vector de los parámetros de posición, escala y forma, $\mathbb{D}(\boldsymbol{\beta},\boldsymbol{\theta}) = -2\ln \Pr(\mathbf{y}|\boldsymbol{\beta},\boldsymbol{\theta})$ es la devianza, $\mathbb{E}[\mathbb{D}(\boldsymbol{\beta},\boldsymbol{\theta})]$ es la esperanza de la devianza a posteriori, $\mathbb{D}[\mathbb{E}(\boldsymbol{\beta}),\mathbb{E}(\boldsymbol{\theta})]$ es la devianza de la esperanza a posteriori y k^* es el número de parámetros. La esperanza de la devianza a posteriori y la devianza de la esperanza a posteriori pueden ser estimadas utilizando los métodos de Monte Carlo vía Cadenas de Markov (*Markov Chain Monte Carlo*: MCMC) a través de

$$\mathrm{D}_{\mathrm{bar}} = \frac{1}{S}\sum_{s=1}^{S} \mathbb{D}(\boldsymbol{\beta}^s, \boldsymbol{\theta}^s) \tag{5.0.3}$$

y

$$\mathrm{D}_{\mathrm{hat}} = \mathbb{D}\left(\frac{1}{S}\sum_{s=1}^{S}\boldsymbol{\beta}^s, \frac{1}{S}\sum_{s=1}^{S}\boldsymbol{\theta}^s\right), \tag{5.0.4}$$

respectivamente, donde S es el número de simulaciones. Tratamientos completos de los criterios aparecen en Spiegelhalter *et al.* (2002), Carlin y Louis (2009), Ntzoufras (2009). Tratamientos dedicados a los métodos MCMC aparecen en Chen *et al.* (2000), Robert y Casella (2010) y Suess y Trumbo (2010), entre otros.

En general, se dirá que un modelo es mejor que otro cuando su DIC o AIC o BIC es sustancialmente menor. En el sitio web de WinBUGS los autores recomiendan que: a) diferencias de más de 10 en el DIC permiten descartar el modelo con el DIC más alto, b) diferencias entre 5 y 10 en el DIC pueden considerarse sustanciales, y c) cuando la diferencia es menor que cinco y los modelos hacen inferencias muy diferentes, reportar sólo el modelo con el menor DIC puede llevar a conclusiones erróneas.

Capítulo 6

Aplicación

¿Me siento satisfecho con la calidad del servicio de salud pública en general?, es una de las preguntas que se hacen miles de paceños y paceñas, pero ¿cómo podemos analizar los motivos de satisfacción o insatisfacción respecto a un amplio espectro de situaciones que condicionan su calidad de vida y su felicidad? Los modelos de regresión binaria son convenientes para este tipo de estudio.

6.1. Ficha Técnica

El Observatorio La Paz Cómo Vamos es una organización ciudadana, sin fines de lucro, es liderada por la Fundación para el Periodismo, en alianza con Solidar Suiza – Programa de Apoyo a la Democracia Municipal (PADEM), la Universidad Nuestra Señora de La Paz y la Cámara Nacional de Comercio. Tiene el propósito de contribuir a mejorar la calidad de vida en la sede de gobierno. Forma parte de la Red Latinoamericana por Ciudades Justas, Democráticas y Sustentables y de la Red Boliviana por Ciudades Justas, Democráticas y Sustentables.

La Paz Cómo Vamos el 2012 realizo la segunda Encuesta de Percepción Ciudadana sobre la Calidad de Vida en la ciudad de La Paz. Un objetivo específico fue, obtener la percepción y el nivel de satisfacción ciudadana sobre la calidad de vida en el municipio de La Paz a través de los tópicos: educación, entretenimiento, igualdad social, medio ambiente, respeto y apoyo a los diferentes grupos de la población, salud, seguridad ciudadana, transparencia y participación política, transporte y tránsito, y vivienda. Entrevistaron a 1000 personas, hombres y mujeres pertenecientes a los distintos niveles socioeconómicos. El rango de edad partió de los 18 años. Los entrevistados corresponden a las cinco grandes zonas de la ciudad (centro, este, norte, oeste y sur), 100 barrios dentro de 21 distritos. La encuesta se realizó entre el 10 de agosto y el 7 de septiembre de 2012. Emplearon un cuestionario estructurado, compuesto por preguntas cerradas, con una duración máxima de 35 minutos. La encuesta fue cara a cara.

6.2. La Salud en la Ciudad de La Paz

Ilustremos el análisis a partir del tópico salud, las variables son: *tiempo*, e indica la insatisfacción (1) o satisfacción (0) con el tiempo de demora para conseguir una consulta médica en el servicio público; *trato*, e indica la insatisfacción (1) o satisfacción (0) con la calidad del trato en los centros de salud y hospitales públicos; *infraequi*, e indica la insatisfacción (1) o satisfacción (0) con la calidad de la infraestructura y equipamiento de los centros de salud y hospitales públicos; *esfurespo*, e indica la insatisfacción (1) o satisfacción (0) con el esfuerzo y responsabilidad de los paceños para cuidar la salud de la familia; y *servicio*, e indica la insatisfacción (1) o satisfacción (0) con la calidad del servicio de salud pública en general. Las variables explicativas serán *tiempo*, *trato*, *infraequi* y *esfurespo*, y la variable dependiente será *servicio*. Debido a la no respuesta en algunas preguntas, la data se reduce a 955.

Cuadro 6.1. Sumario descriptivo de variables

Variable	Insatisfecho	Satisfecho	Media	DE
tiempo	783 (82.0 %)	172 (18.0 %)	0.820	0.384
trato	781 (81.8 %)	174 (18.2 %)	0.818	0.386
infraequi	782 (81.9 %)	173 (18.1 %)	0.819	0.385
esfurespo	730 (76.4 %)	225 (23.6 %)	0.764	0.424
servicio	776 (81.3 %)	179 (18.7 %)	0.813	0.390

DE: Desviación Estándar

Comparemos a siete modelos de regresión binaria, al probit ($\mathcal{M}_1$), al logit ($\mathcal{M}_2$), al cloglog ($\mathcal{M}_3$), al scobit ($\mathcal{M}_4$), al power logit ($\mathcal{M}_5$), al skew probit CDS ($\mathcal{M}_6$) y al skew probit BBB ($\mathcal{M}_7$). Las fdp's a posteriori son difíciles de obtener, por lo tanto las buenas aproximaciones son dadas bajo los métodos MCMC. Las sintaxis de los modelos se encuentran en el apéndice C y pueden ser implementadas en los programas RStudio o R o OpenBUGS o WinBUGS. Tratamientos completos de cómo elaborar y aplicar modelos Bayesianos aparecen en Congdon (2005), Calvetti y Somersalo (2007), Albert (2009), Ntzoufras (2009), Fox (2010), Kruschke (2011), Cowles (2013) y Marin y Robert (2014), entre otros. Con los modelos inicializados para tres cadenas comenzamos descartando las primeras muestras (burn in) y efectuamos las simulaciones con saltos sistemáticos (thin) acordes al grado de complejidad (cuadro 6.2).

