



REPORTE DE INVESTIGACION

Nro. 27

SERIE B

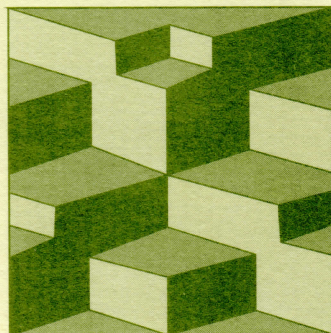
Enfoque Bayesiano en Modelos de Teoría de Respuesta al Ítem

Jorge Bazán,
Luis Valdivieso
Arturo Calderón

San Miguel, febrero del 2010

Series:

- A. INVESTIGACION AVANZADA.**
- B. DIVULGACION - APLICACION.**
- C. EDUCACION MATEMATICA.**
- D. TESIS.**



Nro. 27

SERIE B

**ENFOQUE BAYESIANO EN MODELOS DE
TEORIA DE RESPUESTA AL ITEM**

Jorge Bazán
Luis Valdivieso
Arturo Calderón

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. El evento 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 general

1. Introducción	1
2. Modelos TRI binarios unidimensionales tradicionales	5
2.1. Modelos de variables latentes	5
2.2. Formulación general de la TRI	7
2.3. Casos particulares	9
2.4. El modelo TRI con enlace logit	10
2.5. Características de los modelos TRI	11
2.5.1. Sobreparametrización de parámetros	11
2.5.2. Presencia de parámetros incidentales	11
2.5.3. Identificabilidad	11
2.6. Relación de la TRI con la TCT	13
3. Inferencia Bayesiana	16
3.1. Función de verosimilitud	16
3.2. Inferencia Bayesiana	16
3.2.1. Ejemplo 1: Un intervalo para una proporción	18
3.2.2. Ejemplo 2: Inferencia para una distribución normal	21
3.3. Métodos de MCMC	22
4. Las prioris en TRI	27
4.1. Nociones básicas	27
4.2. Prioris no informativas para el modelo TRI de 1 y 2 parámetros	30
4.3. Prioris no informativas para los modelo TRI de 1 y 2 parámetros con puntajes extremos	32

5. Inferencia Bayesiana en TRI	35
5.1. La distribución a posteriori en TRI	35
5.2. Inferencia Bayesiana en el modelo de ojiva normal	36
5.2.1. Especificación del modelo	36
5.2.2. Esquema MCMC usando adaptative rejection sampling (ARS)	37
5.2.3. Una formulación adecuada para el modelo TRI de ojiva normal	38
5.2.4. MCMC usando un esquema de data augmentation Gibbs sampling (DAGS)	38
5.3. Criterios de comparación de modelos en el contexto Bayesiano	40
6. Aplicaciones	44
6.1. Análisis de una escala de percepción de peso	44
6.2. Análisis de una prueba de Matemáticas	52
6.2.1. Comparando varios modelos TRI para los datos de la prueba de Matemáticas	53
6.2.2. Análisis de sensibilidad usando diferentes prioris para a y b en el modelo 2P	55
6.2.3. Inferencia e interpretación de los parámetros en el modelo 2P	60
7. Extensiones y comentarios finales	65
7.1. Modelos TRI con enlaces asimétricos	66
7.1.1. El modelo de ojiva normal asimétrico	66
7.1.2. Los modelos LPE y RLPE	68
7.2. Modelos TRI bajo restricciones de rapidez	70
7.3. Comentarios finales	72
A. Inferencia Bayesiana en TRI usando WinBUGS	73
A.1. Código WinBUGS para esquema Gibbs Sampling	73
A.2. Código WinBUGS para esquema ARS	74

B. Un Tutorial de WinBUGS para TRI	76
B.1. Fase 1: Definir el modelo	76
B.2. Fase 2: Parámetros para la simulación	82
B.3. Fase 3: Evaluación de las cadenas generadas	86

1 Introducción

Un modelo de variable latente es un modelo estadístico que relaciona un conjunto de variables (las llamadas variables manifiestas) con un conjunto de variables latentes. En general se asume que las respuestas en los indicadores o variables manifiestas son el resultado de la posición individuo en la variable o variables latentes consideradas y que las variables manifiestas no tienen nada en común después de controlar la variable latente, aspecto conocido como independencia local. Para una mayor comprensión acerca de las variables latentes puede revisarse Borsboom, Mellenbergh y van Heerden (2003).

Como proponen Bartholomew y Knoot (1999), existen diferentes tipos de modelos de variables latentes dependiendo si las variables manifiestas y latentes son categóricas o continuas. Así por ejemplo, cuando ambas son continuas tenemos el modelo de Análisis Factorial, cuando ambas son categóricas tenemos los modelos de clase latente. Cuando las variables manifiestas son continuas y las latentes son categóricas tenemos el modelo de análisis de perfiles latentes y de manera inversa, cuando las variables latentes se asumen continuas y las manifiestas discretas tenemos los modelos de rasgos latentes.

Dos de los más importantes modelos de rasgos latentes en orden histórico son la teoría clásica de los test (TCT) (ver por ejemplo, Lord y Novick, 1968) y la teoría de la respuesta al ítem (TRI) (ver por ejemplo, Baker, 1992). Estos modelos se emplean principalmente para el análisis de datos de indicadores o variables manifiestas obtenidas en procesos de medición y o evaluación en los que lo que se pretende medir no puede observarse directamente o es subyacente a estos indicadores y por tanto es conceptualizado como una variable latente.

Históricamente estos modelos han sido desarrollados en el ámbito de la Psicometría pero actualmente y especialmente la TRI es un área importante y creciente no sólo en Estadística si no en áreas de Computación y Probabilidad debido a la importante cantidad de problemas que

se plantean con el desarrollo y aplicaciones crecientes de esta clase de modelos. La importancia del estudio de los modelos TRI subyace también en sus múltiples aplicaciones crecientes a campos tan diversos como la Educación, la Psicología, la Política, la Sociología, la Economía, la Medicina, el Marketing, la Ingeniería Genética entre otras.

Una clase particular de modelos de TRI, son aquellos en las que las variables manifiestas categóricas son representadas en una matriz de respuestas dicotómicas o binarias (éxito y fracaso) y la variable latente es una sola. Un modelo como este se denominada de TRI binario unidimensional. En el Perú, uno de estos modelos, el denominado modelo de Rasch, es usado por el Ministerio de Educación para reportar el rendimiento escolar peruano en pruebas nacionales e internacionales. Otros ejemplos de uso que se pueden citar son, aquel donde las variables manifiestas describen si los postulantes a un examen responden bien o no a sus distintas preguntas y lo que se pretende evaluar es una determinada habilidad cognitiva como el desempeño matemático; si diferentes proyectos de inversión satisfacen o no ciertos criterios y lo que se pretende es generar un ranking de proyectos basado en una variable latente subyacente a estos indicadores que puede ser denominada calidad del proyecto; si los pacientes presentan o no ciertos síntomas donde la variable latente es la condición de salud relacionada con la enfermedad subyacente, etc. En todos estos casos, las variables manifiestas son denominadas de manera genérica como *ítems*.

Si bien a nivel internacional existe un gran desarrollo en la investigación y aplicación de los modelos de TRI y algunas textos estadísticos importantes han sido escritos para la presentación de estos modelos como Baker (1992) y Andrade, Tavares y Valle (2000) así como importantes trabajos compilatorios aparecen en van der Linden y Hambleton (1997) y Fischer y Molenaar (1995), existen pocos textos que tengan un tratamiento estadístico de los modelos TRI binarios unidimensionales. Un esfuerzo en esa línea ha sido desarrollado recientemente por Bazán, Valdivieso y Calderón, (2008). En el presente trabajo, continuando en esa dirección, presentamos el siguiente documento desarrollado en el marco

de la I Jornada Internacional de Probabilidad y Estadística, JIPE 2010.

En este documento, que puede ser considerado una segunda versión del anterior, presentaremos los principales modelos TRI binarios unidimensionales a los que podemos denominar modelos TRI tradicionales. Estos modelos son denominados en la literatura como modelos de 1, 2 y 3 parámetros dependiendo de cuantos parámetros se asocian al comportamiento de los ítems o variables manifiestas e incluyen aquellos donde se emplea un enlace logit o probit entre la probabilidad de acertar un ítem y el predictor latente que contiene la variable latente y los parámetros de ítem.

Nuestra propuesta es realizada desde la perspectiva de la Inferencia Bayesiana, corriente que ha tenido últimamente un gran desarrollo como consecuencia del uso de los denominados métodos MCMC (Markov Chain Monte Carlo, por sus siglas en inglés). Dado que los métodos MCMC son computacionalmente intensivos hemos querido también complementar nuestra propuesta presentando los programas de software libre Bayes@PUCP y WinBUGS, el primero de los cuales está siendo actualmente desarrollado por los autores en el marco de un proyecto de investigación de la Dirección Académica de Investigación de la Pontificia Universidad Católica del Perú.

El resto del documento está organizado de la siguiente manera. En el capítulo 2 presentaremos los modelos de TRI binarios unidimensionales tradicionales. En el capítulo 3 se hará una breve revisión de la Inferencia Bayesiana y de los métodos MCMC. En el capítulo 4 se presenta una importante discusión del uso de prioris en modelos de TRI. En el capítulo 5 se estudia la Inferencia Bayesiana del modelo de TRI binario unidimensional, con especial énfasis en el modelo con enlace probit. En el capítulo 6 se muestran dos aplicaciones, una en la que se emplea un modelo TRI binario con enlace logit y se compara con resultados obtenidos bajo TCT y otra que muestra una aplicación de los modelos TRI para el estudio de una prueba de Matemáticas haciendo uso de los softwares Bayes@PUCP y WinBUGS. En el capítulo 7 daremos una breve introducción a algunas de las posibles extensiones de los modelos TRI

binarios unidimensionales tradicionales y referencias de investigaciones actualmente desarrolladas para estos modelos. Culminaremos este trabajo presentando, como apéndice, un breve tutorial del uso de WinBUGS para la Inferencia Bayesiana en un modelo TRI.

A lo largo de este documento representaremos a la función de probabilidad y/o densidad de un elemento aleatorio Y con la letra f ó f_Y si se necesita enfatizar la dependencia de Y . Denotaremos también en negrita y con mayúsculas a las secuencias o vectores aleatorios, sólo en mayúsculas a las variables aleatorias y en minúsculas a los valores observados de estos elementos aleatorios.

2 Modelos TRI binarios unidimensionales tradicionales

2.1 Modelos de variables latentes

Definición 2.1.1. Definimos un *modelo latente* como el par ordenado (\mathbf{Y}, \mathbf{U}) , donde la secuencia aleatoria $\mathbf{Y} = (Y_1, Y_2, Y_3, \dots)$ y el vector aleatorio $\mathbf{U} = (U_1, U_2, \dots, U_m)$ contienen a las denominadas variables manifiestas y latentes, respectivamente.

Una variable es considerada latente en el sentido que no es empíricamente observable y manifiesta en caso contrario. En general, los modelos de variables latentes se formulan con el propósito de conocer las principales características de \mathbf{U} en base al conocimiento de \mathbf{Y} .

Definición 2.1.2. Un modelo latente (\mathbf{Y}, \mathbf{U}) se dice que satisface:

- La *independencia condicional*, si la distribución condicional de cualquier subconjunto de \mathbf{Y} dado $\mathbf{U} = \mathbf{u}$ puede expresarse como el producto de las distribuciones marginales condicionales de \mathbf{Y} dado $\mathbf{U} = \mathbf{u}$, para todo de $\mathbf{u} \in \mathbb{R}^m$. En este caso diremos que el modelo es *condicionalmente independiente*.
- La *monotonidad*, si

$$1 - P(Y_i \leq y | \mathbf{U} = \mathbf{u}) = P(Y_i > y | \mathbf{U} = \mathbf{u}) \quad (2.1)$$

son funciones no decrecientes de $\mathbf{u} \in \mathbb{R}^m$, para cualquier $y \in \mathbb{R}$ e $i \in \mathbb{N}^+$. En este caso diremos que el modelo es *monótono*.

- La *unidimensionalidad*, si \mathbf{U} es una variable aleatoria unidimensional; es decir, con $m = 1$. En este caso diremos que el modelo es *unidimensional*.

Definición 2.1.3. Un modelo latente (\mathbf{Y}, \mathbf{U}) se denomina *condicionalmente monótono* si satisface la independencia condicional y la monotonicidad. Si además el modelo es unidimensional se dice que (\mathbf{Y}, \mathbf{U}) es un modelo *condicionalmente monótono unidimensional* o que admite una *representación monótona unidimensional*.

Proposición 2.1.1. *Considere un modelo monótono unidimensional (\mathbf{Y}, U) en donde cada $Y_i|U = u$ toma tan solo dos posibles valores: uno llamado éxito con probabilidad u y el otro llamado fracaso con probabilidad $1 - u$, siendo $0 < u < 1$. Entonces existe una única distribución para U , dada por G , tal que toda distribución finito dimensional de \mathbf{Y} puede representarse por:*

$$f(\mathbf{y}) = \int \prod_i u^{y_i} (1 - u)^{1 - y_i} dG(u),$$

donde los y_i denotan a las componentes del vector finito dimensional \mathbf{y} .

Demostración. La existencia está garantizada por el teorema de representación de De Finetti (Schervish, 1995) que indica que esto ocurre si todos las componentes de \mathbf{Y} son una secuencia infinita permutable (las variables independientes e idénticamente distribuidas lo son) siendo

$$G(u) = \lim_{n \rightarrow \infty} P(S_n \leq u)$$

con

$$S_n = Y_1 + Y_2 + \dots, Y_n \quad \text{y} \quad u = \lim_{n \rightarrow \infty} \frac{S_n}{n}.$$

Dado que cada $Y_i|U$ sigue una distribución de Bernoulli, toda distribución conjunta finito dimensional de \mathbf{Y} condicionada a U , resulta ser

$$f(\mathbf{y}|U = u) = \prod_i u^{y_i} (1 - u)^{1 - y_i}$$

y consecuentemente su distribución marginal viene dada por

$$f(\mathbf{y}) = \int f(\mathbf{y}|U = u)dG(u).$$

■

Algunos resultados adicionales para esta clase general de modelos latentes se pueden revisar en Holland y Rosenbaum (1986).

2.2 Formulación general de la TRI

Definición 2.2.1. Considere n sujetos evaluados en una prueba de k ítems. Un modelo de *Teoría de Respuesta al ítem*, o por brevedad modelo TRI, binario unidimensional es un sistema en el que para cada sujeto i se tiene un modelo de *variable latente monótono unidimensional* (\mathbf{Y}, U_i) definido por las expresiones:

$$Y_{ij}|u_i, \eta_j \sim \text{Bernouilli}(p_{ij}) \quad (2.2)$$

$$p_{ij} = P(Y_{ij} = 1 | u_i, \eta_j) = F(m_{ij}) \quad (2.3)$$

$$m_{ij} = a_j(u_i - b_j), \quad (2.4)$$

$$i = 1 \dots, n, \quad j = 1, \dots, k$$

donde

- Y_{ij} es la variable manifiesta que modela la respuesta binaria del sujeto i que responde al ítem j .
- $\eta_j = (a_j, b_j)$ son parámetros que representan, respectivamente, a la discriminación y la dificultad del ítem j .
- u_i es el valor de la variable latente U_i para el sujeto i , que podríamos interpretarla como la habilidad del sujeto i .

- p_{ij} es una probabilidad condicionada a que $U_i = u_i$.
- F se denomina la curva característica del item (CCI) y
- m_{ij} es un predictor latente lineal en relación a la habilidad del sujeto i .

Observaciones

- El modelo TRI satisface la propiedad de *independencia condicional latente* ; es decir, para un sujeto i las respuestas Y_{ij} a los diferentes items son condicionalmente independientes dada la variable latente $U_i, i = 1 \dots, n$.
- El modelo TRI satisface la propiedad de *monotonidad latente* , pues es una función estrictamente no decreciente de $U_i, i = 1 \dots, n$.
- El modelo TRI es *unidimensionalmente latente* .
- En la TRI, $F(m_{ij})$, donde $i = 1 \dots, n$ y $j = 1, \dots, k$, es la misma para cada caso y $F^{-1}(\cdot)$ es llamada la función de enlace.
- Se asume también independencia entre las respuestas de los distintos sujetos.
- Los parámetros de dificultad b_j y de discriminación a_j representan la posición e inclinación del item j , respectivamente, siendo a_j un valor proporcional a la inclinación de la CCI en el punto b_j . Valores $a_j < 0$ no son esperados. El espacio paramétrico para el parámetro b_j es arbitrario y corresponde al mismo que para u_i que por lo usual se toma como la recta real.

Otra parametrización muy común para el predictor lineal latente es $m_{ij} = a_j u_i - b_j$. Esta parametrización es muy importante desde el punto de vista computacional ya que es facilita el tiempo computacional

de convergencia. Cuando se usa esta parametrización, la dificultad en (2.4) puede obtenerse haciendo $\frac{b_j}{a_j}$ en el resultado obtenido con esta parametrización. En general, esta es la parametrización preferida en la Inferencia Bayesiana.

Proposición 2.2.1. *En un modelo TRI binario, la densidad conjunta del vector de respuestas multivariantes $\mathbf{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_n)'$, con $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{ik})$, dado el vector de variables latentes $\mathbf{U} = (U_1, \dots, U_n)'$ y el vector de parámetros de los ítems $\boldsymbol{\eta} = (\eta_1, \dots, \eta_k)'$ puede ser escrito como:*

$$f(\mathbf{y}|\mathbf{u}, \boldsymbol{\eta}) = \prod_{i=1}^n \prod_{j=1}^k F(m_{ij})^{y_{ij}} (1 - F(m_{ij}))^{1-y_{ij}} \quad (2.5)$$

Demostración. La prueba es directa por la independencia condicional latente. ■

El primer modelo TRI binario fue introducido por Lord (1952) con una CCI dada por $F(x) = \Phi(x)$, siendo Φ la función de distribución acumulada de una variable normal estándar. Este modelo es conocido en la literatura psicométrica como un modelo de *ojiva normal* el cual corresponde, en el contexto de los modelos lineales generalizados, a una función de enlace probit. Por otro lado, Birbaum (1968) consideró una CCI dada por $F(x) = L(x)$, donde $L(x) = \frac{e^x}{1+e^x}$ denota a la función de distribución acumulada de una variable logística. Esta induce, en el lenguaje de los modelos lineales generalizados, a una función de enlace logit. Este modelo es conocido como el modelo *logístico*.

2.3 Casos particulares

El modelo (2.2) admite diversas formulaciones, las cuales dependen básicamente de como se considere la CCI. En su versión más simple

podríamos tomar $a_j = 1$ y considerar una CCI de la forma

$$P(Y_{ij} = 1 | u_i, b_j) = F(u_i - b_j). \quad (2.6)$$

Este es llamado un modelo TRI de un parámetro.

De manera más general podríamos considerar una CCI de la forma

$$P(Y_{ij} = 1 | u_i, a_j, b_j, c_j) = c_j + (1 - c_j)F(a_j(u_i - b_j)), \quad (2.7)$$

donde el parámetro $c_j \in [0, 1]$ representa la probabilidad de que un individuo con mínima habilidad responda correctamente al ítem j y F es una función de distribución. Este es conocido como el modelo TRI de tres parámetros. De tomarse $c_j = 0$, el modelo se reduce al denominado modelo TRI de dos parámetros.

2.4 El modelo TRI con enlace logit

El modelo TRI con enlace logit o modelo logístico es el modelo probablemente más usado en TRI. La versión de tres parámetros para este modelo establece que (2.3) viene dada por:

$$P(Y_{ij} = 1 | u_i, a_j, b_j, c_j) = c_j + (1 - c_j) \frac{1}{1 + e^{Da_j(u_i - b_j)}}, \quad (2.8)$$

donde por lo usual se asume que $D = -1$ aunque algunos autores consideran también el valor $D = -1.7$ a fin de aproximar este modelo a uno de ojiva normal. Como casos particulares se tiene

- $$P(Y_{ij} = 1 | u_i, a_j, b_j) = \frac{1}{1 + e^{Da_j(u_i - b_j)}} \quad (\text{con } c_j = 0) \text{ y}$$

- $$P(Y_{ij} = 1 | u_i, b_j) = \frac{1}{1 + e^{D(u_i - b_j)}} \quad (\text{con } c_j = 0 \text{ y } a_j = 1).$$

Al último modelo de un parámetro, se le conoce también como el modelo de Rasch y tiene derivaciones propias (véase por ejemplo Fischer y Molenaar, 1995).

2.5 Características de los modelos TRI

2.5.1 Sobreparametrización de parámetros

Una de las características más saltantes de los modelos TRI es que están sobreparametrizados, el modelo de 3 parámetros por citar tiene $n + 3k$ parámetros desconocidos. Ello hace, como es de esperarse, que la inferencia en estos modelos sea intensiva computacionalmente. Note que a mayor tamaño de muestra y mayor número de ítems aumenta el número de parámetros a ser estimados.

2.5.2 Presencia de parámetros incidentales

Es bastante conocido que los estimados de parámetros estructurales no son necesariamente consistentes cuando son estimados de manera conjunta con parámetros incidentales (Neyman y Scott, 1948). En TRI, para un número fijo de ítems, los parámetros de los ítems son denominados estructurales mientras los parámetros de la variable latente son considerados incidentales debido a que el número de parámetros de la variable latente se incrementa si se incrementa el número de respondientes. Debemos llamar la atención que el procedimiento de estimación conjunta por máxima verosimilitud es un método frecuente en la estimación del modelo de Rasch, y se encuentra implementado en el software Winstep y como es de esperarse tiene resultados no consistentes frente a otros métodos como ha sido mostrado por Chincaro y Bazán (2010).

2.5.3 Identificabilidad

Un parámetro se dice identificado si distintos valores del parámetro implican diferentes distribuciones de probabilidad para los datos. Un

modelo es identificado si todos sus parámetros son identificados. Sin embargo, como indica Rivers (2003) cualquier hipótesis acerca de la no identificabilidad de un parámetro no se puede probar.

Sin embargo en TRI, puede notarse que el predictor lineal en (2.4) puede escribirse como

$$a_j(u_i - b_j) = \frac{a_j}{40}[(40u_i + 200) - (40b_j + 200)] = a_j^*(u_i^* - b_j^*).$$

Así, la verosimilitud con u_i , a_j y b_j es la misma que con

$$u_i^* = 40u_i + 200, \quad a_j^* = \frac{a_j}{40} \quad \text{y} \quad b_j^* = 40b_j + 200.$$

De esta manera, los parámetros del modelo no pueden ser estimados de manera única a menos de que se impongan ciertas restricciones. En este caso distintas combinaciones de valores de los parámetros no necesariamente implican diferentes distribuciones de probabilidad para los datos. Nótese sin embargo que la probabilidad de que un individuo responda correctamente a cierto ítem es siempre la misma, independientemente de la escala utilizada para los parámetros del modelo. Aunque deseable como indica Rivers (2003), identificación es un requerimiento débil para la estimación en el sentido de ser una condición necesaria pero no suficiente para la existencia de buenos estimadores.

