

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

ESCUELA DE GRADUADOS



MODELOS TESTLET LOGÍSTICOS Y LOGÍSTICOS DE
EXPONENTE POSITIVO PARA PRUEBAS DE
COMPRESIÓN DE TEXTOS

TESIS PARA OPTAR POR EL GRADO DE MAGISTER EN
ESTADÍSTICA

Presentado por:

Sandra Elizabeth Flores Ari

Asesor: Dr. Jorge Luis Bazán Guzmán

Miembros del jurado:

Dr. Luis Hilmar Valdivieso Serrano

Dra. Mery Elizabeth Doig Camino

Lima, Julio 2012

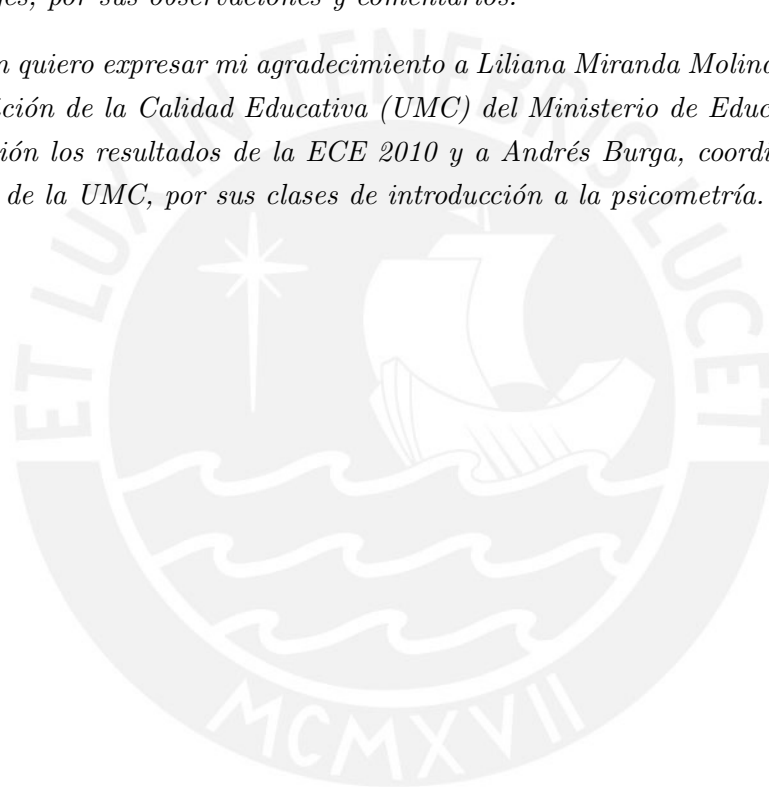
Dedicatoria

A mi hijo Benjamin, mi esposo Jorge y mis padres Donata y Jorge.



Agradecimientos

- *Agradezco de manera especial al Dr. Bazán por haberme guiado en el desarrollo de la presente tesis.*
- *A los docentes de la maestría en Estadística de la PUCP: Dr. Valdivieso, Dra. Doig, Dr. Bayes, por sus observaciones y comentarios.*
- *También quiero expresar mi agradecimiento a Liliana Miranda Molina jefa de la Unidad de Medición de la Calidad Educativa (UMC) del Ministerio de Educación por poner a disposición los resultados de la ECE 2010 y a Andrés Burga, coordinador del área de análisis de la UMC, por sus clases de introducción a la psicometría.*



Resumen

Los modelos de Teoría de Respuesta al Ítem (TRI) para datos binarios multivariados, permiten estimar una medida latente (de habilidad) a partir de información observada, que puede ser respuestas dicotómicas (de éxito y fracaso) a un conjunto de ítems de una determinada prueba. Uno de los supuestos críticos en los modelos TRI es la independencia condicional de los ítems, que permite el cálculo directo de la verosimilitud del modelo. En muchas situaciones de evaluación este supuesto no se cumple, como es el caso de pruebas de comprensión de textos, en la que se presenta un texto y luego varias preguntas relacionadas con ese texto. Este tipo de estructuras son denominadas como testlets. Bradlow et al. (1999) desarrollaron una parametrización adicional para recoger el efecto de esta dependencia. A partir de este trabajo se presenta el modelo Testlet logístico y se propone el modelo Testlet logístico de exponente positivo (2LPET), que es una extensión del modelo LPE propuesto por Samejima (1999) y Bazan y Bolfarine (2010) y considera enlaces asimétricos.

