

Nro. 25

SERIE B

**INFERENCIA BAYESIANA EN MODELOS DE
REGRESION BINARIA USANDO BRMUW**

Jorge Bazán, Cristian Bayes.

San Miguel, febrero del 2010

Departamento de Ciencias
Sección Matemática
Pontificia Universidad Católica del Perú
Apartado 1761
Lima-Perú

PRESENTACIÓN

Con el objetivo de promover el intercambio de experiencias de investigación entre académicos nacionales e internacionales, los programas de posgrado en Estadística, Matemáticas Aplicadas y Matemáticas de la Pontificia Universidad Católica del Perú (PUCP) organizaron la I Jornada Internacional de Probabilidad y Estadística (JIPE-2010).

La jornada contó con sesiones plenarias y conferencias a cargo de reconocidos académicos internacionales, cinco minicursos, una sesión de comunicaciones y presentación de posters. Todos los trabajos fueron sujetos a la aprobación de un comité científico internacional.

Pensado como el primero de una serie de eventos similares, el JIPE se desarrolló en el campus de la PUCP del 3 al 5 de febrero del 2010. Esta jornada fue un proyecto ganador del premio Ulises de la Escuela de Posgrado de la PUCP instituido para promover la internalización de sus programas. El evento ha contado también con el apoyo de la Sección Matemáticas del Departamento de Ciencias.

En este reporte se presenta el material utilizado en uno de los minicursos ofrecidos en el evento. Agradecemos a los autores por su esfuerzo y contribución desinteresada por promover el área y difundir los resultados de sus investigaciones.

Comité Científico JIPE-2010
Comité Organizador JIPE-2010

Índice

1. Introducción	7
2. Inferencia Bayesiana en el Modelo Lineal	10
2.1. Función de Verosimilitud	10
2.2. Inferencia Bayesiana	10
2.3. Inferencia Bayesiana en el Modelo de Regresión Lineal . .	12
2.3.1. Considerando distribuciones a priori no informativas	14
2.3.2. Considerando distribuciones a priori informativas .	15
2.4. Métodos de MCMC	16
3. Inferencia bayesiana en Regresión binaria	21
3.1. Modelo de Regresión Binaria	21
3.2. Enlaces Asimétricos en Regresión Binaria	22
3.3. Inferencia Bayesiana	24
3.4. Inferencia Bayesiana en el Modelo de Regresión Logística	26
4. Otros temas de Inferencia Bayesiana	28
4.1. Condiciones de existencia de la posteriori para los parámetros de regresión	28
4.2. Comparación de Modelos	30
5. Aplicaciones	32
5.1. Aplicación 1: Datos del Challenger	32
5.2. Aplicación 2: Datos de erradicación de cultivos de coca . .	34
6. Uso del BRMUW	38
6.1. Generar la sintaxis del modelo	38
6.2. Generar la sintaxis para los datos	42
6.3. Estimación Bayesiana usando WinBUGS u OPENBUGS .	46
6.4. Estimación Bayesiana dentro de R	54
Referencias	59

1. Introducción

Una variable aleatoria es considerada binaria o dicotómica cuando puede tomar dos posibles valores o categorías, tales como suceso (1) o falla (0), positivo (1) o negativo (0), correcto (1) o incorrecto (0). Ese tipo de datos son comunes en las ciencias sociales, médicas, agricultura, genética, educación y psicología, y son modelados usando la regresión binaria.

Los modelos de regresión binaria son modelos estadísticos usados para predecir la probabilidad de una respuesta binaria en función de diversas variables explicativas o predictores. Conjuntos de datos que requieren este tipo de análisis se encuentran en áreas tan diversas como ingeniería, ciencias naturales, educación, etc. Por ejemplo, el hecho de que un paciente sobreviva o no a una enfermedad puede ser explicado por variables como el tratamiento aplicado, edad, etc; o el resultado de un examen de admisión (aprueba o no aprueba) puede ser influenciado por el nivel de conocimiento del alumno en matemáticas, lenguaje, etc. Este tipo de modelo supone un error que es considerado de distribución simétrica el cual induce un enlace, entre los predictores y las probabilidades, que es simétrico.

Los modelos más conocidos con enlace simétrico son la regresión logística, cuando el enlace es el logit y el error tiene distribución de probabilidad logística estándar, y la regresión probit, cuando dicho enlace es el probit y el error tiene distribución de probabilidad normal estándar.

Sin embargo este tipo de suposiciones son restrictivas y no aplicables cuando se tiene una mayor frecuencia de una de las respuestas binarias. Autores, entre ellos Collet (2003), Czado y Santner (1992), Chen, Dey y Shao (1999) reportan que enlaces asimétricos pueden ser más apropiados que enlaces simétricos en situaciones específicas.

Este ocurre por ejemplo cuando se quiere modelar o explicar el resultado de la probabilidad de tener un paro cardíaco en una población considerando una serie de características individuales como si fuma o no, nivel de colesterol, etc, donde en general la proporción de personas con

este mal es muy baja. Por ello diversas ligaciones asimétricas han sido propuestas en la literatura en los últimos 30 años.

Como ha sido presentado en Bazán y Millones (2008), el estudio de la regresión binaria es una área importante para el modelamiento estadístico y por tanto es conveniente tener modelos alternativos, usando enlaces asimétricos, que puedan ser implementados fácilmente y ofrezcan alternativas a los modelos que emplean enlaces simétricos disponibles en la mayoría de software estadísticos.

BRMUW (Bayesian Regression Model using WinBUGS) es un software para generar sintaxis de varios modelos de Regresión Binaria bajo un enfoque bayesiano usando MCMC que pueden ser ejecutadas en los programas OpenBUGS, WinBUGS o en R a través de las librerías R2WinBUGS o Brugs. BRMUW está pensado en usuarios aplicados que dada una base de datos desean conocer la sintaxis de diversos modelos de Regresión binaria usualmente no disponibles en diversos programas estadísticos incluyendo el programa R.

Este aplicativo básico ha sido desarrollado como parte de los proyectos de la antigua Dirección Académica de Investigación, proyectos DAI 3412, 4031 y 2009-0033, de la Pontificia Universidad Católica del Perú. Actualmente estamos trabajando en un proyecto que implementa estos mismos modelos y otros adicionales en un aplicativo denominado BAYES@PUCP.

El aplicativo junto a algunos modelos disponibles ha sido desarrollado por 3 años y a lo largo de ese tiempo diversas personas han colaborado en el proyecto por lo que deseo expresar mis agradecimientos. Entre ellos a los colegas Oscar Millones y Cristian Bayes por su ayuda durante la revisión de la actual versión del aplicativo y el manual. También deseo agradecer a Adrián Paucar, Enver Tarazona, Luisa Ames, Margareht Sequeiros y Pedro Curich por el apoyo informático en alguna de las etapas del proyecto.

En este documento preparado para un minicurso de la I Jornada Internacional de Probabilidad y Estadística JIPE 2010 puede ser considerado un manual técnico del BRMUW para implementar diversos modelos

de regresión binaria.

El resto del documento está organizado de la siguiente manera. En el capítulo 2 presentaremos una revisión de la Inferencia Bayesiana en Regresión. En el capítulo 3 se presenta la Inferencia Bayesiana en Regresión Binaria. En el capítulo 4 se presenta otros tópicos de Inferencia Bayesiana como las condiciones para la existencia de la distribución posterior de los parámetros de la regresión y se presentan diversas alternativas para la comparación de modelos desde la perspectiva bayesiana. En el capítulo 5 se muestran dos aplicaciones, una en la que se emplea la regresión bayesiana logística para calcular la probabilidad de falla de un transbordador espacial. La segunda aplicación corresponde al uso de diferentes modelos de regresión binaria para predecir la probabilidad de erradicar el cultivo de coca entre agricultores peruanos. Finalmente en Anexo presentamos un tutorial de uso del BRMUW incluyendo su uso en combinación con WinBUGS y R.

El BRMUW puede ser requerido al primer autor al correo jlbazan@pucp.edu.pe.

2. Inferencia Bayesiana en el Modelo Lineal

En este capítulo abordamos la inferencia bayesiana para el modelo de regresión lineal que es parte de los modelos lineales y repasamos brevemente diferentes aspectos de la inferencia bayesiana.

2.1. Función de Verosimilitud

Sea \mathbf{Y} un vector aleatorio cuya distribución dependa de un parámetro θ . Definimos la función de verosimilitud L para un vector de datos observados \mathbf{y} de \mathbf{Y} como

$$L(\mathbf{y}|\theta) = f(\mathbf{y} | \theta)$$

La verosimilitud es la probabilidad que usted pueda encontrar el valor observado dado el modelo.

Considerando la función de verosimilitud en la inferencia estadística se está interesado en estimar el parámetro del modelo. En la inferencia clásica, un estimador de *máxima verosimilitud* para el parámetro θ es el valor $\hat{\theta}$ para el cual la función de verosimilitud $L(\theta|\mathbf{y})$ es máxima. La solución de Inferencia Clásica consiste en maximizar la función L y obtener la solución correspondiente $\hat{\theta}$. A menudo es más conveniente maximizar el logaritmo natural de la función de verosimilitud

$$\ell(\mathbf{y}|\theta) = \log L(\mathbf{y}|\theta).$$

Cuando la solución de máxima verosimilitud no puede obtenerse analíticamente se recurre a métodos numéricos. Para mayores detalles de los diferentes métodos numéricos puede revisarse por ejemplo Casella y Berger (2002).

2.2. Inferencia Bayesiana

En la Inferencia Bayesiana hay importantes diferencias con la Inferencia Clásica (de Máxima Verosimilitud). Para una revisión rápida

puede verse Casella y Berger (2002) y para una revisión más detallada puede verse Robert (2002).

A manera de resumen consideremos el vector aleatorio $\mathbf{Y}|\theta$ donde θ es un parámetro que caracteriza a la distribución de \mathbf{Y} .

En la inferencia Bayesiana

1. θ es un vector aleatorio y no un número, por lo tanto tiene una distribución de probabilidades asociada.
2. Es posible usar información preliminar acerca de θ , la cual se puede sintetizar proponiendo una distribución “a priori” para $\theta : f(\theta)$.
3. Los datos se organizan en la función de verosimilitud: $L(\mathbf{y}|\theta)$.
4. Usando el teorema de Bayes es posible obtener la distribución “a posteriori” de θ dados los datos:

$$f(\theta|\mathbf{y}) = \frac{f(\theta, \mathbf{y})}{f(\mathbf{y})} = \frac{L(\mathbf{y}|\theta)f(\theta)}{f(\mathbf{y})}$$

Nótese que la distribución a posteriori es proporcional a la verosimilitud y la priori; es decir,

$$f(\theta|\mathbf{y}) \propto L(\mathbf{y}|\theta) \times f(\theta)$$

desde que la distribución marginal o no condicional de \mathbf{Y} , $f(\mathbf{y})$, no depende del parámetro θ .

Observación:

- Si tomamos logaritmo a la expresión anterior tenemos

$$\log f(\theta|\mathbf{y}) \approx \ell(\mathbf{y}|\theta) + \log f(\theta).$$

Por tanto, si se considera una priori no informativa para θ ; es decir,

$$f(\theta) = c$$

maximizar $\log f(\theta|\mathbf{y})$ para θ resulta equivalente a maximizar $\ell(\mathbf{y}|\theta)$. En otras palabras, la estimación de Máxima Verosimilitud (la que maximiza $\ell(\mathbf{y}|\theta)$) se puede considerar como un caso particular de inferencia Bayesiana sin información a priori. Naturalmente cuando $f(\theta) \neq c$, la maximización de la posteriori será distinta a la de máxima verosimilitud.

- Una vez identificada la distribución a posteriori para θ , $f(\theta|\mathbf{y})$, es posible realizar la inferencia requiriendo para esta distribución las medidas que se deseen como por ejemplo:

$$E(\theta|\mathbf{y}), V(\theta|\mathbf{y}), Med(\theta|\mathbf{y}), Q_\gamma(\theta|\mathbf{y})$$

que corresponden a la media, varianza, media y cuantil γ de la distribución a posteriori de θ dado $\mathbf{Y} = \mathbf{y}$, respectivamente. Así por ejemplo se puede definir un intervalo para $\theta|\mathbf{y}$ dado de esta manera

$$P(A \leq \theta|\mathbf{y} \leq B) = 1 - \gamma$$

donde

$$A = Q_{\frac{\gamma}{2}}(\theta|\mathbf{y}), B = Q_{1-\frac{\gamma}{2}}(\theta|\mathbf{y})$$

son los cuantiles correspondientes.