Cuadro 6.2. Sumario para realizar la inferencia

	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_5$	$\mathcal{M}_6$	$\mathcal{M}_7$
cadenas	3	3	3	3	3	3	3
burn in	10000	10000	10000	30000	30000	150000	100000
simulaciones	20000	20000	20000	30000	30000	50000	50000
thin	5	5	5	50	50	100	100

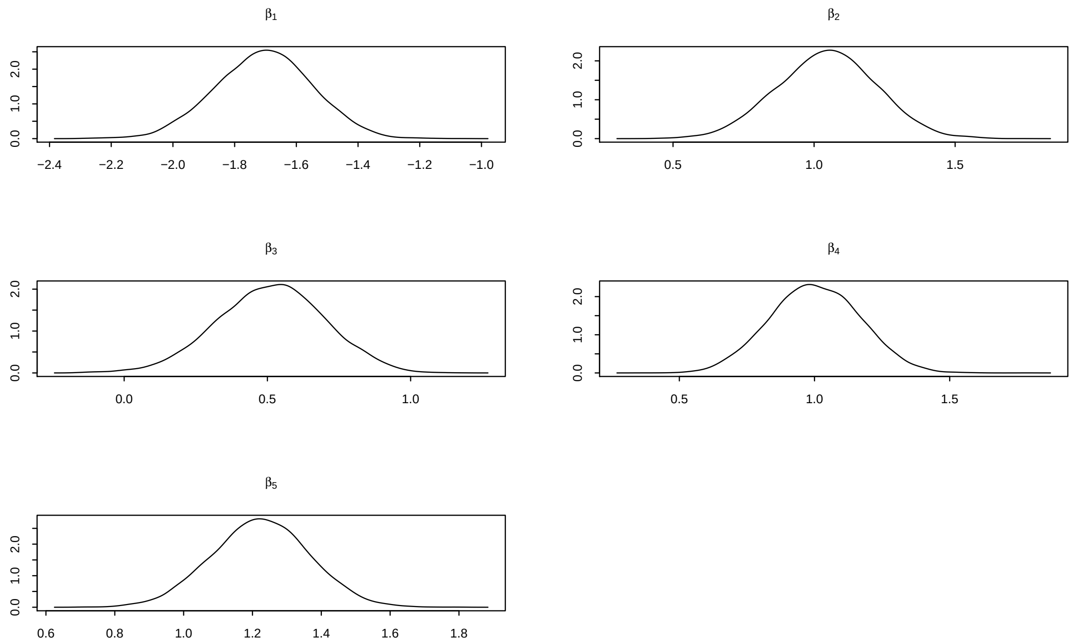

Figura 6.2.1. Plots de las fdp's a posteriori de $\mathcal{M}_1$ (1.ª cadena)

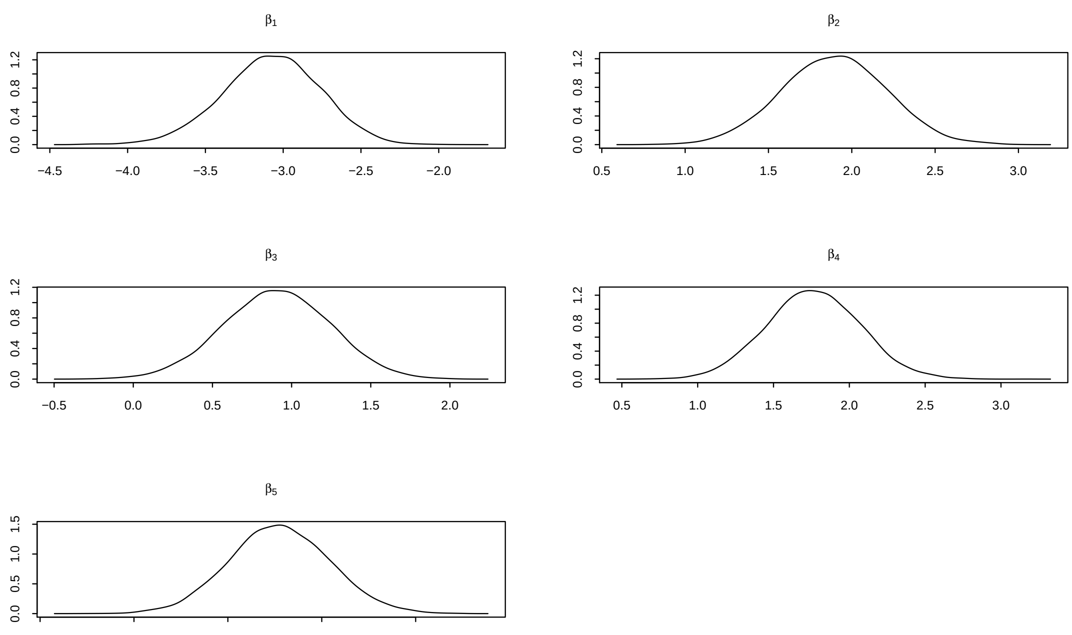

Figura 6.2.2. Plots de las fdp's a posteriori de $\mathcal{M}_2$ (1.ª cadena)

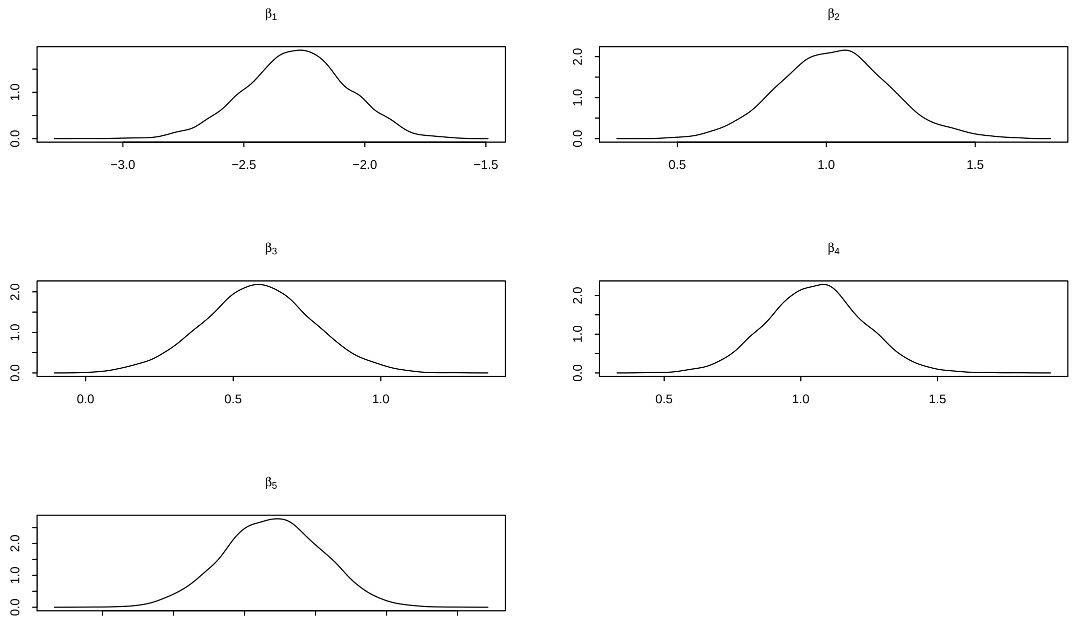

Figura 6.2.3. Plots de las fdp's a posteriori de $\mathcal{M}_3$ (1.ª cadena)

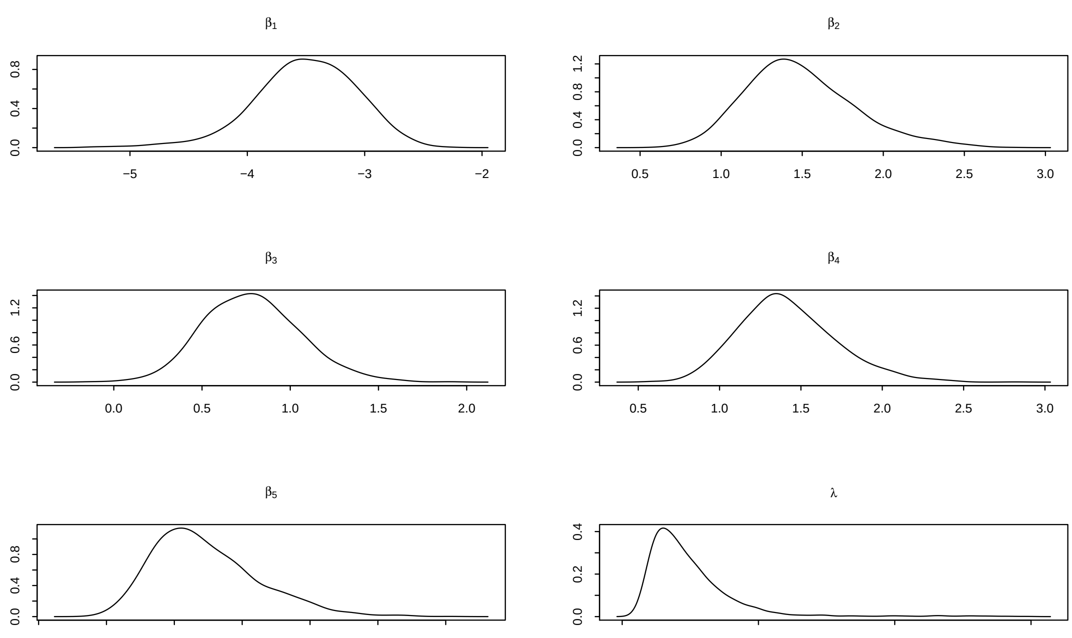

Figura 6.2.4. Plots de las fdp's a posteriori de $\mathcal{M}_4$ (1.ª cadena)