Se desarrollaron varios estudios de simulación en los que se muestra que cuando se tiene testlets, los modelos Testlet recuperan mejor los parámetros respecto a los modelos TRI.

Finalmente se realizó una aplicación con datos del Ministerio de Educación, específicamente con los resultados de la prueba de comprensión de textos de la Evaluación Censal de Estudiantes (ECE) dirigido a estudiantes de segundo grado de primaria, en un conjunto de escuelas de Lima metropolitana. De los resultados obtenidos se concluye que los modelos TRI sobreestiman la medida de habilidad respecto a los modelos Testlets y además la información de la prueba es mejor distribuida por el modelo propuesto.

Palabras-clave: Testlet, Asimetría, Teoría de respuesta al ítem, Análisis Bayesiano.

Abstract

Item response theory (IRT) models to multivariate dichotomous responses from a test with a set of items, permit to estimate a latent measure (ability). A critical assumption in IRT is conditional independence item responses, in a lot of test situation this assumption fail, like reading comprehension, where a set of items are related to a reading passage (testlet). An additional random effect in standard IRT model was developed by [Bradlow et al. \(1999\)](#), inspired in this paper, logistic Testlet model is presented and a logistic positive exponent Testlet model (2LPET) is suggested, this is an extension for LPE model proposed by [Samejima \(1999\)](#) and [Bazan y Bolfarine \(2010\)](#) and has asymmetric item characteristic curves (ICC).

Simulation studies were developed, when a set of responses are in testlets, parameters are better estimated for Testlet than TRI models, and skewed ICC could be adequate in some situations.

An application with Ministry of Education data set was developed, specifically the results for reading comprehension from Census Student Testing, conducted to students for second grade of primary education in a set of schools from Lima metropolitana, the results lead to come the conclusion that TRI models overestimate the ability measure and the proposal model divide the test information in the best way.

Keywords: Testlet, Skewness, Item response theory, Bayesian modeling.

Índice general

Lista de Abreviaturas	IX
Índice de figuras	X
Índice de cuadros	XII
1. Introducción	1
1.1. Consideraciones Preliminares	1
1.2. Objetivos	3
1.3. Organización del Trabajo	3
2. Distribución Logística y Logística Generalizada	5
2.1. Distribución Logística	5
2.1.1. Función de densidad de probabilidad	5
2.1.2. Función de distribución acumulada	5
2.1.3. Función cuantil	5
2.1.4. Momentos de orden superior	6
2.1.5. Propiedades	7
2.1.6. Parametrización alternativa	7
2.2. Distribución Logística Generalizada	7
2.2.1. Función de distribución acumulada	8
2.2.2. Función de densidad de probabilidad	8
2.2.3. Función generadora de momentos	9
2.2.4. Función de distribución completa	10
3. Teoría de Respuesta al Item y Testlet	12
3.1. Definiciones generales	12
3.1.1. Definición general de un modelo TRI	13
3.1.2. Supuestos de un modelo TRI	13
3.1.3. Características del modelo TRI	14
3.1.4. Modelos Testlet	15
3.2. El modelo TRI Logístico	16
3.2.1. Definición	16
3.2.2. Información de Fisher	17
3.2.3. El modelo de un parámetro	20
3.3. El modelo TRI Logístico de exponente positivo	21

3.3.1.	Definición	21
3.3.2.	Función de Información	22
3.3.3.	El modelo LPE de un parámetro	23
3.4.	El modelo Testlet Logístico	23
3.4.1.	Definición	24
3.4.2.	Función de Información	26
3.4.3.	El modelo Testlet logístico de un parámetro	26
3.5.	El modelo Testlet Logístico de Exponente Positivo	26
3.5.1.	Definición	26
3.5.2.	Función de Información	27
3.5.3.	Casos particulares del Modelo Testlet logístico de exponente positivo	28
4.	Estimación Bayesiana del Modelo Testlet LPE	29
4.1.	Definiciones Generales	29
4.2.	Prioris en TRI	30
4.3.	Función de Verosimilitud para el modelo 2LPET	31
4.4.	Prioris en el modelo 2LPET	31
4.4.1.	Especificación de la priori	31
4.5.	Inferencia bayesiana	32
4.6.	Criterios de comparación de modelos	33
5.	Estudio de Simulación	35
5.1.	Estudio 1	35
5.1.1.	Objetivo	35
5.1.2.	Algoritmo para simular los datos	35
5.1.3.	Método para estimar los parámetros	36
5.1.4.	Criterios para evaluar la simulación	36
5.1.5.	Escenario	36
5.1.6.	Resultados	37
5.2.	Estudio 2	40
5.2.1.	Objetivo	40
5.2.2.	Algoritmo para simular los datos	40
5.2.3.	Método para estimar los parámetros	41
5.2.4.	Criterios para evaluar la simulación	42
5.2.5.	Escenario	42
5.2.6.	Resultados	42
5.3.	Estudio 3	45
5.3.1.	Objetivo	45
5.3.2.	Algoritmo para simular los datos	45
5.3.3.	Método para estimar los parámetros	47
5.3.4.	Criterios para evaluar la simulación	47
5.3.5.	Escenario	47
5.3.6.	Resultados	47