Dada las características de este tipo de modelos enunciadas, el problema de ajustar los parámetros del modelo es por tanto bastante complejo. Desde un punto de vista clásico la solución más utilizada es enfrentar el problema dividiéndolo en dos etapas. En una primera etapa, denominada de *calibración*, se asume una distribución aleatoria conocida para las variables latentes U y se estiman los parámetros de ítem usando una “muestra de calibración”. En una segunda etapa, denominada de *estimación*, se asumen conocidas las estimaciones de los parámetros de los ítems y se estiman las variables latentes para la muestra de estudio. Con esto el problema de sobreparametrización disminuye y el problema

de estimación de parámetros incidentales se evita. Esta es la estrategia denominada por Patz y Junker (1989) como la de “divide y conquistarás” se aplica cuando se obtienen estimativas de Máxima Verosimilitud Marginal vía un algoritmo EM donde es común imponer restricciones para los parámetros de ítemes como los considerados por ejemplo en Bock y Aitkin (1981) aspecto que permite que el problema de identificabilidad sea evitado. Este método de estimación es bastante usado y se encuentra implementado en softwares como *Conquest* y *BILOG – MG* y en la librería *ltm* de *R*.

Desde la perspectiva bayesiana, que es la que nos interesa en este documento, la sobreparametrización se mantiene y por tanto el tiempo computacional es generalmente mayor que la estrategia anterior debido a la necesidad de estimar más parámetros de manera conjunta. Pero la estimación conjunta resulta conveniente porque refleja mejor el problema original de la TRI que es estimar las características individuales y de los ítemes de manera simultánea y completa. Adicionalmente, en la inferencia bayesiana la distinción entre parámetros incidentales y estructurales carece de sentido pues tanto los parámetros de ítem como de sujetos son tratados como variables aleatorias y la incertidumbre acerca de su valor es recogida por una especificación de una distribución a priori para ellas. Finalmente el problema de identificabilidad también es resuelto considerando la especificación de distribuciones a priori para los parámetros del modelo TRI binario unidimensional porque estas prioris puede proporcionar un origen o punto cero, una métrica y una dirección. Un análisis de las condiciones de las prioris en TRI para obtener distribuciones a posteriori propias $f(\mathbf{u}, \boldsymbol{\eta}|\mathbf{y})$ se estudiará en el subsiguiente capítulo.

2.6 Relación de la TRI con la TCT

Como ha sido mencionado en la introducción otro modelo de rasgos latentes, históricamente anterior a la TRI, para el estudio de tests o pruebas es la teoría clásica de los test (TCT) y que se basa en el modelo lineal de Spearman ampliamente documentado por Lord y Novick (1968),

ver también por ejemplo Bazán (2004). En la TCT, el modelo usado se basa en la ecuación:

$$S_i = U_i + \epsilon_i,$$

donde $S_i = \sum_{j=1}^k Y_{ij}$ es la puntuación total (score) o puntaje observado obtenido por el individuo i para una prueba de k ítemes, U_i es el rasgo latente o puntaje verdadero medido por el test y ϵ_i es el error aleatorio asociado al individuo i . Otra forma como es conocida esta teoría es teoría del puntaje verdadero o teoría de error latente.

En la TCT una tarea complementaria e importante es el análisis clásico de ítemes, que se apoya principalmente en dos indicadores de amplio uso que pueden ser revisados en Lord y Novick (1968):

1. La proporción de respuesta positiva o probabilidad empírica en el ítem j :

$$\hat{p}_j = P(\widehat{Y}_{ij} = 1).$$

2. La discriminación clásica del ítem, denotada por d_j , que nos indica la capacidad del ítem para discriminar entre casos con alto y bajo nivel del rasgo latente medido por los k ítemes de la prueba. Hay varias alternativas para d_j , siendo la más usada

$$d_j = \hat{p}_j^A - \hat{p}_j^B,$$

donde \hat{p}_j^A es la proporción de respuesta positiva en el tercio superior de la distribución de los puntajes totales y \hat{p}_j^B es la proporción de respuesta positiva en el tercio inferior, indicador que es usado en el capítulo de aplicaciones. Otra alternativa es la correlación de Pearson entre el puntaje del ítem y la puntuación total depurada del puntaje de este ítem. Esta, que en tal contexto se denomina la correlación biserial puntual, no es la empleada en el capítulo empleará en este trabajo; pero vale indicar que con cualquiera de estas medidas debe ocurrir que $d_j > 0$ y cuanto mayor sea d_j , mayor poder de discriminación tendrá el ítem j .

La teoría clásica tiene problemas no resueltos, a saber, el primero es que no provee una medida de ajuste de los datos al modelo. El segundo es que la capacidad del ítem para medir el rasgo no está en un mismo eje con las puntuaciones o scores S de la prueba, pues en un caso se trata de una probabilidad (empírica) y en el segundo de un puntaje total en otra escala. Asociado con lo anterior, está el hecho cierto que dos personas pueden tener igual puntuación S y sin embargo haber obtenido este mismo valor a partir de respuestas positivas a ítems correspondientes a diferentes niveles de intensidad del rasgo medido por la prueba o test. Los problemas mencionados son resueltos precisamente en la TRI.

Una explicación de la TCT se encuentra en Calderón, A. (2005), Bazán, J.L. (1997) y Lord y Novick (1968). Algunos programas libres para el análisis de ítems que, desde el punto de vista de la TCT, pueden destacarse son Vista (<http://www.visualstats.org/>), Openstat (<http://statpages.org/miller/openstat/>) y varias librerías en R .

En la práctica el análisis clásico de ítems es complementario y muy relacionado a la TRI como lo ilustraremos en el capítulo 6, aplicación 1, donde se realiza una comparación empírica entre la TCT y la TRI para el análisis de una escala de percepción de peso.

3 Inferencia Bayesiana

3.1 Función de verosimilitud

Definición 3.1.1. 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.

Definición 3.1.2. Un estimador de *máxima verosimilitud* para el parámetro θ es el valor $\hat{\theta}$ para el cual la función de verosimilitud $L(\mathbf{y}|\theta)$ es máxima.

La solución de inferencia clásica consiste en maximizar 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.

3.2 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). Una revisión más detallada puede encontrarse en Robert (2001).

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 no determinístico y 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 θ , $f(\theta)$.
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 a la priori; es decir

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

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

Observación:

- Tomando logaritmo en la expresión anterior obtenemos:

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

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

$$f(\theta) = c \quad (\text{con } c \text{ una constante})$$

maximizar $\log f(\theta|\mathbf{y})$ para θ resulta equivalente a maximizar $\ell(\mathbf{y}|\theta)$. En otras palabras, la estimación de máxima verosimilitud (que maximiza $\ell(\mathbf{y}|\theta)$) puede considerarse 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, respectivamente, a la media, varianza, mediana y cuantil γ de la distribución a posteriori de θ dado $\mathbf{Y} = \mathbf{y}$. Así por ejemplo, se puede definir un intervalo para $\theta|\mathbf{y}$ dado de esta manera

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

donde

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

son los cuantiles correspondientes.

Nótese 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*.

3.2.1 Ejemplo 1: Un intervalo para una proporción

Un problema bastante frecuente es el de la inferencia sobre una proporción p . Por ejemplo, supongamos que estamos interesados en construir un intervalo de confianza para la proporción p de aprobación presidencial en Lima Metropolitana con base a la información obtenida por una encuestadora en una muestra aleatoria de $n = 400$ personas de Lima Metropolitana. Si la proporción de aprobación obtenida en la muestra fue de $\bar{p} = 0.3$, ¿cómo entonces podríamos hacer inferencias sobre p , tanto desde un punto de vista clásico como Bayesiano.

Solución clásica

En la inferencia clásica es conocido que una aproximación para el intervalo de confianza al $100(1 - \alpha)\%$ para p viene dado por:

$$IC(p) = \bar{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}.$$

En nuestro caso para un nivel de confianza del 95% obtendríamos $z_{1-\frac{\alpha}{2}} = z_{0,975} = 1.96$ y por tanto el intervalo

$$IC(p) = 0.3 \pm 1.96 \sqrt{\frac{0.3 \times 0.7}{400}} = [0.2550908, 0.3449092].$$

Podemos concluir entonces, con un nivel de confianza del 95%, que la aprobación presidencial se encuentre en aproximadamente un rango de entre el 26 y 35 %.

Solución Bayesiana

En la solución Bayesiana se reconoce a p como una variable aleatoria y por tanto puede establecerse una distribución a priori. Luego, usando la evidencia muestral en la verosimilitud de los datos se busca obtener la distribución a posteriori para p dados estos; es decir, $f(p|\mathbf{y})$. Para realizar esto podemos seguir los siguientes pasos:

- Fase 1. Especificación de la distribución a priori. Dado que $0 \leq p \leq 1$ una distribución natural para p sería considerar una distribución Beta del tipo

$$p \sim \text{Beta}(\alpha, \beta).$$

- Fase 2. Especificación de los parámetros de la distribución a priori. De estudios anteriores, digamos considerando el resultado de los últimos meses, se puede considerar que

$$E(p) = 0.35 \quad \text{y} \quad V(p) = 0.01.$$

Sobre esta base podemos encontrar que

$$\alpha = 1.6125 \quad \text{y} \quad \beta = 14.375$$

y por tanto

$$f(p) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1} \propto p^{1.6125-1} (1-p)^{14.375-1},$$

donde la última es una igualdad excepto por una constante que no depende de p y por tanto no es de interés.

- Fase 3. Obteniendo la verosimilitud. En este caso se trata de un modelo Binomial $\mathbf{Y}|p \sim B(n, p)$ y por tanto tenemos que

$$L(\mathbf{y}|p) = \prod_{i=1}^n C_{y_i}^n p^{y_i} (1-p)^{1-y_i},$$

donde las combinatorias $C_{y_i}^n$ podrían omitirse ya que son constantes en relación a p . Como $\bar{p} = 0.30$ tenemos que $\sum_{i=1}^n y_i = 0.30 \times 400 = 120$ y por tanto

$$L(\mathbf{y}|p) \propto p^{\sum_{i=1}^n y_i} (1-p)^{n - \sum_{i=1}^n y_i} = p^{120} (1-p)^{400-120}.$$

- Fase 4. Hallando la distribución a posteriori. Combinando la priori con la posteriori tenemos

$$f(p|\mathbf{y}) \propto p^{127.6125-1} (1-p)^{294.375-1}$$

y por lo tanto la distribución a posteriori es

$$p|\mathbf{y} \sim \text{Beta}(127.6125, 294.375)$$

- Fase 5. Haciendo inferencia en base a la distribución a posteriori. Algunas medidas son

$$p^{est} = E(p|\mathbf{y}) = 0.30$$

que corresponde a la media a posteriori. Los percentiles son $P_{97.5} = 0.35$, $P_{2.5} = 0.26$ por lo que la probabilidad de estar entre 0.26 y 0.35 es del 90 %.

En el ejemplo anterior observamos que la distribución a posteriori fue relativamente fácil de obtener e identificar. En muchas ocasiones, sin embargo, tal tarea no es del todo automática ya que dependiendo de la priori elegida la distribución a posteriori puede ser bastante compleja e incluso no pertenecer a algunas de las familias de distribuciones conocidas. En tal situación, como ejemplificaremos en el siguiente problema, tenemos por fortuna la posibilidad de recurrir a un conjunto de métodos de simulación, globalmente conocidos como de Montecarlo, que nos permitirán obtener de manera aproximada la deseada distribución a posteriori.

3.2.2 Ejemplo 2: Inferencia para una distribución normal

Supongamos que $Y|\mu \sim N(\mu, \sigma^2)$ es un modelo probabilístico para una variable de interés con σ^2 conocido y μ un parámetro a estimar. Si se toma una muestra aleatoria Y_1, \dots, Y_n de $Y|\mu$, la función de verosimilitud es:

$$\begin{aligned} L(\mathbf{y}|\mu) &= \prod_{i=1}^n f(y_i|\mu) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(y_i-\mu)^2} \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i-\mu)^2}. \end{aligned}$$

Dado que σ^2 es conocido, en la inferencia clásica se buscaría obtener el estimador de máxima verosimilitud de μ al optimizar

$$\ell(\mathbf{y}|\mu) = \log L(\mathbf{y}|\mu),$$

cuya solución viene dada por la media de \mathbf{y} .

Por otro lado, desde un punto de vista Bayesiano, hay que considerar una distribución a priori para μ . Como $-\infty < \mu < \infty$ se puede proponer cualquier modelo probabilístico en ese rango, como una distribución t-Student, normal, logística, Cauchy, etc.

Si consideramos por ejemplo $\mu \sim Cauchy(0, 1)$, entonces

$$f(\mu) \propto \frac{1}{1 + \mu^2}$$

y consecuentemente la distribución a posteriori posee el núcleo siguiente:

$$f(\mu|\mathbf{y}) \propto L(\mathbf{y}|\mu)f(\mu) = \frac{e^{-\frac{1}{2} \sum_{i=1}^n (y_i - \mu)^2}}{1 + \mu^2},$$

donde como es usual hemos omitimos todo lo que no dependa de μ . Esta no corresponde a distribución conocida alguna y por lo tanto la posibilidad de realizar inferencias en base a ella se hace complicada. Para salvar esta situación haremos uso de métodos de simulación estocástica conocidos como cadenas de Markov Montecarlo o brevemente MCMC.

3.3 Métodos de MCMC

Como hemos visto en este capítulo, el paradigma Bayesiano centra su interés en la distribución a posteriori $f(\theta|\mathbf{y})$. Ella contiene toda la información relevante del parámetro desconocido θ dada la data observada \mathbf{y} . Así, toda inferencia estadística puede deducirse de la distribución a posteriori al considerarse algún resumen adecuado. Tales resúmenes toman típicamente la siguiente forma integral:

$$I = \int g(\theta)f(\theta|\mathbf{y})d\theta. \quad (3.1)$$

Por ejemplo, si se tiene interés en estimadores puntuales del parámetro desconocido θ uno puede utilizar la media a posteriori, que no es sino un caso particular de (3.1) de tomarse $g(x) = x$. Otro interés podría centrarse en predecir algún valor futuro \tilde{y} en base a la distribución predictiva a posteriori

$$f(\tilde{y}|\mathbf{y}) = \int f(\tilde{y}|\theta, \mathbf{y})f(\theta|\mathbf{y})d\theta,$$

la cual no es sino un caso particular de (3.1) con $g(x) = f(\tilde{y}|x, \mathbf{y})$.

El problema con (3.1) es que, como en el caso del ejemplo 2, es usualmente muy complicado o imposible evaluar I . Incluso las técnicas numéricas de cuadratura u otras para aproximar (3.1) podrían presentar problemas, más aún si el parámetro θ es multidimensional.

Durante los últimos años una gran cantidad de artículos han aparecido en relación a la evaluación de (3.1) por métodos de simulación colectivamente conocidos como cadenas de Markov de Montecarlo (MCMC). 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

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

converge a la integral deseada en (3.1). 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. El siguiente ejemplo nos ilustra su aplicación.

Ejemplo 3.3.1. (Gelfand y Smith, 1990) Consideremos un modelo de análisis de varianza de efectos aleatorios:

$$y_{ij} = \tau_i + \epsilon_{ij},$$

donde $i = 1, 2, \dots, k$, $j = 1, 2, \dots, n$, los errores $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$ se asumen independientes entre si y los efectos aleatorios $\tau_i \sim N(\mu, \sigma_\tau^2)$ se asumen también independientes entre si y estos de los errores. Si asumimos 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 también inversas Gaussianas y las distribuciones a posteriori de $\mu | \mathbf{y}, \tau, \sigma_\tau^2, \sigma_\epsilon^2$ y $\tau | \mathbf{y}, \mu, \sigma_\tau^2, \sigma_\epsilon^2$ son normales, donde $\mathbf{y} = (y_{ij})$ denota al vector de la data y τ al vector de efectos de los k tratamientos.

Para estimar los momentos a posteriori del tipo (3.1) definiremos una cadena de Markov para el parámetro de interés $\theta = (\mu, \tau, \sigma_\tau^2, \sigma_\epsilon^2)$. 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 θ dada la data a través de los siguientes pasos, donde por \sim estamos denotando que las variables condicionadas en el lado izquierdo tiene las mismas distribución normales e inversas Gaussianas que las ya comentadas del lado derecho:

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

Los pasos 1 al 4 definen una cadena de Markov $\{\theta_n\}$ cuya distribución converge a la deseada $f(\theta | \mathbf{y})$. Los promedios ergódicos del tipo

$\hat{I} = \frac{1}{M} \sum_{j=1}^M g(\theta_j)$ nos proveen luego de una aproximación numérica de la integral a posteriori (3.1).

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 dimensión $j = 1, 2, \dots, p$ genera las distribuciones condicionales a posteriori de

$$\theta_{j,n+1} \sim \theta_j | \mathbf{y}, \theta_{1,n+1}, \dots, \theta_{j-1,n+1}, \theta_{j+1,n}, \dots, \theta_{p,n}. \quad (3.2)$$

El método de Gibbs debe su popularidad al hecho de que en muchos modelos estadísticos la distribución condicional a posteriori completa $f(\theta_j | \mathbf{y}, \theta_k, k \neq j)$ se puede simular al ser explícita. Ocurren, sin embargo, casos en donde esto no es posible y por ello se hace necesario 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 lo mantiene sin cambios.

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 el parámetro 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 probabilidad $\pi(B) > 0$, 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 (3.1). 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 que se pueden indicar son Chen, Shao, Ibrahim (2000), Garmerman y Freitas (2006).

4 Las priors en TRI

4.1 Nociones básicas

Denotemos por θ a un vector no observable de cantidades o *parámetros* poblacionales de interés y sea \mathbf{y} el vector de valores observados.

Las conclusiones estadísticas bayesianas acerca de θ , son hechas en términos de *proposiciones probabilísticas*, las cuales están condicionadas a los valores observados de \mathbf{y} . Es en este nivel fundamental de condicionamiento de los datos observados que la inferencia bayesiana se aparta de la inferencia estadística clásica que está basada en una evaluación retrospectiva del procedimiento usado para estimar θ sobre la distribución de posibles valores de \mathbf{y} condicionada al valor verdadero pero desconocido de θ .

Con el propósito de hacer proposiciones probabilísticas acerca de θ dado \mathbf{y} , debemos empezar con *un modelo* que proporciona la *distribución de probabilidad conjunta* para θ e \mathbf{y} . La función de densidad (y/o probabilidad) conjunta puede escribirse como el producto de dos densidades que a menudo se refieren como la *densidad a priori* $f(\theta)$ y la *densidad muestral* (o *verosimilitud*) $f(\mathbf{y} | \theta)$:

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

Condicionando simplemente a los valores conocidos de los datos \mathbf{y} y usando la propiedad básica de probabilidad condicional conocida como regla de Bayes, obtenemos la densidad *a posteriori*

$$f(\theta | \mathbf{y}) = \frac{f(\theta, \mathbf{y})}{f(\mathbf{y})} = \frac{f(\mathbf{y} | \theta)f(\theta)}{f(\mathbf{y})} \quad (4.1)$$

donde $f(\mathbf{y}) = \int_{\theta} f(\mathbf{y} | \theta)f(\theta)d\theta$. Como anteriormente vimos, una forma equivalente de la expresión anterior es omitir el factor $f(\mathbf{y})$, que no depende de θ . Esta se denomina la *densidad a posteriori no normalizada*. En tal caso escribiremos:

$$f(\theta|\mathbf{y}) \propto f(\mathbf{y} | \theta)f(\theta) \quad (4.2)$$

Cuando la distribución a priori no tiene base poblacional, ella pueden ser difícil de construir. En tal sentido si uno desea que esta priori juegue un rol mínimo en la distribución a posteriori, uno llega a las llamadas *distribuciones a priori de referencia*, donde su densidad es descrita como vaga, plana, difusa o *no informativa*. La racionalidad para usar prioris no informativas es a menudo justificada diciendo que los *datos hablen por sí mismos* desde que la distribución a posteriori de θ depende apenas de $f(\mathbf{y} | \theta)$ y no de información externa a los datos recientes dada por la priori $f(\theta)$.

Si $f(\theta) \propto c$, con c constante y $\theta \in] - \infty, \infty[$, tal distribución no es estrictamente posible, pues su integración es infinita. En general, llamaremos a la densidad de la priori $f(\theta)$ como *propia* si no depende de los datos y su integral da 1. En el caso de una integral no finita diremos que la priori es *impropia*. No obstante, una priori impropia puede llevar a una distribución a posteriori propia.

En casos más complejos, donde por citar $\theta = (\theta_1, \theta_2)$ es un vector bivariado, es importante estudiar bajo que condiciones la distribución a posteriori $f(\theta | \mathbf{y})$ es propia. Para esto consideraremos las siguientes definiciones preliminares.

Definición 4.1.1. Sea \mathbf{Y} un vector aleatorio con función de verosimilitud $f(\mathbf{y} | \theta_1, \theta_2)$. Diremos que \mathbf{Y} es *no identificable para θ_2* si $f(\mathbf{y} | \theta_1, \theta_2) = f(\mathbf{y} | \theta_1), \forall \mathbf{y}$.

Definición 4.1.2. Si \mathbf{Y} es un vector aleatorio con función de verosimilitud $f(\mathbf{y} | \theta_1, \theta_2)$. Diremos que θ_2 es *no identificable* si $f(\theta_2 | \theta_1, \mathbf{y}) = f(\theta_2 | \theta_1), \forall \mathbf{y}$.