Note que en este caso A y B son valores de la distribución a posteriori de θ y por tanto el intervalo corresponde a un intervalo de probabilidad y no a un nivel de confianza como en la inferencia clásica. Por esta razón a este intervalo se le conoce como un *intervalo de credibilidad*.

2.3. Inferencia Bayesiana en el Modelo de Regresión Lineal

En esta sección estudiaremos como realizar inferencia bayesiana en el modelo de regresión lineal normal

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \epsilon$$

en donde

- \mathbf{y} es un vector n -dimensional,
- \mathbf{X} es una matriz de diseño $n \times k$,
- $\boldsymbol{\beta}$ es un vector k -dimensional de parámetros, y
- $\boldsymbol{\epsilon}$ es un vector n -dimensional de errores aleatorios

consideraremos que

$$\boldsymbol{\epsilon} \sim N(0, \sigma^2 I)$$

donde I es una matriz identidad $n \times n$.

Así, en este modelo la distribución de \mathbf{y} condicional a los parámetros $\boldsymbol{\beta}$ y σ^2 es

$$\mathbf{y} \mid \boldsymbol{\beta}, \sigma^2 \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 I).$$

Luego, la verosimilitud es dada por

$$L(\boldsymbol{\beta}, \sigma^2) = (2\sigma^2)^{-n/2} \exp \left[-\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right]$$

Entonces podemos reescribir la forma cuadrática

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} - \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{y} - \mathbf{y}^T \mathbf{X} \boldsymbol{\beta} + \mathbf{y}^T \mathbf{y}$$

Ahora si consideramos que $\mathbf{X}^T \mathbf{X}$ es una matriz no singular, podemos reescribir la expresión anterior como

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T \mathbf{X}^T \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + S$$

donde

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

el estimador de máxima verosimilitud de $\boldsymbol{\beta}$ y

$$S = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

es la suma de cuadrados residual, que se relaciona con el estimador de la varianza

$$\widehat{\sigma^2} = \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n - k} = \frac{S}{n - k}$$

De este modo

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T \mathbf{X}^T \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + (n - k)\widehat{\sigma^2}$$

2.3.1. Considerando distribuciones a priori no informativas

Si consideramos las siguientes prioris no informativas para los parámetros

$$\begin{aligned} f(\boldsymbol{\beta}) &\propto 1 \\ f(\sigma^2) &\propto \frac{1}{\sigma^2} \end{aligned} ,$$

la distribución a posteriori conjunta es dada por

$$f(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}) \propto (\sigma^2)^{-\left(\frac{n}{2}-1\right)} \exp \left[-\frac{1}{\sigma^2} \left((\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T \mathbf{X}^T \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + (n - k)\widehat{\sigma^2} \right) \right]$$

Entonces para encontrar la distribución marginal de $\boldsymbol{\beta}$ necesitamos integrar en σ^2 . Para simplificar los cálculos consideramos el siguiente cambio de variable

$$s = \sigma^{-2} \Rightarrow \sigma^2 = s^{-1}$$

$$J = \left\| \frac{\partial \sigma^2}{\partial s} \right\| = \left\| \frac{\partial}{\partial s} s^{-1} \right\| = s^{-2}$$

Entonces

$$\begin{aligned} f(\boldsymbol{\beta}, s | \mathbf{y}) &\propto s^{\frac{n}{2}+1} \exp \left[(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T \mathbf{X}^T \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + (n - k)\widehat{\sigma^2} \right] s^{-2} \\ &\propto s^{\frac{n}{2}-1} \exp \left[(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T \mathbf{X}^T \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + (n - k)\widehat{\sigma^2} \right] \end{aligned}$$

Integrando en s obtenemos la distribución marginal

$$f(\boldsymbol{\beta} | \mathbf{y}) \propto \int_0^\infty s^{\frac{n}{2}-1} \exp \left[(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})^T \mathbf{X}^T \mathbf{X} (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) + (n-k)\widehat{\sigma}^2 \right] ds$$

de donde obtenemos que

$$f(\boldsymbol{\beta} | \mathbf{y}) \propto \left[1 + \frac{1}{n-k} (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})^T \frac{1}{\widehat{\sigma}^2} \mathbf{X}^T \mathbf{X} (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) + (n-k)\widehat{\sigma}^2 \right]^{\frac{n}{2}}$$

que es el núcleo de una distribución t -Student multivariada tal que

$$E(\boldsymbol{\beta}) = \widehat{\boldsymbol{\beta}}$$

$$Var(\boldsymbol{\beta}) = \frac{(n-k)\widehat{\sigma}^2(\mathbf{X}^T \mathbf{X})^{-1}}{n-k-2}$$

y integrando en $\boldsymbol{\beta}$ obtenemos que la distribución marginal de σ^2 es una gama inversa con parámetros $\frac{1}{2}(n-k) - 1$ y $\frac{1}{2}(n-k)\widehat{\sigma}^2$.

2.3.2. Considerando distribuciones a priori informativas

Si consideramos una distribución a priori conjugada

$$\boldsymbol{\beta} | \sigma^2 \sim N(m, \sigma^2 V)$$

$$\sigma^2 \sim \text{Gamma} - \text{Inv}(a, d)$$

entonces, después de algunas manipulaciones, la distribución a posteriori de $\boldsymbol{\beta}$ y σ^2 es dada por

$$\boldsymbol{\beta} | \sigma^2 \sim N(m^*, \sigma^2 V^*)$$

$$\sigma^2 \sim \text{Gamma} - \text{Inv}(a^*, d^*)$$

donde

$$\begin{aligned}
 m^* &= (V^{-1} + \mathbf{X}^T \mathbf{X})^{-1} (V^{-1} m + \mathbf{X}^T \mathbf{y}) \\
 V^* &= (V^{-1} + \mathbf{X}^T \mathbf{X})^{-1} \\
 a^* &= a + m^T V^{-1} m + m^{*T} V^{*-1} m^* \\
 d^* &= d + n
 \end{aligned}$$

Debemos notar que podemos escribir

$$m^* = (I - A)m + A\hat{\boldsymbol{\beta}}$$

donde $A = (V^{-1} + \mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X}$. Entonces la media a posteriori m^* es una media ponderada entre m la media a priori y $\hat{\boldsymbol{\beta}}$ el estimador clásico por mínimos cuadrados.

2.4. Métodos de MCMC

Como hemos revisado hasta ahora en este capítulo, la inferencia Bayesiana esta basada en el análisis de la distribución a posteriori porque está contiene toda la información sobre el parámetro a ser estimado θ condicional a los datos observados \mathbf{y} . En general en inferencia Bayesiana es útil resumir la información que está contenida en la distribución a posteriori, y estos resúmenes típicamente toman la forma de esperanzas de funciones particulares de los parámetros, esto es,

$$I = E[g(\theta)] = \int g(\theta) f(\theta | \mathbf{y}) d\theta. \quad (1)$$

Así, el problema general de inferencia bayesiana consiste en calcular estos valores esperados según la distribución posteriori de θ . Algunos ejemplos son

1. Si $g(\theta) = \theta$ entonces obtenemos $E[\theta | \mathbf{y}]$ la media a posteriori, que generalmente es utilizada como estimador puntual de θ
2. Si $g(\theta) = f(x | \theta)$ donde x es independiente de \mathbf{y} . En estas condiciones obtenemos $E[f(x | \mathbf{y})]$ es la distribución predictiva de x una observación futura.

El problema en calcular este tipo de esperanzas, es que usualmente muy complicado o imposible evaluar I en forma analítica. Incluso las técnicas numéricas de cuadratura u otras para aproximar podrían presentar problemas, más aún si el parámetro θ es multidimensional. Por este motivo se hace necesario utilizar métodos aproximados para obtener estas integrales.

En estos últimos años la inferencia bayesiana ha experimentado un gran avance debido a la introducción de técnicas de simulación que permiten en forma relativamente simple obtener muestra de una distribución objetivo. En particular, los métodos de conocidos como Cadenas de Markov de Monte Carlo (MCMC) que son actualmente muy utilizados. La racionalidad de estos métodos subyace en diseñar iterativamente una cadena de Markov para θ de tal manera que $f(\theta|\mathbf{y})$ sea su distribución ergódica estacionaria. Empezando en algún estado inicial θ_0 la idea es simular un número suficientemente grande M de transiciones bajo la cadena de Markov y registrar los correspondientes estados simulados θ_j . Luego, bajo ciertas condiciones de regularidad, es posible mostrar que la media muestral ergódica

$$\hat{I} = \frac{1}{M} \sum_{j=1}^M g(\theta_j) \quad (2)$$

converge a la integral I deseada. En otras palabras, \hat{I} nos provee de una buena aproximación para I .

El reto de los métodos MCMC consiste entonces en precisar una cadena de Markov adecuada con la distribución a posteriori $f(\theta|\mathbf{y})$ como su distribución estacionaria y decidir cuando detener la simulación. Una excelente introducción a los procesos de Markov y al teorema ergódico puede encontrarse en Ross (1995). Para un enfoque más formal en relación a la inferencia Bayesiana puede consultarse Tierney(1994).

Describamos ahora uno de los métodos MCMC más populares conocido como el muestreador de Gibbs. Este método fue propuesto por Geman y Geman (1984) y difundido por Gelfand y Smith (1990). El siguiente ejemplo nos ilustra su aplicación.

Ejemplo 2.1. (Gelfand y Smith, 1990) Consideremos un modelo de análisis de varianza de efectos aleatorios, que es un caso particular de modelos lineales:

$$y_{ij} = \tau_i + \epsilon_{ij}, i = 1, 2, \dots, k; j = 1, 2, \dots, n,$$

donde los errores $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$ se asumen independientes y los efectos aleatorios $\tau_i \sim N(\mu, \sigma_\tau^2)$. Si se asumen también independientes entre sí y con los errores. Si asumimos las siguientes prioris inversas Gaussianas y normales del tipo:

- $\sigma_\tau^2 \sim IG(a_1, b_1)$
- $\mu | \sigma_\tau^2 \sim N(\mu_0, \sigma_\tau^2)$
- $\sigma_\epsilon^2 \sim IG(a_2, b_2)$

se puede mostrar que la distribuciones a posteriori de $\sigma_\tau^2 | \mathbf{y}, \mu, \tau, \sigma_\epsilon^2$ y $\sigma_\epsilon^2 | \mathbf{y}, \mu, \tau, \sigma_\tau^2$ son inversa Gamma y las distribuciones a posteriori de $\mu | \mathbf{y}, \tau, \sigma_\tau^2$ y $\tau | \mathbf{y}, \mu, \sigma_\tau^2, \sigma_\epsilon^2$ son normales, donde $\mathbf{y} = (y_{ij}; i = 1, 2, \dots, k, j = 1, 2, \dots, n)$ denota al vector de la data y τ es el vector de efectos de los k tratamientos.

Para estimar los momentos a posteriori del tipo $I = \int g(\theta) f(\theta | \mathbf{y}) d\theta$ definiremos una cadena de Markov para el parámetro $\theta = (\mu, \tau, \sigma_\tau^2, \sigma_\epsilon^2)$. Denotaremos por $\theta_n = (\mu_n, \tau_n, \sigma_{\tau,n}^2, \sigma_{\epsilon,n}^2)$ al vector de estados de la cadena en la n -ésima iteración. Dada la naturaleza de una cadena de Markov, todo lo que necesitamos es definir las probabilidad condicionales de transición de la cadena entre las iteraciones n y $n + 1$. Haremos esto, muestreando de la distribución condicional a posteriori completa para μ, τ, σ_τ^2 y σ_ϵ^2 dada la data a través de los siguientes pasos:

1. $\mu_{n+1} \sim \mu | \mathbf{y}, \tau_n, \sigma_{\tau,n}^2, \sigma_{\epsilon,n}^2,$
2. $\tau_{n+1} \sim \mu | \mathbf{y}, \mu_{n+1}, \sigma_{\tau,n}^2, \sigma_{\epsilon,n}^2,$
3. $\sigma_{\epsilon,n+1}^2 \sim \sigma_\epsilon | \mathbf{y}, \mu_{n+1}, \tau_{n+1}, \sigma_{\tau,n}^2,$

$$4. \sigma_{\tau, n+1}^2 \sim \sigma_\epsilon | \mathbf{y}, \mu_{n+1}, \tau_{n+1}, \sigma_{\epsilon, n+1}^2,$$

Los pasos 1 al 4 definen una cadena de Markov $\{\theta_n\}$ cuya distribución converge a la deseada $f(\mu, \tau, \sigma_\epsilon, \sigma_\tau | \mathbf{y})$. Los promedios ergódicos del tipo $\hat{I} = \frac{1}{M} \sum_{j=1}^M g(\theta_j)$ proveen luego de una evaluación numérica de la integral a posteriori $I = \int g(\theta) f(\theta | \mathbf{y}) d\theta$.