6. Aplicación	54
6.1. Instrumento de Medición	54
6.2. Muestra	56
6.3. Recolección de datos	56
6.4. Análisis de los datos	56
6.4.1. Análisis clásico de los datos	57
6.5. Estimación de parámetros mediante los modelos desarrollados	58
6.5.1. Modelo TRI Logístico (2L)	58
6.5.2. Modelo Testlet Logístico (2LT)	58
6.5.3. Modelo TRI Logístico de Exponente Positivo (2LPE)	60
6.5.4. Modelo Testlet Logístico de Exponente Positivo (2LPET)	62
6.6. Comparación de modelos	62
6.7. Función de Información	64
6.8. Resultados del modelo 2LPET	66
7. Conclusiones	69
7.1. Conclusiones	69
7.2. Sugerencias para investigaciones futuras	70
A. Markov Chain Monte Carlo Methods (MCMC)	71
A.1. Metropolis-Hastings Sampling	71
A.2. Muestreador de Gibbs (MGibbs)	72
B. Programas en OpenBUGS y R	73
B.1. Modelo TRI Logístico de dos parámetros (2L)	73
B.2. Modelo Testlet Logístico de dos parámetros (2LT)	73
B.3. Modelo TRI Logístico de Exponente Positivo (LPE)	74
B.4. Modelo Testlet Logístico de Exponente Positivo (LPET)	74
B.5. Código en R para el proceso de simulación - Estudio 1	74
B.6. Código en R para el proceso de simulación - Estudio 2	80
B.7. Código en R para el proceso de simulación - Estudio 3	84
B.8. Código en R para calcular la Función de Información	88
C. Prueba de Comprensión de Textos - ECE2010	92
Bibliografía	97

Lista de Abreviaturas

2L	Modelo TRI logístico de dos parámetros.
2LT	Modelo Testlet logístico de dos parámetros .
2LPE	Modelo TRI logístico de exponente positivo de dos parámetros.
2LPET	Modelo Testlet logístico de exponente positivo de dos parámetros.
CCI	Curva característica del ítem.
CCIs	Curvas características de los ítems.
C1	Cuadernillo no.1 en la ECE.
ECE	Evaluación censal de estudiantes.
fdp	Función de densidad de probabilidad.
fa	Función de distribución acumulada.
fgm	Función generadora de momentos.
LPE	Modelo logístico de exponente positivo.
RMSE	Raíz del error cuadrático medio.
MAE	Media absoluta del error.
DIC	Criterio de información de desvío.
TRI	Teoría de respuesta al ítem.