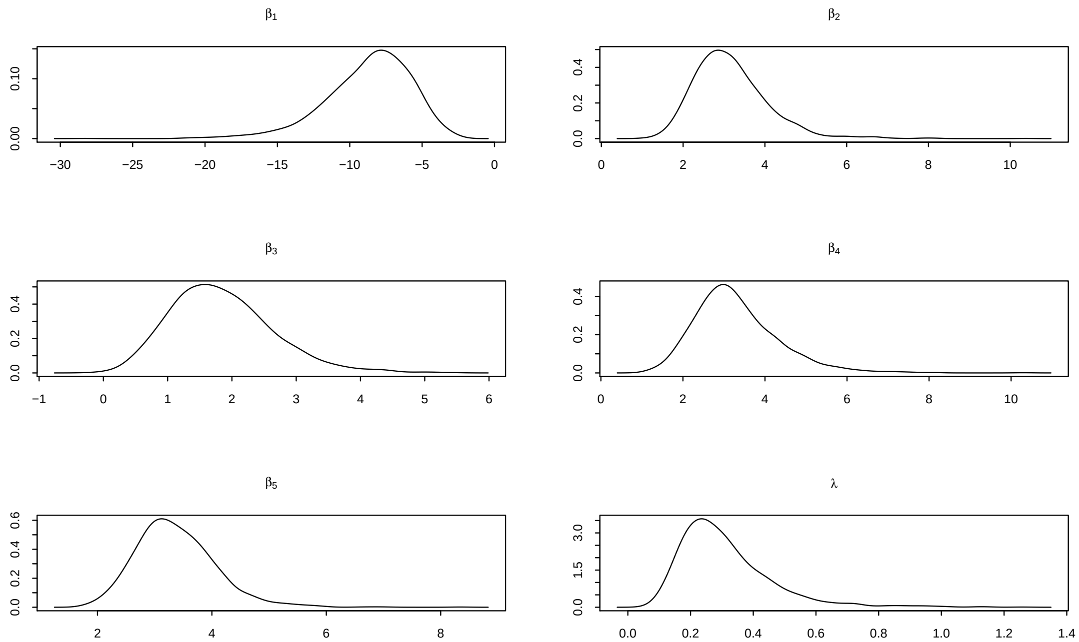

Figura 6.2.5. Plots de las fdp's a posteriori de $\mathcal{M}_5$ (1.ª cadena)

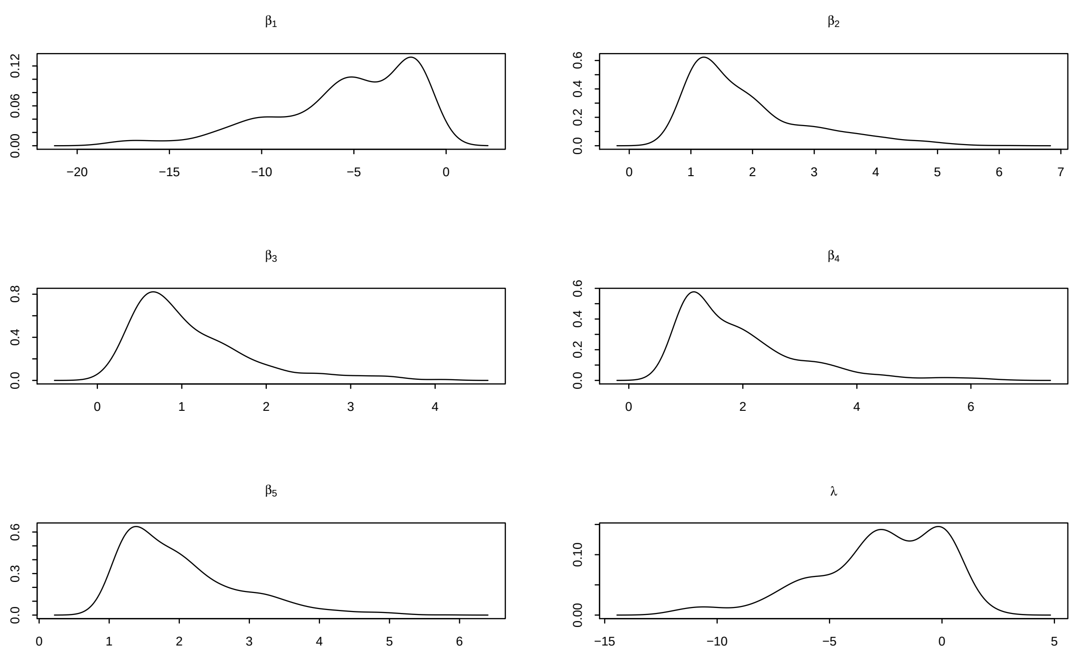

Figura 6.2.6. Plots de las fdp's a posteriori de $\mathcal{M}_6$ (1.ª cadena)

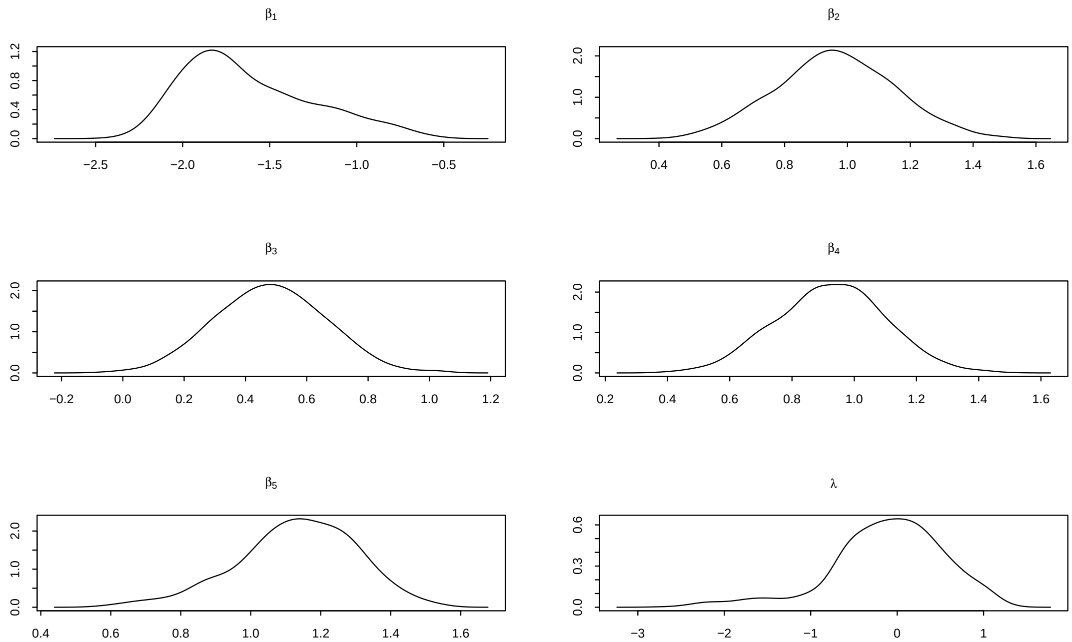

Figura 6.2.7. Plots de las fdp's a posteriori de $\mathcal{M}_7$ (1.ª cadena)

Cuadro 6.3. Sumario de la inferencia para los parámetros de regresión

Parámetro	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_5$	$\mathcal{M}_6$	$\mathcal{M}_7$
β_1	-1.705	-3.076	-2.280	-3.502	-8.698	-5.421	-1.621
β_2	1.046	1.895	1.028	1.490	3.167	1.904	0.962
β_3	0.512	0.906	0.585	0.784	1.841	1.097	0.482
β_4	1.003	1.763	1.054	1.420	3.304	1.939	0.925
β_5	1.223	2.275	1.081	1.732	3.369	2.021	1.131
λ				2.474	0.311	-2.790	-0.110

Cuadro 6.4. Valores de los criterios de bondad de ajuste

Modelo	D_{bar}	D_{hat}	DIC	AIC	BIC
$\mathcal{M}_1$	411.5	406.5	416.5	421.5	426.4
$\mathcal{M}_2$	416.1	411.1	421.1	426.1	431.0
$\mathcal{M}_3$	412.3	407.2	417.4	422.3	427.2
$\mathcal{M}_4$	412.4	421.2	403.6	424.4	430.3
$\mathcal{M}_5$	414.6	412.2	417.0	426.6	432.5
$\mathcal{M}_6$	275.1	268.1	282.1	287.1	293.0
$\mathcal{M}_7$	388.2	407.3	369.1	400.2	406.1

De manera unánime, los valores de los tres criterios de bondad de ajuste (cuadro 6.4) determinan que, el modelo con mejor desempeño o el más apropiado es el skew probit CDS ($\mathcal{M}_6$).