Proposición 4.1.1. *Las definiciones arriba son equivalentes.*

Demostración. Usando la definición de probabilidad condicional y el hecho de que θ_2 es no identificable podemos escribir

$$f(\mathbf{y} | \theta_1, \theta_2) = \frac{f(\mathbf{y}, \theta_1, \theta_2)}{f(\theta_1, \theta_2)} = \frac{f(\theta_2 | \mathbf{y}, \theta_1)f(\mathbf{y}, \theta_1)}{f(\theta_2 | \theta_1)f(\theta_1)} = f(\mathbf{y} | \theta_1).$$

Análogamente, usando la definición de probabilidad condicional y el hecho que \mathbf{Y} tiene una función de densidad no identificable para θ_2 podemos escribir

$$f(\theta_2 | \mathbf{y}, \theta_1) = \frac{f(\mathbf{y}, \theta_1, \theta_2)}{f(\theta_1, \mathbf{y})} = \frac{f(\mathbf{y} | \theta_1, \theta_2)f(\theta_1, \theta_2)}{f(\mathbf{y} | \theta_1)f(\theta_1)} = f(\theta_2 | \theta_1).$$

■

El significado de la no identificabilidad para θ_2 es que los datos observados no incrementan el conocimiento a priori acerca de $\theta_2 | \theta_1$ o que la función muestral no dependa de θ_2 .

Lema 4.1.1. *(Ghosh et al. 2000)*

Sea \mathbf{Y} un vector de observaciones con verosimilitud $f(\mathbf{y}|\theta_1, \theta_2)$. Si \mathbf{Y} es no identificable para θ_2 , entonces la distribución a posteriori de $\theta = (\theta_1, \theta_2)$ es propia si y solamente si las distribuciones $\theta_1|\mathbf{y}$ y $\theta_2|\theta_1$ son ambas propias.

Demostración. Usando (4.2), la hipótesis y la definición de probabilidad condicional tenemos

$$f(\theta_1, \theta_2|\mathbf{y}) \propto f(\mathbf{y}|\theta_1, \theta_2)f(\theta_1, \theta_2) = f(\mathbf{y}|\theta_1)f(\theta_2|\theta_1)f(\theta_1) = f(\theta_2|\theta_1)f(\theta_1|\mathbf{y}).$$

De aquí sigue que para que $\theta|\mathbf{y}$ sea propia $\theta_1|\mathbf{y}$ y $\theta_2|\theta_1$ deben ser propias. Basta que una de ellas sea impropia para obtener que $(\theta_1, \theta_2)|\mathbf{y}$ es impropia. ■

4.2 Prioris no informativas para el modelo TRI de 1 y 2 parámetros

Teorema 4.2.1. (Ghosh et al. 2000)

Considere un modelo TRI de un parámetro. Si la distribución a priori para las habilidades de los n sujetos y los parámetros de dificultad de los k ítems (\mathbf{U}, \mathbf{b}) es no informativa y proporcional a una constante, entonces la distribución a posteriori para (\mathbf{U}, \mathbf{b}) es impropia.

Demostración. Defina $U_i^c = U_i - b_1$, $\forall i = 1, \dots, n$, y $b_j^c = b_j - b_1$, $\forall j = 2, \dots, k$. Estas son transformaciones lineales 1-1. Si escribimos $\mathbf{U}^c = (U_1^c, \dots, U_n^c)$ y $\mathbf{b}^c = (b_2^c, \dots, b_k^c)$, entonces $(\mathbf{U}^c, \mathbf{b}^c, b_1)$ es una transformación 1-1 de (\mathbf{U}, \mathbf{b}) . Como la transformación es lineal el Jacobiano de la transformación de (\mathbf{U}, \mathbf{b}) a $(\mathbf{U}^c, \mathbf{b}^c, b_1)$ es constante libre de cualquier parámetro. Entonces $f(\mathbf{u}^c, \mathbf{b}^c, b_1) \propto c$ con c una constante. De esta manera la distribución a posterior de $f(b_1 | \mathbf{u}^c, \mathbf{b}^c) \propto c$. Como b_1 tiene distribución sobre \mathbb{R} , se sigue que $f(b_1 | \mathbf{u}^c, \mathbf{b}^c)$ es impropia.

Por otro lado, la función de verosimilitud del modelo de un parámetro puede ser reformulada como:

$$f(\mathbf{y} | \mathbf{u}^c, \mathbf{b}^c, b_1) = \prod_{i=1}^n \prod_{j=2}^k F(u_i^c - b_j^c)^{y_{ij}} \bar{F}(u_i^c - b_j^c)^{1-y_{ij}} \prod_{i=1}^n F(u_i^c)^{y_{i1}} \bar{F}(u_i^c)^{1-y_{i1}},$$

donde $\bar{F}(\cdot) = 1 - F(\cdot)$.

Nótese que para este modelo b_1 es no identificable; es decir,

$$f(\mathbf{y} | \mathbf{u}^c, \mathbf{b}^c, b_1) = f(\mathbf{y} | \mathbf{u}^c, \mathbf{b}^c).$$

Considere $\theta_2 = b_1$ y $\theta_1 = (\mathbf{u}^c, \mathbf{b}^c)$. Aplicando el lema 4.1.1 tenemos finalmente el resultado deseado desde que $f(\theta_2 | \theta_1)$ es impropia. ■

Teorema 4.2.2. Considere un modelo TRI de 2 parámetros. Si la distribución a priori para $(\mathbf{U}, \mathbf{a}, \mathbf{b})$ es no informativa y proporcional a una constante, entonces la distribución a posteriori de $(\mathbf{U}, \mathbf{a}, \mathbf{b})$ es impropia.

Demostración. Defina $U_i^c = U_i - b_1$, $i = 1, \dots, n$, y $b_j^c = b_j - b_1$, $j = 2, \dots, k$ transformaciones lineales 1-1. Si escribimos $\mathbf{U}^c = (U_1^c, \dots, U_n^c)$ y $\mathbf{b}^c = (b_2^c, \dots, b_k^c)$, entonces $(\mathbf{U}^c, \mathbf{a}, \mathbf{b}^c, b_1)$ es una transformación 1-1 de $(\mathbf{U}, \mathbf{a}, \mathbf{b})$. Como la transformación es lineal el Jacobiano de la transformación de $(\mathbf{U}, \mathbf{a}, \mathbf{b})$ a $(\mathbf{U}^c, \mathbf{a}, \mathbf{b}^c, b_1)$ es constante libre de cualquier parámetro. Entonces $f(\mathbf{u}^c, \mathbf{a}, \mathbf{b}^c, b_1) \propto c$ con c constante. De esta manera la distribución a posterior de $f(b_1 | \mathbf{u}^c, \mathbf{a}, \mathbf{b}^c) \propto c$. Como b_1 tiene distribución sobre \mathbb{R} , se sigue que $f(b_1 | \mathbf{u}^c, \mathbf{a}, \mathbf{b}^c)$ es impropia.

Por otro lado, la función de verosimilitud del modelo de 2 parámetros puede ser reescrita como:

$$f(\mathbf{y} | \mathbf{u}^c, \mathbf{a}, \mathbf{b}^c, b_1) = \prod_{i=1}^n \prod_{j=2}^k F[a_j(u_i^c - b_j^c)]^{y_{ij}} \bar{F}[a_j(u_i^c - b_j^c)]^{1-y_{ij}} \times \prod_{i=1}^k F[a_1 u_i^c]^{y_{i1}} \bar{F}[a_1 u_i^c]^{1-y_{i1}},$$

donde $\bar{F}(\cdot) = 1 - F(\cdot)$.

Nótese que para el modelo TRI de 2 parámetros, b_1 es no identificable. Es decir

$$f(\mathbf{y} | \mathbf{u}^c, \mathbf{a}, \mathbf{b}^c, b_1) = f(\mathbf{y} | \mathbf{a}, \mathbf{u}^c, \mathbf{b}^c).$$

Considere $\theta_2 = b_1$ y $\theta_1 = (\mathbf{U}^c, \mathbf{a}, \mathbf{b}^c)$. Aplicando el lema 4.1.1 tenemos finalmente el resultado deseado desde que $f(\theta_2 | \theta_1)$ es impropia. ■

Teorema 4.2.3. (Albert y Ghosh, 2000) *Considere un modelo TRI de 2 parámetros. Si la distribución a priori para al menos un (a_j) es impropia, entonces la distribución a posteriori de $(\mathbf{U}, \mathbf{a}, \mathbf{b})$ es impropia.*

Demostración. La prueba en detalle es presentada en Albert y Ghosh (2000) y consiste en probar en primer lugar que para algún a_j con distribución impropia la distribución posterior de $f(\mathbf{a}, \mathbf{b} | \mathbf{y})$ es impropia. De aquí $f(\mathbf{a}, \mathbf{b}, \mathbf{u} | \mathbf{y})$ es también impropia. ■

Es interesante notar, sin embargo, que todas las posteriores condicionales completas son propias cuando se asume al menos un a_j impropia. De esta manera esta propiedad de la distribución a posteriori no podría ser detectada en la implementación del algoritmo de Gibbs.

Por otro lado, algunos autores sugieren que prioris no informativas para \mathbf{U} , \mathbf{a} y \mathbf{b} hacen que el análisis Bayesiano sea comparable al análisis de máxima verosimilitud. Pero en el análisis Bayesiano esto es altamente cuestionable, porque la posterior es impropia en este caso. Esto es consecuencia de un resultado más general establecido abajo por Albert y Ghosh (2000) pero no probado.

Teorema 4.2.4. *(Albert y Ghosh, 2000) Considere un modelo TRI de 2 parámetros. Si la distribución a priori $f(\mathbf{U}, \mathbf{a}) \propto g(\mathbf{a})$, donde g es una función positiva arbitraria de \mathbf{a} (por ejemplo g puede ser una función de densidad propia para \mathbf{a}). Entonces la distribución posterior es siempre impropia.*

4.3 Prioris no informativas para los modelo TRI de 1 y 2 parámetros con puntajes extremos

Definición 4.3.1. Considerando una prueba de k items respondida por n sujetos, definimos $y_{i\bullet} = \sum_{j=1}^k y_{ij}$ y $y_{\bullet j} = \sum_{i=1}^n y_{ij}$ como respectivamente los puntajes del sujeto i y del ítem j . El *puntaje extremo de sujeto* ocurre cuando existe un sujeto i para el cual se cumple que $y_{i\bullet} = k$ ó $y_{i\bullet} = 0$. Así también el *puntaje extremo de ítem* ocurre cuando existe un ítem j para el cual se cumple que $y_{\bullet j} = n$ ó $y_{\bullet j} = 0$.

Teorema 4.3.1. *(Ghosh et al. 2000)*

Considere un modelo TRI de un parámetro. Si la distribución a priori para $(\mathbf{U}^c, \mathbf{b}^c)$ es no informativa, proporcional a una constante y existe al menos un puntaje extremo (de ítem o sujeto), entonces la distribución posterior de $(\mathbf{U}^c, \mathbf{b}^c)$ es impropia.

Demostración. Sin pérdida de generalidad asumamos que el sujeto h tiene puntaje extremo $y_{h\bullet} = 0$ lo cual implica que su puntaje en cada ítem es cero ($y_{hj} = 0$ para todo $j = 1, \dots, k$). Ahora

$$L(\mathbf{b}^c|\mathbf{y}) = \int_{-\infty}^{\infty} \prod_{j=2}^k \overline{F}(u_h^c + b_j^c) \overline{F}(u_h^c) du_h^c \geq \int_{-\infty}^0 \prod_{j=2}^k \overline{F}(b_j^c) \overline{F}(0) du_h^c = \infty.$$

Similarmente si el sujeto h tiene puntaje extremo $y_{h\bullet} = k$, lo cual implica que su puntaje en cada ítem es uno ($y_{hj} = 1$ para todo $j = 1, \dots, k$), se tiene que

$$L(\mathbf{b}^c|\mathbf{y}) = \int_{-\infty}^{\infty} \prod_{j=2}^k F(u_h^c + b_j^c) F(u_h^c) du_h^c \geq \int_0^{\infty} \prod_{j=2}^k F(b_j^c) F(0) du_h^c = \infty.$$

Por otro lado, considere que el ítem k tiene puntaje extremo $y_{\bullet k} = 0$ lo cual implica que el puntaje de cada sujeto en el ítem k es cero ($y_{ik} = 0$ para todo $i = 1, \dots, n$). Ahora

$$L(\mathbf{u}^c|\mathbf{y}) = \int_{-\infty}^{\infty} \prod_{i=1}^n \overline{F}(u_i^c + b_k^c) db_k^c \geq \prod_{i=1}^n \overline{F}(u_i^c) \int_{-\infty}^0 db_k^c = \infty.$$

Finalmente si $y_{\bullet k} = n$ lo cual implica que el puntaje de cada sujeto en el ítem k es uno ($y_{ik} = 1$ para todo $i = 1, \dots, n$), se tiene que

$$L(\mathbf{u}^c|\mathbf{y}) = \int_{-\infty}^{\infty} \prod_{i=1}^n F(u_i^c + b_k^c) db_k^c \geq \prod_{i=1}^n F(u_i^c) \int_0^{\infty} db_k^c = \infty.$$

En consecuencia $f(\mathbf{u}^c, \mathbf{b}^c|\mathbf{y})$ es impropia. ■

Teorema 4.3.2. (Ghosh et al. 2000)

Considere un modelo TRI de un parámetro. Si la distribución a priori para $(\mathbf{U}^c, \mathbf{b}^c)$ es no informativa, proporcional a una constante, no existe puntajes extremos (de ítem o sujeto) y si la la función de distribución usada para la curva característica del ítem satisface $\int_{-\infty}^{\infty} |z|^{n+k-1} dF(z) \leq \infty$, entonces la distribución posterior de $(\mathbf{U}^c, \mathbf{b}^c)$ es propia.

Demostración. Remitimos al lector a revisar la prueba realizada en Ghosh et al (2000) ■

El tercer supuesto del teorema anterior, relativa a las funciones usadas para las curvas características de ítem, ocurre automáticamente para cualquier distribución con momentos finitos. De esta manera, el teorema se cumple para los enlaces logit y probit. El teorema es también aplicable a distribuciones t con grados de libertad que exceden $n + k + 1$. Para otros enlaces como los discutidos en el capítulo final de este documento así como para la de los modelos de 2 y 3 parámetros se debe establecer nuevos teoremas acerca de la existencia de distribuciones a posteriori propias bajo prioris impropias. En general los teoremas mostrados indican que es recomendable considerar prioris propias para los parámetros de ítem y la variable latente, especialmente en el caso de la priori para el parámetro de discriminación. Además como ha sido comentado en el capítulo 2, prioris propias aseguran la identificabilidad del modelo y evitan el problema de presencia de parámetros incidentales.

5 Inferencia Bayesiana en TRI

5.1 La distribución a posteriori en TRI

Como recordamos, la función de verosimilitud para un modelo TRI estándar viene dado por

$$f(\mathbf{y}|\mathbf{u}, \boldsymbol{\eta}) = \prod_{i=1}^n \prod_{j=1}^k F(m_{ij})^{y_{ij}} (1 - F(m_{ij}))^{1-y_{ij}}, \quad (5.1)$$

donde $F(\cdot)$ podría denotar a la función de distribución acumulada de una variable normal estándar o logística.

En este capítulo consideraremos la siguiente clase general de distribuciones a priori para los parámetros del modelo:

$$f(\mathbf{u}, \boldsymbol{\eta}) = \prod_{i=1}^n g_{1i}(u_i) \prod_{j=1}^k g_{2j}(\eta_j) \quad (5.2)$$

donde $g_{1i}(u_i)$, con $i = 1, \dots, n$, denotará a la función de densidad de una variable normal estándar y $g_{2j}(\eta_j) = g_{21j}(a_j)g_{22j}(b_j)$, con $j = 1, \dots, k$, estará constituida por distribuciones propias tanto para g_{21j} como para g_{22j} a fin de garantizar, como se vio en el capítulo anterior, distribuciones a posteriori propias.

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

$$f(\mathbf{u}, \boldsymbol{\eta}|\mathbf{y}) \propto f(\mathbf{y}|\mathbf{u}, \boldsymbol{\eta}) \times f(\mathbf{u}, \boldsymbol{\eta}). \quad (5.3)$$

No obstante, como se estudió antes, esta distribución no es en general identificable. De esta manera se requiere de métodos MCMC. En lo que resta ilustraremos detalladamente la inferencia Bayesiana usando métodos MCMC para TRI para el caso del modelo de ojiva normal.

5.2 Inferencia Bayesiana en el modelo de ojiva normal

5.2.1 Especificación del modelo

La función de verosimilitud para el modelo *de ojiva normal* viene dada por:

$$f(\mathbf{y}|\mathbf{u}, \boldsymbol{\eta}) = \prod_{i=1}^n \prod_{j=1}^k \Phi(m_{ij})^{y_{ij}} (1 - \Phi(m_{ij}))^{1-y_{ij}},$$

donde Φ es la función de distribución acumulada normal estándar.

Si en la clase de prioris fijada consideramos (como por ejemplo en Rupp et al. , 2004) tomar g_{21j} como la densidad correspondiente a una variable normal con media μ_a y varianza σ_a^2 y a g_{22j} como la densidad correspondiente a una variable normal con media 0 y varianza σ_b^2 para todo $j = 1, \dots, k$, se tiene que g_{2j} sigue una distribución normal bivariada con vector de medias μ_η y matriz de varianza-covarianza Σ_η dadas por:

$$\mu_\eta = (\mu_a, 0)' \quad \text{y} \quad \Sigma_\eta = \begin{bmatrix} \sigma_a^2 & 0 \\ 0 & \sigma_b^2 \end{bmatrix}.$$

La distribución conjunta a posteriori viene dada entonces por:

$$f(\mathbf{u}, \boldsymbol{\eta}|\mathbf{y}) \propto \prod_{i=1}^n \prod_{j=1}^k \Phi(m_{ij})^{y_{ij}} (1 - \Phi(m_{ij}))^{1-y_{ij}} \prod_{j=1}^k \left(g_{1i}(u_i) \prod_{i=1}^n g_{2j}(a_j, b_j) \right)$$

o más explícitamente por:

$$f(\mathbf{u}, \boldsymbol{\eta}|\mathbf{y}) \propto \prod_{i=1}^n \prod_{j=1}^k \Phi(m_{ij})^{y_{ij}} (1 - \Phi(m_{ij}))^{1-y_{ij}} \times \\ \exp \left[-\frac{1}{2} \left(\sum_{i=1}^n u_i^2 + \frac{1}{\sigma_b^2} \sum_{j=1}^k b_j^2 + \frac{1}{\sigma_a^2} \sum_{j=1}^k (a_j - \mu_a)^2 \right) \right].$$

En base a esta distribución conjunta no es difícil encontrar las siguientes distribuciones condicionales completas:

$$f(u_i|\boldsymbol{\eta}, \mathbf{y}) = \prod_{j=1}^k \Phi(m_{ij})^{y_{ij}} (1 - \Phi(m_{ij}))^{1-y_{ij}} \phi(u_i; 0, 1)$$

y

$$\pi(\boldsymbol{\eta}_j|\mathbf{u}, \mathbf{y}) = \prod_{i=1}^n \Phi(m_{ij})^{y_{ij}} (1 - \Phi(m_{ij}))^{1-y_{ij}} \phi(a_j; \mu_a, \sigma_a^2) \phi(b_j; 0, \sigma_b^2),$$

donde por $\phi(\cdot; \mu, \sigma^2)$ estamos denotando a la función de densidad de una variable aleatoria normal con media μ y varianza σ^2 .

5.2.2 Esquema MCMC usando adaptative rejection sampling (ARS)

El modelo de ojiva normal puede ser ajustado usando MCMC. Note que todas las distribuciones condicionales previamente indicadas no son estándares. Por este hecho es difícil implementar un muestreo de Gibbs. Sin embargo, dado que las distribuciones condicionales completas del modelo son log-concávas (es decir, el logaritmo de las densidades correspondientes son funciones cóncavas), el muestreo exacto de ellas puede ser ejecutado usando rejection sampling (ver e.g. Gilks and Wald (1982)). Esos autores desarrollan también un esquema adaptative rejection sampling (ARS). El ARS construye dinámicamente dos envolventes (una superior y una inferior) para la distribución a ser muestreada de evaluaciones sucesivas de la densidad en los puntos rechazados. El algoritmo se detiene cuando el punto propuesto ha sido aceptado. Este es el procedimiento por defecto en el software WinBUGS 1.3 para el modelo de ojiva normal.

5.2.3 Una formulación adecuada para el modelo TRI de ojiva normal

Teorema 5.2.1. *Para $i = 1 \dots, n$ sujetos que responden a $j = 1, \dots, k$ ítems de una prueba, el modelo TRI de ojiva normal puede representarse por:*

$$Y_{ij} = \begin{cases} 1, & \text{si } Z_{ij} > 0 \\ 0, & \text{si } Z_{ij} \leq 0, \end{cases} \quad (5.4)$$

donde

$$Z_{ij} = m_{ij} + e_{ij}$$

y $e_{ij} \sim N(0, 1)$.