Índice de figuras

2.1. Densidades de la distribución logística para $\mu = 0$ y diferentes valores de s . . .	6
2.2. Distribución logística acumulada para $\mu = 0$ y diferentes valores de s	6
2.3. Función de distribución acumulada de la logística generalizada para diferentes valores de d	8
2.4. Densidades de la distribución logística generalizada para diferentes valores de d	9
2.5. Coeficiente de asimetría de la distribución logística generalizada tipo I para d	11
2.6. Kurtosis de la distribución logística generalizada tipo I para d	11
3.1. Probabilidad de responder correctamente a un ítem para $a = 1$ y diferentes valores de b	17
3.2. Probabilidad de responder correctamente a un ítem para $b = 0$ y diferentes valores de a	18
3.3. Función de información del ítem para diferentes valores de a (izquierda) y diferentes valores de b (derecha)	20
3.4. Función de información de test para los 3 ítems presentados en la figura 3.3 .	20
3.5. CCI en la familia de modelos logísticos de exponente positivo con parámetro de discriminación único $a_j = 1$ y con parámetro de dificultad $b_j = 0$ y diferente parámetro de penalidad $d = 0.3, 0.5, 1, 2$ y 5	22
3.6. Función de información del ítem de la familia de modelos logísticos de exponente positivo con parámetro de discriminación único $a_j = 1$, parámetro de dificultad único $b_j = 0$ y siete valores para el parámetro de penalidad $d = 0.3, 0.5, 0.8, 1, 1.5, 2$ y 3	24
5.1. Raíz del error cuadrático medio (RMSE) de los estimadores simulados para la discriminación (a) y dificultad (b) de cada ítem para un ajuste Testlet logístico (2LT) y un ajuste TRI logístico (2L), considerando $m = 20$ réplicas	38
5.2. Media absoluta del error (MAE) de los estimadores simulados para la discriminación (a) y dificultad (b) de cada ítem para un ajuste Testlet logístico (2LT) y un ajuste TRI logístico (2L), considerando $m = 20$ réplicas	39
5.3. Sesgo en la estimación de los parámetros a y b obtenidos por los modelos Testlet logístico (2LT) y TRI logístico (2L), considerando $m = 20$ réplicas . .	40
5.4. Estimación de la habilidad para un ajuste Testlet logístico (2LT) y un ajuste TRI logístico (2L)	41

5.5. Raiz del Error Cuadrático Medio (RMSE) de los estimadores simulados para la discriminación (a) y dificultad (b) de cada ítem con un modelo TRI logístico de exponente positivo (2LPE) y TRI logístico (2L), considerando $m = 20$ réplicas	44
5.6. Media Absoluta del Error de los estimadores simulados para la discriminación (a) y dificultad (b) de cada ítem con un modelo TRI logístico de exponente positivo (2LPE) y TRI logístico (2L), considerando $m = 20$ réplicas	45
5.7. Estimación de la habilidad (θ) por el modelo TRI logístico de exponente positivo (2LPE) y TRI logístico (2L), considerando $m = 20$ réplicas	46
5.8. Raiz del Error Cuadrático Medio (RMSE) de los estimadores simulados, para la discriminación (a), dificultad (b) y el parámetro de penalidad (d) de cada ítem, con el modelo Testlet logístico de exponente positivo (2LPE) y TRI logístico de exponente positivo (2LPE), considerando $m = 20$ réplicas	50
5.9. Media Absoluta del Error de los estimadores simulados, para la discriminación (a), dificultad (b) y penalidad (d) de cada ítem, con un modelo Testlet logístico de exponente positivo (2LPET) y TRI logístico de exponente positivo (2LPE), considerando $m = 20$ réplicas	51
5.10. Sesgo de los estimadores simulados, para la discriminación (a), dificultad (b) y penalidad (d) de cada ítem, con un modelo Testlet logístico de exponente positivo (2LPET) y TRI logístico de exponente positivo (2LPE), considerando $m = 20$ réplicas	52
5.11. Estimación promedio de la habilidad (θ) por el modelo Testlet logístico de exponente positivo (2LPET) y TRI logístico de exponente positivo (2LPE), considerando $m = 20$ réplicas	53
6.1. Distribución de las puntuaciones totales	57
6.2. Curvas características del ítem para los modelos 2L y 2LPE	62
6.3. Dispersión de las habilidades estimadas con el modelo 2L y 2LT	64
6.4. Dispersión de las habilidades estimadas con el modelo 2L y 2LPE Testlet	64
6.5. Función de información de la prueba para los diferentes modelos planteados	65
6.6. Medida de habilidad (θ) estimada por el modelo 2LPET	66
6.7. Discriminación y dificultad de los ítems, en el modelo 2LPET	67
6.8. Dificultad y penalidad de los ítems en el modelo 2LPET	67
6.9. Penalidad y dificultad de los ítems en el modelo 2LPET	68