Cuadro 6.5. Sumario de la inferencia para los parámetros de regresión de $\mathcal{M}_6$

Parámetro	A posteriori			Intervalo HPD 95 %	
	Media	Mediana	DE	Inferior	Superior
β_1	-5.421	-4.794	3.834	-15.350	-1.113
β_2	1.904	1.608	0.986	0.794	4.516
β_3	1.097	0.890	0.713	0.243	3.123
β_4	1.939	1.645	1.083	0.779	4.915
β_5	2.021	1.813	0.839	1.041	4.199
λ	-2.790	-2.507	2.891	-10.320	1.094

El modelo seleccionado para el análisis, está dado por

$$p_i = \Pr(servicio_i = 1) = \Phi(\widehat{\eta}_i; 0, 1 + 2.790^2, -2.790), \qquad i = 1, 2, \ldots, 955.$$

donde

$$\widehat{\eta}_i = -5.421 + 1.904 tiempo_i + 1.097 trato_i + 1.939 infraequi_i + 2.021 esfurespo_i$$

y $\Phi(\cdot; 0, \sigma^2, \zeta)$ denota a la fda de la $\mathcal{SN}(0, \sigma^2, \zeta)$.

La insatisfacción con la calidad del servicio de salud pública en general se debe en primer lugar por la insatisfacción con el esfuerzo y responsabilidad de los paceños para cuidar la salud de la familia, en segundo lugar por la insatisfacción con la calidad de la infraestructura y equipamiento de los centros de salud y hospitales públicos, en penúltimo lugar por la insatisfacción con el tiempo de demora para conseguir una consulta médica en el servicio público, y en último lugar por la insatisfacción con la calidad del trato en los centros de salud y hospitales públicos.

Capítulo 7

Consideraciones Finales

La teoría exquisita desplegada especialmente en la distribución skew normal permitió desarrollar dos modelos de regresión binaria, el skew probit CDS y el skew probit BBB, y al aplicarlos resaltaron por tener mejor desempeño, exclusivamente el skew probit CDS. También y sin olvidarnos como buenos modelos alternativos están el scobit y el power logit.

7.1. Futuras Líneas de Investigación

Está claro que, la actualización del investigador va coherente con los nuevos aplicativos estadísticos para tomar mejores decisiones. Sería muy despectivo al decir en donde se puede utilizar la distribución skew normal, porque está reemplazando el sitial de la distribución normal. Mirando al modelo skew probit CDS, conocido por skew probit se puede generar un algoritmo más eficiente y analizar sus residuos de data latente, tratamientos completos aparecen Farias y Branco (2011, 2012). La nueva propuesta en modelos de regresión binaria es el skew logit muy bien fundamentado en Chen *et al.* (1999), y sus aplicaciones exquisitas se encuentran en Bermúdez *et al.* (2008), Sáez–Castillo *et al.* (2010) y Pérez–Sánchez *et al.* (2014). Las extensiones van hacia los modelos de teoría de la respuesta al ítem (TRI), recomendamos ver Bazán *et al.* (2006, 2014), Bolfarine y Bazán (2010) y Fox (2010).

Apéndice A

Distribuciones de Probabilidad

Distribución Bernoulli, $\mathcal{B}er(\pi)$.
$\pi \in [0,1]$,

$$f(y;\pi) = \pi^y (1-\pi)^{1-y} I_{\{0,1\}}(y).$$

Distribución Burr tipo II, $\mathcal{B}urr - II(\mu, \sigma^2, \zeta)$.
$\mu \in \mathbb{R}$ y $\sigma, \zeta \in \mathbb{R}_+$,

$$f(y;\mu,\sigma^2,\zeta) = \frac{\zeta}{\sigma} \exp\left(\frac{\mu - y}{\sigma}\right) \left[1 + \exp\left(\frac{\mu - y}{\sigma}\right)\right]^{-\zeta - 1}.$$

Distribución Gamma, $\mathcal{G}(\alpha, \beta)$.
$\alpha, \beta \in \mathbb{R}_+$,

$$f(y;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} \exp(-\beta y) I_{[0,+\infty)}(y).$$

Cuando $\alpha = \nu/2$ y $\beta = 1/2$ se definirá que, la variable aleatoria y tiene distribución *chi cuadrado* con parámetro de grados de libertad $\nu \in \mathbb{R}_+$ y representaremos por $y \sim \chi^2_\nu$. También $y^{-1} \sim \mathcal{GI}(\alpha, \beta)$.
Distribución Gamma Inversa, $\mathcal{GI}(\alpha, \beta)$.
$\alpha, \beta \in \mathbb{R}_+$,

$$f(y;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{-(\alpha+1)} \exp\left(-\frac{\beta}{y}\right) I_{[0,+\infty)}(y).$$

Cuando $\alpha = \nu/2$ y $\beta = (\nu\sigma^2)/2$ se definirá que, la variable aleatoria y tiene distribución *chi cuadrado inversa escalada* con parámetros de escala $\sigma \in \mathbb{R}_+$ y grados de libertad $\nu \in \mathbb{R}_+$ y representaremos por $y \sim \mathcal{I}nv - \chi^2_\nu(\sigma^2)$. También $y^{-1} \sim \mathcal{G}(\alpha, \beta)$.
Distribución Gumbel, $\mathcal{G}u(\mu, \sigma^2)$.
$\mu \in \mathbb{R}$ y $\sigma \in \mathbb{R}_+$,

$$f(y;\mu,\sigma^2) = \frac{1}{\sigma} \exp\left[\left(\frac{y-\mu}{\sigma}\right) - \exp\left(\frac{y-\mu}{\sigma}\right)\right]$$
$$F(y;\mu,\sigma^2) = 1 - \exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right].$$

Distribución Half Normal, $\mathcal{HN}(\mu, \sigma^2)$.
$\mu \in \mathbb{R}$ y $\sigma \in \mathbb{R}_+$,

$$\psi(y; \mu, \sigma^2) = \frac{2}{\sigma}\phi\left(\frac{y-\mu}{\sigma}\right) I_{[\mu,+\infty)}(y).$$

Los momentos de la $\mathcal{HN}(0,1)$ están dados por

$$\mathbb{E}[y^n] = \frac{1}{\sqrt{\pi}} 2^{n/2}\Gamma\left(\frac{n+1}{2}\right), \quad n = 1, 2, \ldots$$

Distribución Logística, $\mathcal{L}o(\mu, \sigma^2)$.
$\mu \in \mathbb{R}$ y $\sigma \in \mathbb{R}_+$,

$$f(y; \mu, \sigma^2) = \frac{1}{\sigma}\exp\left(\frac{y-\mu}{\sigma}\right)\left[1 + \exp\left(\frac{y-\mu}{\sigma}\right)\right]^{-2}$$
$$F(y; \mu, \sigma^2) = \exp\left(\frac{y-\mu}{\sigma}\right)\left[1 + \exp\left(\frac{y-\mu}{\sigma}\right)\right]^{-1}.$$

Distribución Normal, $\mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
$\boldsymbol{\mu} \in \mathbb{R}^p$ y $\boldsymbol{\Sigma} \in \mathbb{R}^p \times \mathbb{R}^p$ es definida–positiva simétrica,

$$\phi_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^p}\exp\left[-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right].$$

Los momentos de orden par de la $\mathcal{N}(0,1)$ están dados por

$$\mathbb{E}[y^{2n}] = \frac{(2n)!}{2^n n!}, \quad n = 1, 2, \ldots \tag{A.0.1}$$