Demostración. Basta observar que $p_{ij} = P(Y_{ij} = 1) = P(Z_{ij} > 0) = \Phi(m_{ij})$, $\forall i = 1 \dots, n$, $j = 1, \dots, k$. Luego, la estructura normal lineal de la variable latente auxiliar produce un modelo equivalente al modelo de ojiva normal. ■

5.2.4 MCMC usando un esquema de data augmentation Gibbs sampling (DAGS)

Teorema 5.2.2. *Para $i = 1 \dots, n$ sujetos que responden a $j = 1, \dots, k$ ítems de una prueba, el modelo TRI de ojiva normal (5.4) presenta las siguientes distribuciones condicionales completas:*

- $f(z_{ij} | u_i, \boldsymbol{\eta}_j, \mathbf{y}) \propto \phi(z_{ij}; m_{ij}, 1) I(z_{ij}, y_{ij})$.
- $f(u_i | \mathbf{z}_i, \boldsymbol{\eta}, \mathbf{y}) \propto \phi(u_i; m_{u_i}, v_{u_i})$, donde:

$$m_{u_i} = \frac{\sum_{j=1}^k a_j (z_{ij} + b_j)}{\sum_{j=1}^k a_j^2 + 1} \quad \text{y} \quad v_{u_i} = \frac{1}{\sum_{j=1}^k a_j^2 + 1}.$$

- $f(\boldsymbol{\eta}_j | \mathbf{u}, \mathbf{z}_j, \mathbf{y}) \propto \phi_2(\boldsymbol{\eta}_j; \mathbf{m}_{\boldsymbol{\eta}_j}, \mathbf{v}_{\boldsymbol{\eta}_j})$, donde

$$\mathbf{m}_{\boldsymbol{\eta}_j} = [\mathbf{W}'\mathbf{W} + \Sigma_{\boldsymbol{\eta}}^{-1}]^{-1} [\mathbf{W}'\mathbf{Z}_j + \Sigma_{\boldsymbol{\eta}}^{-1} \boldsymbol{\mu}_{\boldsymbol{\eta}}], \quad \mathbf{v}_{\boldsymbol{\eta}_j} = [\mathbf{W}'\mathbf{W} + \Sigma_{\boldsymbol{\eta}}^{-1}]^{-1},$$

$$\boldsymbol{\mu}_\eta = \begin{bmatrix} \mu_a \\ 0 \end{bmatrix}, \Sigma_\eta = \begin{bmatrix} \sigma_a^2 & 0 \\ 0 & \sigma_b^2 \end{bmatrix} \text{ y } \mathbf{W} = (\mathbf{u}, -\mathbf{1}) \text{ con } \mathbf{W}'_i = (u_i, -1).$$

Demostración. Usando la nueva formulación tenemos que la función de verosimilitud de datos aumentados, considerando a estos como $D = (\mathbf{Z}, \mathbf{y})$, viene dada por:

$$f(D|\boldsymbol{\mu}, \boldsymbol{\eta}) = \prod_{i=1}^n \prod_{j=1}^k \phi(z_{ij}; m_{ij}, 1) I(z_{ij}, y_{ij}),$$

donde $I(z_{ij}, y_{ij}) = I(z_{ij} > 0)I(y_{ij} = 1) + I(z_{ij} \leq 0)I(y_{ij} = 0)$, $\forall i = 1, \dots, n$ y $j = 1, \dots, k$. Así, la *distribución conjunta a posteriori completa* $f(\boldsymbol{\mu}, \boldsymbol{\eta}|D)$ es proporcional a:

$$\left[-\frac{1}{2} \left(\sum_{i=1}^n \sum_{j=1}^k (z_{ij} - m_{ij})^2 + \sum_{i=1}^n u_i^2 + \frac{1}{\sigma_b^2} \sum_{j=1}^k b_j^2 + \frac{1}{\sigma_a^2} \sum_{j=1}^k a_j^2 \right) \right] I(z_{ij}, y_{ij})$$

y por tanto las distribuciones condicionales completas siguen. ■

Con estas condicionales uno puede implementar fácilmente el muestreo de Gibbs. Rutinas en R (en MCMCpack de Martin y Quinn, 2003) y Matlab (Johnson e Albert, 1999) se disponen en la Web. En WinBUGS, la implementación de este procedimiento no es directa pues precisa de una correcta especificación de las variables indicadoras. Mayores detalles pueden seguirse en Bazán, Bolfarine, Leandro (2006). Un detalle de los códigos para estos procedimientos se muestra en el apéndice A y serán usados en la subsección 6.2.2. Los esquemas de simulación MCMC para los otros modelos tradicionales son similares al caso presentado. Para mayores detalles, véase Sahu (2002).

5.3 Criterios de comparación de modelos en el contexto Bayesiano

El desarrollo de métodos MCMC permite ajustar una gran cantidad de modelos para un mismo conjunto de datos y por tal motivo es necesario comparar modelos alternativos con el propósito de identificar un modelo apropiado que pueda describir adecuadamente estos datos. Existe una variedad de metodologías para comparar el ajuste de modelos Bayesianos alternativos. Se puede ver por ejemplo el trabajo Gilks *et al.* (1996). El principal criterio usado en este trabajo es el *criterio de información de desvío* (ó *DIC* de Deviance Information Criterion) propuesto por Spiegelhalter *et al.* (2002). EL *DIC* es útil aquí porque es relativamente fácil de calcular, interpretar y es satisfactorio para modelos jerárquicos complicados (Johnson, 2003) y está basado en la noción de “desvío Bayesiano” (Dempster, 1977).

El *desvío Bayesiano* (Dempster, 1977) para modelos TRI es definido, de manera general, usando la verosimilitud del modelo como

$$D(\boldsymbol{\eta}, \mathbf{u}) = -2\ln(f(\mathbf{y}|\mathbf{u}, \boldsymbol{\eta})) = -2 \sum_{i=1}^n \sum_{j=1}^k \ln P(Y_{ij} = y_{ij}|\mathbf{u}, \boldsymbol{\eta}). \quad (5.5)$$

Recordemos aquí que \mathbf{y} denota a las respuestas observadas (0 ó 1) de los n sujetos para los k ítems, $\boldsymbol{\eta}$ se asocia a los parámetros de los ítems y \mathbf{u} a las habilidades de los sujetos.

El desvío bayesiano es también una medida de bondad de ajuste. Dado un conjunto de modelos alternativo a ser comparados, el que mejor ajuste a los datos será aquel que tiene menor valor de desvío Bayesiano. Es interesante destacar que la distribución a posteriori puede también ser usada para evaluar el ajuste (relativo) de un modelo dado. Esto ha sido propuesto en varios trabajos, entre ellos Gilks *et al.* (1996) y Spiegelhalter *et al.* (1996). Aquí, la cantidad $E(D(\boldsymbol{\eta}, \mathbf{u}))$, está basada en la distribución a posteriori de $\boldsymbol{\eta}$ y \mathbf{u} y , es denominada el *esperado*

del desvío bayesiano a posteriori o desvío medio a posteriori (posterior mean of the deviance) también es una medida de bondad de ajuste. Esta esperanza puede ser aproximada usando el algoritmo MCMC a través de

$$E(D(\boldsymbol{\eta}, \mathbf{u})) \approx \frac{1}{G} \sum_{g=1}^G D(\boldsymbol{\eta}^g, \mathbf{u}^g), \quad (5.6)$$

donde el índice g representa la g -ésima realización simulada de un total de G realizaciones simuladas. Este estimador es denominado como $Dbar$ por Spiegelhalter *et al.* (2002) y es obtenido por defecto en WinBugs.

Por otro lado, Carlin y Louis (2000) y Brooks (2002) han sugerido penalizar $E(D(\boldsymbol{\eta}, \mathbf{u}))$ como en los casos del criterio de Información Bayesiana (Schwarz) o Criteiro de Akaike (Akaike, 1973). Esto lleva al Esperado del Criterio de Información de Akaike ($EAIC$) definido por

$$EAIC = E(D(\boldsymbol{\eta}, \mathbf{u})) + 2p, \quad (5.7)$$

y al esperado del Criterio de Información Bayesiana de Schwarz ($EBIC$) definido por

$$EBIC = E(D(\boldsymbol{\eta}, \mathbf{u})) + p \ln N \quad (5.8)$$

en el que p es el número de parámetros del modelo y N es el número total de observaciones. Estos criterios se pueden obtener de manera aproximada usando el algoritmo MCMC a través de

$$\widehat{EAIC} = Dbar + 2p, \quad \widehat{EBIC} = Dbar + p \ln N \quad (5.9)$$

Lamentablemente, en el caso de los modelos jerárquicos se tiene problemas con estos criterios, porque no es fácil definir p y N . Por ejemplo, en la TRI tenemos respuestas de k ítems para los examinados $i = 1, \dots, n$, entonces, $N = nk$ el total de observaciones ó $N = n$ es el total de examinados. Si las respuesta a los ítems en cada examinado son independientes, entonces la primera elección sería más apropiada; pero si las respuestas fueran correlacionadas en cada examinado, se puede escoger la última

alternativa. En la TRI se debería considerar el primer caso tomándose en cuenta que la independencia condicional se cumple. Sin embargo la situación real es algo intermedia. Similarmente, si consideramos una colección de efectos aleatorios como es el caso de las variables latentes en TRI, una para cada examinado, la pregunta es ¿ cómo esto puede contribuir con el número de parámetros p ?. Si los efectos aleatorios no tienen nada en común (i.e., estos se comportan esencialmente como efectos fijos), ellos deberían contribuir con n parámetros para cada p , pero si los datos (o las prioris) indican que todos ellos son esencialmente idénticos, ellos deberían contribuir con un poco más o un “número efectivo de parámetros” para el tamaño del modelo p .

Debido a este tipo de problemas Spiegelhalter *et al.* (2002) propusieron ρ_D como un valor apropiado para penalizar $E(D(\boldsymbol{\eta}, \mathbf{u}))$. El *DIC* es una modificación del esperado del desvío y es definido como

$$DIC = E(D(\boldsymbol{\eta}, \mathbf{u})) + \rho_D, \quad (5.10)$$

donde

$$\rho_D = E(D(\boldsymbol{\eta}, \mathbf{u})) - D(E(\boldsymbol{\eta}), E(\mathbf{u})). \quad (5.11)$$

es denominado como el *número de parámetros efectivos* (Spiegelhalter *et al.*, 1996) y $D(E(\boldsymbol{\eta}), E(\mathbf{u}))$ es el *desvío bayesiano de las medias a posteriori* o *desvío bayesiano de los esperados a posteriori* (deviance of posterior mean) obtenido considerando los valores medios o esperados a posteriori de los parámetros del modelo. Las esperanzas arriba están basadas en las distribuciones a posteriori de $\boldsymbol{\eta}$ y \mathbf{u} las que se pueden aproximar usando el algoritmo MCMC a través de

$$D(E(\boldsymbol{\eta}), E(\mathbf{u})) \approx D\left(\frac{1}{G} \sum_{i=1}^G \boldsymbol{\eta}^g, \frac{1}{G} \sum_{i=1}^G, \frac{1}{G} \sum_{i=1}^G \mathbf{u}^g\right), \quad (5.12)$$

donde como se ha indicado antes el índice g representa a g -ésima realización simulada de un total de G realizaciones simuladas. Esta estimación es denominado como *Dhat* en Spiegelhalter *et al.* (2002) y es obtenida

por defecto en el WinBugs.

El DIC penaliza el esperado del desvío bayesiano por la complejidad del modelo o número de parámetros efectivos representado por ρ_D . El DIC es aproximadamente análogo al AIC y $EAIC$ y puede aproximarse por:

$$\widehat{DIC} = Dbar - Dhat = Dbar + 2\rho_D, \quad (5.13)$$

donde valores bajos del DIC indican el mejor ajuste. Otra ventaja del DIC es que como está basado en la función de verosimilitud, esta puede ser descompuesta para cada observación, por lo que se puede implementar residuales bayesianos (Gelfand, 1996) para evaluar la bondad de ajuste local. La estimación del DIC se obtiene por defecto en el WinBugs.

Spiegelhalter *et al.* (2002) indican que el DIC implementado en WinBUGS puede ser usado para comparar modelos complejos (vea por ejemplo Johnson, 2003) y grandes diferencias en los criterios se pueden atribuir a diferencias predictivas reales en los modelos.

6 Aplicaciones

En este capítulo ilustraremos el enfoque Bayesiano para el modelo de TRI a través de dos ejemplos. En ellos discutiremos cómo especificar las priors, los valores iniciales para definir el estado inicial de la cadena de Markov y cómo realizar diagnósticos de convergencia en la implementación del Algoritmo MCMC. Más aún, en nuestra primera aplicación haremos una comparación con la teoría clásica de los tests.

6.1 Análisis de una escala de percepción de peso

Esta aplicación está basada en los datos de la investigación de Martínez, et.al. (2003), relativa a trastornos alimentarios, donde se aplicó el Inventario de Conductas Alimenticias de Zusman (2000) que mide el riesgo de trastorno alimentario (Anorexia, Bulimia, etc.) y factores asociados en adolescentes mujeres. Los datos provienen de una muestra probabilística de 2,141 alumnas distribuidas entre 18 colegios públicos y 8 privados, obtenidos a partir del Padrón de Centros Educativos 1999 del Ministerio de Educación.

La prueba tiene varias áreas o escalas y de ellas se ha seleccionado la llamada escala de peso, que registra la "percepción y preocupación por el peso, rutinas, temores, controles y conductas patológicas asociadas". Su importancia radica en que si bien para la adolescente en riesgo el peso se convierte en la manifestación cuantitativa del problema que la perturba, por razones culturales y de globalización, el peso es también una preocupación general de las adolescentes, tengan un trastorno o no. Para efecto de análisis las respuestas de las participantes han sido pasadas a formato de variable binaria Y , donde el valor 0 indica poca o nula frecuencia (Nunca, Casi nunca) y el valor 1 revela una mayor frecuencia (A veces, Casi siempre, Siempre) de la conducta registrada por el ítem. Los 15 ítems son:

01	Me preocupa mi peso.
02	Me peso todos los días.
03	Pienso que mi rendimiento escolar ha mejorado desde que he comenzado a bajar de peso.
04	Subo y bajo de peso con facilidad.
05	Acudo a médicos y/o centros especializados para bajar de peso.
06	Me gusta que mi ropa me quede suelta.
07	Siento angustia cuando subo de peso.
08	Tengo miedo de pesarme.
09	Siento que mis padres exageran cuando se preocupan por mi peso.
10	Mi familia está pendiente de lo que como.
11	Quiero bajar de peso.
12	Pienso que es mejor ser delgada que inteligente.
13	Controlo mi peso.
14	Las mujeres queremos estar más flacas.
15	Mis amigas me dicen que estoy muy delgada.

Para este tipo de data, donde no hay una "habilidad" por medir y la persona lo sabe, la concepción de acertar al azar no es aplicable. Por otra parte, no se puede garantizar que todos los ítems tienen similar discriminación a_j , y naturalmente, la "dificultad" b_j de cada ítem representa el mayor o menor riesgo asociado a la conducta específica que se mide. El modelo debe ser uno de dos parámetros y para efectos de este ejemplo, usaremos el modelo logístico.

Para la estimación de los parámetros de este modelo usaremos la metodología Bayesiana con ayuda del software WinBUGS como fue especificado en el capítulo 5. Siguiendo las recomendaciones de Patz y Junker (1999) y de Albert y Ghosh (2000), la secuencia jerárquica de distribuciones que se carga en WinBUGS es:

$$Y_{ij} \mid u_j, a_i, b_i \sim \text{Bernoulli}(p_{ij}),$$

donde p_{ij} viene dado por el modelo 2PL, $u_j \sim N(0, 1)$, $a_i \sim N(0, \sigma_a^2)$ y $b_i \sim \text{LogN}(0, \sigma_b^2)$. Aquí se tomó

$$\sigma_a = 2 \quad \text{y} \quad \sigma_b = \frac{1}{2},$$

pues el tamaño de muestra de 2,141 casos hacía poco sensibles los resultados a las especificaciones de σ_a y σ_b (Patz y Junker, 1999).

En la tabla 6.1 se presenta una comparación de los indicadores de riesgo y discriminación estimados según la teoría clásica (TCT) y la teoría de respuesta al ítem (TRI).

Item	TCT		IRT			
	Riesgo ítem	Índice de discriminación	Riesgo Item b_i		discriminación Item a_j	
			Media	D. Estándar	Media	D. Estándar
Item 01	0.67	0.61	-0.64	0.04	1.02	0.06
Item 02	0.12	0.20	1.23	0.04	0.37	0.04
Item 03	0.12	0.28	1.39	0.05	0.64	0.05
Item 04	0.47	0.52	0.08	0.03	0.53	0.04
Item 05	0.05	0.11	1.86	0.07	0.50	0.06
Item 06	0.51	0.42	-0.03	0.03	0.37	0.03
Item 07	0.54	0.83	-0.24	0.06	1.91	0.14
Item 08	0.42	0.72	0.31	0.04	1.08	0.06
Item 09	0.41	0.53	0.25	0.03	0.51	0.04
Item 10	0.59	0.38	-0.24	0.03	0.29	0.03
Item 11	0.62	0.76	-0.65	0.06	1.78	0.12
Item 12	0.19	0.35	1.01	0.04	0.59	0.04
Item 13	0.45	0.53	0.15	0.03	0.50	0.04
Item 14	0.86	0.27	-1.28	0.05	0.63	0.05
Item 15	0.78	0.24	-0.82	0.03	0.32	0.04

Tabla 6.1

Nótese que en el caso del modelo TRI es posible obtener además una medida de dispersión para cada estimación, porque la TRI provee de una distribución para cada parámetro. Una consecuencia es que con la TRI es posible obtener intervalos de confianza para esos valores, algo que facilita las comparaciones entre ítems, que es una tarea importante cuando se está en la fase de construcción de una prueba.

El gráfico 6.1 es una diagrama de dispersión que compara visualmente los indicadores de riesgo (proporción de positivos en la TCT y parámetro b_j en TRI) en sus respectivas escalas.

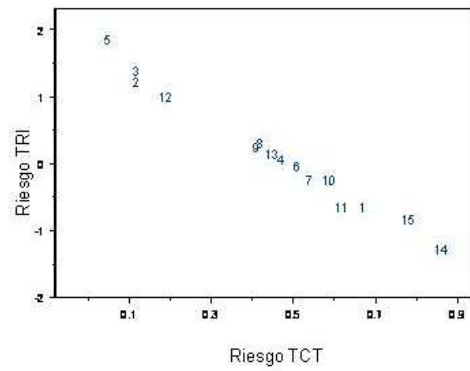


Gráfico 6.1
Riesgo según TRI vs riesgo según TCT

Análogamente, el gráfico 6.2 compara los índices de discriminación (d_j en TCT y a_j en TRI).

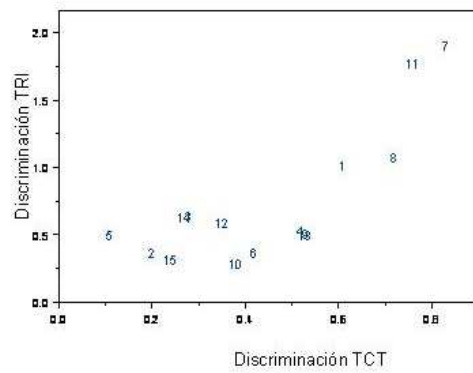


Gráfico 6.2
Discriminación según TRI vs discriminación según TCT

El gráfico 6.1 muestra que los valores del indicador de riesgo en TCT y TRI producen la misma jerarquización de los ítems y por tanto son igualmente interpretables. Destaca el ítem 5 (**Acudo a médicos y/o centros especializados para bajar de peso**) como un ítem referido a un mayor riesgo. El ítem 14 (**Las mujeres queremos estar más flacas**) es un ítem correspondiente a un menor riesgo de todos. Nótese que el modelo TRI permite diferenciar valores de riesgo que la TCT no logra. Así por ejemplo, los ítems 2 y 3 tienen igual riesgo TCT pero distinta media b_1 en TRI (ver los valores en la tabla 6.1)

La tabla 6.1 y el gráfico 6.2 muestran que los valores del indicador de discriminación en TCT y IRT no necesariamente producen la misma jerarquización de los ítems, así por ejemplo, en TCT el ítem 13 (**Controlo mi peso**) es más discriminador que el ítem 12 (**Pienso que es mejor ser delgada que inteligente**) pero en TRI es al revés. No se interpreta aquí la significación psicológica de estas frases, eso corresponde a un profesional del área, pero si se observan los riesgos, resulta que el ítem 12 también es más riesgoso y eso cuadra más con la idea extrema de preferir la delgadez a la inteligencia.

En general, aunque el modelo TRI presente la misma información del modelo TCT, provee al constructor del test de otras informaciones que TCT no tiene, por ejemplo, Intervalos de confianza para riesgos y capacidad de analizar la separación entre ítems.

La tabla 6.2 muestra las estadísticas de las puntuaciones en la Escala de percepción de peso, estimadas según la teoría clásica (denominadas scores en la psicometría) y según el modelo TRI (U). En ambos casos un mayor valor indica una mayor propensión al riesgo. Aunque las dos teorías generan puntuaciones, por construcción las del TRI están en la misma escala que la de los parámetros b_j de los ítems, cosa que no sucede en la TCT, donde los parámetros de los ítems son proporciones.

Estadísticas de percepción de peso		
	TCT	TRI
	Score	U
Mínimo	0	-2.00
Máximo	15	2.48
Media	6.81	0.00
D. Estándar	3.01	0.88
Asimetría	0.06	0.02
Kurtosis	-0.79	-0.78
Correlación	-0.96	

Tabla 6.2
Estadísticas de puntuaciones según modelos

El gráfico 6.3 compara las puntuaciones obtenidas según los dos modelos, proporcionando además intervalos de confianza de 95 % para las puntuaciones TRI.

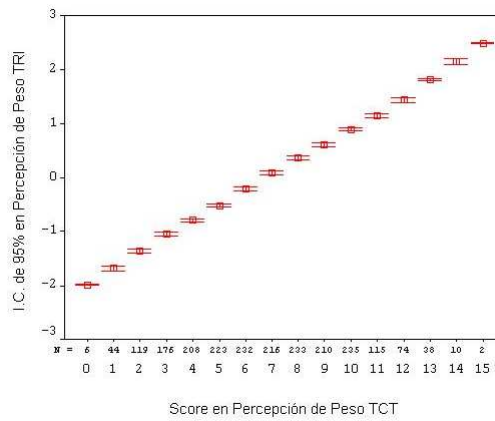


Gráfico 6.3
Puntuaciones en TCT vs puntuaciones TRI

De la tabla 6.2, se ve que los scores TCT son números enteros que van de 0 a 15, pues se obtienen como la suma simple de los 15 ítems binarios. En cambio las puntuaciones TRI forman un continuo. Lo anterior es una diferencia fundamental entre ambas teorías, pues la TCT obliga a que muchas personas obtengan igual puntuación, aún cuando tengan riesgos distintos, cosa que no sucede en la TRI.

El gráfico 6.2 muestra cómo personas con igual puntuación TCT sí llegan a ser diferenciadas con las puntuaciones TRI, por eso lo de los intervalos de confianza que ilustran lo que pasa. Aunque la correlación alta entre escalas indica que proporcionan prácticamente los mismos órdenes entre personas, la continuidad en el caso TRI permite diferenciar casos de modo muy fino, que es algo importante en el diagnóstico clínico.

La comparación de ítems puede hacerse visualmente via intervalos de confianza en diagramas de caja. Así el gráfico 6.4 muestra las distribuciones a posteriori de los parámetros de riesgo b_1 de los ítems (el riesgo va en el eje vertical) y los ubica mostrando su extensión y posibles traslapes. El gráfico 6.5 hace algo análogo con las discriminaciones.

Finalmente el gráfico 6.6 es como un mapa de ítems que los identifica en sus dos características importantes. Así tenemos por ejemplo, que el Item 14 es de menor riesgo y menor discriminación, que el ítem 5 es de mayor riesgo y poco discriminativo, mientras que el Item 7 es de riesgo mediano y bastante discriminativo.

Para esta prueba otros modelos alternativos pueden ser estudiados, como por ejemplo el modelo de Rasch. También los modelos estudiados en la sección 7 podrían utilizarse en esta aplicación. En general se pueden comparar los diferentes modelos propuesto en este trabajo desde la perspectiva Bayesiana empleando los diferentes criterios mostrados en la sección 5.3, como es el caso de la siguiente aplicación.