El ejemplo descrito es un caso particular del muestreador de Gibbs. En general dado el parámetro $\theta = (\theta_1, \dots, \theta_p)$, el muestreador de Gibbs trabaja en forma iterativa. Para cada $j = 1, 2, \dots, p$ genera las distribuciones condicionales a posteriori de

$$\theta_{j, n+1} \sim \theta_j | \mathbf{y}, \theta_{1, n+1}, \dots, \theta_{j-1, n+1}, \theta_{j+1, n}, \dots, \theta_{p, n}. \quad (3)$$

El método de Gibbs debe su popularidad al hecho de que en muchos modelos estadísticos la distribución condicional a posteriori completa $f(\theta_j | \mathbf{y}, \theta_k, k \neq j)$ es posible de simular. Ocurren, sin embargo, casos en donde esto no es posible y por ello se hace necesarios contar con otros métodos MCMC alternativos. Posiblemente el más genérico de estos esquemas es el de Metropolis. Para generar la distribución a posteriori, este método define una cadena de Markov en el que una transición sigue los pasos siguientes:

1. Se genera un valor de $\tilde{\theta}$ a partir de alguna distribución $h(\tilde{\theta} | \theta)$ propuesta que la detallaremos más adelante.
2. Se calcula

$$a(\theta, \tilde{\theta}) = \min \left\{ 1, \frac{f(\tilde{\theta} | \mathbf{y})}{f(\theta | \mathbf{y})} \cdot \frac{h(\theta | \tilde{\theta})}{h(\tilde{\theta} | \theta)} \right\}$$

3. Se reemplaza θ por $\tilde{\theta}$ con probabilidad a y en caso contrario se mantiene igual.

La selección de la distribución propuesta h es esencialmente arbitraria sujeta a ciertas restricciones técnicas. Utilizándose por ejemplo una distribución simétrica con $h(\tilde{\theta} | \theta) = h(\theta | \tilde{\theta})$ como por decir la normal

centrada en θ se tiene la ventaja práctica de que el radio $\frac{h(\theta|\tilde{\theta})}{h(\tilde{\theta}|\theta)}$ se cancela en a . Otra variante práctica de interés es el uso de distribuciones propuestas independientes $h(\tilde{\theta})$. Tierney (1994) refiere a estos algoritmos como cadenas independientes. Hasting (1970) propone una larga clase de algoritmos similares basados en una expresión más general para la probabilidad de aceptación a .

Las cadenas de Markov que son utilizados en los esquemas MCMC poseen generalmente un espacio continuo de estados. Tierney (1994) muestra que estos algoritmos convergen a una distribución ergódica estacionaria $\pi(\theta) = f(\theta|\mathbf{y})$ sujeta a tres condiciones de regularidad: irreducibilidad, aperiodicidad e invarianza. La noción de irreducibilidad manifiesta que para cualquier estado θ y cualquier conjunto de estados B con $\pi(B) > 0$, existe $n \in \mathbb{N}$ tal que al cabo de n iteraciones la cadena pueda hacer una transición de θ a B con probabilidad positiva. La invarianza se refiere por otro lado, a la propiedad de que si empezamos con un vector de estados generado por π , entonces futuras transiciones en la cadena dejarán la distribución marginal de θ inalterada; es decir, $\theta_n \sim \pi$, para cualquier $n \in \mathbb{N}^+$.

El muestreador de Gibbs y el esquema de Metropolis-Hastings son por construcción invariantes con respecto a la distribución a posteriori buscada. Lo que uno debe de verificar entonces son la aperiodicidad e irreducibilidad de la cadena, siendo esta última la más crítica pues en ocasiones es posible encontrar un subconjunto de estados tales que cuando la cadena simulada entre en ella sea improbable salir y el algoritmo por tanto se entrampe en ese punto sin llegar a converger. En la práctica más importante que establecer convergencias teóricas es reconocer la convergencia práctica; es decir, juzgar cuantas transiciones M debe de ser suficientes como para obtener promedios ergódicos \hat{I} que esten cerca de (1). El procedimiento más simple radica en graficar las trayectorias θ_n contra el número de iteraciones n y juzgar por inspección que la convergencia se da de no presentarse tendencia alguna obvia.

Algunas referencias adicionales de métodos MCMC en inferencia bayesiana son Chen, Shao, Ibrahim (2000), Gamerman y Freitas (2006).

3. Inferencia bayesiana en Regresión binaria

En este capítulo abordamos la inferencia bayesiana de regresión binaria así como presentamos diferentes modelos basados en diferentes enlaces simétricos y asimétricos implementados en BRMUW.

3.1. Modelo de Regresión Binaria

Considere un modelo de regresión binaria

$$Y_i \sim \text{Bernoulli}(p_i) \\ p_i = F(\mathbf{x}_i^T \boldsymbol{\beta})$$

donde

- Y_i una variable binaria tal que $Y_i = 1$ ocurre con probabilidad p_i .
- $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ik})^T$ un vector con los valores de k variables explicativas.
- $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_k)^T$ un vector de k coeficientes de regresión.
- $F(\cdot)$ denota una función de distribución acumulada (fda). La función inversa $F^{-1}(\cdot)$ es comúnmente denominada función de enlace.
- $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$ es el i -ésimo predictor lineal.

Cuando F es una fda de una distribución simétrica la función de enlace resultante es simétrica y tiene una forma simétrica alrededor de $p_i = 0,5$. En el caso que F sea la fda de una normal estándar tenemos el enlace probit,

$$F(t) = \Phi(t)$$

y en el caso de F sea la fda de una distribución logística obtenemos el enlace logit,

$$F(t) = \frac{e^t}{1 + e^t}.$$

3.2. Enlaces Asimétricos en Regresión Binaria

Chen et al (1999) sostienen que cuando la probabilidad de una respuesta binaria se aproxima a 0 en una tasa diferente que cuando se aproxima a 1, los enlaces simétricos para el ajuste de datos pueden ser inadecuados. En este caso, hay que considerar enlaces asimétricos. En este caso se considera la fda de distribuciones asimétricas para construir enlaces asimétricos. Un ejemplo muy popular es el enlace log-log complementario o cloglog, donde la fda usada en el enlace corresponde a la Distribución de Gumbel.

En este caso, la fda está completamente especificada, no depende de ningún parámetro adicional desconocido y no presenta como caso particular un enlace simétrico.

Información de como implementar la estimación bayesiana de la regresión binaria usando los enlaces cloglog, probit y logit en WinBUGS u OpenBUGS puede ser vista en el Ejemplo Beetles:logistic,probit and extreme value models del Manual. Sin embargo la regresión binaria bayesiana considerando otros enlaces como los discutidos en Bazán, Bolfarine y Branco (2006 y 2010) actualmente no se encuentran disponibles. Otros enlaces considerados en el BRMUW es aquel que se obtiene considerando otras funciones de distribución acumuladas como las siguientes:

$$F(t) = 1 - (1 + e^t)^{-\lambda} \text{ y } F(t) = (1 + e^{-t})^{-\lambda} \quad \lambda > 0$$

estos enlaces son logit asimetrizados y son conocidos como scobit y power logit, respectivamente, e incluyen al enlace logit como caso especial cuando el parámetro. Para una revisión de estos enlaces ver Prentice (1976) y Nagler (1994).

En BRMUW también son implementados tres enlaces que se basan en la fda de una distribución normal asimétrica, esta acumulada puede ser representada de manera general la siguiente manera:

$$F(t; \boldsymbol{\theta}) = 2\Phi_2(\boldsymbol{x} \mid \boldsymbol{\mu}, \boldsymbol{\Omega})$$

donde

- $\mathbf{x} = (t, 0)'$;
- $\boldsymbol{\theta} = (\mu, \sigma^2, \lambda)'$;
- $\Phi_2(\cdot)$ representa la distribución acumulada de una distribución normal bivariada con parámetros $\boldsymbol{\mu} = (\mu, 0)'$ y $\boldsymbol{\Omega} = \begin{bmatrix} \sigma^2 & -\delta \\ -\delta & 1 \end{bmatrix}$; y
- $\delta = \frac{\lambda}{\sqrt{1 + \lambda^2}}$,

para más detalles en este tipo de enlaces ver Bazán, Bolfarine y Branco (2006 y 2010).

Luego, los enlaces implementados en el BRMUW son los siguientes:

- Si $\boldsymbol{\theta} = (0, 1 + \lambda^2, -\lambda)$, se obtiene el enlace probit asimetrizado propuesto en Chen et al (1999) denominado aquí CDS skew probit.
- Si $\boldsymbol{\theta} = (0, 1, \lambda)$, se obtiene el enlace propuesto por Bazán, Branco y Bolfarine (2006) denominado aquí BBB skew probit.
- Si $\boldsymbol{\theta} = \left(-\frac{\sqrt{2}\delta}{\sqrt{\pi - 2\delta^2}}, \frac{\pi}{\pi - 2\delta^2}, \lambda \right)$, obtenemos el enlace denominado estándar probit asimetrizado (Bazán, Bolfarine y Branco 2006 y 2010), denominado aquí Standard skew probit.

En estos tres enlaces, λ es el parámetro que controla la asimetría, así tenemos que para valores negativos (positivos) de λ tenemos asimetría negativa (positiva).

Esta clase de modelos puede verse también como perteneciente a la clase de mezclas de distribuciones elípticas propuestas por Basu y Mukhopadhyay (2000) dada por:

$$F = \left\{ F(\cdot) = \int_{[0, \infty]} H(\cdot | v) dG(v) \right\}$$

, donde G es la función de distribución acumulada $[0, \infty >$ y H es una distribución elíptica. En este caso el CDS skew probit considera una clase

de mezclas de normales donde la medida de mezcla es la distribución normal positiva con función de densidad dada por $g(x) = 2\phi(x)$, $x > 0$, con $\phi(\cdot)$ siendo la función de densidad de la normal estándar. Otro caso interesante cuando se mezcla la normal positiva con H la función de distribución acumulada de la distribución logística es conocida como skew logistic o skew logit (ver Chen, Dey y Shao, 2001) que también está implementado en BRMUW.

En resumen, los modelos de regresión binaria implementados en BRMUW clasificados según sus enlaces son

- simétricos: probit, logit.
- Asimétricos: cloglog, scobit, power logit, skew logit, skew probit (CDS, BBB y standard).

3.3. Inferencia Bayesiana

Considerando la distribución Bernoulli para la variable respuesta, la función de verosimilitud es dada por

$$L(\beta, \theta, \gamma) \prod_{i=1}^n [F_{\theta}(m(x_i^T \beta, \gamma))]^{y_i} [1 - F_{\theta}(m(x_i^T \beta, \gamma))]^{1-y_i}$$

donde $F_{\theta}(m(x_i^T \beta, \gamma))$ es la función de densidad de una distribución que puede ser indexada por el parámetro θ que no necesariamente es unidimensional y $m(\cdot)$ es una función continua del predictor lineal $x_i^T \beta$ que también incluye la función identidad con γ como un parámetro de forma. Para los enlaces logit, probit, cloglog, scobit y power logit se ha considerado esta función de verosimilitud, para los enlaces skew probit y skew logit se ha considerado otras versiones de la función de verosimilitud considerando versiones aumentadas que son discutidas en las referencias específicas de estos modelos.

En la Inferencia bayesiana, a diferencia de la inferencia clásica, los parámetros de interés β, θ, γ se asumen como variables aleatorias y así se

establecen diferentes distribuciones de probabilidad a priori que reflejan nuestro conocimiento previo de su conducta. Combinando la función de verosimilitud y las distribuciones a priori podemos obtener la distribución posterior de los parámetros de interés.