Distribución Scobit, $\mathcal{S}c(\mu, \sigma^2, \zeta)$.
$\mu \in \mathbb{R}$ y $\sigma, \zeta \in \mathbb{R}_+$,

$$f(y; \mu, \sigma^2, \zeta) = \frac{\zeta}{\sigma}\exp\left(\frac{y-\mu}{\sigma}\right)\left[1 + \exp\left(\frac{y-\mu}{\sigma}\right)\right]^{-\zeta-1}.$$

Distribución Skew Normal, $\mathcal{SN}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\zeta})$.
$\boldsymbol{\mu}, \boldsymbol{\zeta} \in \mathbb{R}^p$, $\boldsymbol{\Sigma} \in \mathbb{R}^p \times \mathbb{R}^p$ es definida–positiva simétrica y $\boldsymbol{\Xi} = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_p)$,

$$\phi_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\zeta}) = 2\phi_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi(\boldsymbol{\zeta}'\boldsymbol{\Xi}^{-1}(\mathbf{y}-\boldsymbol{\mu})).$$

Distribución t de Student, $\mathcal{T}(\mu, \sigma^2)$.
$\mu \in \mathbb{R}$ y $\sigma, \nu \in \mathbb{R}_+$,

$$t(y; \mu, \sigma^2, \nu) = \frac{\Gamma((\nu+1)/2)}{\sigma\Gamma(\nu/2)\Gamma(1/2)\sqrt{\nu}}\left[1 + \frac{1}{\nu}\left(\frac{y-\mu}{\sigma}\right)^2\right]^{-(\nu+1)/2}.$$

Distribución Uniforme $\mathcal{U}(a,b)$.

$a, b \in \mathbb{R}$,

$$f(y; a, b) = \frac{1}{b-a} I_{[a,b]}(y).$$

Apéndice B

Complementos del Capítulo 3

Proposición 3.1

Demostración. Nótese que, $g(\mathbf{x}) = 2[G_0\{h(\mathbf{x}) - \frac{1}{2}\}]f_0(\mathbf{x})$ es una función impar y es integrable porque $|g(\mathbf{x})| \leq f_0(\mathbf{x})$. Entonces

$$0 = \int_{\mathbb{R}^n} g(\mathbf{x})d\mathbf{x} = 2\int_{\mathbb{R}^n} f_0(\mathbf{x})G_0\{h(\mathbf{x})\}d\mathbf{x} - 1.$$

□

Lema 3.5

Demostración. Consideremos que, $u_0 \sim \mathcal{N}(\mu, \sigma^2)$ y $v_0 \sim \mathcal{N}(0,1)$. Entonces

$$\epsilon = u_0 - v_0 \sim \mathcal{N}(\mu, \sigma^2 + 1) \tag{B.0.1}$$

La probabilidad de que la variable aleatoria ϵ sea inferior a cero es

$$\begin{aligned} \Pr(\epsilon < 0) &= \Pr(u_0 < v_0) \\ &= \mathbb{E}_{v_0}[\Pr(u_0 < v_0 | v_0)] \\ &= \int_{-\infty}^{\infty} \Phi\left(\frac{v_0 - \mu}{\sigma}\right)\phi(v_0)dv_0 \\ &= \mathbb{E}\left[\Phi\left(\frac{v_0 - \mu}{\sigma}\right)\right], \end{aligned} \tag{B.0.2}$$

también es

$$\Pr(\epsilon < 0) = \Phi\left(-\frac{\mu}{\sqrt{\sigma^2 + 1}}\right). \tag{B.0.3}$$

Igualando (B.0.2) con (B.0.3) y realizando $\lambda_1 = -\mu/\sigma$ y $\lambda_2 = 1/\sigma$ se prueba la proposición. □

Plots

La definición e interpretación de la distancia media entre dos fdp's aparece en Weiss (1996) y Vidal *et al.* (2006).

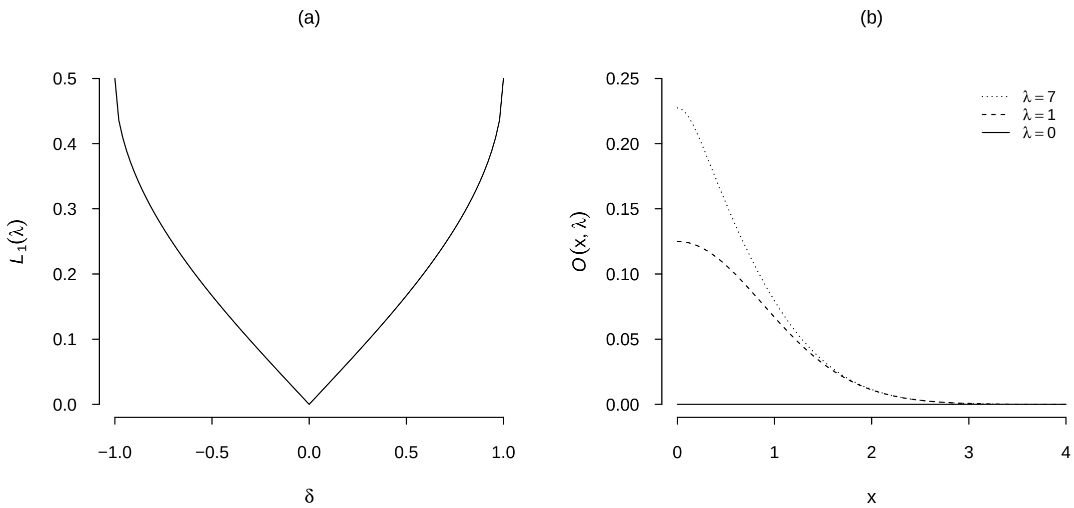

Figura B.0.1. Plots de las funciones (a) $\mathcal{L}_1(\lambda)$ para $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}} \in [-1, 1]$ y (b) $\mathcal{O}(x; \lambda)$

$\mathcal{O}(x, \lambda)$ es decreciente en x. $\mathcal{O}(-x, \lambda) = \mathcal{O}(x, \lambda)$, $\mathcal{O}(x, -\lambda) = -\mathcal{O}(x, \lambda)$ y $2\mathcal{O}(x, 1) = \Phi(x)\Phi(-x)$ (Owen, 1956; Azzalini, 1985, 2005, 2014).

Apéndice C

Sintaxis y Comandos para RStudio o R

Las siguientes sintaxis están elaboradas para ser implementadas en los programas RStudio o R a través de sus librerías BRugs o R2OpenBUGS o R2WinBUGS. Su presentación es general i.e. se debe asignar las variables a tratar y el tipo de archivo a utilizar debe ser txt.

De acuerdo a la aplicación, se debe realizan los reemplazos: y[i] por servicio[i], x2[i] por tiempo[i], x3[i] por trato[i], x4[i] por infraequi[i] y x5[i] por esfurespo[i]. La bodega de datos también debe encontrarse en txt. Los comandos de todos los modelos irán conforme a los requerimientos de la librería BRugs, por presentar un sumario a detalle.

Modelo Probit ($\mathcal{M}_1$)

```
model
    {
    for(i in 1:n)
            {
            eta[i] <- beta[1]+beta[2]*x2[i]+beta[3]*x3[i]+···+beta[k]*xk[i]
            p[i] <- phi(eta[i])
            y[i] ~ dbern(p[i])
            }
    for(j in 1:k)
            {
            beta[j] ~ dnorm(0.0,1.0E-4)
            }
    }
```

Comandos del Modelo Probit

```
library("BRugs")
modelFile <- "D:/Documents/Aplicación tesis salud/Probit Salud OLCV 2012 R.txt"
bodega <- read.table("D:/Documents/Aplicación tesis salud/bodega Salud OLCV 2012 R.txt", header=TRUE, sep="", na.strings="NA", dec=".", strip.white=TRUE)
data <- c(as.list(bodega), list(n=955, k=5))
inits <- function(){list(beta=c(0,0,0,0,0))}
```

parametersToSave <- “beta”
BRugsFit(modelFile, data, inits, numChains=3, parametersToSave, nBurnin=10000, nIter=20000, nThin=5, coda=FALSE, DIC=TRUE, working.directory=NULL, digits=3, seed=NULL, BRugsVerbose=getOption(“BRugsVerbose”))