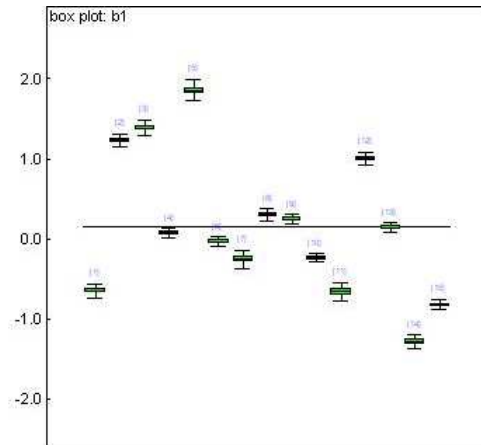


Gráfico 6.4
Diagrama de cajas de parámetros TRI de riesgo

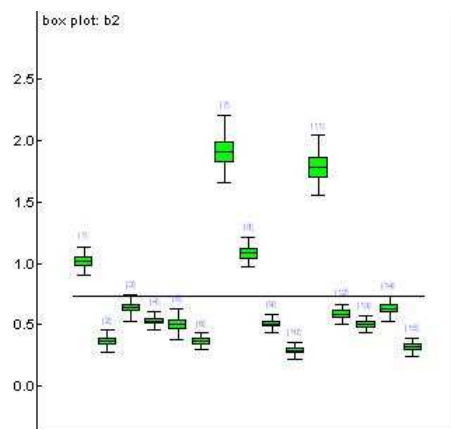


Gráfico 6.5
Diagrama de cajas de parámetros TRI de discriminación

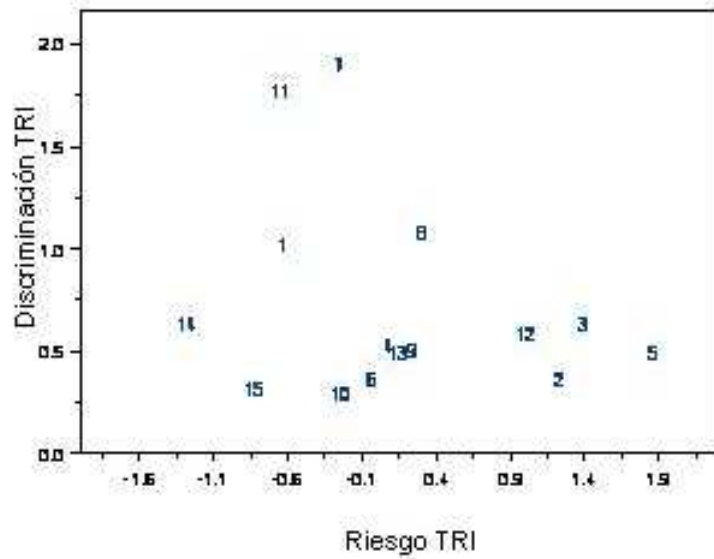


Gráfico 6.6
Diagrama de dispersión riesgo vs discriminación en TRI

6.2 Análisis de una prueba de Matemáticas

En esta aplicación, 14 ítems de una prueba de Matemáticas disponibles en www2.minedu.gob.pe/umc/admin/images/publicaciones/boletines/Boletin-13.pdf se aplicaron a 131 estudiantes de nivel socio económico alto de sexto grado de primaria. Las estadísticas de los puntajes de esta prueba se presentan en la tabla 6.3.

Estadísticas	valores	Estadísticas	valores
Media	10.84	Mediana	11
Varianza	3.432	Desv. Estd.	1.853
Asimetría	-0.795	Curtosis	0.449
Mínimo	5	Máximo	14
Alpha	0.481	Media P	0.774
Media Item-Tot.	0.364	Mean Biserial	0.572
Max Puntaje	10	N (Grupo alto)	52
Min Puntaje	12	N (Grupo bajo)	45

Tabla 6.3

Estadísticas de los puntajes de la prueba de Matemáticas con 14 ítems y 131 examinados

Los datos presentan un puntaje medio de 11 puntos (sobre un puntaje máximo de 20 puntos) y una desviación estándar de casi 2 puntos. De las estadísticas, se puede indicar que los puntajes presentan asimetría con dominio de puntajes altos. La prueba presenta una confiabilidad dada por el coeficiente alfa de Cronbach de 0.48.

En el anexo B de este trabajo, el lector podrá encontrar un tutorial del WinBUGS para el análisis de estos datos.

6.2.1 Comparando varios modelos TRI para los datos de la prueba de Matemáticas

Para comparar los diferentes modelos propuestos en este trabajo e ilustrar el uso del DIC, generamos 202000 iteraciones y descartamos los 2000 valores iniciales. Usando un thin de 100, se obtuvo un tamaño de muestra efectiva de 2000. Estimadores de los parámetros del modelo se calcularon de estas iteraciones. Varios criterios para evaluar la convergencia se calcularon, usando la librería CODA, entre ellos los propuestos por Geweke (1992).

Los valores de DIC se muestran en la tabla 6.4 para seis modelos TRI: 1P, 1L, 2P, 2L, 3P, 3L, donde el primer número indica si el modelo es de uno, dos ó tres parámetros, respectivamente; mientras que la última letra denota si el modelo es de ojiva normal (P) ó logístico (L). El modelo 2P se ajustó usando ARS y Gibbs Sampling. Así para este modelo observamos que la aproximación basada en datos aumentados (Gibbs Sampling) lleva menos tiempo que la aproximación MCMC basada en la verosimilitud original (ARS) lo que es coherente con un resultado similar obtenido por Sahu, 2002, con el modelo 3P. Note también que en general los modelos 1P, 2P y 3P son más rápidos que sus correspondientes modelos 1L, 2L y 3L. Además en ambos casos el tiempo de simulación se incrementa conforme el modelo incluye más parámetros. Considerando el valor de DIC, el mejor modelo ajustado resultó el modelo 3P pero sin embargo el modelo 2P es bastante cercano. Nosotros preferimos el modelo 2P por ser un modelo más simple.

models	type	time	parameters	$Dbar$	$Dhat$	ρ_D	DIC
(1)	1L	53	145	1467	1372	94.46	1561
(2)	1P	42	145	1460	1340	120.3	1581
(3)	2L	78	159	1461	1378	82.74	1544
(4)	2P	71	159	1447	1359	88.41	1536
(5)	3L	148	173	1464	1384	80.08	1544
(6)	3P	87	173	1443	1356	87.8	1531

Tabla 6.4
Comparación de modelos usando el criterio DIC

Vale aclarar, que el tiempo en la tabla se refiere al número de segundos necesarios para correr 2000 iteraciones en un PC con procesador Pentium IV de 1800 MHZ y 256 de memoria RAM.

6.2.2 Análisis de sensibilidad usando diferentes prioris para a y b en el modelo 2P

Para evaluar la sensibilidad de la estimación Bayesiana para el modelo de ojiva normal 2P se han tomado en cuenta diferentes prioris y se ha realizado un análisis considerando los datos descritos anteriormente. Los códigos MCMC requeridos son presentados en el anexo A.

La estimación bayesiana basada en MCMC fue implementada en WinBUGS. Se generaron cadenas con 50,000 iteraciones considerando saltos (thin)=1, 5, 10 y descartando las primeras 500 iteraciones (Burnin), así que los tamaños de muestra son de 49,500, 9,900 y 4,950, respectivamente. Cuando se usa MCMC, los valores de muestra para iteraciones iniciales de la cadena son descartados por su dependencia con los valores iniciales y para garantizar la convergencia. También, en este modelo TRI, se espera una presencia de autocorrelaciones entre los valores de la cadena debido a la presencia de variables latentes como lo apunta Chen et al. (2000). Debido a esto se recomienda usar valores de thin superiores a 10.

Como ha sido mencionado en el capítulo 4, prioris propias para a_j y b_j garantizan que las distribuciones a posteriori completas de los parámetros del modelo sean propias. Albert y Ghosh (2000) mencionan que la elección de prioris propias en las variables latentes resuelven el problema de identificación de estos modelos, y, adicionalmente, distribuciones a priori informativas para a_j y b_j se pueden usar para reflejar la creencia de que los valores de los parámetros de ítem no son extremos (no están en la frontera del espacio paramétrico). Si se está en la situación donde se dispone de poca información a priori acerca de los parámetros de dificultad, uno puede elegir varianzas para estos parámetros grandes. Esta elección puede tener un efecto modesto en la distribución a posteriori para datos no extremos, y puede resultar en una distribución a posteriori propia cuando hay datos extremos (cuando se observa estudiantes que tienen todos sus ítems correctos o incorrectos). (Albert y Ghosh, 2000) y también, Sahu (2002) establecen que valores grandes de la varianza

llevan a estimados no estables.

En la tabla 6.5 se muestra algunas prioris consideradas en la literatura para los parámetros de ítem en el modelo de ojiva normal. Aquí por $N(0,1)I(0,)$ estamos denotando a la distribución normal estándar que está truncada para valores negativos.

Priori	Autor	a prior	b prior
A	Jhonson y Albert (2000)	$N(2,1)$	$N(0,1)$
B	Congdon (2001)	$N(1,1)$	$N(0,1)$
C	Albert y Ghosh (2000)	$N(0,1)$	$N(0,1)$
D	Sahu (2002), Albert e Ghosh (2000)	$N(0,1)$	$N(0,10000)$
E	Spiegelhalter et al (1996)	$N(0,1)I(0,)$	$N(0,10000)$
F	Sahu (2002), Patz e Junker (1999)	$N(1,0.5)I(0,)$	$N(0,2)$

Tabla 6.5
Especificación de prioris para parámetros de ítem en el modelo de ojiva normal

En el gráfico 6.7 se presentan las estimaciones de los parámetros de discriminación y de dificultad para el modelo de ojiva normal para las diferentes prioris dadas en la tabla 6.5. Como se observa en este gráfico las prioris A, B y C son precisas y las prioris D y E tienen prioris difusas o prioris no informativas en el parámetro de dificultad. Las prioris E y F son truncadas en el parámetro de discriminación.

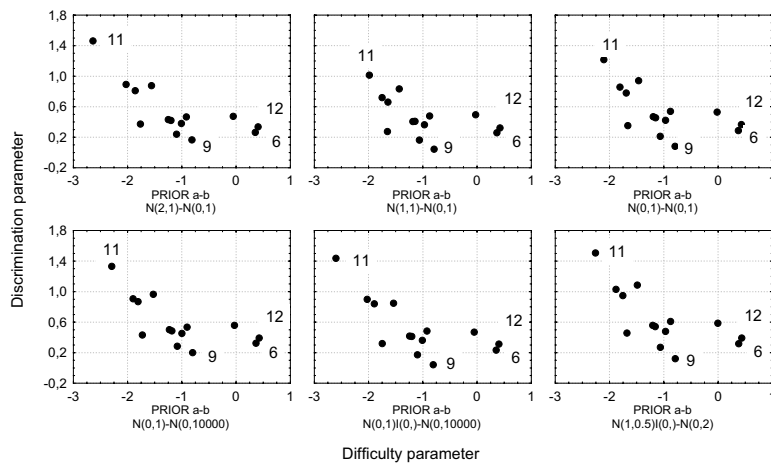


Gráfico 6.7

Media a posteriori de los parámetros de dificultad y de discriminación para diferentes priors en el modelo de ojiva normal

De acuerdo al gráfico 6.7, el modelo 2PLP es insensible a la especificación de priors para los parámetros de dificultad y de discriminación. Un análisis de una vía para evaluar si los parámetros de ítem son diferentes de acuerdo a las priors resultó no significativo.

En el análisis de la sensibilidad se ajustó varios modelos de probabilidad para los mismos datos. Para comparar los seis modelos correspondientes a cada priori, se calculó la “expected deviance a posterior” (Dbar), la “deviance information criterion” (DIC) y el número efectivo de parámetros ρ_D que son descritos en el capítulo 5 y en Spiegelhalter et al. (2002).

Spiegelhalter et al. (2002) indican que el DIC implementado en el software WinBUGS se puede usar para comparar modelos complejos y

grandes diferencias en el criterios se pueden atribuir a diferencias predictivas reales en los modelos, aunque persisten algunas críticas. En el modelamiento jerárquico con variables latentes auxiliares como en los modelos TRI, la verosimilitud o complejidad del modelo no es única así que la deviance del modelo (el DIC y $\rho_D D$ que se basan en ella) cuando hay variables latentes no es única y puede ser calculada de varias maneras (Delorio and Roberts, 2002). Con variables latentes auxiliares, el software WinBUGS usa la verosimilitud completa de las variables observadas y de la variable latente introducida (como efectos fijos y aleatorios respectivamente en el modelamiento jerárquico) para obtener la distribución posterior para los parámetros de interés. Cuando este es el caso, WinBUGS presenta DICs marginales para la variable observada (efectos fijos) y para las variables latentes auxiliares (efecto aleatorio). Para una comparación de los modelos propuestos, nosotros consideramos DIC marginal para la variable observada porque el foco del análisis está en $f(\mathbf{y}|\mathbf{u}, \eta)$ y aunque variables aleatorias auxiliares se introducen (en dos etapas, para variables latentes y parámetros de ítem) estas no son el focos del análisis.

El análisis del DIC en la tabla 6.6, confirma que todas las prioris tienen el mismo ajuste. El gráfico 6.8 hace comparaciones de estas diferentes prioris considerando un gráfico de la media y desviación estándar para los parámetros de ítem.

prior	a media	b media	Dbar	Dhat	pD	DIC
A	0.6355	-1.021	1450.14	1369.86	80.28	1530.43
B	0.5356	-0.999	1461.09	1383.61	77.48	1538.57
C	0.4599	-0.979	1453.16	1368.36	84.80	1537.96
D	0.5168	-1.099	1446.60	1358.26	88.34	1534.94
E	0.5414	-1.101	1453.22	1375.48	77.75	1530.97
F	0.5879	-1.048	1445.00	1352.90	92.11	1537.11

Tabla 6.6
Comparación Dbar y DIC de prioris para el modelo de ojiva normal

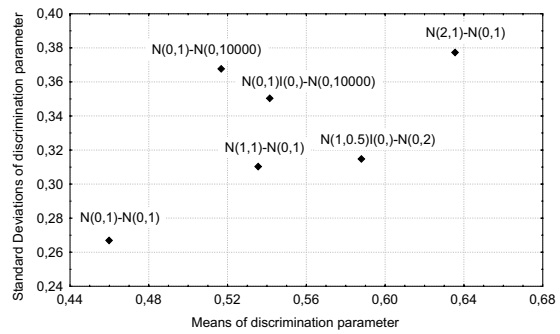
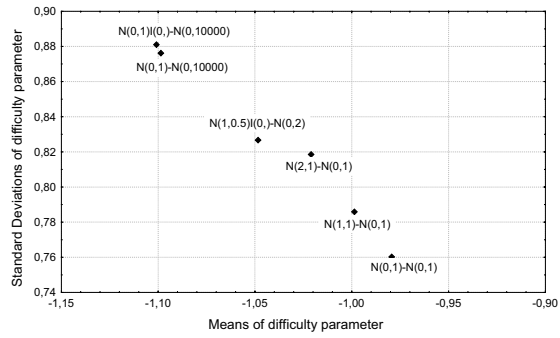


Gráfico 6.8
Comparación del modelo de ojiva normal con diferentes prioris

Para los parámetros de dificultad, se tiene que las prioris están organizadas desde las prioris no informativas o difusas hasta prioris precisas. A mayor precisión de la priori especificada menor varianza en la posteriori y mayor valor de la media a posteriori de la dificultad. En el caso del parámetro de discriminación, tenemos que las prioris están organizadas, en el eje x , de menor a mayor media a posteriori, y en el eje y de mayor precisión a mayor valor de la variancia especificada a priori excepto para el caso de mayor de la media a priori.

A pesar de que observamos que las prioris son similares en el ajuste del modelo consideramos que la priori F es más apropiada porque presenta menor coeficiente de variabilidad en relación a la especificación de un gran varianza a priori para la dificultad. Este resultado prueba lo indicado por Sahu (2002) que indica que ha investigado otras prioris con otros hiperparámetros pero esta es la más estable.

6.2.3 Inferencia e interpretación de los parámetros en el modelo 2P

Considerando las prioris $a \sim N(1, 0.5)I(0,)$ y $b \sim N(0, 2)$ dada por Sahu (2002), o priori F , se realizó un análisis final de los datos con el propósito de inferencia. Se consideró un Bur-in de 1000 y un tamaño de muestra efectivo de 1000 iteraciones considerando un valor de `thin=10`. Estas estimaciones son esencialmente estables si se consideran otros tamaños de muestra. La convergencia fue monitoreada usando hasta 5 cadenas de tamaños diferentes después de descartar las primeras 1000 iteraciones. Fue usado la librería CODA de R para obtener la estadística de Gelman y Rubin (1992) que indicó que un total de 159 cadenas generadas convergen.

En media, para generar un tamaño de bur-in de 1000 el programa demora 90 segundos, y para 10000 iteraciones adicionales usa 957 segundos. En el gráfico 6.9 se presenta el histórico de la cadena y la densidad empírica a posteriori para el item 11 y la media y desviación estándar de la variable latente.

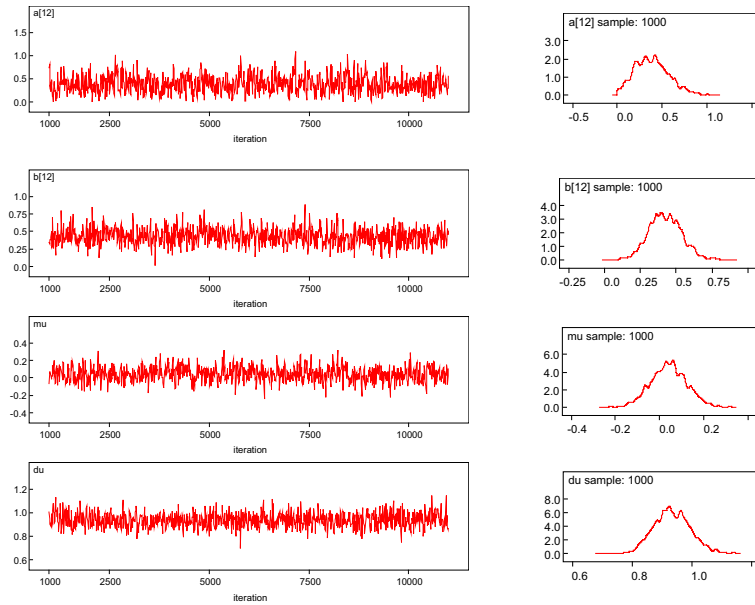


Gráfico 6.9

Histórico de la cadena y densidad para los parámetros del ítem 11 y media y desviación estándar a posteriori de la variable latente

Estimaciones de los parámetros de discriminación y de dificultad para el modelo de ojiva normal se presentan en la tabla 6.7. El ítem 11 es el más discriminativo mientras el ítem 9 es el menor. También, el ítem 11 es el más fácil mientras el ítem 12 es el más difícil.

		Media	sd	$P_{2,5}$	Mediana	$P_{97,5}$
Parámetros de discriminación	a_1	0.54	0.24	0.13	0.52	1.09
	a_2	0.29	0.18	0.03	0.27	0.70
	a_3	0.55	0.24	0.16	0.53	1.09
	a_4	0.91	0.33	0.33	0.88	1.61
	a_5	0.49	0.24	0.09	0.48	1.02
	a_6	0.32	0.18	0.03	0.30	0.73
	a_7	0.88	0.34	0.29	0.84	1.60
	a_8	0.97	0.35	0.39	0.92	1.80
	a_9	0.20	0.14	0.01	0.18	0.51
	a_{10}	0.49	0.23	0.08	0.48	1.02
	a_{11}	1.35	0.41	0.64	1.32	2.18
	a_{12}	0.39	0.19	0.06	0.38	0.80
	a_{13}	0.45	0.23	0.05	0.43	0.93
	a_{14}	0.41	0.26	0.03	0.37	0.97
Parámetros de dificultad	b_1	-0.90	0.16	-1.25	-0.90	-0.60
	b_2	-1.09	0.15	-1.37	-1.08	-0.80
	b_3	-0.03	0.13	-0.27	-0.03	0.23
	b_4	-1.91	0.33	-2.62	-1.88	-1.36
	b_5	-1.23	0.18	-1.62	-1.22	-0.90
	b_6	0.37	0.11	0.16	0.37	0.61
	b_7	-1.82	0.33	-2.63	-1.78	-1.28
	b_8	-1.53	0.31	-2.30	-1.50	-1.03
	b_9	-0.80	0.13	-1.06	-0.80	-0.57
	b_{10}	-1.19	0.17	-1.55	-1.18	-0.87
	b_{11}	-2.31	0.46	-3.32	-2.25	-1.53
	b_{12}	0.43	0.12	0.21	0.42	0.67
	b_{13}	-1.01	0.16	-1.35	-1.00	-0.71
	b_{14}	-1.71	0.23	-2.20	-1.70	-1.32
Variable latente	u media	0.94	0.06	0.82	0.93	1.06
	d.e u	0.04	0.08	-0.12	0.04	0.21

Tabla 6.7: Media y desviación estándar a posteriori e intervalo de credibilidad al 95 % para los parámetros del modelo de ojiva normal

El ítem 11 dice: “Luisa, Dora y María compran tela. Luisa compró la mitad de un metro, Dora compró 75 centímetros y María compró cincuenta centímetros. ¿Quiénes compraron la misma cantidad de tela?”. El ítem 12 dice: “Un recipiente recibe 4.5 litros de agua cada minuto. ¿Cuántos litros de agua tendrá el recipiente después de un hora y media?”. Por otro lado, el ítem 6 dice: “Resuelve las siguientes operaciones con decimales: $0.75 - 0.2 + 1.2 - 0.30$ ”.

Es claro, del texto de los ítems, que la habilidad latente requerida para encontrar la solución del ítem 11 es menor que la del ítem 12. Para un mismo valor de habilidad, el alumno tiene mayor probabilidad de suceso en el ítem 11 antes que en el ítem 12. Por eso el ítem 11 es más fácil que el ítem 12 como el modelo de ojiva normal postula.

También es importante percibir para el ítem 11 que un pequeño cambio en la habilidad del alumno (digamos, un conocimiento de lo que significa mitad de un metro) produce un rápido incremento de la probabilidad de suceso del ítem. Pero en el ítem 9 un pequeño cambio de habilidad del alumno (digamos el conocimiento de decimales) no se traduce en un incremento rápido de la probabilidad de suceso del ítem. Así el ítem 11 es más discriminativo porque permite distinguir mejor entre los alumnos que saben o no saben un determinado conocimiento específico al ítem. Diagramas de caja para las medias a posteriori de los ítems se presentan en el gráfico 6.10. Para mayor detalle de interpretación de los parámetros de ítem puede verse Johnson y Albert (2000).