Estos parámetros tienen significados diferentes. Los parámetros θ y γ están asociados con el enlace, y el parámetro β corresponde a los datos observados y no depende del modelo escogido.

En nuestro trabajo, consideramos prioris vagas (prioris propias con distribuciones conocidas con varianza grande)

Asumimos independencia entre las prioris, esto es:

$$f(\beta, \theta, \gamma) = f(\beta)f(\theta)f(\gamma)$$

Usamos prioris para β comunes en la literatura incluyendo prioris normales. Especificaciones para $f(\theta)$ y $f(\gamma)$ dependen de la elección particular del modelo considerando un intervalo de variación. En muchas situaciones esos intervalos son determinados de acuerdo a la literatura.

Una vez especificada la distribución a priori tenemos que la posteriori viene dada por:

$$f(\beta, \theta, \gamma | \mathbf{y}) = L(\beta)f(\beta)f(\theta)f(\gamma).$$

También es posible obtener versiones aumentadas introduciendo variables auxiliares. La inferencia (Bayesiana) para los modelos de regresión binaria, es especialmente para los modelos citados antes, y puede ser facilitada por la simulación MCMC implementada en el programa WinBUGS. Usando una programación mínima es posible implementar todos los modelos presentados. Para una revisión de los métodos MCMC revisar Chen, Shao y Ibrahim, J. G (2000) y Gamerman y Lopes (2006).

En la siguiente sección será presentada la estimación bayesiana del modelo de regresión binaria considerando como función de enlace la función logit y daremos la sintaxis generada en BRMUW.

3.4. Inferencia Bayesiana en el Modelo de Regresión Logística

Considere un modelo de regresión binaria

$$Y_i \sim \text{Bernoulli}(p_i) \\ p_i = F(\mathbf{x}_i^T \boldsymbol{\beta})$$

donde

$$F(t) = \frac{e^t}{1 + e^t}$$

entonces tenemos que la función de verosimilitud será dada por

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n \left(\frac{e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right)^{y_i} \left(\frac{1}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right)^{1-y_i} \quad (4)$$

y si consideramos que a priori $\boldsymbol{\beta}$ sigue una distribución normal multivariada con vector de medias $\mathbf{0}$ y matriz de covarianza $\boldsymbol{\Sigma}$,

$$f(\boldsymbol{\beta}) = \left(\frac{1}{2\pi} \right)^{n/2} e^{\boldsymbol{\beta}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\beta}}$$

Luego la distribución a posteriori viene dada por

$$f(\boldsymbol{\beta} | \mathbf{y}) \propto \frac{e^{\sum_{i=1}^n y_i \mathbf{x}_i^T \boldsymbol{\beta}}}{\prod_{i=1}^n (1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}})} e^{\boldsymbol{\beta}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\beta}}$$

Como podemos observar en esta caso la distribución a posteriori no presenta una forma conocida por lo que se hace necesario utilizar métodos MCMC como los implementados en los programas WinBUGS y OpenBUGS por lo que usando una programación mínima es posible implementar estos métodos. En este caso, la programación en WinBUGS u OPENBUGS para $N = 200$ sujetos y dos variables predictoras ($k = 3$ coeficientes de regresión) queda dado por:

```
model
{
  for(i in 1:N) {
    y[i] ~ dbern(p[i])
    logit(p[i]) <- m[i]
    m[i] <- beta[1]+beta[2]*x1[i]+beta[3]*x2[i]
  }
  for (j in 1:k) {beta[j] ~ dnorm(0.0,1.0E-3)}
}

Inits
list(beta=c(0,0,0,0,0))

Data
list(N=200,k=3)
```

Para una revisión de las sintaxis de otros modelos de regresión binaria tradicionales como el probit y cloglog, sugerimos el libro de Congdon (2005). Las diferentes sintaxis de los modelos de regresión binaria implementados en BRMUW pueden ser revisadas en el propio programa. En algunos casos se requiere una revisión adicional para la especificación de prioris en los modelos asimétricos que incluyen un parámetro diferente del coeficiente de regresión. En este caso, para cada modelo implementado la sintaxis del modelo indica las referencias pertinentes en la especificación de prioris del modelo.

4. Otros temas de Inferencia Bayesiana

En este capítulo estudiamos las condiciones de existencia de la posteriori para los parámetros de regresión β cuando se consideran diferentes prioris impropias. Adicionalmente se presentan diferentes indicadores para la comparación de modelos en la perspectiva bayesiana.

4.1. Condiciones de existencia de la posteriori para los parámetros de regresión

En el caso que consideremos una priori no informativa para β del tipo uniforme esto es

$$f(\beta) \propto 1$$

debemos comprobar la existencia de la distribución a posteriori. Estos aspectos han sido abordados por Chen y Shao (2000) y son reseñados a continuación.

Consideraremos el caso en que la función de enlace $F(\cdot)$ no está indexada por otros parámetros, en este caso la función de verosimilitud solo es una función de β y es dada por

$$L(\beta) = \prod_{i=1}^n [F(x_i^T \beta)]^{y_i} [1 - F(x_i^T \beta)]^{1-y_i}$$

Así tenemos que la distribución a posteriori de β es dada por

$$f(\beta | \mathbf{y}) \propto \prod_{i=1}^n [F(x_i^T \beta)]^{y_i} [1 - F(x_i^T \beta)]^{1-y_i}$$

Desde que una distribución a posteriori impropia hace que la inferencia bayesiana sea imposible, es importante que estudiemos si la distribución a posteriori $f(\beta | \mathbf{y})$ es propia. Debe quedar claro que esta

distribución será propia solamente si

$$\int_{\mathfrak{R}^k} \prod_{i=1}^n [F(x_i^T \beta)]^{y_i} [1 - F(x_i^T \beta)]^{1-y_i} < \infty$$

Para obtener condiciones necesarias y suficientes para que la posteriori de β sea propia, sea $z_i = -1$ si $y_i = 0$ y $z_i = 1$ si $y_i = 1$, \mathbf{X} es una matriz de diseño $n \times k$ con filas x_i^T y definamos \mathbf{X}^* una matriz con filas $z_i x_i^T$.

Teorema 4.1. *Asumiendo que las siguientes condiciones son verdaderas:*

(C1) *La matriz de diseño \mathbf{X} es de rango completo*

(C2) *Existe un vector positivo $a = (a_1, \dots, a_n)^T \in \mathfrak{R}^n$, esto es, que cada componente $a_i > 0$, tal que*

$$\mathbf{X}^{*T} a = 0$$

(C3) $\int_{-\infty}^{\infty} |u|^k dF(u) > \infty$

entonces

$$\int_{\mathfrak{R}^k} L(\beta) d\beta < \infty$$

El teorema 1 garantiza la existencia de la distribución a posteriori desde que las condiciones (C1)-(C3) sean satisfechas. El siguiente teorema extiende el teorema 2.1 para la existencia de momentos a posteriori.

Teorema 4.2. *Asumiendo que las condiciones (C1) y (C2) son verdaderas. Si $\int_{-\infty}^{\infty} |u|^{k+p} dF(u) > \infty$ para algún $p \geq 0$, entonces*

$$\int_{\mathfrak{R}^k} \|\beta\|^p L(\beta) d\beta < \infty$$

donde $\|\cdot\|$ denota la norma euclídeana

Los modelos logístico, probit y cloglog cumplen la condición (C3) del teorema 1 entonces solo se deben verificar las condiciones (C1) y (C2) referente a la matriz de diseño para saber si la posteriori β será propia.

Este tipo de estudio también ha sido realizado para distribuciones que tengan otros parámetros además de β , por ejemplo en Chen y Shao (1999) y recientemente para modelos que consideran enlaces asimétricos basados en la distribución normal asimétrica Bazán, Branco y Bolfarine (2006).

4.2. Comparación de Modelos

Como hemos visto podemos considerar diferentes modelos para un conjunto de datos binarios con solo cambiar la función de enlace, en esta sección revisaremos diferentes criterios para la comparación de modelos que nos ayudaran a decidir que modelo es más apropiado.

Existen una seria de metodologías para comparar modelos alternativos, entre los principales criterios para comparación de modelos en la inferencia bayesiana tenemos: (deviance information criterion) (DIC) propuesto por Spiegelhalter et al. (2002), el esperado del criterio de información de Akaike (EAIC) and y el esperado del criterio de información de Schwarz o Bayesiano (EBIC) estos dos últimos propuestos en Carlin and Louis (2000) y Brooks (2002). Estos criterios son basados en media a posteriori del *desvío* $E\left[D(\mathbf{a}, \mathbf{b}, \lambda, \theta)\right]$, donde

$$D(\mathbf{a}, \mathbf{b}, \lambda, \theta) = -2\ln(p(\mathbf{y}|\mathbf{a}, \mathbf{b}, \lambda, \theta)) = -2 \sum_{i=1}^n \ln P(Y_{ij} = y_{ij}|\mathbf{a}, \mathbf{b}, \lambda, \theta),$$

que es una medida de ajuste que puede ser aproximada utilizando la salida de la simulación MCMC de la distribución a posteriori, esta aproximación es dada por

$$Dbar = \frac{1}{G} \sum_{i=1}^G D(\mathbf{a}^g, \mathbf{b}^g, \lambda^g, \theta^g),$$

donde el índice g indica el g -ésimo valor simulado de un total de G simulaciones.

El EAIC, EBIC y DIC pueden ser estimados de la siguiente manera

$$\widehat{EAIC} = Dbar + 2p,$$

$$\widehat{EBIC} = Dbar + p \log N,$$

y

$$\widehat{DIC} = Dbar + \widehat{\rho}_D = 2Dbar - Dhat,$$

respectivamente donde p es el número de parámetros en el modelo, N es el total de observaciones y ρ_D es el *número efectivo de parámetros* y es definido como

$$\rho_D = E \left[D(\mathbf{a}, \mathbf{b}, \lambda, \theta) \right] - D \left[E(\mathbf{a}), E(\mathbf{b}), E(\lambda), E(\theta) \right],$$

donde $D \left[E(\mathbf{a}), E(\mathbf{b}), E(\lambda), E(\theta) \right]$ es el *desvío de la media a posteriori* obtenido cuando evaluamos la función desvío en la *média a posteriori* de los parámetros, el cual es estimado por

$$Dhat = D \left(\frac{1}{G} \sum_{i=1}^G \mathbf{a}^g, \sum_{i=1}^G \mathbf{b}^g, \frac{1}{G} \sum_{i=1}^G \lambda^g, \frac{1}{G} \sum_{i=1}^G \theta^g \right)$$

Para comparar dos o más modelos alternativos, el modelo que presente mejor ajuste al conjunto de datos será el modelo que presente el menor valor de DIC, EAIC y EBIC. En el EAIC y EBIC $2p$ y $p \log N$ son valores fijos que penalizan a la media a posteriori del *desvío*. Desde que, no existe consenso en la literatura de que criterio sea el mejor, el uso de más de un criterio parece ser apropiado para realizar comparación de modelos.

$Dbar$ y DIC son reportados en WinBUGS directamente cuando se requiere durante el proceso de simulación. $EAIC$ y $EBIC$ pueden ser derivados a partir del valor obtenido de $Dbar$ considerando las expresiones presentadas.

5. Aplicaciones

5.1. Aplicación 1: Datos del Challenger

El 28 de enero de 1986 el transbordador *Challenger*, tuvo una falla catastrófica debido a la falla de uno de los sellos especiales sellados con anillos de goma (O-rings), diseñados para prevenir el escape de combustible muy caliente producido durante la ignición. En total hay seis de tales anillos en cada nave (tres en cada uno de los dos cohetes).

El lanzamiento del *Challenger* fue el número 25 de un transbordador. Después de cada lanzamiento se inspeccionan estos sellos de goma. De los 24 lanzamientos previos, 6 tuvieron incidentes con daños en los anillos, 17 no tuvieron incidentes y uno no se pudo recuperar. Los datos provienen de Lavine, M. (1991).

Entonces tenemos una clásica respuesta binaria, hubo o no hubo daño en los anillos, que en este caso puede ser explicada por la temperatura ambiente en el momento del lanzamiento.