Modelo Logit ($\mathcal{M}_2$)

```
model
    {
    for(i in 1:n)
            {
            eta[i] <- beta[1]+beta[2]*x2[i]+beta[3]*x3[i]+···+beta[k]*xk[i]
            p[i] <- ilogit(eta[i])
            y[i] ~ dbern(p[i])
            }
    for(j in 1:k)
            {
            beta[j] ~ dnorm(0.0,1.0E-4)
            }
    }
```

Comandos del Modelo Logit

library(“BRugs”)
modelFile <- “D:/Documents/Aplicación tesis salud/Logit Salud OLCV 2012 R.txt”
bodega <- read.table(“D:/Documents/Aplicación tesis salud/bodega Salud OLCV 2012 R.txt”, header=TRUE, sep=“”, na.strings=“NA”, dec=“.”, strip.white=TRUE)
data <- c(as.list(bodega), list(n=955, k=5))
inits <- function(){list(beta=c(0,0,0,0,0))}
parametersToSave <- “beta”
BRugsFit(modelFile, data, inits, numChains=3, parametersToSave, nBurnin=10000, nIter=20000, nThin=5, coda=FALSE, DIC=TRUE, working.directory=NULL, digits=3, seed=NULL, BRugsVerbose=getOption(“BRugsVerbose”))

Modelo Cloglog ($\mathcal{M}_3$)

```
model
    {
    for(i in 1:n)
            {
            eta[i] <- beta[1]+beta[2]*x2[i]+beta[3]*x3[i]+···+beta[k]*xk[i]
            cloglog(p[i]) <- eta[i]
            y[i] ~ dbern(p[i])
            }
```

```
    for(j in 1:k)
            {
            beta[j] ~ dnorm(0.0,1.0E-4)
            }
    }
```

Comandos del Modelo Cloglog

```
library("BRugs")
modelFile <- "D:/Documents/Aplicación tesis salud/Cloglog Salud OLCV 2012 R.txt"
bodega <- read.table("D:/Documents/Aplicación tesis salud/bodega Salud OLCV 2012 R.txt", header=TRUE, sep="", na.strings="NA", dec=".", strip.white=TRUE)
data <- c(as.list(bodega), list(n=955, k=5))
inits <- function(){list(beta=c(0,0,0,0,0))}
parametersToSave <- "beta"
BRugsFit(modelFile, data, inits, numChains=3, parametersToSave, nBurnin=10000, nIter=20000, nThin=5, coda=FALSE, DIC=TRUE, working.directory=NULL, digits=3, seed=NULL, BRugsVerbose=getOption("BRugsVerbose"))
```

Modelo Scobit ($\mathcal{M}_4$)

```
model
    {
    for(i in 1:n)
            {
            eta[i] <- beta[1]+beta[2]*x2[i]+beta[3]*x3[i]+···+beta[k]*xk[i]
            p[i] <- 1-pow(ilogit(-eta[i]),lambda)
            y[i] ~ dbern(p[i])
            }
    for(j in 1:k)
            {
            beta[j] ~ dnorm(0.0,1.0E-4)
            lambda ~ dgamma(0.25,0.25)
            }
    }
```

Comandos del Modelo Scobit

```
library("BRugs")
modelFile <- "D:/Documents/Aplicación tesis salud/Scobit Salud OLCV 2012 R.txt"
bodega <- read.table("D:/Documents/Aplicación tesis salud/bodega Salud OLCV 2012 R.txt", header=TRUE, sep="", na.strings="NA", dec=".", strip.white=TRUE)
data <- c(as.list(bodega), list(n=955, k=5))
inits <- function(){list(beta=c(0,0,0,0,0), lambda=0.125)}
```

parametersToSave <- c("beta", "lambda")
BRugsFit(modelFile, data, inits, numChains=3, parametersToSave, nBurnin=30000, nIter=30000, nThin=50, coda=FALSE, DIC=TRUE, working.directory=NULL, digits=3, seed=NULL, BRugsVerbose=getOption("BRugsVerbose"))

Modelo Power Logit ($\mathcal{M}_5$)

```
model
    {
    for(i in 1:n)
            {
            eta[i] <- beta[1]+beta[2]*x2[i]+beta[3]*x3[i]+···+beta[k]*xk[i]
            p[i] <- pow(ilogit(eta[i]),lambda)
            y[i] ~ dbern(p[i])
            }
    for(j in 1:k)
            {
            beta[j] ~ dnorm(0.0,1.0E-4)
            lambda ~ dgamma(0.25,0.25)
            }
    }
```

Comandos del Modelo Power Logit

library("BRugs")
modelFile <- "D:/Documents/Aplicación tesis salud/Power Logit Salud OLCV 2012 R.txt"
bodega <- read.table("D:/Documents/Aplicación tesis salud/bodega Salud OLCV 2012 R.txt", header=TRUE, sep="", na.strings="NA", dec=".", strip.white=TRUE)
data <- c(as.list(bodega), list(n=955, k=5))
inits <- function(){list(beta=c(0,0,0,0,0), lambda=0.125)}
parametersToSave <- c("beta", "lambda")
BRugsFit(modelFile, data, inits, numChains=3, parametersToSave, nBurnin=30000, nIter=30000, nThin=50, coda=FALSE, DIC=TRUE, working.directory=NULL, digits=3, seed=NULL, BRugsVerbose=getOption("BRugsVerbose"))

Modelo Skew Probit CDS ($\mathcal{M}_6$)

```
model
    {
    for(i in 1:n)
            {
            eta[i] <- beta[1]+beta[2]*x2[i]+beta[3]*x3[i]+···+beta[k]*xk[i]
            a[i] <- eta[i]-lambda*tau[i]
            z[i] ~ dnorm(a[i],1)I(lower[y[i]+1],upper[y[i]+1])
```

```
		tau[i] ~ dnorm(0,1)I(0,)
		}
	for(j in 1:k)
		{
		beta[j] ~ dnorm(0.0,1.0E-4)
		}
	lambda ~ dt(0,2,2)
	lower[1] <- -50
	lower[2] <- 0
	upper[1] <- 0
	upper[2] <- 50
	}
```

Comandos del Modelo Skew Probit CDS

```
library("BRugs")
modelFile <- "D:/Documents/Aplicación tesis salud/Skew Probit CDS Salud OLCV 2012 R.txt"
bodega <- read.table("D:/Documents/Aplicación tesis salud/bodega Salud OLCV 2012 R.txt", header=TRUE, sep="", na.strings="NA", dec=".", strip.white=TRUE)
data <- c(as.list(bodega), list(n=955, k=5))
inits <- function(){list(beta=c(0,0,0,0,0), lambda=0.25)}
parametersToSave <- c("beta", "lambda")
BRugsFit(modelFile, data, inits, numChains=3, parametersToSave, nBurnin=150000, nIter=50000, nThin=100, coda=FALSE, DIC=TRUE, working.directory=NULL, digits=3, seed=NULL, BRugsVerbose=getOption("BRugsVerbose"))
```

Modelo Skew Probit BBB ($\mathcal{M}_7$)

```
model
	{
	for(i in 1:n)
		{
		eta[i] <- beta[1]+beta[2]*x2[i]+beta[3]*x3[i]+···+beta[k]*xk[i]
		a[i] <- eta[i]-lambda*pow(b,-0.5)*tau[i]
		z[i] ~ dnorm(a[i],b)I(lower[y[i]+1],upper[y[i]+1])
		tau[i] ~ dnorm(0,1)I(0,)
		}
	for(j in 1:k)
		{
		beta[j] ~ dnorm(0.0,1.0E-4)
		}
```

b <- 1+pow(lambda,2)
lambda ~ dt(0,2,2)
lower[1] <- -50
lower[2] <- 0
upper[1] <- 0
upper[2] <- 50
}

Comandos del Modelo Skew Probit BBB

library("BRugs")
modelFile <- "D:/Documents/Aplicación tesis salud/Skew Probit BBB Salud OLCV 2012 R.txt"
bodega <- read.table("D:/Documents/Aplicación tesis salud/bodega Salud OLCV 2012 R.txt", header=TRUE, sep="", na.strings="NA", dec=".", strip.white=TRUE)
data <- c(as.list(bodega), list(n=955, k=5))
inits <- function(){list(beta=c(0,0,0,0,0), lambda=0.25)}
parametersToSave <- c("beta", "lambda")
BRugsFit(modelFile, data, inits, numChains=3, parametersToSave, nBurnin=100000, nIter=50000, nThin=100, coda=FALSE, DIC=TRUE, working.directory=NULL, digits=3, seed=NULL, BRugsVerbose=getOption("BRugsVerbose"))

Bibliografía

[1] Achen, C. H. (2002) Toward a New Political Methodology: Microfoundations and Art. *Annual Review of Political Science* **5**, 423–450.