Con respecto a la variable latente, la media de las medias a posteriori para los examinados es 0.94 lo que indica que el grupo de estudiantes presenta habilidades con asimetría negativa como fue encontrado en los puntajes (véase la tabla 6.7). De esta manera encontramos que las habilidades estimadas tienen correspondencia con el puntaje. Modelos que dan cuenta de esta asimetría han sido desarrollados recientemente por Bazán et al (2006), Bolfarine y Bazán (2007). Un detalle de estos modelos puede encontrarse en el capítulo siguiente.

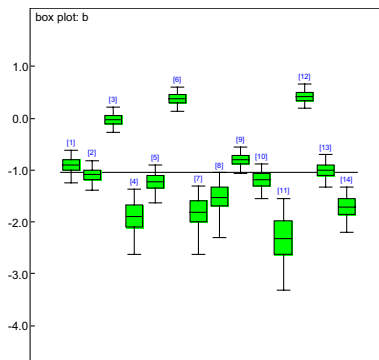
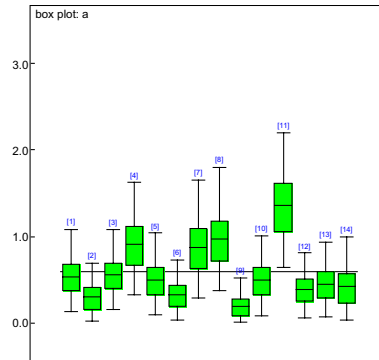


Gráfico 6.10
 Box-plots de las medias a posteriori de los parámetros de discriminación (a) y dificultad (b) en el modelo de ojiva normal

7 Extensiones y comentarios finales

En este trabajo hemos presentado los modelos de teoría de respuesta al ítem binarios tradicionales desde la perspectiva de la inferencia Bayesiana. Una presentación parecida para el caso del Brasil, pero desde el punto de vista clásico, puede revisarse en Andrade, Tavares y Valle (2000).

El modelo al que hemos hecho mayor referencia es el modelo de dos parámetros, el cual vimos ajustaba muy bien al análisis de los datos en nuestras aplicaciones. Este modelo es bastante interesante porque diversas extensiones han sido propuestas a partir de él. Una de ellas es la inclusión de parámetros adicionales de ítems como, por citar, el modelo tres parámetros donde se incluye un parámetro de adivinación, el modelo skew-probit (Bazán, et al, 2006) que incluye un parámetro de penalidad o bonificación del ítem o el modelo de Teslets que incluye un parámetro para la interacción entre determinados ítems (Wang, Bradlow, y Wainer, 2003).

Otras extensiones de estos modelos se obtienen considerando a la variable latente U como multidimensional y no unidimensional (ver por ejemplo, Beguin y Glas, 2001, Linardakis y Dellaportas, 2002). También existen extensiones donde se consideran variables predictoras o variables explicativas. Se da también el caso de modelos multinivel (Fox y Glas, 2001) y de modelos con errores de medición (Fox y Glas, 2003). Otro grupo de extensiones se obtiene cuando se toma como variable manifiesta valores discretos mínimamente ordinales y no respuesta binaria. En ese caso se está frente a modelos de TRI para respuesta policotómica (las escalas de actitudes y las pruebas de desempeño son ejemplos de estos casos).

Con esto queremos resaltar que la TRI no es sólo de interés desde el punto de vista aplicativo sino también por los diferentes desarrollos de tipo formal y computacional que generan. Si bien la lista de extensiones dada bosqueja una visión muy panorámica del área, invitamos a los

interesados a comunicarse con los autores para conocer mas detalle de otras extensiones y problemas abiertos que persisten en este campo. A manera de ilustración mostraremos seguidamente con un poco mayor detalle tres modelos TRI no estándares en los cuales los autores han venido trabajando.

7.1 Modelos TRI con enlaces asimétricos

En los modelos tradicionales de TRI, la CCI es simétrica; este es el caso del modelo logístico y de ojiva normal. Sin embargo, como ha observado Samejima (2000), Bazán et. al. (2006) y Bolfarine y Bazán (2007) CCI asimétricas pueden ser incorporadas considerando un nuevo parámetro de ítem que controla la forma de dicha curva. Esta asimetría es en muchos casos necesaria para una mejor modelización de respuestas con una baja proporción de 0's o 1's. Seguidamente mostraremos tres de estos modelos.

7.1.1 El modelo de ojiva normal asimétrico

Este modelo postula que la probabilidad de acierto dados los parámetros de habilidad y de ítems es:

$$p_{ij} = P(Y_{ij} = 1 | u_i, \eta_j) = \Phi_{SN}(m_{ij}; \lambda_j),$$

donde $\lambda_j > 0$ es un parámetro de asimetría, $m_{ij} = a_j u_i - b_j$ es el predictor latente lineal y Φ_{SN} denota a la función de distribución acumulada de una variable skew normal con función de densidad

$$\phi_{SN}(x; \lambda) = 2\Phi(x)\Phi(\lambda x),$$

siendo ϕ la función de densidad de una variable normal estándar.

Nótese que si $\lambda = 0$, el modelo de ojiva normal asimétrico es un modelo estándar de ojiva normal. El gráfico 7.1 muestra la sensibilidad del modelo con respecto al nuevo parámetro λ presentando diferentes CCI

para diferentes valores de la variable latente asociada con la habilidad U , y considerando parámetros de ítem $a = 1$, $b = 0$ y $\lambda = -2, -1, 0, 1, 2$ en el modelo de ojiva normal asimétrico. Note que cuando $\lambda > 0$, la probabilidad de respuesta correcta tienen un bajo crecimiento para bajos valores de variable latente U . De otra manera, cuando $\lambda < 0$, la probabilidad de respuesta correcta tiene un rápido crecimiento para bajos valores de la variable latente U . Es por este comportamiento que este parámetro es interpretado como un parámetro de penalidad o de bonificación.

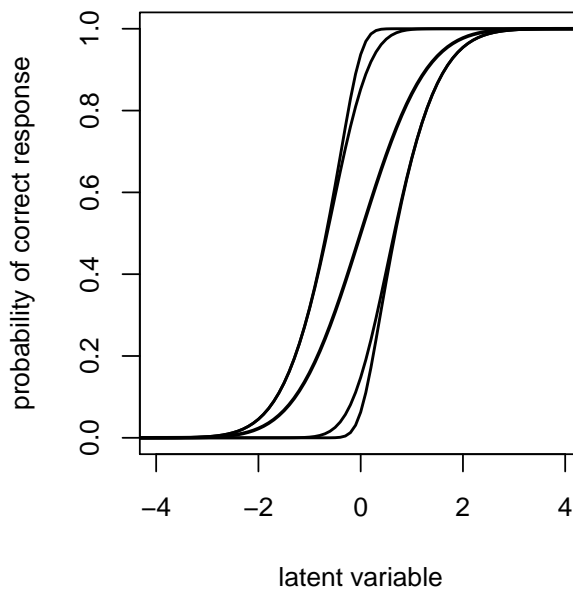


Gráfico 7.1
CCI para el modelo de ojiva skew normal

La verosimilitud $L(\mathbf{y}|\mathbf{u}, \mathbf{a}, \mathbf{b}, \lambda)$ del modelo es proporcional a

$$\prod_{i=1}^n \prod_{j=1}^k [\Phi_{SN}(m_{ij}; \lambda_j)]^{y_{ij}} [1 - \Phi_{SN}(m_{ij}; \lambda_j)]^{1-y_{ij}},$$

pero similarmente a lo visto en (5.4) el modelo TRI de ojiva normal asimétrico puede ser definido de manera alternativa por

$$Y_{ij} = \begin{cases} 1, & \text{si } Z_{ij}^* > 0 \\ 0, & \text{si } Z_{ij}^* \leq 0, \end{cases}$$

donde Z_{ij}^* denota a la variable Z_{ij} dada en (5.4) condicionada a que $V_{ij} = v_{ij}$,

$$Z_{ij}^* \sim N(-\delta_j v_{ij} + m_{ij}, 1 - \delta_j^2), \quad V_{ij} \sim HN(0, 1)$$

$\delta_j = \frac{\lambda_j}{\sqrt{1+\lambda_j^2}} \in [0, 1]$ es una reparametrización más conveniente que λ_j y HN denota a la distribución normal positiva o half normal.

Bajo esta nueva formulación, la verosimilitud puede reescribirse proporcional a

$$\prod_{i=1}^n \prod_{j=1}^k \phi(Z_{ij}^*; -\delta_j v_{ij} + m_{ij}, 1 - \delta_j^2) \phi(V_{i,j}; 0, 1) I(v_{ij} > 0) p(y_{ij} | z_{ij}^*),$$

donde

$$p(y_{ij} | z_{ij}^*) = I(z_{ij}^* > 0) I(y_{ij} = 1) + I(z_{ij}^* \leq 0) I(y_{ij} = 0).$$

Mayores detalles acerca de este modelo puede ser revisado en Bazán et al. (2006)

7.1.2 Los modelos LPE y RLPE

Estos modelos postulan que la probabilidad de acierto dados los parámetros de habilidad y de ítemes es:

$$p_{ij} = P(Y_{ij} = 1 | u_i, a_j, b_j) = F_{\lambda_j}(m_{ij}),$$

donde F_{λ_j} es una función de distribución acumulada indexada por el parámetro $\lambda_j > 0$ y evaluada en $m_{ij} = a_j u_i b_j$. Dependiendo de que función de distribución se especifique tendremos los modelos Logístico positivo exponencial (LPE) y reflejo del logístico positivo exponencial (RLPE). En el primer caso, este se caracteriza por

$$F_1(x) = 1 - (1 + e^x)^{-\lambda_j}$$

y en el segundo por:

$$F_2(x) = (1 + e^{-x})^{-\lambda_j}$$

Estos corresponden a las funciones de distribución acumuladas Scobit y Burr de tipo II, respectivamente.

Note que $F_1(-x) \neq 1 - F_1(x)$ ó $F_2(-x) \neq 1 - F_2(x)$ y que F_1 y F_2 son asimétricas y se cumple que

$$F_2(x) = 1 - F_1(-x)$$

ó

$$F_1(x) = 1 - F_2(-x).$$

El modelo LPE fue propuesto por Samejima (2000) pero ha sido implementado en Bolfarine y Bazán (2007) junto al modelo RLPE propuesto por ellos.

En ambos modelos λ_j puede también interpretarse como un parámetro de penalidad o bonificación de manera similar al caso del modelo de ojiva skew normal. Ello puede observarse en las CCI del gráfico 7.2. Mayores detalles de este modelo puede revisarse en Bolfarine y Bazán (2007).

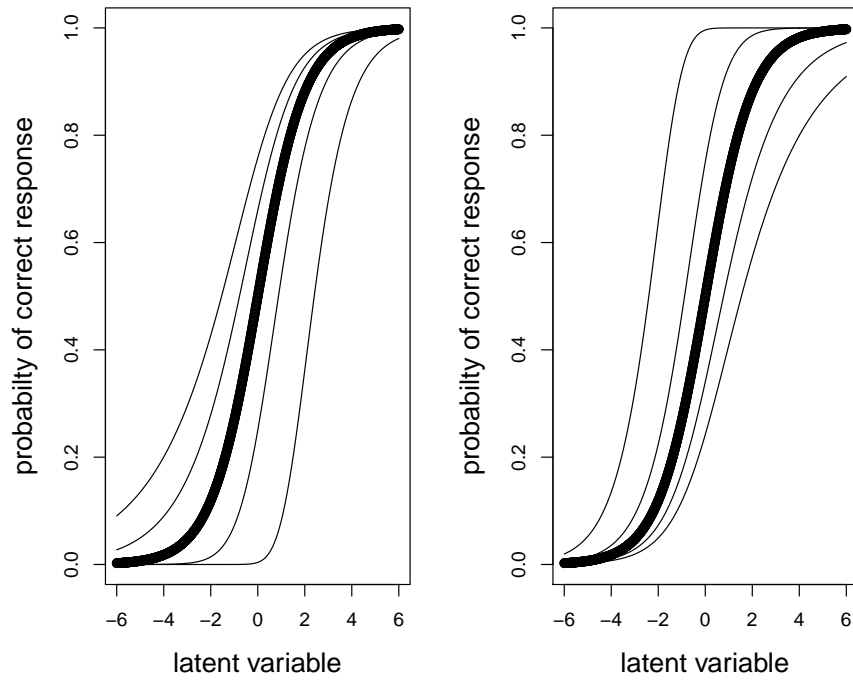


Gráfico 7.2
CCI para los modelos LPE y RLPE

7.2 Modelos TRI bajo restricciones de rapidez

El problema de la rapidez surge cuando una prueba tiene restricciones de tiempo. Ello hace que los sujetos que la respondan limiten sus análisis en especial para los últimos ítems de la prueba.

La posibilidad de construir un modelo que tome en cuenta esta restricción fue considerada por Yamamoto y Everson (1997) y Bolt, et.

al. (2002). El primero asume básicamente un modelo logístico de dos parámetros para la probabilidad de responder correctamente a los primeros ítems y un proceso de adivinanza para los últimos. De otro lado, Bolt planteó más bien un modelo Rash para todos los ítems, pero en donde los parámetros de dificultad de los ítems varían según sean estos los primeros o últimos, siendo la dificultad de los últimos mayor a la de los primeros. Basados en estos trabajos Goegebeur, et. al.(2008) han recientemente propuesto un modelo de proceso de cambio gradual. Este modelo postula que la probabilidad de acierto dados los parámetros de habilidad y de ítems es:

$$p_{ij} = P[Y_{ij} = 1 | \theta_i, \eta_i, \lambda_i, a_j, b_j, c_j] = c_j + (1 - c_j)G(m_{ij}),$$

donde

$$G(m_{ij}) = \tilde{p}_{ij}P_j(\eta_i, \lambda_i), \quad P_j(\eta_i, \lambda_i) = \min\left\{1, \left[1 - \left(\frac{j}{k} - \eta_i\right)\right]^{\lambda_i}\right\}$$

y \tilde{p}_{ij} se escribe bajo el modelo logístico de dos parámetros.

El parámetro $\eta_i \in [0, 1]$ es introducido al modelo para expresar el punto, representado como una fracción del número de ítems, donde el examinado experimenta por primera vez el efecto de rapidez. De otro lado, $\lambda_i \geq 0$ controla la tasa de decrecimiento hacia la situación de adivinación.

La racionalidad del modelo es como sigue. Cuando un examinado i encuentra el ítem j , él responde de acuerdo a un modelo TRI de 3 parámetros o a un proceso de adivinación aleatorio con probabilidad $P_j(\eta_i, \lambda_i)$ y $1 - P_j(\eta_i, \lambda_i)$, respectivamente. Bajo el proceso de solución de problemas el examinado conoce la respuesta con probabilidad \tilde{p}_{ij} ; si es ignorante, el examinado adivina al azar.

El lector interesado en la estimación de los parámetros de este modelo, desde un punto de vista Bayesiano, puede consultar Valdivieso y Bazán (2009).

7.3 Comentarios finales

En este trabajo hemos intentado presentar la estimación Bayesiana de los modelos TRI binarios tradicionales, tanto desde un punto de vista teórico como práctico. Nuestra intención ha sido la de interesar a la comunidad académica como profesional en la aplicación adecuada de estos modelos. Esto supone mejorar las propuestas de evaluación y medición disponibles en nuestro medio y acercarnos a los modelos que ya se usan en otros países. Asimismo hemos presentado diversas alternativas desarrolladas más allá del modelo Rasch facilitando la interpretación de sus resultados.

Actualmente estamos implementando todos los modelos detallados en un software generador de códigos. Estos códigos permitirán al interesado realizar de manera directa las estimaciones u otros procedimientos Bayesianos en WinBUGS. Este programa, al que hemos denominado Bayes@PUCP, estará disponible para los interesados que lo requieran.

Este tipo de trabajos son parte de los desarrollos que realiza el grupo de investigación de Modelos Matemáticos y Estadísticos para evaluación MMEPE de la Sección Matemáticas de la PUCP.

A Inferencia Bayesiana en TRI usando WinBUGS

En este anexo presentamos los códigos para implementar en WinBUGS los dos procedimientos de estimación bayesiana en TRI, esquemas ARS y Gibbs Sampling, desarrollados en el capítulo. Estos códigos, para los modelos de uno, dos o tres parámetros con ojiva normal o logística pueden ser obtenidos también considerando el aplicativo Bayes@PUCP, el cual implementa los modelos indicados más otros nuevos modelos propuestos en Bazán et al (2006) y Bolfarine y Bazan (2007).

A.1 Código WinBUGS para esquema Gibbs Sampling

A continuación se presenta el código WinBUGS para el modelo de ojiva normal de un parámetro (1P) y de dos parámetros (2P) considerando aumentación de datos, es decir considerando el esquema de Gibbs Sampling.

```
model { #likelihood function
  for (i in 1 : n) {
    for (j in 1 : I) {
#2P
      #      m[i,j] <- a[j]*u[i] - b[j]
#1P
      m[i,j] <- u[i] - b[j]
      z[i,j] ~ dnorm(m[i,j],1)I(1o[y[i,j]+1],up[y[i,j]+1])
    }
  }
#priors F for item parameters
  for (j in 1:k) {
    b[j] ~ dnorm(0,0.5);
  }
}
```

```

# 1P
      #      a[j] ~ dnorm(1,2)I(0,);
      }
#prior for latent variable
      for (i in 1:n) { u[i] ~ dnorm(0,1) }
# auxiliary latent variable
lo[1] <- -50; lo[2] <- 0; \# i.e., z| y=0 ~ N(m,1)I(-50,0)
up[1] <- 0; up[2] <- 50; \# i.e., z| y=1 ~ N(m,1)I(0,50)
# mean and standard deviation for latent variable
mu<-mean(u[ ])
du<-sd(u[ ])
      }

```

Note que si se quiere obtener las versiones correspondientes para los modelos de Rasch y 2L logístico solamente se debe modificar la línea

```
dnorm(m[i,j],1)I(lo[y[i,j]+1],up[y[i,j]+1])
```

por la línea

```
dlogis(m[i,j],1)I(lo[y[i,j]+1],up[y[i,j]+1]).
```

A.2 Código WinBUGS para esquema ARS

Otra posibilidad de códigos para WinBUGS en los modelos logísticos de un, dos y tres parámetros se presenta a continuación

```

model{
for (i in 1:n) { for (j in 1:k )
      { y[i,j]~dbern(p[i,j])

#1L
m[i,j]<-theta[i]-b[j]
#2L Model
#m[i,j]<-a[j]*(theta[i]-b[j])
p[i,j]<-exp(m[i,j])/(1+exp(m[i,j]))
#3L Model

```

```

    #p1[i,j]<-exp(m[i,j])/(1+exp(m[i,j]))
    # p[i,j]<-c[j]+(1-c[j])*p1[i,j]
    }
  }
#abilities priors
for (i in 1:n) { theta[i]~dnorm(0,1)}
#items priors
for (j in 1:k) {
# Sinharay (2004)
    b[j]~dnorm(0,1)
#    a[j]~dlnorm(0,1)
#    c[j] ~ dbeta(5,17)
    }
}

```

Note que en este caso no se requiere la versión no aumentada del modelo para implementar el esquema ARS.

Aprovechando que el programa WinBUGS ya contiene la función de enlace `logit`, también es posible escribir `logit(p[i,j])<-m[i,j]` en lugar de `p[i,j]<-exp(m[i,j])/(1+exp(m[i,j]))`

Si se desea obtener las correspondientes versiones del modelo de oji-va normal, solamente ha de modificarse las expresiones correspondientes de las probabilidades por `p[i,j]<-Phi(m[i,j])` aprovechando que el WinBUGS tiene la función de la acumulada de la normal `Phi`.

B Un Tutorial de WinBUGS para TRI

Para usar la inferencia Bayesiana usando WinBUGS asegúrese de contar con el programa. El programa puede ser obtenido de

<http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml>

Para un uso efectivo del programa con todas sus funciones es necesario registrarse y obtener un archivo que es enviado anualmente. Otra posibilidad, sin registro previo es usar OpenBugs disponible en

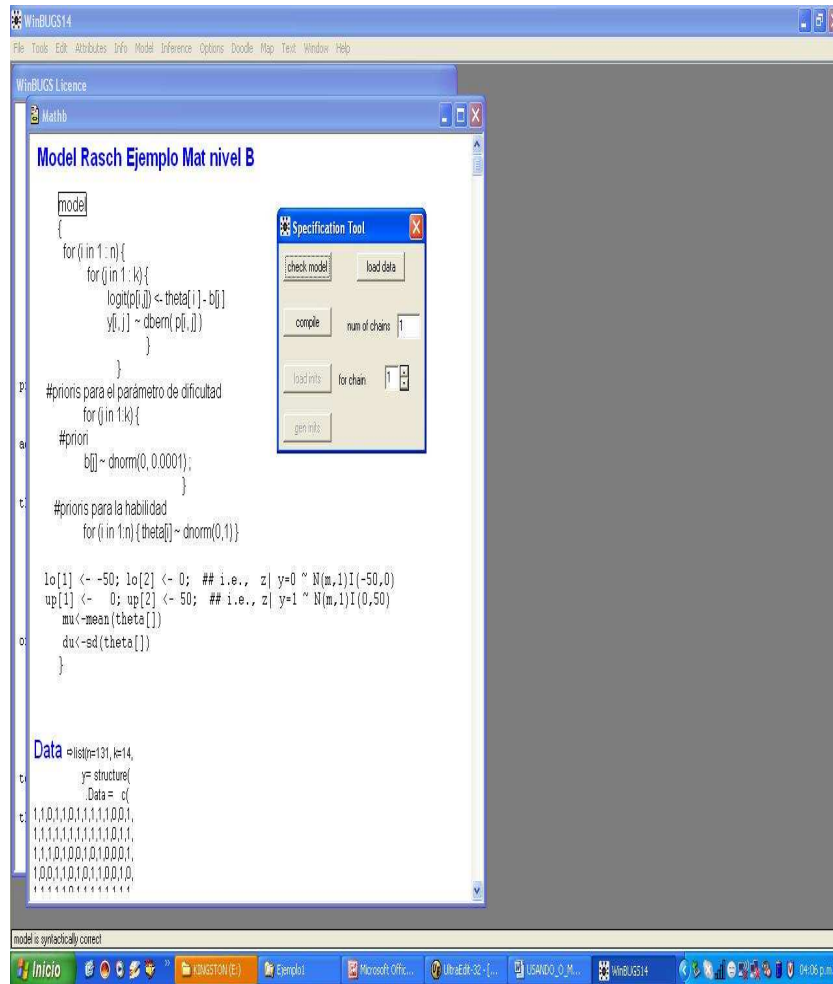
<http://mathstat.helsinki.fi/openbugs/>

En este tutorial usaremos el conjunto de datos de la prueba de Matemáticas (Math Data) analizado en el capítulo 6 de aplicaciones, pero bajo el modelo de Rasch. Si usted desea usar el WinBUGS para sus propios datos, no necesita modificar la sintaxis del modelo. Solamente debe introducir sus propios datos especificando el número de sujetos (n) y el número de ítems de su prueba (k). Para mejorar la interpretación de los resultados obtenidos en el modelo revise a Bond y Fox (2005).