Índice de cuadros

5.1. Resultados de medidas promedio de la simulación para comparar los modelos Testlet logístico (2LT) y TRI logístico (2L), con $m = 20$ réplicas	37
5.2. Raíz del error cuadrático medio y Media absoluta del error de los estimadores en la simulación - Estudio 1, considerando $m = 20$ réplicas	37
5.3. Resultados de medidas de ajuste de la simulación para comparar los modelos Testlet logístico (2LT) y TRI logístico (2L)	38
5.4. Medidas de correlación entre los valores simulados y los estimados promedio, de los modelos Testlet logístico (2LT) y TRI logístico (2L), considerando $m = 20$ réplicas	39
5.5. Resultados de medidas promedio de la simulación para comparar el modelo TRI logístico de exponente positivo (2LPE) y TRI logístico (2L), considerando $m = 20$ réplicas	42
5.6. Raíz del error cuadrático medio y Media absoluta del error de los estimadores en la simulación - Estudio 2, considerando $m = 20$ réplicas	43
5.7. Resultados de medidas de ajuste de la simulación para comparar los modelos TRI Logístico de Exponente Positivo (2LPE) y TRI Logístico (2L), considerando $m = 20$ réplicas	43
5.8. Medidas de correlación entre los valores verdaderos y los estimados promedio de los modelos TRI logístico de exponente positivo (2LPE) y TRI logístico (2L)	43
5.9. Resultados del promedio de RMSE y MAE en el estudio de simulación 3, para comparar los modelos Testlet logístico de exponente positivo (2LPET) y TRI logístico de exponente positivo (2LPE), considerando $m = 20$ réplicas	48
5.10. Raíz del error cuadrático medio y Media absoluta del error de los estimadores en la simulación - Estudio 3, considerando $m = 20$ réplicas.	49
5.11. Resultados de medidas promedio de ajuste de la simulación para comparar los modelos Testlet logístico de exponente positivo (2LPET) y TRI logístico de exponente positivo (2LPE)	50
5.12. Medidas de correlación entre los valores verdaderos y los estimados promedio de los modelos Testlet logístico de exponente positivo (2LPET) y TRI logístico de exponente positivo (2LPE), considerando $m = 20$ réplicas	51
6.1. Cuadernillos ECE 2010	55
6.2. Estructura de los testlet en el instrumento de medición	55
6.3. Estadísticos del score	57
6.4. Estimación de los parámetros del modelo 2L	59

6.5. Estimación de los parámetros de ítem del modelo 2LT	60
6.6. Varianza estimada de los testlets en el modelo 2LT	60
6.7. Estimación de los parámetros de ítem del modelo 2LPE	61
6.8. Estimación de los parámetros de ítem del modelo 2LPET	63
6.9. Varianza estimada de los testlets en el modelo 2LPET	63
6.10. Medidas de comparación de los modelos	63
6.11. Estadísticos de la medida de habilidad (θ) estimada por el modelo 2LPET . .	66



Capítulo 1

Introducción

1.1. Consideraciones Preliminares

La Teoría de Respuesta al Ítem (TRI) considera modelos para describir cómo la probabilidad de responder correctamente a un ítem de una prueba depende de la habilidad del sujeto o habilidad latente y, de los parámetros relacionados a los ítems (Bazan, Bolfarine y Aparecida, 2006).

Los modelos TRI de respuesta binaria o dicotómica son aquellos en los que las variables manifiestas u observadas son matrices de respuestas binarias (éxito y fracaso). Los más comunes, son denominados modelos logísticos y de ojiva normal y tienen como casos particulares los modelos TRI de 1, 2 y 3 parámetros en función al número de parámetros de ítem que afecta la probabilidad de respuesta al ítem (Bazan et al., 2009).

Un modelo TRI logístico de respuesta binaria puede ser definido, considerando las respuestas de n sujetos a k ítems condicionalmente independientes, donde $Y_{ij} = 1$ denota la respuesta correcta del sujeto i al ítem j e $Y_{ij} = 0$ denota la respuesta incorrecta, es decir:

$$Y_{ij} \sim \text{Bernoulli}(p_{ij}) \quad (1.1)$$

$$p_{ij} = F(m_{ij}) \quad (1.2)$$

$$i = 1, 2, \dots, n; \quad j = 1, 2, \dots, k$$

donde, p_{ij} , es la probabilidad de obtener una respuesta correcta y F , denominada curva característica del ítem (CCI), es la función de distribución acumulada (fda) de la distribución logística, es decir $F(t) = (1+e^{-t})^{-1}$. La función inversa de F es denominada función de enlace. Además se asume que,

$$m_{ij} = a_j(\theta_i - b_j) \quad (1.3)$$

es un predictor lineal latente, donde $\theta_i \in R$ es el parámetro de habilidad del sujeto i , $a_j > 0$ es el parámetro de discriminación del ítem j , $b_j \in R$ es la dificultad del ítem j . Esta forma es conocida como el modelo logístico de dos parámetros, un caso particular se da cuando se fija el valor de la pendiente $a_j = 1$, que nos da como resultado el modelo logístico de un parámetro, conocido como el modelo de Rash.