[2] Agresti, A., y Kateri, M. (2014) Some Remarks on Latent Variable Models in Categorical Data Analysis. *Communications in Statistics – Theory and Methods* **43**, 801–814.

[3] Ahsanullah, M., Kibria, B. M. G., y Shakil, M. (2014) *Normal and Student's t Distributions and Their Applications*. Atlantis Press, Paris.

[4] Albert, J. (2009) *Bayesian Computation with R* (2.ª ed.). Springer, New York.

[5] Albert, J. H., y Chib, S. (1995) Bayesian Residual Analysis for Binary Response Regression Models. *Biometrika* **82**, 747–759.

[6] Arellano–Valle, R. B., Ozan, S., Bolfarine, H., y Lachos V. H. (2005) Skew Normal Measurement Error Models. *Journal of Multivariate Analysis* **96**, 265–281.

[7] Azzalini, A. (1985) A Class of Distributions which Includes the Normal Ones. *Scandinavian Journal of Statistics* **12**, 171–178.

[8] ——— (1986) Further Results on a Class of Distribution which Includes the Normal Ones. *Statistica* **46**, 199–208.

[9] ——— (2005) The Skew–Normal Distribution and Related Multivariate Families. *Scandinavian Journal of Statistics* **32**, 159–188.

[10] ——— (2014) *The Skew–Normal and Related Families*. Cambridge University Press, New York.

[11] Azzalini, A., y Capitanio, A. (1999) Statistical Applications of the Multivariate Skew Normal Distribution. *Journal of the Royal Statistical Society, Series B* **61**(3), 579–602.

[12] ——— (2003) Distributions Generated by Perturbation of Symmetry with Emphasis on a Multivariate Skew t–Distribution. *Journal of the Royal Statistical Society, Series B* **65**(2), 367–389.

[13] Azzalini, A., y Chiogna, M. (2004) Some Results on the Stress–Strength Model for Skew–Normal Variates. *METRON* **LXII**(3), 315–326.

[14] Bagui, S., y Bagui, S. (2006) Computing Percentiles of Skew–Normal Distributions. *Journal of Modern Applied Statistical Methods* **5**(2), 283–595.

[15] Balakrishnan, N., y Lai, C.–D. (2009) *Continuous Bivariate Distributions* (2.ª ed.). Springer, New York.

[16] Basu, D. (1977) On the Elimination of Nuisance Parameters. *Journal of the American Statistical Association* **72**(358), 355–366.

[17] Bayes, C. L., y Branco, M. D. (2007) Bayesian Inference for the Skewness Parameter of

the Scalar Skew–Normal Distribution. *Brazilian Journal of Probability and Statistics* **21**, 141–163.

[18] Bayes, T. (1763) An Essay Towards Solving a Problem in the Doctrine of Chances. *Philosophical Transactions of the Royal Society of London* **53**, 370–416; **54**, 296–325.

[19] Bazán, J. L., Bolfarine, H., y Branco, M. D. (2010) A Framework for Skew–Probit Links in Binary Regression. *Communications in Statistics – Theory and Methods* **39**, 678–697.

[20] Bazán, J. L., Branco, M. D., y Bolfarine, H. (2006) A Skew Item Response Model. *Bayesian Analysis* **1**, 861–892.

[21] ——— (2014) Extensions of the Skew–Normal Ogive Item Response Model. *Brazilian Journal of Probability and Statistics* **28**(1), 1–23.

[22] Berger, J. O., y Wolpert, R. (1988) *The Likelihood Principle* (2.ª ed.). IMS Lecture Notes – Monograph Series, 9. Hayward, California.

[23] Berkson, J. (1944) Application of the Logistic Function to Bio–Assay. *Journal of the American Statistical Association* **39**, 357–365.

[24] Bermúdez, Ll., Pérez–Sánchez, J. M., Ayuso, M., Gómez–Déniz, E., y Vázquez, F. J. (2008) A Bayesian Dichotomous Model with Asymmetric Link for Fraud in Insurance. *Insurance: Mathematics and Economics* **42**, 779–786.

[25] Bernardo, J. M., y Smith, A. F. M. (2000) *Bayesian Theory* (2.ª ed.). Wiley, New York.

[26] Bingham, N. H., y Fry, J. M. (2010) *Regression: Linear Models in Statistics*. Springer, London.

[27] Bliss, C. I. (1935) The Calculation of the Dosage–Mortality Curve. *Annals of Applied Biology* **22**, 134–167.

[28] Bolfarine, H., y Bazán, J. L. (2010) Bayesian Estimation of the Logistic Positive Exponent IRT Model. *Journal of Educational and Behavioral Statistics* **35**, 693–713.

[29] Branco, M. D., Genton, M. G., y Liseo, B. (2012) Objective Bayesian Analysis of Skew–t Distributions. *Scandinavian Journal of Statistics* **40**, 63–85.

[30] Bryc, W. (1995) *The Normal Distribution, Characterizations with Applications*. Springer, New York.

[31] Cabras, S., y Castellanos, M. E. (2009) Default Bayesian Goodness–of–Fit Tests for the Skew–Normal Model. *Journal of Applied Statistics* **36**(2), 223–232.

[32] Calvetti, D., y Somersalo, E. (2007) *Introduction to Bayesian Scientific Computing*. Springer, New York.

[33] Carlin, B. P., y Louis, T. A. (2009) *Bayesian Methods for Data Analysis* (3.ª ed.). Chapman & Hall/CRC, Boca Raton.

[34] Chaibub–Neto, E., y Branco, M. D. (2003) Bayesian Reference Analysis for Binomial Calibration Problem. Technical Report, IME–USP, 12.

[35] 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**(7), 487–494.

[36] Chen, M.–H. (2004) Skewed Link Models for Categorical Response Data. En *Skew–Elliptical Distributions and Their Applications: A Journey Beyond Normality* (M. G. Genton, ed.). Chapman & Hall/CRC, Boca Raton.

[37] Chen, M.–H., Dey, D. K., y Shao, Q.–M. (1999) A New Skewed Link Model for Dicho-

tomous Quantal Response Data. *Journal of the American Statistical Association* **94**(448), 1172–1186.

[38] Chen, M.–H., Shao, Q.–M., y Ibrahim, J. G. (2000) *Monte Carlo Methods in Bayesian Computation*. Springer, New York.

[39] Chiogna, M. (1998) Some Results on the Scalar Skew–Normal Distribution. *Journal of the Italian Statistical Society* **7**, 1–13.

[40] Congdon, P. (2005) *Bayesian Models for Categorical Data*. Wiley, NewYork.

[41] Cowles, M. K. (2013) *Applied Bayesian Statistics with R and OpenBUGS Examples*. Springer, New York.

[42] Czado, C. (1994) Bayesian Inference of Binary Regression Models with Parametric Link. *Journal of Statistical Planning and Inference* **41**, 121–140.

[43] Dalla–Valle, A. (2004) The Skew–Normal Distribution. En *Skew–Elliptical Distributions and Their Applications: A Journey Beyond Normality* (M. G. Genton, ed.). Chapman & Hall/CRC, Boca Raton.

[44] Dobson, A. J. (2002) *An Introduction to Generalized Linear Models* (2.ª ed.). Chapman & Hall/CRC, London.

[45] Fahrmeir, L., Kneib, T., Lang, S., y Marx, B. (2013) *Regression: Models, Methods and Applications*. Springer, New York.

[46] Farias, R. B. A., y Branco, M. D. (2011) Efficient Algorithms for Bayesian Binary Regression Model with Skew–Probit Link. En *Recent Advances in Biostatistics* (M. Bhattacharjee, S. K. Dhar, y S. Subramanian, eds.). World Scientific, Singapore.

[47] ——— (2012) Latent Residual Analysis in Binary Regression with Skewed Link. *Brazilian Journal of Probability and Statistics* **26**(4), 344–357.

[48] Fisher, R. A. (1930) Inverse probability. *Mathematical Proceedings of the Cambridge Philosophical Society* **26**, 528–535.