En general, el procedimiento para realizar una inferencia Bayesiana mediante WinBUGS se pueden sintetizar en 3 fases que detallamos a continuación.

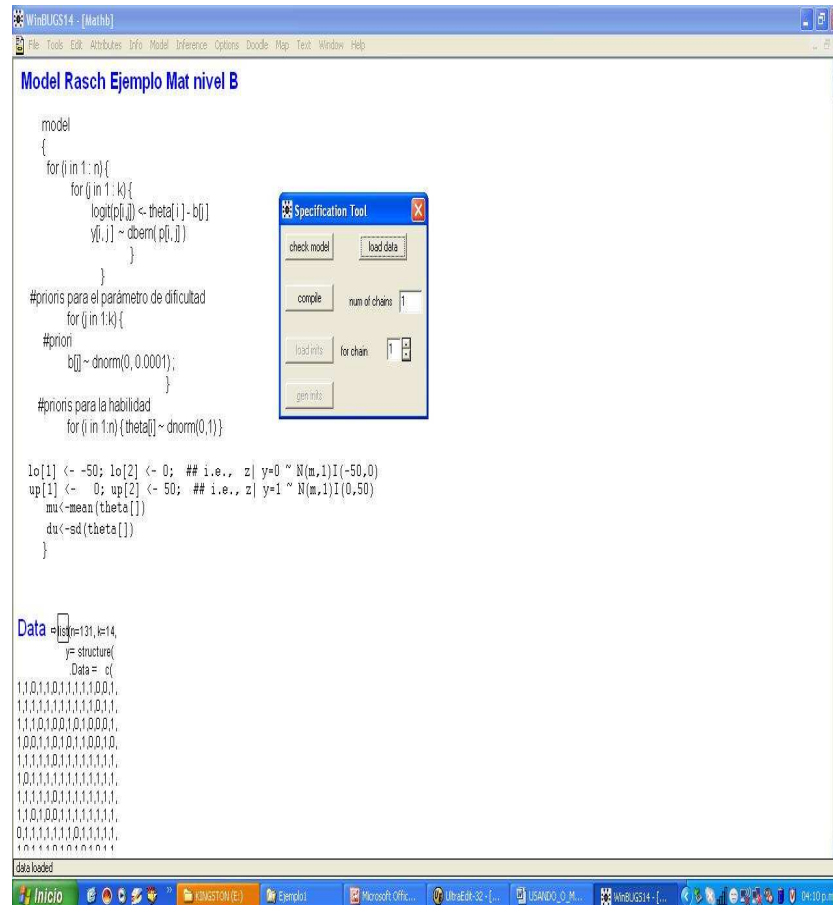
B.1 Fase 1: Definir el modelo

Paso 1. Seleccionar el modelo. Señalar con el mouse **model** como aparece en la figura abajo. En el menú seleccionar *Model*, luego *Specification*. Hacer click en *Check Model*



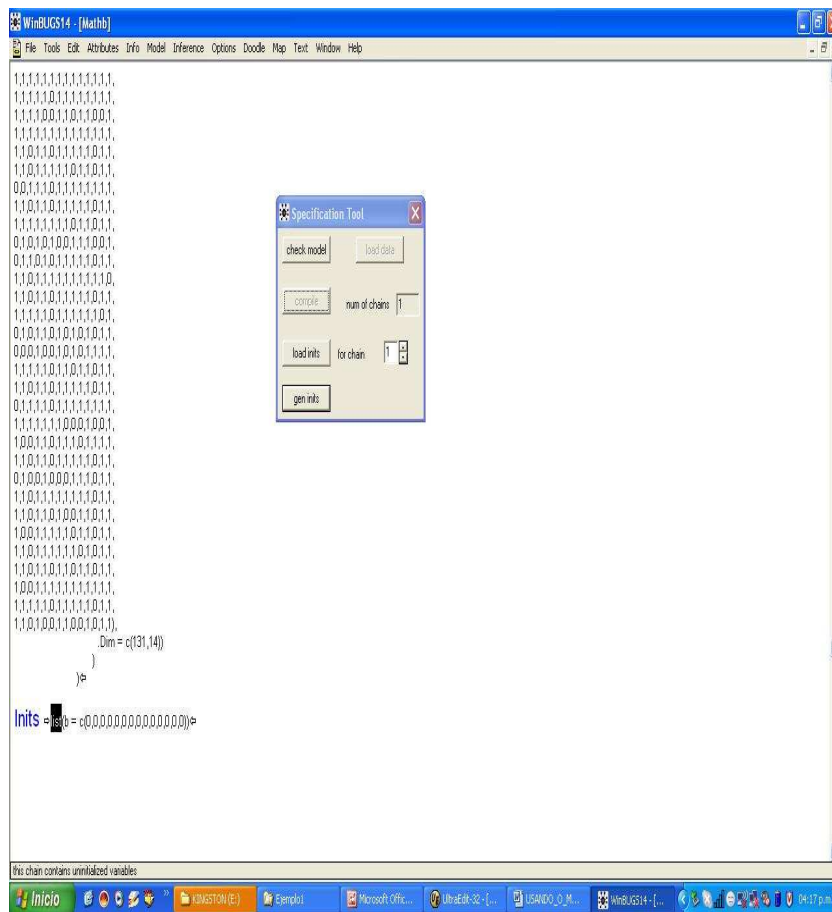
Si todo estuviera correcto, el mensaje abajo, en la parte izquierda es *model is syntactically correct*

Paso 2. Llamar los datos. Seleccionar **list** (señalar con el Mouse) como aparece en la figura abajo. Hacer click en *Load Data*.



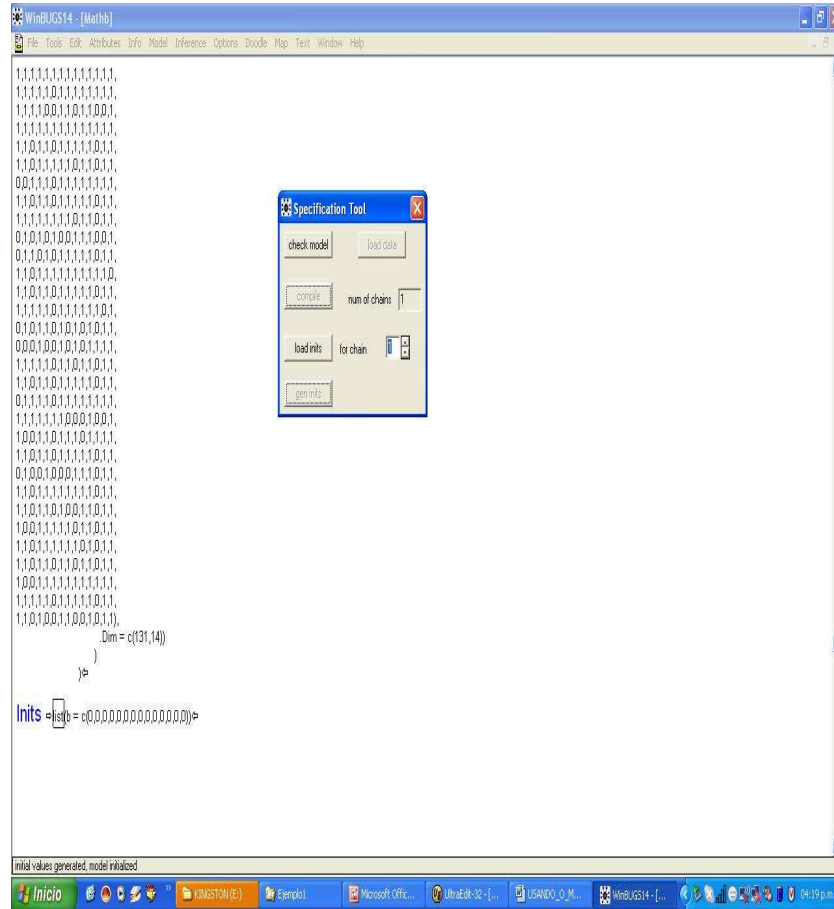
Si todo estuviera correcto, el mensaje abajo, en la parte izquierda es *data loaded*

Paso 4. Llamar los valores iniciales. Seleccionar **list** de los valores iniciales (señalar con el mouse) como aparece en la figura abajo. Hacer click en *Load Inits*



Si todo estuviera correcto, el mensaje abajo, en la parte izquierda es *this chain contains uninitialized variables*.

Paso 5. Simular valores iniciales de las variables faltantes. Hacer click en **Gen Inits**



Si todo estuviera correcto, el mensaje abajo, en la parte izquierda es *initial values generated, model initialized*.

B.2 Fase 2: Parámetros para la simulación

Paso 6. Definir parámetros para monitorear. Seleccionar en el Menú, *Inference*, luego *Samples*. En la ventana, *node*, escribir los parámetros a monitorear. En el modelo *Rasch* son *b* y *theta*. Cada vez que coloca un nodo haga click en *Set*. Al final escriba en la ventana: *

The screenshot shows the WinBUGS14 interface with a model editor window titled "Model Rasch Ejemplo Mat nivel B". The code in the editor is as follows:

```
model
{
  for (i in 1:n){
    for (j in 1:k){
      logit(p[i,j]) <- theta[i] - b[j]
      y[i,j] ~ dbern(p[i,j])
    }
  }
  #prioris para el parámetro de dificultad
  for (j in 1:k){
    #priori
    b[j] ~ dnorm(0, 0.0001);
  }
  #prioris para la habilidad
  for (i in 1:n){theta[i] ~ dnorm(0,1)}

  lo[1] <- -50; lo[2] <- 0; ## i.e., z | y=0 ~ N(m,1)I(-50,0)
  up[1] <- 0; up[2] <- 50; ## i.e., z | y=1 ~ N(m,1)I(0,50)
  mu <- mean(theta[])
  du <- sd(theta[])
}
```

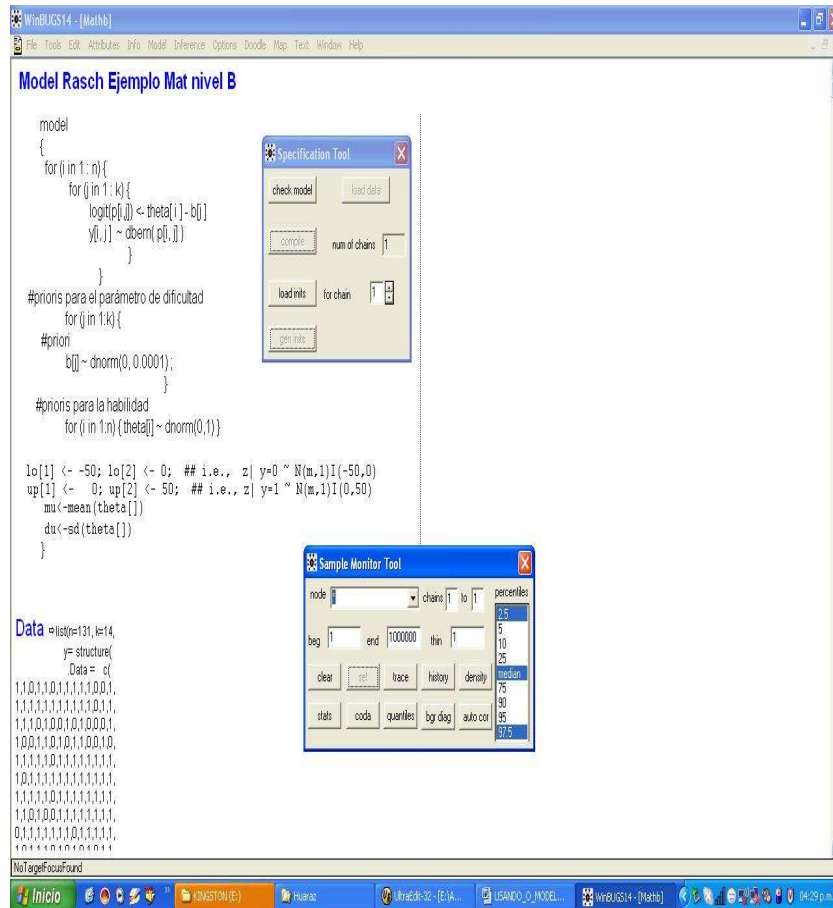
Below the code, there is a "Data =" section with a list of 14 data points:

```
Data = list(n=14, k=14,
  y= structure(
    Data = c(
      1,1,0,1,1,0,1,1,1,1,0,0,1,
      1,1,1,1,1,1,1,1,1,0,1,1,
      1,1,1,0,0,1,0,1,0,0,0,1,
      1,0,0,1,1,0,1,1,1,0,0,1,0,
      1,1,1,1,1,0,1,1,1,1,1,1,
      1,0,1,1,1,1,1,1,1,1,1,1,
      1,1,1,1,0,1,1,1,1,1,1,1,
      1,1,0,1,0,0,1,1,1,1,1,1,
      0,1,1,1,1,1,1,0,1,1,1,1,
      1,0,1,1,1,1,1,1,0,0,1,1
```

Two tool windows are open over the code:

- Specification Tool:** Contains buttons for "check model", "load data", "compile", "load inits", and "generate". It also has a "num of chains" field set to 1 and a "for chain" dropdown set to 1.
- Sample Monitor Tool:** Shows a list of nodes to monitor. The "node" dropdown is set to "theta". The "chars" field is 1, and "to" is 1. A table of percentiles is visible:

percentiles
2.5
5
10
25
50
75
90
95
97.5

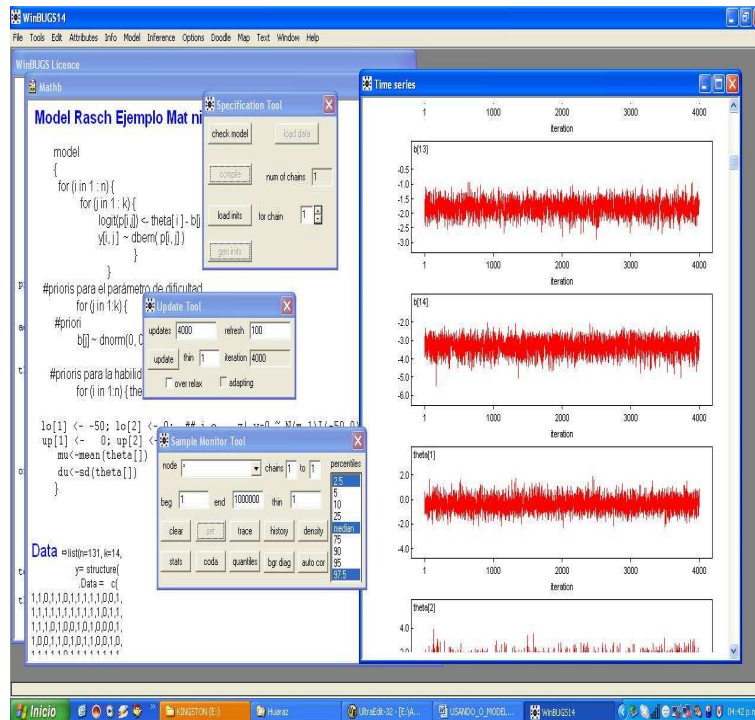


Paso 7. Hacer la simulación MCMC. En el Menú seleccionar *Model*, luego *Update*. En la ventana, updates, escriba el número de simulaciones que desea hacer. El default es 1,000. Para el modelo de Rasch la sugerencia es 4,000 las cuales serán descartadas para la inferencia (proceso Bur-In).

B.3 Fase 3: Evaluación de las cadenas generadas

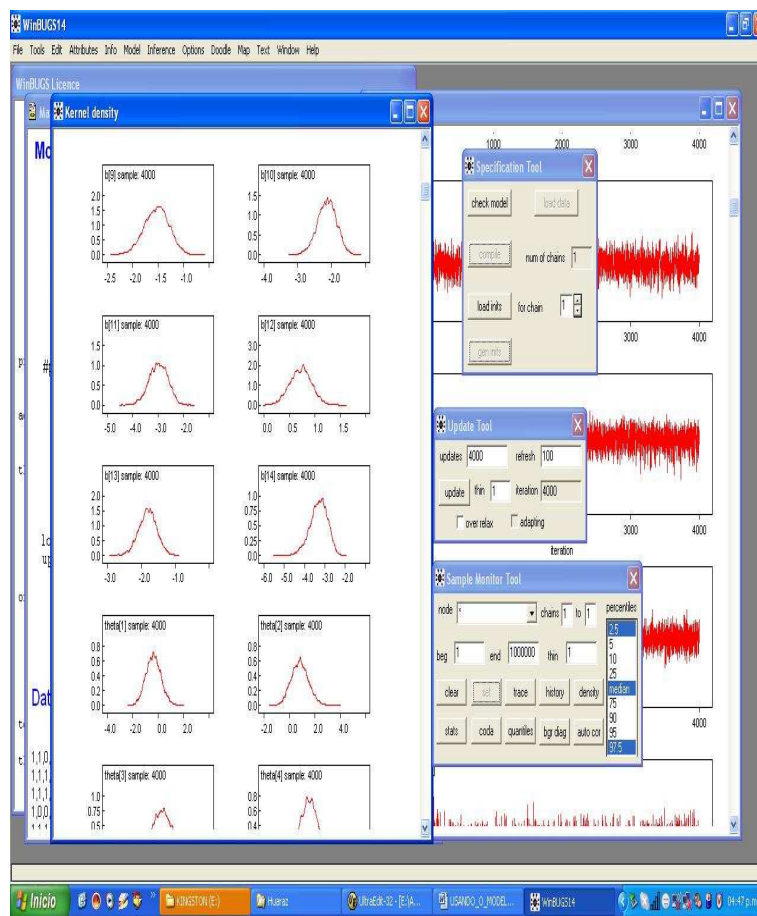
Paso 8. Evaluar la cadenas generadas.

1. En la ventana *Sample Monitor Tool*, hacer clic en *history*. En este caso ud. obtiene una gráfica para cada parámetro del modelo que corresponde a los diferentes valores que toma el parámetro en cada iteración. En este caso, son 4,000 valores. Lo que debe analizarse es si las cadenas (series de valores generados) son convergentes.



En el caso de los datos de Matemática, por inspección todas las cadenas son convergentes o estacionarias.

- En la ventana *Sample Monitor Tool*, hacer clic en *density*. En este caso ud. obtiene una gráfica para cada parámetro del modelo que corresponde a los diferentes valores que toma el parámetro en cada iteración. La gráfica corresponde a la distribución empírica obtenida de los valores generados.



Por inspección las cadenas mostradas en el gráfico anterior son unimodales y simétricas.

3. En la ventana *Sample Monitor Tool*, hacer clic en *coda*. En este caso ud. obtiene dos nuevos archivos: Coda index y Coda for chain 1. Los cuales deben ser grabados como nombre **.ind** y nombre **.out**. Estos archivos permiten realizar un análisis mediante algunas medidas de evaluación de las cadenas disponibles en las librerías CODA y BOA del programa R.

Estas librerías pueden obtenerse respectivamente en

<http://cran.r-project.org/src/contrib/Descriptions/coda.html>

y

<http://www.public-health.uiowa.edu/boa/Home.html>

Por otro lado, el programa R de software libre está disponible en

<http://www.r-project.org/>

WinBUGS14

File Tools Edit Attributes Info Model Inference Options Doodle Map Text Window Help

WinBUGS Licence

Mathbb CODA for chain 1

CODA index

1	-1.834	b[11]	1	4000
2	-1.495	b[12]	4001	8000
3	-1.571	b[13]	8001	12000
4	-1.615	b[14]	12001	16000
5	-1.412	b[15]	16001	20000
6	-1.654	b[16]	20001	24000
7	-1.557	b[17]	24001	28000
8	-1.398	b[18]	28001	32000
9	-1.819	b[19]	32001	36000
10	-1.328	b[110]	36001	40000
11	-1.53	b[111]	40001	44000
12	-1.501	b[112]	44001	48000
13	-1.972	b[113]	48001	52000
14	-1.493	b[114]	52001	56000
15	-1.401	theta(1)	56001	60000
16	-1.622	theta(2)	60001	64000
17	-1.617	theta(3)	64001	68000
18	-1.507	theta(4)	68001	72000
19	-1.517	theta(5)	72001	76000
20	-1.949	theta(6)	76001	80000
21	-1.534	theta(7)	80001	84000
22	-1.831	theta(8)	84001	88000
23	-1.702	theta(9)	88001	92000
24	-1.499	theta(10)	92001	96000
25	-1.697	theta(11)	96001	100000
26	-1.695	theta(12)	100001	104000
27	-1.151	theta(13)	104001	108000
28	-1.436	theta(14)	108001	112000
29	-1.484	theta(15)	112001	116000
30	-1.699	theta(16)	116001	120000
31	-1.561	theta(17)	120001	124000
32	-1.293	theta(18)	124001	128000
33	-1.77	theta(19)	128001	132000
34	-1.851	theta(20)	132001	136000
35	-1.642	theta(21)	136001	140000
36	-1.704	theta(22)	140001	144000
37	-1.616	theta(23)	144001	148000
38	-1.724	theta(24)	148001	152000
39	-1.633	theta(25)	152001	156000
40	-1.909	theta(26)	156001	160000
41	-1.721	theta(27)	160001	164000
42	-1.644	theta(28)	164001	168000
43	-1.509	theta(29)	168001	172000
44	-1.661	theta(30)	172001	176000

Specification Tool

check model load data

compile num of chains 1

load mvs for chain 1

generate

Update Tool

updates 4000 refresh 100

update thin 1 iteration 4000

over relax adapting

Sample Monitor Tool

node chain 1 to 1 percentiles

beg 1 end 100000 thin 1 median

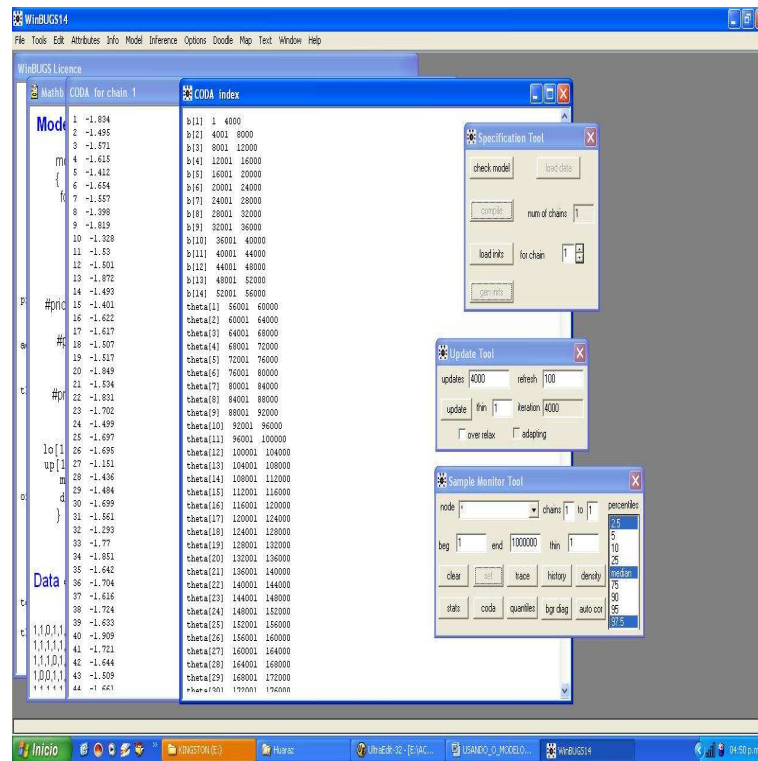
clear trace history density

stats coda quantiles bpr diag auto cov

2.5
5
10
25
75
90
95
97.5

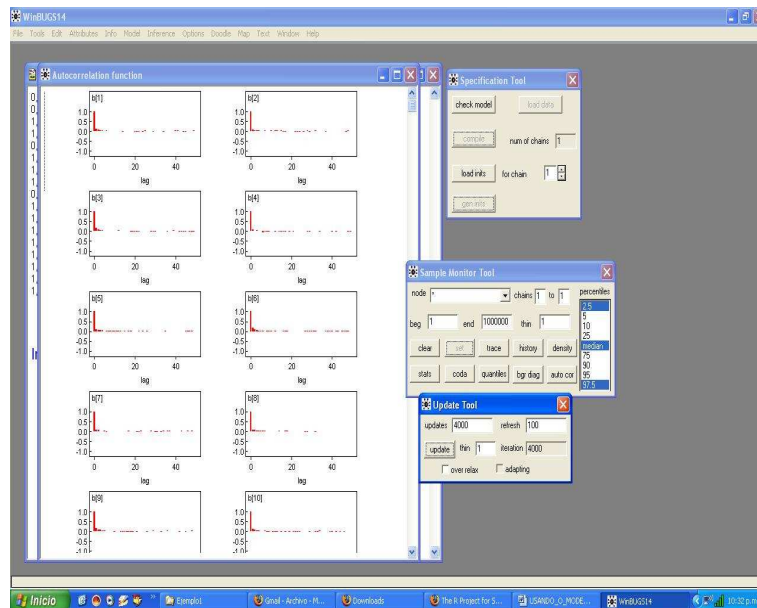
Inicio Windows (E) WinBUGS14 14:50 p.m.