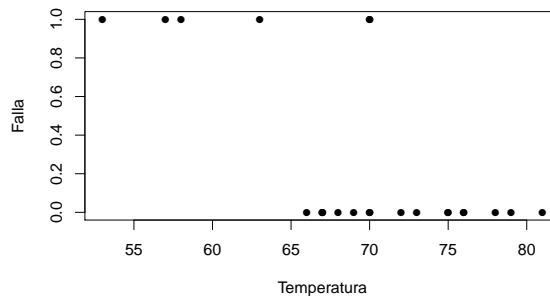


Figura 1: Datos del *Challenger*

La Figura 1 parece indicar que para temperaturas menores que 65 grados Fahrenheit hay una gran probabilidad de que ocurra una falla en los anillos de goma. A continuación veremos como utilizar un modelo de regresión

binaria para calcular la probabilidad de falla del transbordador.

Consideremos el siguiente modelo

$$Falla_i \sim \text{Bernoulli}(\pi_i)$$

$$\pi_i = F(\eta_i)$$

$$\eta_i = \beta_0 + \beta_1 \text{Temperatura}_i, \quad i = 1, 2, \dots, 23$$

donde para F consideraremos el enlace logístico. Utilizando el BRMUW para generar la sintaxis y el WinBUGS para simular de la distribución a posteriori (en anexo se tiene un tutorial sobre el uso del BRMUW y el WinBUGS) obtuvimos los siguiente resultados presentados en la tabla 1.

Tabla 1: Estimativas del los parámetros del modelo logístico para los datos del Challenger

Parámetro	Média a posteriori	Desv. Est.	2.5 %	mediana	97.5 %
β_0	31.86	13.2	9.711	30.8	61.02
β_1	-0.488	0.196	-0.9216	-0.4723	-0.1589

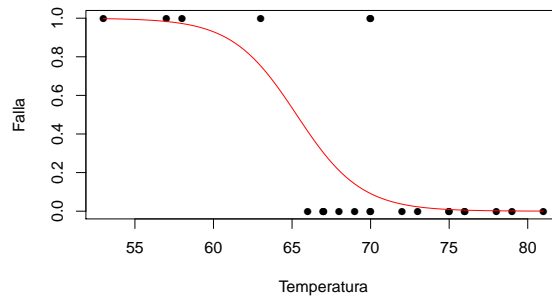


Figura 2: Datos del *Challenger* con curva estimada del modelo logístico

El día del lanzamiento del *Challenger* la temperatura era de 31 grados Fahrenheit (-0.5 grados centígrados). Si consideramos la media a posteriori como estimador de β la probabilidad de falla de los anillos de goma en el transbordador era de 0.9998.

5.2. Aplicación 2: Datos de erradicación de cultivos de coca

El programa BRMUW genera la sintaxis necesaria para la estimación bayesiana de varios modelos de regresión binaria, en el programa WinBUGS (ver Spiegelhalter et al 1996) o OpenBUGS (Spiegelhalter et al 2007), usando diversos métodos MCMC. Para ello solo es necesario contar con un archivo de texto con los datos, generado de cualquier programa estadístico o de Excel. En las columnas usualmente aparece los nombres de las variables en la primera línea y la primera columna deberá contener la variable respuesta.

Como ejemplo consideramos una data que contiene algunas variables de un estudio con agricultores beneficiarios de un programa favorable a la erradicación de cultivos de coca (Bazán y Millones, 2008). La data se denomina `concoca.txt` y se encuentra dentro de la descarga del programa. Las variables en `concoca.txt` son

<i>sierr</i>	si se muestra favorable a erradicar el cultivo de coca
<i>permedyc</i>	índice de percepción de que el cultivo de coca produce daño al medio ambiente
<i>partco</i>	índice de participación comunal
<i>concoca</i>	índice acerca de si consume coca
<i>pobrez</i>	niveles de pobreza

El archivo de datos tiene la siguiente estructura

sierr	permedyc	partco	concoca	pobrez
1	2	2	1	2
0	0	6	1	2
⋮	⋮	⋮	⋮	⋮
1	2	9	0	3

Como un ejemplo de aplicación consideremos el siguiente modelo

$$\begin{aligned} \text{sierr}_i &\sim \text{Bernoulli}(\pi_i) \\ \pi_i &= F(\eta_i) \end{aligned}$$

$$\eta_i = \beta_1 + \beta_2 \text{permedyc}_i + \beta_3 \text{partco}_i + \beta_4 \text{concoc}_i + \beta_5 \text{pobrez}_i, \quad i=1,2,\dots,1947$$

donde para F consideraremos diferentes enlaces. Utilizando el BRMUW para generar la sintaxis y el WinBUGS para simular de la distribución a posteriori (en la siguiente sección se tiene un tutorial sobre el uso del BRMUW y el WinBUGS) obtuvimos los resultados presentados en las tablas 2 y 3.

En este caso observamos que de acuerdo al DIC que todos los enlaces asimétricos con excepción del cloglog presentan mejor ajuste que los enlaces simétricos. El modelo que mejor se ajusta a los datos es aquel que considera el enlace *skew logit*. Mayores detalles sobre esta aplicación en Bazán y Millones (2008).

En las tablas 3 y 4 presentamos los estimadores de los parámetros de los modelos logístico y *skew logit*. Aunque los resultados de los coeficientes de regresión son similares en ambos modelos, la presencia del parámetro de forma $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$ que toma valores en el intervalo $[0, 1]$, indica que una adecuada interpretación corresponde al modelo *skew logit*.

De acuerdo a los resultados del modelo *skew logit*, podemos comentar que la variable *percepción de que la producción ilegal de hoja de coca daña al medio ambiente* (*permedyc*) impacta en forma positiva al incrementar la posibilidad de la erradicación de este cultivo. Similarmente la variable *participación de los agricultores en actividades de la comunidad* (*partco*) por ejemplo, la ejecución de tareas en la comunidad, reuniones de la comunidad, participación en proyectos con grupos de trabajo de la comunidad, influyen favorablemente en la erradicación. Dos resultados interesantes que hemos encontrado en este análisis es que la *pobreza* (*pobrez*) incrementa la probabilidad de erradicación de los cultivo ilícitos, así como a mayor sea el *consumo personal de hojas de coca* (*concoc*) menor es la probabilidad de esta erradicación.

Tabla 2: Comparación de los modelos de regresión binaria para los datos de erradicación de cultivos de coca

Enlaces	Modelos	Bur-in	Thin	Dbar	DIC
Simétricos	Probit	4000	5	2451.5	2456.8
	Logístico	4000	5	2450.9	2455.8
Asimétricos	Cloglog	4000	5	2451.6	2457
	Scobit	4000	25	2462.1	2441.2
	Power Logit	54000	100	2458.5	1794.1
	Skew logit	4000	25	2458.1	1708.4
	BBB sp	4000	35	2345.2	2252.5
	Standard sp	4000	15	1538.1	1751.7

Tabla 3: Estimativas del los parámetros del modelo logístico para los datos de erradicación de cultivos de coca

Parámetro	media	Desv. Est.	2.5 %	mediana	97.5 %
β_1	-2.84	1.08	-5.10	-2.70	-0.98
β_2	0.65	0.09	0.52	0.64	0.86
β_3	0.08	0.02	0.05	0.08	0.11
β_4	-0.21	0.05	-0.32	-0.21	-0.11
β_5	0.73	0.20	0.33	0.73	1.15
δ	0.14	0.59	-0.90	0.21	0.91

Tabla 4: Estimativas del los parámetros del modelo *skew logit* para los datos de erradicación de cultivos de coca

Parámetro	media	Desv. Est.	2.5 %	mediana	97.5 %
β_1	-2.80	0.55	-3.77	-2.80	-1.70
β_2	0.61	0.06	0.50	0.61	0.72
β_3	0.07	0.01	0.04	0.07	0.10
β_4	-0.20	0.05	-0.29	-0.20	-0.11
β_5	0.79	0.18	0.43	0.79	1.12

Inferencia Bayesiana en modelos de regresión binaria usando BRMUW

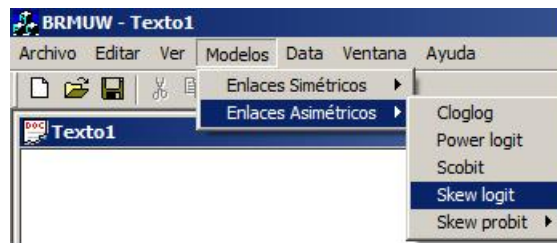
Adicionalmente, llevamos a cabo un análisis de la capacidad predictiva de ambos modelos siendo que el modelo logístico presenta un 64% de buena clasificación, en contraste con el 95% que se obtiene cuando se utiliza el modelo *skew logit*.

6. Uso del BRMUW

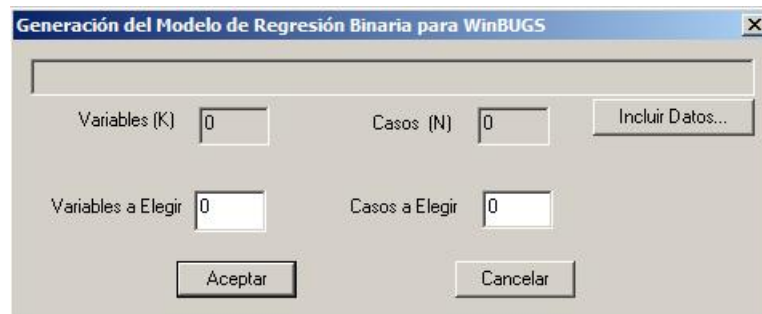
Vamos a usar el BRMUW para implementar el modelo de regresión binaria con enlace skew logit para los datos de `concoca.txt` descrito en la sección 5.2. Para mayores detalles de la aplicación revisar Bazán y Millones (2008) y el help del programa.

6.1. Generar la sintaxis del modelo

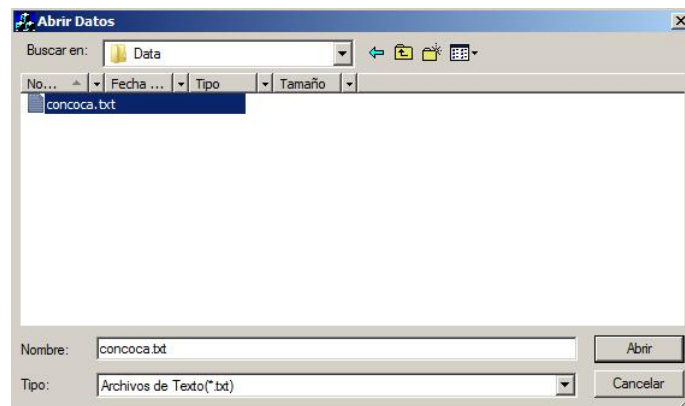
1. Escoger skew logit de Modelos > Enlaces Asimétricos.



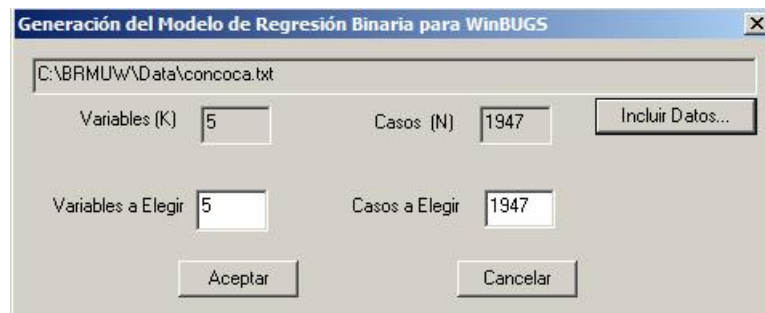
2. Esto abrirá la caja de dialogo *Generación del Modelo de Regresión Binaria para WinBUGS*



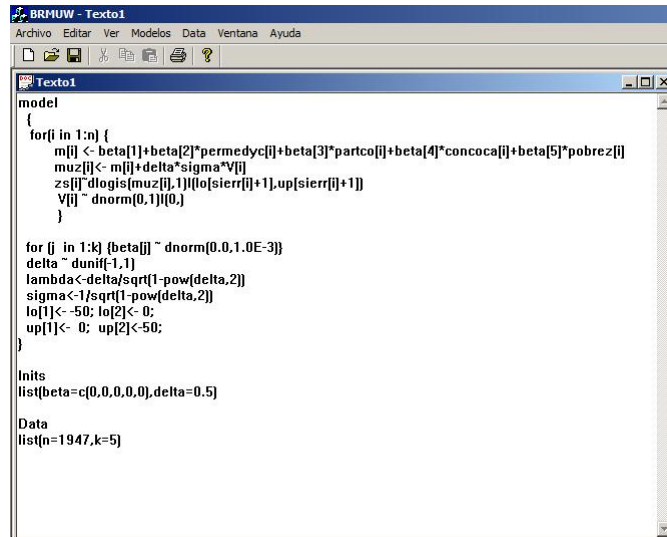
3. Hacer click en *Incluir Datos* y abrir el archivo con los datos.