[49] Flury, B. (1997) *A First Course in Multivariate Statistics*. Springer, New York.

[50] Fox, J.–P. (2010) *Bayesian Item Response Modeling: Theory and Applications*. Springer, New York.

[51] Genz, A., y Bretz, F. (2009) *Computation of Multivariate Normal and t Probabilities*. Springer, New York.

[52] Gumbel, E. J. (1935) Les valeurs extrêmes des distributions statistiques. *Annales de l'institut Henri Poincaré* **5**(2), 115–158.

[53] Gupta, A. K. (2003) Multivariate Skew t–Distribution. *Statistics* **37**(4), 359–363.

[54] Gupta, A. K., Nguyen, T. T., y Sanqui, J. A. T. (2004) Characterization of the Skew–Normal Distribution. *Annals of the Institute of Statistical Mathematics* **56**(2), 351–360.

[55] Gupta, A. K., Varga, T., y Bodnar, T. (2013) *Elliptically Contoured Models in Statistics and Portfolio Theory* (2.ª ed.). Wiley, New York.

[56] Gupta, A. K., y Chen, T. (2001) Goodness–of–Fit Tests for the Skew–Normal Distribution. *Communications in Statistics – Simulation and Computation* **30**(4), 907–930.

[57] ——— (2003) On the Sample Characterization Criterion for Normal Distributions. *Journal of Statistical Computation and Simulation* **73**(3), 155–163.

[58] Held, L., y Bové, D. S. (2014) *Applied Statistical Inference: Likelihood and Bayes*. Springer, New York.

[59] Henze, N. (1986) A Probabilistic Representation of the “Skew–Normal” Distribution. *Scandinavian Journal of Statistics* **13**, 271–275.

[60] Jeffreys, H. (1946) An Invariant Form for the Prior Probability in Estimation Problems. *Proceedings of the Royal Society of London, Series A* **186**, 453–461.

[61] Kim, H.–J. (2002) Binary Regression with a Class of Skewed t Link Models. *Communications in Statistics – Theory and Methods* **31**(10), 1863–1886.

[62] Kotz, S., Balakrishnan, N., y Johnson, N. L. (2000) *Continuous Multivariate Distributions* (2.ª ed.). Wiley, New York.

[63] Kroese, D. P., y Chan, J. C. C. (2014) *Statistical Modeling and Computation.* Springer, New York.

[64] Kruschke, J. K. (2011) *Doing Bayesian Data Analysis, A Tutorial with R and BUGS.* Academic Press/Elsevier, Burlington, MA.

[65] Lancaster, H. O. (1965) The Helmert Matrices. *The American Mathematical Monthly* **72**(1), 4–12.

[66] Laplace, P. (1812) Théorie Analytique des Probabilités. Courcier.

[67] Leonard, T. H. (2014) A personal history of Bayesian statistics. *WIREs Computational Statistics* **6**, 80–115.

[68] Lindsey, J. K. (1997) Applying Generalized Linear Models. Springer, New York.

[69] Liseo, B. (2004) Skew–Elliptical Distributions in Bayesian Inference. En *Skew–Elliptical Distributions and Their Applications: A Journey Beyond Normality* (M. G. Genton, ed.). Chapman & Hall/CRC, Boca Raton.

[70] 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.

[71] Mameli, V., y Musio, M. (2013) A Generalization of the Skew–Normal Distribution: The Beta Skew–Normal. *Communications in Statistics – Theory and Methods* **42**, 2229-2244.

[72] Marin, J.–M., y Robert, C. P. (2014) *Bayesian Essentials with R* (2.ª ed.). Springer, New York.

[73] Martínez, E. H., Varela, H., Gómez, H. W., y Bolfarine, H. (2008) A Note on the Likelihood and Moments of the Skew–Normal Distribution. *Statistics and Operations Research Transactions* **32**(1), 57–65.

[74] McCullagh, P., y Nelder, J. (1989) *Generalized Linear Models* (2.ª ed.). Chapman & Hall/CRC, London.

[75] Myers, R. H., Montgomery, D. C., Vining, G. G., y Robinson, T. J. (2010) *Generalized Linear Models with Applications in Engineering and the Sciences* (2.ª ed.). Wiley, New Jersey.

[76] Nagler, J. (1994) Scobit: An Alternative Estimator to Logit and Probit. *American Journal Political Science* **38**, 230–255.

[77] Nelder, J., y Wedderburn, R. W. M. (1972) Generalized linear models. *Journal of Royal Statistical Society, Series A* **135**, 370–384.

[78] Ntzoufras, I. (2009) *Bayesian Modeling Using WinBUGS.* Wiley, New Jersey.

[79] Owen, D. B. (1956) Tables for Computing Bivariate Normal Probabilities. *The Annals of Mathematical Statistics* **27**(4), 1075–1090.

[80] Pérez–Sánchez, J. M., Negrín–Hernández, M. A., García–García, C., y Gómez–Déniz, E.

(2014) Bayesian Asymmetric Logit Model for Detecting Risk Factors in Motor Ratemaking. *ASTIN Bulletin* **44**(2), 445–457.

[81] Pericchi, L. R. (1998) *Análisis de Decisiones, Inferencia y Predicción Estadística Bayesiana*. CESMa–USB, Caracas.

[82] Prentice, R. L. (1976) A Generalization of the Probit and Logit Methods for Dose Response Curves. *Biometrics* **32**, 761–768.

[83] Press, S. J. (2003) *Subjective and Objective Bayesian Statistics: Principles, Models, and Applications* (2.ª ed.). Wiley, New Jersey.

[84] Rencher, A. C., y Schaalje, G. B. (2008) *Linear Models in Statistics* (2.ª ed.). Wiley, New Jersey.

[85] Robert, C. P. (2007) *The Bayesian Choice* (2.ª ed.). Springer, New York.

[86] Robert, C. P., y Casella, G. (2010) *Introducing Monte Carlo Methods with R*. Springer, New York.

[87] Rubio, F. J., y Liseo, B. (2014) On the Independence Jeffreys Prior for Skew–Symmetric Models. *Statistics and Probability Letters* **85**, 91–97.

[88] Sáez–Castillo, A. J., Olmo–Jiménez, M. J., Pérez–Sánchez, J. M., Negrín–Hernández, M. A., Arcos–Navarro, Á., y Díaz–Oller, J. (2010) Bayesian Analysis of Nosocomial Infection Risk and Length of Stay in a Department of General and Digestive Surgery. *Value in Health* **13**(4), 431–439.

[89] Sarısoy, E. E., Potas, N., y Kara, M. (2014) A Simulation Study Goodness–of–Fit Tests for the Skewed Normal Distribution. En *Chaos, Complexity and Leadership 2012* (S. Banerjee, y Ş. Ş. Erçetin, eds.). Springer, New York.

[90] Savchuk, V. P., y Tsokos, C. P. (2011) *Bayesian Theory and Methods with Applications*. Atlantis Press, Paris.

[91] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., y van der Linde, A. (2002) Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society, Series B* **64**, 583–639.

[92] Stukel, T. A. (1988) Generalized Logistic Models. *Journal of the American Statistical Association* **83**, 426–431.

[93] Suess, E. A., y Trumbo, B. E. (2010) *Introduction to Probability Simulation and Gibbs Sampling with R*. Springer, New York.

[94] Tong, Y. L. (1990) *The Multivariate Normal Distribution*. Springer, New York.

[95] Vidal, I., Iglesias, P., Branco, M. D., y Arellano–Valle, R. B. (2006) Bayesian Sensitivity Analysis and Model Comparison for Skew Elliptical Models. *Journal of Statistical Planning and Inference* **136**, 3435–3457.

[96] Wakefield, J. (2013) *Bayesian and Frequentist Regression Methods*. Springer, New York.

[97] Weiss, R. (1996) An Approach to Bayesian Sensitivity Analysis. *Journal of the Royal Statistical Society, Series B* **58**, 739–750.

[98] Yan, X., y Su, S. G. (2009) *Linear Regression Analysis: Theory and Computing*. World Scientific, Singapore.

[99] Zacks, S. (1981) *Parametric Statistical Inference*. Pergamon Press, Oxford.

[100] Ziegler, A. (2011) *Generalized estimating equations*. Springer, New York.