Paso 9. Obtener estadísticas de resumen. En la ventana *Sample Monitor Tool*, hacer click en *Stat*

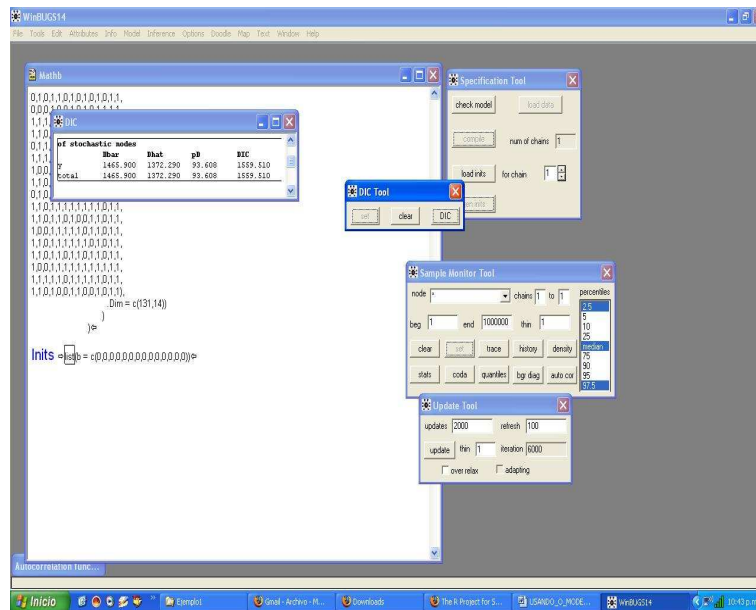


Note que se obtienen medidas de resumen de la distribución a posteriori de los parámetros. La media a posteriori es comparable con las estimativas obtenidas usando inferencia clásica. Pero a diferencia de dicha inferencia donde es necesario asumir ciertas suposiciones asintóticas para obtener los errores estándar asociados, en la inferencia Bayesiana puede obtenerse cualquier medida a posteriori como mediana y percentiles. En este caso se obtiene un resumen de 4,000 valores generados usando MCMC.

Paso 10. Obtener medidas definitivas y medida de comparación de modelos. En la inferencia Bayesiana es importante generar una cadena lo suficientemente grande para obtener la convergencia de la misma. Para ello es importante estudiar la auto correlación de la serie, lo cual ayuda a determinar el número de saltos o thin que debe hacerse para obtener las estadísticas definitivas de los parámetros del modelo. Si hay una auto correlación alta es necesario un thin (salto) mayor para seleccionar los valores. También hay que considerar el Bur-in o valores iniciales a ser descartados. Esos aspectos son discutidos en los artículos de las referencias. Nosotros vamos a considerar las primeras 4,000 iteraciones como Bur-in y como la autocorrelación mostrada en el gráfico para el modelo es baja, el thin será de 1 como por defecto.



Para tener una medida de comparación de modelos frente a modelos alternativos es importante usar la medida DIC (Deviance Information Criteria). En el Menú seleccionar *Inference*, luego *DIC*. En la ventana DIC tool seleccione set. Luego finalmente repita el paso 7 considerando 2,000 iteraciones adicionales. Finalmente vuelva a la ventana DIC tool y presione DIC.



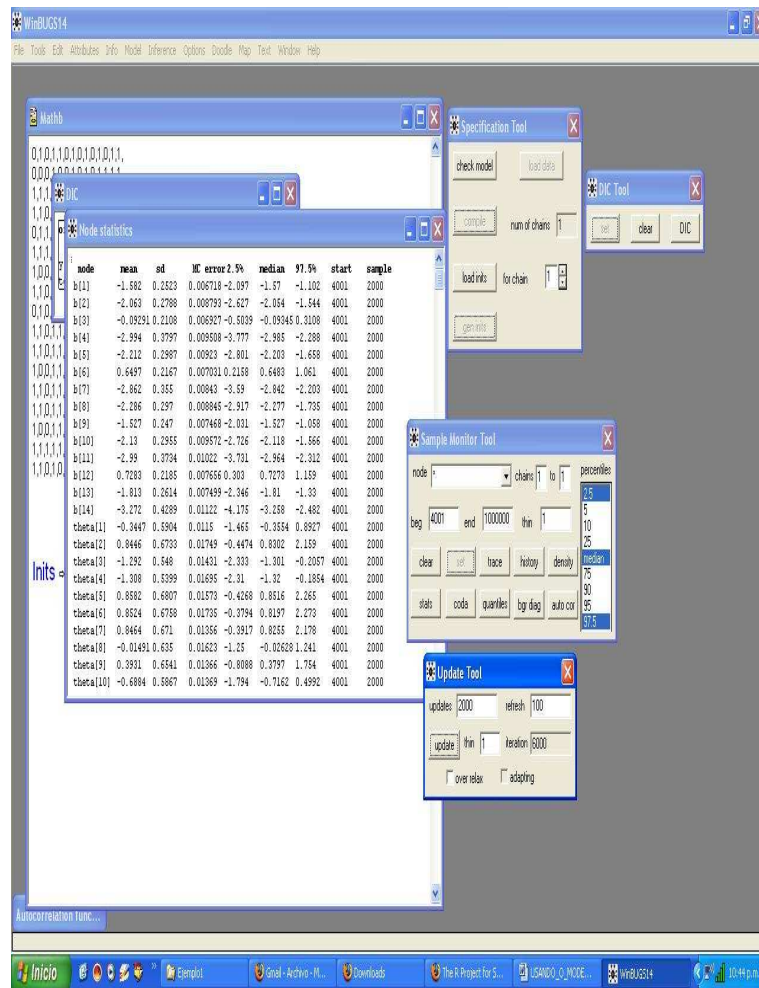
El resultado es

$$Dbar = post.meanof - 2 \log L;$$

$$Dhat = -2 \log Lat \text{ post.mean of stochastic nodes}$$

	Dbar	Dhat	pD	DIC
y	1465.900	1372.290	93.608	1559.5109
total	1465.900	1372.290	93.608	1559.5109

Finalmente repetimos el paso 9 para las nuevas 2,000 iteraciones. Asegúrese de colocar en el espacio beg 4,001 para que las estadísticas sean de las últimas 2,000 iteraciones.



Los primeros resultados aparecen a continuación

node	mean	sd	MC error	2,5 %	median	97,5 %	start	sample
<i>b</i> [1]	-1.582	0.2523	0.006718	-2.097	-1.57	-1.102	4001	2000
<i>b</i> [2]	-2.063	0.2788	0.008793	-2.627	-2.054	-1.544	4001	2000
<i>b</i> [3]	-0.0929	0.2108	0.006927	-0.5039	-0.09345	0.3108	4001	2000
<i>b</i> [4]	-2.994	0.3797	0.009508	-3.777	-2.985	-2.288	4001	2000
<i>b</i> [5]	-2.212	0.2987	0.00923	-2.801	-2.203	-1.658	4001	2000
<i>b</i> [6]	0.6497	0.2167	0.007031	0.2158	0.6483	1.061	4001	2000
<i>b</i> [7]	-2.862	0.355	0.00843	-3.59	-2.842	-2.203	4001	2000
<i>b</i> [8]	-2.286	0.297	0.008845	-2.917	-2.277	-1.735	4001	2000
<i>b</i> [9]	-1.527	0.247	0.007468	-2.031	-1.527	-1.058	4001	2000
<i>b</i> [10]	-2.13	0.2955	0.009572	-2.726	-2.118	-1.566	4001	2000
<i>b</i> [11]	-2.99	0.3734	0.01022	-3.731	-2.964	-2.312	4001	2000
<i>b</i> [12]	0.7283	0.2185	0.007656	0.303	0.7273	1.159	4001	2000
<i>b</i> [13]	-1.813	0.2614	0.007499	-2.346	-1.81	-1.33	4001	2000
<i>b</i> [14]	-3.272	0.4289	0.01122	-4.175	-3.258	-2.482	4001	2000
<i>theta</i> [1]	-0.3447	0.5904	0.0115	-1.465	-0.3554	0.8927	4001	2000
<i>theta</i> [2]	0.8446	0.6733	0.01749	-0.4474	0.8302	2.159	4001	2000
<i>theta</i> [3]	-1.292	0.548	0.01431	-2.333	-1.301	-0.2057	4001	2000
<i>theta</i> [4]	-1.308	0.5399	0.01695	-2.31	-1.32	-0.1854	4001	2000
<i>theta</i> [5]	0.8582	0.6807	0.01573	-0.4268	0.8516	2.265	4001	2000

Bibliografía

- [1] Albert, J. H. (1992). Bayesian Estimation of Normal Ogive Item Response Curves Using Gibbs Sampling. *Journal of Educational Statistics*, 17, 251 - 269.
- [2] Albert, J.H. y Ghosh, M. (2000). Item response modeling. *Generalized Linear Models: A Bayesian Perspective*(D. Dey, S. Ghosh & Mallick, eds.), Marcel-Dekker, New York, 173-193.
- [3] Andrade, D. F. , Tavares, H. R. y Valle, R. C. (2000). Introdução à Teoria da resposta ao Item : Conceitos e Aplicações. 14^o SINAPE : Caxambu, MG.
- [4] Baker, F.B. (1992). *Item Response Theory - Parameter Estimation Techniques*. New York: Marcel Dekker, Inc.
- [5] Bartholomew,D.J. y Knuot, M. (1999). *Latent variable models and factor analysis*. (2nd ed.). London: Arnold. (Kendall 's Library of Statistics 7.
- [6] Bazán, J., Bolfarine, H. y Branco, M. (2004b). A skew item response model. ISBA 2004 World Meeting. Viña del Mar. Chile, May 23-27, 2004. ISBA (International Society for Bayesian Analysis).
- [7] Bazán, J. L., Bolfarine, H. y Leandro, A. R. (2006). Sensitivity analysis of prior specification for the probit-normal IRT model: an empirical study. *Estadística*. Journal of The Inter-American Statistical Institute 58, 17-42.
- [8] Bazán, J. L., Branco, D. M. y Bolfarine, H. (2006). A skew item response model. *Bayesian Analysis*, 1 861- 892.
- [9] Bazán, J. L., (2004). Introducción al modelo psicométrico de la Teoría Clásica de los Test (parte I). *Pro Matemática*. PUCP. 18 (35-36), 79-107.

- [10] Bazán, J. (1997) Metodología estadística de construcción de pruebas. Una aplicación al estudio de las actitudes hacia la matemática en la Unalm. Tesis para optar el título de Ingeniero Estadístico. Universidad Nacional Agraria La Molina.
- [11] Bolfarine, H. y Bazán, J. L. (2007). Skewed Logit Item Response Models. Presentación oral. *Escuela de Modelos de Regresión*. Bahía, Brasil. Sesión de Poster en el *Congreso latinoamericano de probabilidad y estadística matemática*. CLAPEM. Lima-Peru. Febrero.
- [12] Bazán, J., Calderón, A. y Valdivieso, L (2009). Modelos de teoría de respuesta al ítem bajo enfoque Bayesiano- 1era. Edición- Sociedad Matemática Peruana - Lima.
- [13] Béguin, A. A.,y Glas, C. A. W. (2001). MCMC estimation of multidimensional IRT models. *Psychometrika*, 66, 541-562.
- [14] Birnbaum, A. (1968). Some Latent Trait Models and Their Use in Infering an Examinee's Ability. In Lord, F.M. y Novick, M.R. *Statistical Theories of Mental Test Scores*. eading, MA : Addison-Wesley.
- [15] Bock, R. D. y Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika* 64,153-168.
- [16] Bolt, D.M., Cohen, A.S.; y Wollack, J.A. (2002). Item parameter estimation under conditions of test speededness: Application of a mixture Rasch model with ordinal constraints. *Journal of Educational Measurement*, 39,331-348.
- [17] Bond, T. G. y Fox, C. M. (2007). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Philadelphia: Lawrence Erlbaum Associates.
- [18] Borsboom, D., Mellenbergh, G. J. y van Heerden, J.(2003). The Theoretical Status of Latent variables. *Psychological Review*, 110, 203-219.

- [19] Brooks, S. P. (2002). Discussion on the paper by Spiegelhalter, Best, Carlin, and van de Linde. *Journal of the Royal Statistical Society Series B*, 64, 3,616-618.
- [20] Calderón, A. (2005). Una aplicación del Modelo de Respuesta al Item en el diagnóstico de Trastornos Alimentarios. Monografía para optar al título de Ingeniero Estadístico. Universidad Nacional Agraria La Molina. Lima
- [21] 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.
- [22] Casella, G. y Berger, R. L (2002). *Statistical Inference*, Duxbury: Pacific Grove, CA.
- [23] Chen, M-H, Shao, Q. M. y Ibrahim, J. G (2000). *Monte Carlo Methods in Bayesian Computation*. New York: Springer Verlag.
- [24] Chincaro, O. y Bazán, J. (2010). Una comparación de métodos de estimación del modelo de Rasch. Poster presentados en la *I Jornada Internacional de Probabilidad y Estadística JIPE 2010*. Lima: PUCP, Febrero.
- [25] Delorio, M. y Robert, C. P. (2002). Discussion on the Paper by Spiegelhalter, Best, Carlin and van der linde (2002). *Journal Royal Statistics Society. Serie B*. 64, 629-630.
- [26] Dempster, A. P., Laird N.M. y Rubin, D.B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society B*, 39, 1, 1-38.
- [27] Fischer, G. y Molenaar, I. (1995). *Rasch Models. Foundations, recent development, and applications*. The Nerthelands: Springer-Verlag.
- [28] Fox, J. P. y Glas, C. A.W. (2001). Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika*, 66, 271-288.

- [29] Fox, J. P. y Glas, C. A.W. (2003). Bayesian Modeling of measurement error in predictor variables using item response theory. *Psychometrika*, 68, 169-191.
- [30] Gamerman, D. y Lopes, H. F (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, Chapman and Hall/CRC” .
- [31] Gelfand, A. E. (1996). Model Determination using Sampling-based methods. En *Markov Chain Monte Carlo in Practice*, Gilks, W.R. y Richardson, S. y Spiegelhalter, D. J. (editor) capítulo 9, 145-161.
- [32] Gelfand, A.E. y Smith, A.F.M. (1990). Sampling-based approaches to calculating marginal densities. *J. Am. Stat. Assoc.* 85, 398-409.
- [33] Gelman, A. y Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7, 457-472.
- [34] Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. En: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (eds.) *Bayesian Statistics*, vol. 4, pp. 169-193. Oxford University Press, Oxford .
- [35] Ghosh, M., Ghosh, A., Chen, Ming-Hui y Agresti, A. (2000). Noninformative priors for one parameter item response models. *Journal of Statistical Planning and Inference*. 88, 99-115.
- [36] Gilks, W. R., Richardson, S. e Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in practice*. London : Chapman & Hall.
- [37] Gilks, W. R. y Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*, 41, 337-348.
- [38] Goegebeur, Y., De Boeck, P., Wollack, J.A. y Cohen, A.S. (2008). A Speeded Item Response Model with Gradual Process Change. *Psychometrika*, 73,65-87.

- [39] Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains. *Biometrika* 57, 97-109.
- [40] Holland, P. y Rosenbaum, P. (1986). Conditional Association and Unidimensionality in Monotone Latent variable models. *The Annals of Statistics*. 14 1523-1543.
- [41] Johnson, V. y Albert, J. (2000). *Ordinal Data Modeling*. New York, MA: Springer-Verlag.
- [42] Johnson, T. R. (2003). On the use of heterogeneous thresholds ordinal regression models to account for individual differences in response style. *Psychometrika*, 68(4), 563-583.
- [43] Linardakis, M. y Dellaportas, P. (2002). An approach to multidimensional item response modeling. E. I. George. (Ed.). *Bayesian Methods with applications to science policy and oficial statistics*. 331-340.
- [44] Lord, F. M. (1952). *A theory of test scores*. New York: Psychometric Society.
- [45] Lord, F. y Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Adisson-Wesley.
- [46] Martin, A.D. y Quinn, K. M. (2002). [MCMCpack.scythe.wustl.edu/mcmcpack.html](http://mcmcpack.scythe.wustl.edu/mcmcpack.html).
- [47] Martinez, P., Zusman, L., Hartley, J., Morote R. y Calderón A.D. (2003). Estudio epidemiológico de los erantornos elementarios y factores esociados en Lima Metropolitana. *Revista de Psicología*, 2, 234-269.
- [48] Neyman, J. y Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrika*, 16(1), 1-32.
- [49] Patz, R. J., y Junker, B. W. (1999). A straightforward approach to Markov Chain Monte Carlo methods for item response models. *Journal of Educactional and Behavioral Statistics*, 24, 146-178.

- [50] Rivers, D. (2003). *Identification of Multidimensional Spatial Voting Models*. Technical Report: Political Methodology Working Papers. Stanford University.
- [51] Roberts, C., P.(2001) *The Bayesian Choice: from decision-theoretic foundations to computational implementation*. 2nd ed. New york: Springer-Verlag.
- [52] Rupp, A., Dey, D. K., y Zumbo, B. (2004). To Bayes or Not to Bayes, from Whether to When: Applications of Bayesian Methodology To Item Response Modeling. *Structural Equations Modeling*. 11, 424-451.
- [53] Ross, S. (1995). *Stochastic Processes*, Wiley: New York, NY.
- [54] Schervish, M. J. (1995). *Theory of Statistics*, Springer-Verlag: New York, NY.
- [55] Sahu, S. K. (2002). Bayesian estimation and model choice in item response models. *Journal Statistical Computing Simulation*, 72,217-232.
- [56] Samejima, F (2000). Logistic positive exponent family of models: Virtue of asymmetric item characteristic curves. *Psychometrika*, 65:3: 319-335.
- [57] Schervish, M. J. (1995). *Theory of Statistics*, Springer-Verlag: New York, NY.
- [58] Spiegelhalter, D. J., Thomas, A., Best, N. G. y Gilks, W.R.(1996). *BUGS 0.5 examples* (Vol. 1 Version i). Cambridge, UK: University of Cambridge.
- [59] Tierney, L. (1994). Markov chains for exploring posterior distributions. *Ann. Stat.* 22, 1701-1762.
- [60] Valdivieso, L. y Bazán, J. L. (2009). Bayesian Estimation in a Speed Item Response Model. Presentación oral. *1er Congresso Brasileiro de Teoria de Resposta ao Item*. Florianopolis Brasil. Diciembre.

- [61] van der Linden, W.J. y Hambleton, R. K. (1997). *Handbook of Modern Item response Theory*. New York: Springer-Verlag.
- [62] Wang, X, Bradlow, E. T. y Wainer, H. (2003). A General Bayesian Model for Testlets: Theory and Applications. *Applied Psychological Measurement*, 26, 109-128.
- [63] Yamamoto, K. y Everson, H. (1997). Modeling the effects of test length and test time on parameter estimation using the hybrid model. Applications of latent trait and latent class models in the social sciences. In J. Rost & R. Langeheine (Eds.), New York: Waxmann, 89-99.
- [64] Zusman, L (2000). Las conductas alimenticias en adolescentes mujeres de Lima Metropolitana: sus manifestaciones y las características personales y familiares asociadas. Tesis de Maestría PUCP.

Dr. Jorge Luis Bazán
Pontificia Universidad Católica del Perú
email: jlbazan@pucp.edu.pe

Dr. Luis Valdivieso
Pontificia Universidad Católica del Perú
email: lvaldiv@pucp.edu.pe

Lic. Arturo Calderón
Pontificia Universidad Católica del Perú
email: acalder@pucp.edu.pe