4. Hacer click en Aceptar en *Generación del Modelo de Regresión Binaria para WinBUGS*.



5. Esto genera la sintaxis del modelo.



```
BRMUW - Texto1
Archivo Editar Ver Modelos Data Ventana Ayuda
Texto1
model
{
  for(i in 1:n) {
    m[i] <- beta[1]+beta[2]*permedyc[i]+beta[3]*partco[i]+beta[4]*concoca[i]+beta[5]*pobrez[i]
    muz[i]<- m[i]+delta*sigma*V[i]
    zs[i]~dlogis(muz[i],1)/[lo[sierr[i]+1],up[sierr[i]+1]]
    V[i]~dnorm(0,1)/[0,]
  }

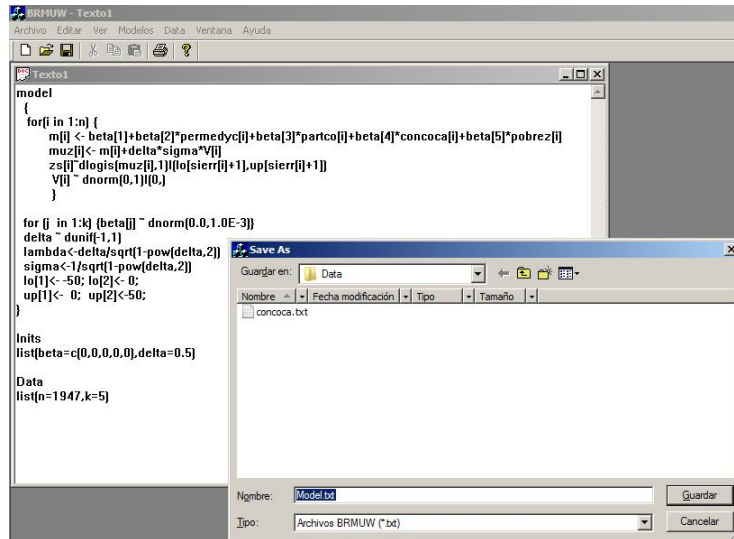
  for(j in 1:k) {beta[j]~dnorm(0.0,1.0E-3)}
  delta~dunif(-1,1)
  lambda<-delta/sqrt(1-pow(delta,2))
  sigma<-1/sqrt(1-pow(delta,2))
  lo[1]<- -50; lo[2]<- 0;
  up[1]<- 0; up[2]<-50;
}

Inits
list(beta=c(0,0,0,0),delta=0.5)

Data
list(n=1947,k=5)
```

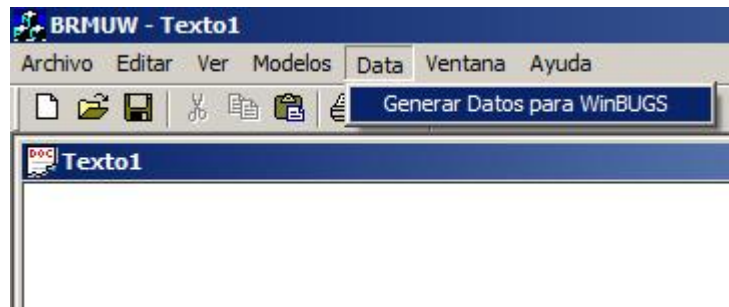

Inferencia Bayesiana en modelos de regresión binaria usando BRMUW

6. Guardar la sintaxis del modelo, Archivo > Guardar.



6.2. Generar la sintaxis para los datos

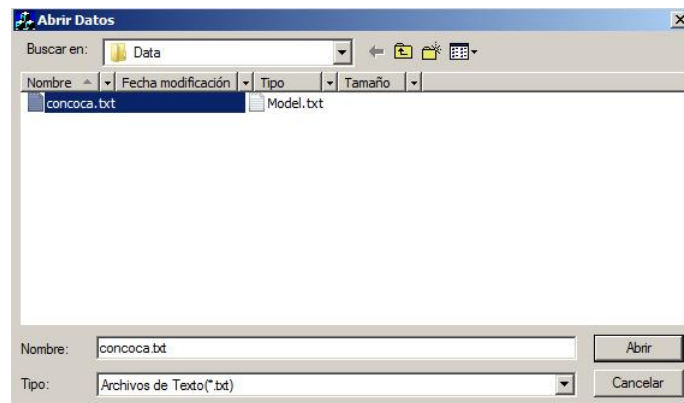
1. Hacer click en Data > Generar data para WinBUGS.



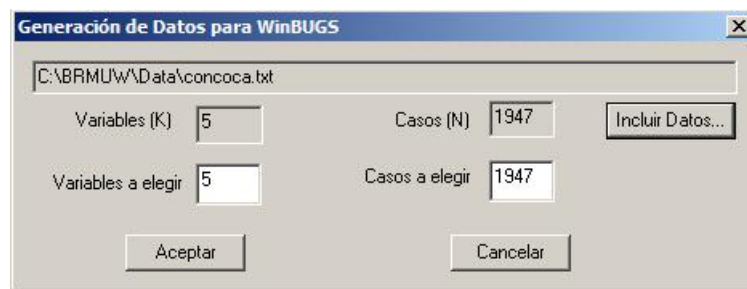
2. Esto abrirá la caja de dialogo *Generación de Datos para WinBUGS*.



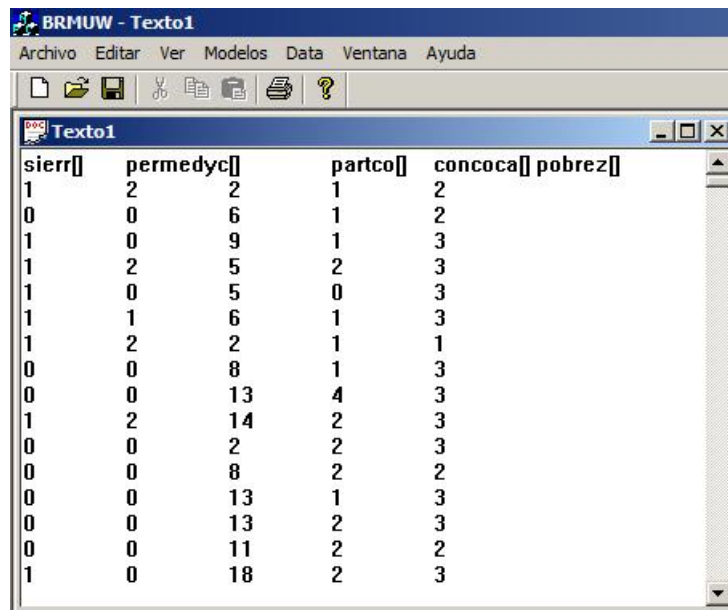
3. Hacer click en *Incluir Datos* y abrir el archivo con los datos.



4. Hacer click en Aceptar en *Generación de Datos para WinBUGS*.



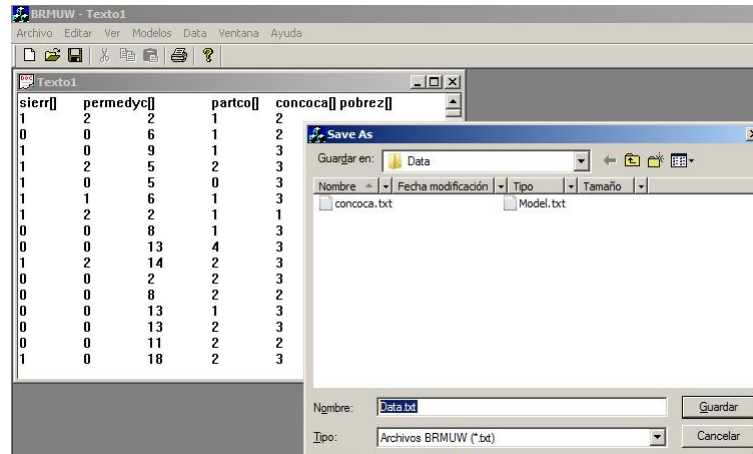
5. Esto genera la sintaxis para los datos.



The screenshot shows a text editor window with the following data:

sierr[]	permedyc[]	partco[]	concoca[]	pobrez[]
1	2	2	1	2
0	0	6	1	2
1	0	9	1	3
1	2	5	2	3
1	0	5	0	3
1	1	6	1	3
1	2	2	1	1
0	0	8	1	3
0	0	13	4	3
1	2	14	2	3
0	0	2	2	3
0	0	8	2	2
0	0	13	1	3
0	0	13	2	3
0	0	11	2	2
1	0	18	2	3

6. Guardar la sintaxis para los datos, Archivo > Guardar.

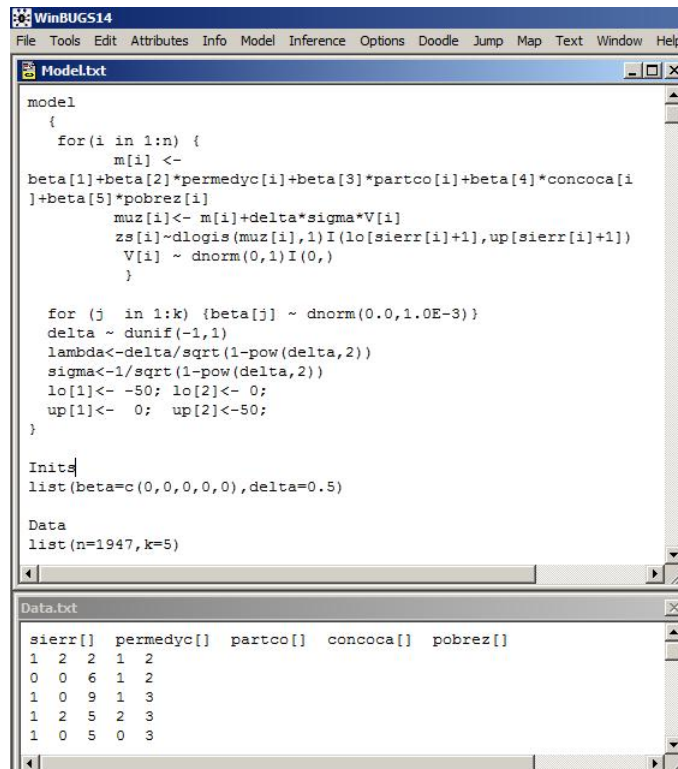


Un video tutorial de como BRMUW puede ser usado para generar sintaxis de diferentes modelos de regresión binaria así como generar sintaxis para datos leídos puede verse <http://videos.pucp.edu.pe/videos/ver/b55ab3b7633c6dab0cad8eec47066e40>.

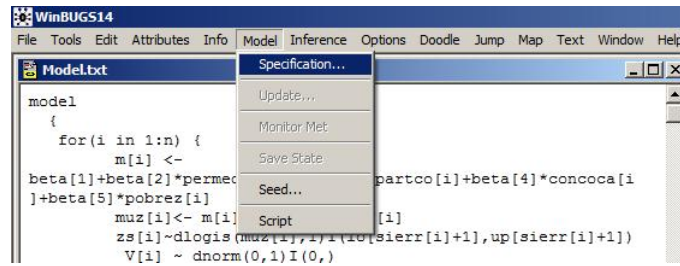
6.3. Estimación Bayesiana usando WinBUGS u OPEN-BUGS

Como se ha presentado el BRMUW genera dos archivos, uno que contiene el modelo de regresión binaria con el enlace seleccionado y otro que contiene el conjunto de datos. Ambos archivos en formato txt deben ser abiertos dentro del programa WinBUGS u OpenBUGS para realizar el análisis de inferencia bayesiana correspondiente.

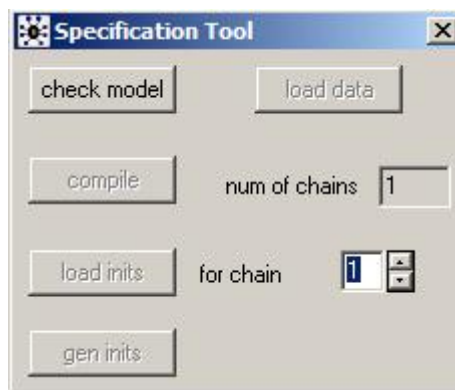
1. Abrir los archivos con la sintaxis del modelo y de los datos previamente generados por el BMRWU en WinBUGS u OpenBUGS.



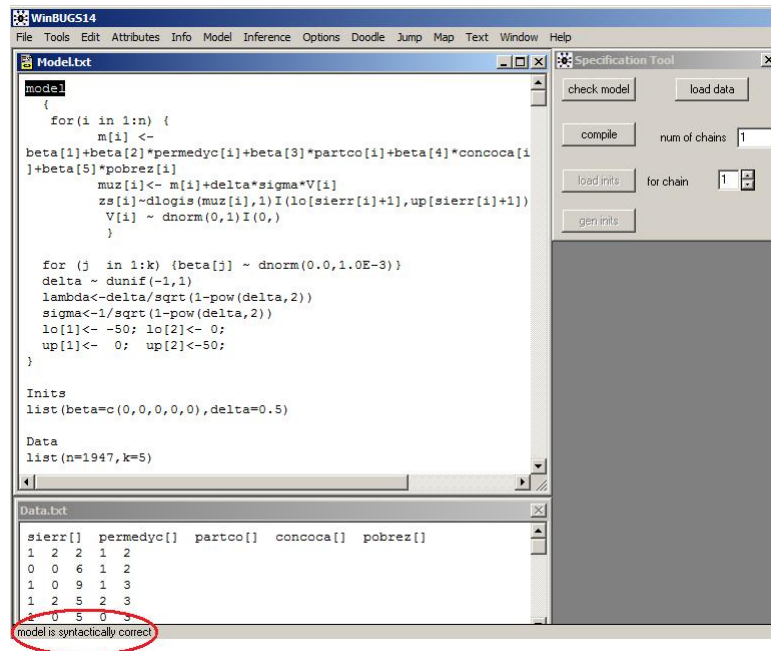
2. Hacer click en Model > Especification.



3. Esto abrirá la caja de dialogo *Specification Tool*.

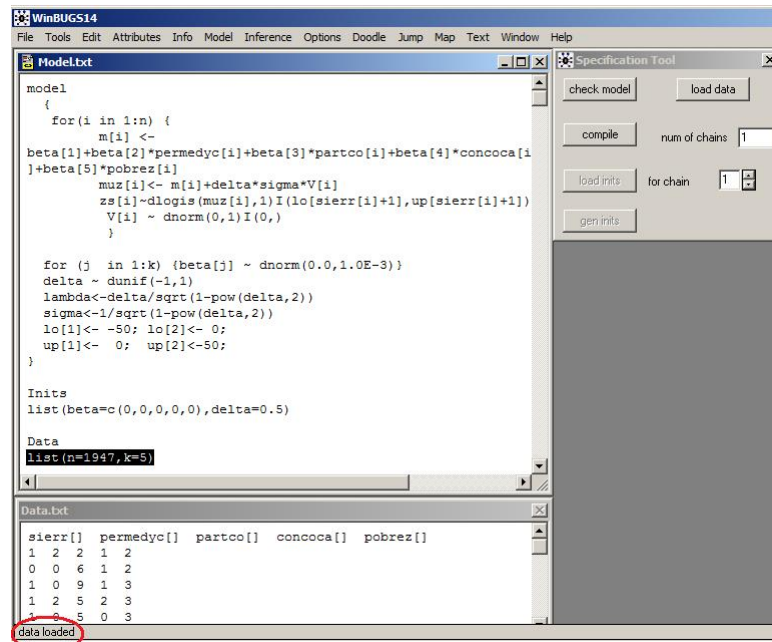


4. Seleccionar el modelo y hacer click en *check model*. En la parte inferior izquierda debe aparecer *model is syntactically correct* que indica que la sintaxis del modelo esta correctamente formulada.

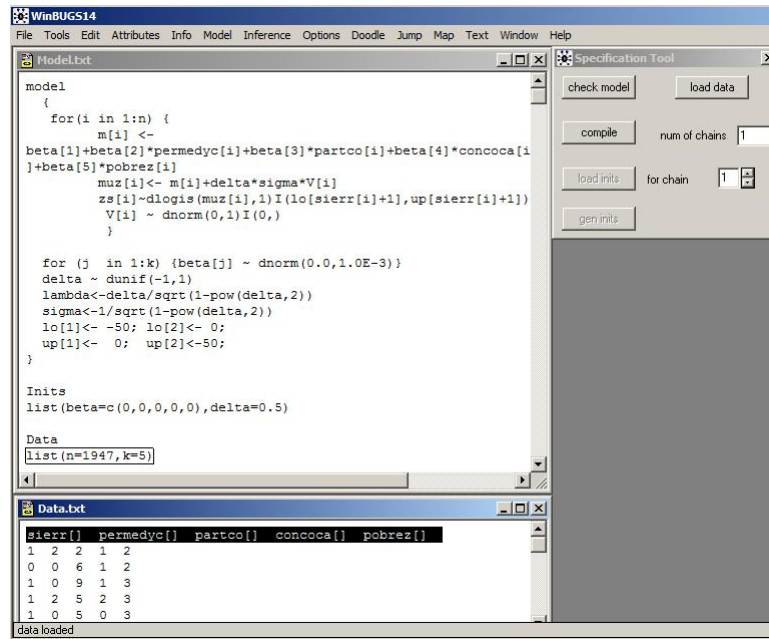


Inferencia Bayesiana en modelos de regresión binaria usando BRMUW

5. Seleccionar en el archivo del modelo la línea debajo de Data y hacer click en *load data*. En la parte inferior izquierda aparece *data loaded* indicando que los datos han sido leídos.

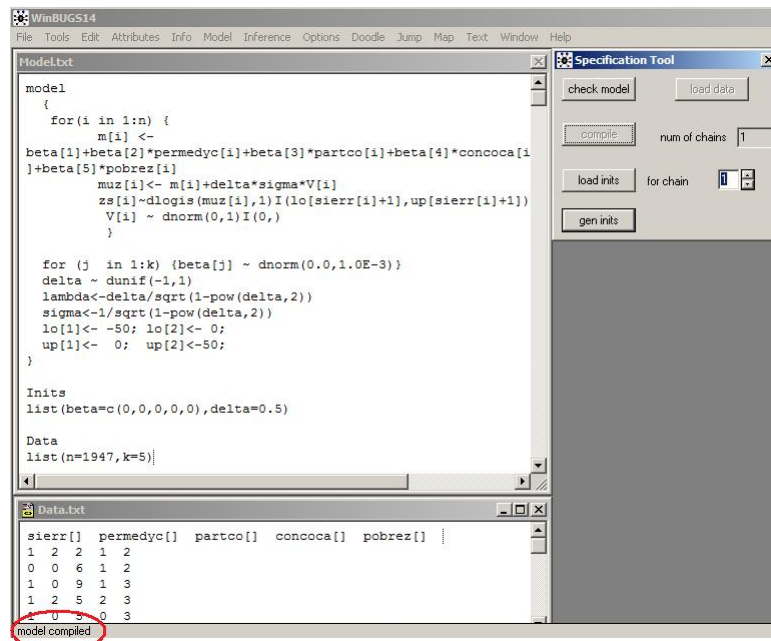


6. En el archivo de los datos seleccionar los nombres de las variables y hacer click en *load data*.

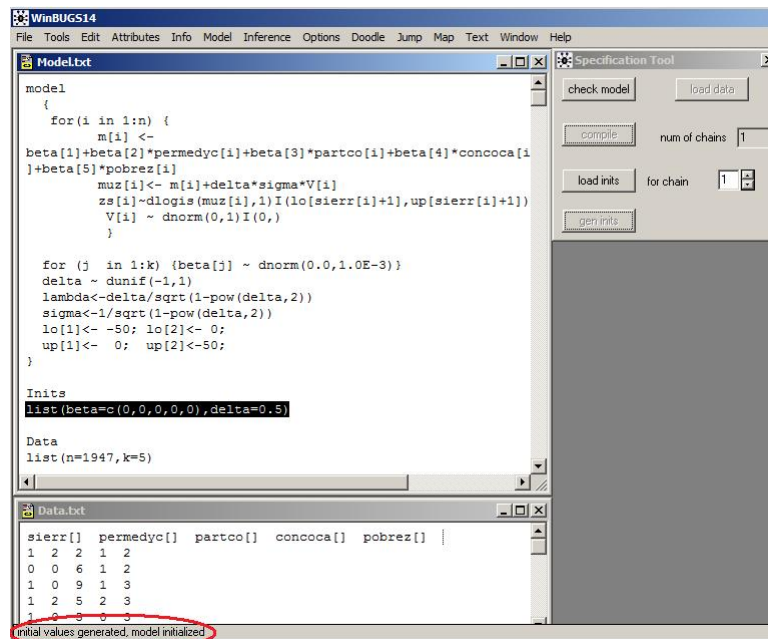


Inferencia Bayesiana en modelos de regresión binaria usando BRMUW

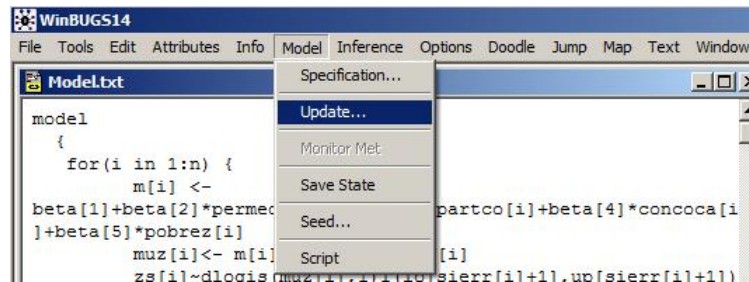
7. En la caja de dialogo *Specification tool*, se especifica el numero de cadenas que queremos generar en la caja de texto en *num of chains*. Una vez especificado el número de cadenas a generar (en este ejemplo 1 cadena) hacer click en *compile*. En la parte inferior debe aparecer *model compiled*.



8. Seleccionar la línea debajo de Inits en el archivo del modelo y hacer click en *load inits*. Luego hacer click en *gen inits*. Esto genera los valores iniciales para la Estimación Bayesiana. En la parte inferior debe aparece *initial values generated, model initialized*



9. Hacer click en Model > Update



10. Esto abrirá la caja de dialogo *Update tool*. En la caja de texto *updates* se ingresa el número de iteraciones que uno requiera y luego se hace click en *update*.



11. Luego debemos especificar que parámetros necesitamos que el programa guarde, para esto vamos a **Inference > Samples**, lo cual abrirá la caja de dialogo *Sample monitor tool*. En la caja de texto node se escribe el nombre del parámetro y luego se hace click en *set*, esto debe ser hecho para cada parámetro.



12. Repetimos el paso 10 generando mas iteraciones que ahora están siendo guardadas por el WinBUGS u OpenBugs. En la caja de dialogo *Sample Monitor Tool*, podemos calcular estadísticas a posteriori de los parámetros haciendo click en *stats*, un histórico de las cadenas haciendo click en *history*, una estimativa de la densidad a posteriori haciendo click en *density* y otras estadísticas de las cadenas pueden ser calculadas a través de esta caja de dialogo.

Un video tutorial de como WinBUGS puede ser usado para ejecutar la estimación bayesiana de modelos de regresión binaria puede verse en <http://videos.pucp.edu.pe/videos/ver/db8373ad4703990c51fd196ef2500c9f>.

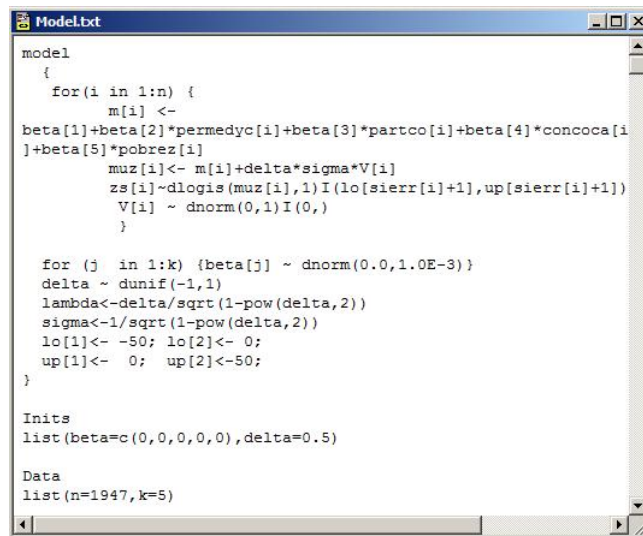
6.4. Estimación Bayesiana dentro de R

Como hemos visto en la sección anterior con los dos archivos que genera el BRMUW podemos implementar la Estimación Bayesiana con el programa WinBUGS u OpenBugs. Alternativamente también se puede implementar la estimación bayesiana utilizando interfaces del R con WinBUGS u OpenBugs. Para esto necesitaremos el archivo de texto original con los datos en columna:

Inferencia Bayesiana en modelos de regresión binaria usando BRMUW

```
sierr permedyc partco concoca pobrez
1 2 2 1 2
0 0 6 1 2
. . . . .
. . . . .
. . . . .
1 2 9 0 3
```

este archivo lo guardaremos con el nombre `datos.txt` y del archivo de sintaxis del modelo generado en BRMUW deberemos copiar solo la sintaxis del modelo. En el ejemplo abajo implementamos el modelo de regresión binaria *Skew logit*,



```
model
{
  for(i in 1:n) {
    m[i] <-
beta[1]+beta[2]*permedyc[i]+beta[3]*partco[i]+beta[4]*concoca[i
]+beta[5]*pobrez[i]
    muz[i]<- m[i]+delta*sigma*V[i]
    zs[i]~dlogis(muz[i],1)I(lo[sierr[i]+1],up[sierr[i]+1])
    V[i] ~ dnorm(0,1)I(0,)
  }

  for(j in 1:k) {beta[j] ~ dnorm(0.0,1.0E-3)}
  delta ~ dunif(-1,1)
  lambda<-delta/sqrt(1-pow(delta,2))
  sigma<-1/sqrt(1-pow(delta,2))
  lo[1]<- -50; lo[2]<- 0;
  up[1]<- 0; up[2]<-50;
}

Inits
list(beta=c(0,0,0,0,0),delta=0.5)

Data
list(n=1947,k=5)
```

esto es debemos copiar toda la sintaxis antes de *Inits* y guardarla en un archivo, en este ejemplo `modelo.txt`. Entonces el archivo `modelo.txt` quedaría como

```
model
{
  for(i in 1:n) {
    m[i] <-
beta[1]+beta[2]*permedyc[i]+beta[3]*partco[i]+beta[4]*concoca[i
]+beta[5]*pobrez[i]
    muz[i]<- m[i]+delta*sigma*V[i]
    zs[i]~dlogis(muz[i],1)I(lo[sierr[i]+1],up[sierr[i]+1])
    V[i] ~ dnorm(0,1)I(0,)
  }

  for (j in 1:k) {beta[j] ~ dnorm(0.0,1.0E-3)}
  delta ~ dunif(-1,1)
  lambda<-delta/sqrt(1-pow(delta,2))
  sigma<-1/sqrt(1-pow(delta,2))
  lo[1]<- -50; lo[2]<- 0;
  up[1]<- 0; up[2]<-50;
}
```

Luego para implementar la estimación bayesiana en el R seguiremos los siguientes pasos para usar la librería R2WinBUGS.

1. Dentro de R, cargar la librería R2WinBUGS con el siguiente comando que ha sido instalada previamente `library(R2WinBUGS)`
2. Leer los datos (el archivo datos.txt en este ejemplo esta en la carpeta C:\BRMUW\

```
datos <- read.table("C:/BRMUW/datos.txt", header=TRUE, sep="",
na.strings="NA", dec=".",strip.white=TRUE)
```
3. Crear una lista que contenga los datos y la información que esta debajo de Data en el archivo generado por el BRMUW

```
Data
list(n=1947,k=5)
```

Esto la hacemos con el siguiente comando

```
data<-c(as.list(datos),list(n=1947,k=5))
```

4. Crear un programa que genere los valores iniciales.


```
Inits  
list(beta=c(0,0,0,0,0),delta=0.5)
```

```
inits<-function(){list(beta=c(0,0,0,0,0),delta=0.5)}
```

5. Finalmente con el comando bugs se implementa la estimación bayesiana. Aquí explicaremos brevemente la sintaxis del comando Bugs

<i>parameters.to.save</i>	es un vector con los nombres de los parámetros del modelo cuyas simulaciones deseamos guardar.
<i>model.file</i>	es la dirección donde se encuentra el archivo del modelo.
<i>n.chains</i>	es el número de cadenas a ser generadas.
<i>n.iter</i>	es el número iteraciones totales de cada cadena.
<i>n.burnin</i>	es el número de iteraciones a ser descartadas como burn-in.
<i>program</i>	es el programa que se utilizará para implementar la estimación bayesiana.
<i>n.burnin</i>	es el número de iteraciones a ser descartadas como burn-in.

Luego con el siguiente comando se implementa la estimación bayesiana y las simulaciones son guardadas en el objeto salida.

```
salida<-bugs(data,inits,parameters.to.save=c("beta","lambda"),  
model.file="C:/BRMUW/modelo.txt", n.chains=1, n.iter=2000,  
n.burnin=1000,program="OpenBUGS")
```

Si ponemos `salida` en la línea de comandos del R obtenemos un resumen de la simulación

```
> salida  
Inference for Bugs model at "C:/Users/vaio/Documents/Jorge/  
modelo.txt", fit using OpenBUGS, 1 chains, each with 20000  
iterations (first 10000 discarded), n.thin = 1 n.sims = 1000  
iterations saved
```

Bazán y Bayes

	mean	sd	2.5%	25%	50%	75%	97.5%
beta[1]	-2.8	0.9	-4.2	-3.5	-2.9	-1.8	-1.5
beta[2]	0.7	0.1	0.5	0.6	0.7	0.7	0.8
beta[3]	0.1	0.0	0.1	0.1	0.1	0.1	0.1
beta[4]	-0.2	0.1	-0.3	-0.2	-0.2	-0.2	-0.1
beta[5]	0.5	0.1	0.4	0.4	0.5	0.7	0.8
lambda	0.8	0.9	-0.8	0.0	0.8	1.6	2.3
deviance	4984.5	155.5	4664.7	4862.6	5043.7	5123.0	5151.8

DIC info (using the rule, $pD = \bar{D} - \hat{D}$)

$pD = -90.0$ and $DIC = 4894.0$

DIC is an estimate of expected predictive error (lower deviance is better).

6. Finalmente, para mayores detalles sobre el comando bugs se puede consultar la ayuda escribiendo en la línea de comandos `?bugs`.

Referencias

- [1] Albert, J. H. (2009). *Bayesian Computation with R*. Springer Verlag
- [2] Basu, S. and Mukhopadhyay, S. (2000). Binary response regression with normal scale mixtures links, in *Generalized Linear Models: A Bayesian Perspective*, eds. D.K. Dey, S.K. Ghosh, and B.K: Mallick, New York: Marcel Dekker.
- [3] Bazán, J. L., Bolfarine, H. y Branco, M. D. (2006) A generalized skew probit class link for binary regression. *Technical report* (RT-MAE-2006-05). Department of Statistics. University of São Paulo.
- [4] Bazán, J. L. , Bolfarine, H. y Branco, D. M. (2010) A framework for skew-probit links in Binary regression (aceptado para publicación *Communications in Statistics - Theory and Methods*)
- [5] Bazán, J. L., Millones, O. (2008). A classification of binary asymmetric regression models: The use of BRMUW in an application to the decision to eradicate illegal crops of coca leaf. *Simposio Nacional de Probabilidade e Estatística. SINAPE*, São Pedro, Julio 2008.
- [6] Bazán, J. L., Millones, O (2008). Una clasificación de modelos de regresión binaria asimétrica: el uso del BAYES-PUCP en una aplicación sobre la decisión del cultivo ilícito de hoja de coca. *Economía 29(62)*, 17-32. PUCP.
- [7] Brooks, S. P. (2002). Discussion on the paper by Spiegelhalter, Best, Carlin, and van de Linde (2002). *Journal of the Royal Statistical Society Series B*, 64, 3,616-618.
- [8] Carlin, B.P. y Louis, T.A. (2001). *Bayes and Empirical Bayes Methods for Data Analysis Essays on Item Response Theory*. Second edition. New York: Chapman & Hall.
- [9] Casella, G. y Berger, R. L (2002). *Statistical Inference*, Duxbury: Pacific Grove, CA.
- [10] Chen, M. H., Dey, D. K., y Shao, Q-M. (1999) A new skewed link model for dichotomous quantal response data. *Journal of the American Statistical Association*, 94, 1172-1186.

- [11] Chen, M. H., Dey, D., and Shao, Q-M. (2001). Bayesian analysis of binary data using skewed logit models. *Calcutta Statistical Association Bulletin*, 51, 201-202.
- [12] Chen, M-H, Shao, Q. M, & Ibrahim, J. G (2000). Propriety of Posterior Distribution for Dichotomous Quantal Response Models. *Proceedings of the American Mathematical Society*, 129, 283-302.
- [13] Chen, M-H, Shao, Q. M, & Ibrahim, J. G (2000). *Monte Carlo Methods in Bayesian Computation*. New York: Springer Verlag.
- [14] Collet, D. (2003). *Modelling binary data*. Chapman & Hall/CRC, Second Edition, Boca Raton, USA.
- [15] Congdon, P. (2005). *Bayesian Models for Categorical Data*, Wiley.
- [16] Czado, C., and Santner, T. J. (1992). The effect of link misspecification on binary regression inference. *Journal of Statistical Planning and Inference*, 33, 213-231
Dey, D. K., Ghosh, S. K., Mallick, B. K. (eds) (2000). *Generalized Linear Models: a Bayesian Perspective*, New York, Dekker.
- [17] Dalal, S.R., Fowlkes, E.B., and Hoadley, B. (1989). Risk analysis of space shuttle : Pre-Challenger Prediction of Failure, *Journal of the American Statistical Association*, 84, 945-957.
- [18] Gamerman, D. y Lopes, H. F (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, Chapman and Hall/CRC".
- [19] Gelfand, A.E. y Smith, A.F.M. (1990). Sampling-based approaches to calculating marginal densities. *J. Am. Stat. Assoc.* 85, 398-409.
- [20] Geman, S. y Geman, D. (1984). Stochastic Relaxation, Gibbs Distribution and the Bayesian Restoration of Images *IEEE Transactions of Pattern Analysis an Machine Intelligence*, 6, 721-741.
- [21] Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains. *Biometrika* 57, 97-109.
- [22] Lavine, M. (1991). Problems in extrapolation illustrated with space shuttle o-ring data, *Journal of the American Statistical Association*, 86, 919-922.
- [23] Nagler J. (1994) Scobit: an alternative estimator to logit and probit. *American Journal Political Science*, 38, 230-255.

- [24] Prentice, R. L. (1976) A Generalization of the probit and logit methods for dose-response curves. *Biometrika*, 32, 761-768.
- [25] Roberts, C., P.(2002) *The Bayesian Choice: from decision-theoretic foundations to computational implementation*. 2nd ed. New york: Springer-Verlag.
- [26] Ross, S. (1995). *Stochastic Processes*, Wiley: New York, NY.
- [27] 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.
- [28] Spiegelhalter, D. J., Thomas, A., Best, N. G., & Gilks, W.R.(1996). *BUGS 0.5 examples* (Vol. 1 Version i). Cambrigde, UK: University of Cambride.
- [29] Spiegelhalter, D. J., Thomas, A., Best, N. G., Lunn, D (2007) *OpenBUGS User Manual version 3.0.2*. MRC Biostatistics Unit, Cambridge.
- [30] Tierney, L. (1994). Markov chains for exploring posterior distributions. *Ann. Stat.* 22, 1701-1762.

Jorge Bazán, Cristian Bayes.
Departamento de Ciencias,
Sección Matemática.
Pontificia Universidad Católica del Perú
e-mail: jlbazan@pucp.edu.pe, cbayes@pucp.edu.pe