Una crítica asumida en la TRI, como lo mencionan Bock y Moustaki (2007), es que las

respuestas a los ítems, dada la localización de los sujetos en el espacio latente, son estadísticamente independientes. Esta suposición es importante en la TRI, porque permite escribir de manera directa la densidad conjunta del vector de respuestas dado el vector de variables latentes y el vector de parámetros de los ítems (Bazan, Calderon y Valdivieso, 2010) y en consecuencia permite el cálculo directo de la verosimilitud de los parámetros del modelo. Sin embargo, en muchas situaciones de evaluación este supuesto no se cumple. Por ejemplo, cuando existe un conjunto de ítems que está relacionado a un estímulo común, como es el caso de las pruebas de comprensión de textos, donde se presenta un texto y luego varias preguntas relacionadas con ese texto. Esta estructura de la prueba genera dependencia en los ítems, en el sentido de que los ítems comunes de una lectura se encuentran asociados.

Según Bradlow, Wainer y Wang (1999) la dependencia de los ítems a un estímulo común, denominado testlet, produce dependencia en las respuestas del ítem para un sujeto dado. Si se asume independencia condicional en estos casos, se tiene como resultado la sobreestimación o subestimación de la precisión de la habilidad estimada, y un sesgo en la estimación de los parámetros. Bradlow, Wainer y Wang (1999) sugieren una parametrización aproximada que modifica el modelo TRI estándar de ojiva normal y permite la detección, modelamiento explícito, y evaluación del grado de dependencia condicional del ítem. El modelo testlet modifica el predictor lineal latente (1.3), como:

$$m_{ij} = a_j(\theta_i - b_j - g_{ih(j)}) \quad (1.4)$$

donde $g_{ih(j)}$ recoge el efecto testlet específico del sujeto i , independientemente de la habilidad y de los parámetros de los ítems, $h(j)$ denota el testlet del ítem j .

Por otro lado, en algunas situaciones sucede que la CCI no necesariamente debe ser una distribución simétrica, como es el caso de la distribución logística. Una característica de una CCI simétrica, es que esta trata ambas, las respuestas correctas e incorrectas simétricamente, que es una contradicción lógica en el ordenamiento de los examinados en el tratamiento latente o escala de habilidad (Samejima, 1999). Samejima propuso como CCI la función de distribución logística de exponente positivo (LPE) y recientemente Bazan y Bolfarine (2010) desarrollaron el modelo LPE y un modelo de reflexión, manifestando que en el contexto de regresión binaria, enlaces simétricos no siempre proveen buenos estimadores para algunos conjuntos de datos, en especial cuando la probabilidad de tener una respuesta binaria se aproxima a cero. Este último modelo considera que la CCI esta definida como $F(t) = (1 + e^{-t})^{-\lambda}$, donde λ es considerado como un nuevo parámetro de ítem, llamado de penalidad.

En nuestro país, Chincaro (2010) desarrolló el modelo Rash y su estimación desde el punto de vista clásico y bayesiano mediante diferentes softwares. El modelo Rash es utilizado por el Ministerio de Educación para reportar los niveles de logro de los estudiantes. Una de las conclusiones del trabajo de Chincaro es que la estimación por el método de máxima verosimilitud marginal y conjunta presentan problemas cuando los estudiantes han respondido

correcta o incorrectamente todos los ítems de la prueba, y concluye también que el mejor método de estimación es mediante el análisis bayesiano.

Por estas consideraciones en el presente trabajo se estudia los modelos Testlet logísticos de uno y dos parámetros, comparándolos con los modelos TRI de uno y dos parámetros, y se considera además enlaces asimétricos como una propuesta en los modelos Testlet, todo esto analizado desde la perspectiva bayesiana.

1.2. Objetivos

El objetivo general del presente trabajo es estudiar los modelos Testlet de uno y dos parámetros con enlace logístico simétrico y asimétrico, comparándolos con los modelos TRI de uno y dos parámetros con enlace logístico simétrico y asimétrico, desde la perspectiva bayesiana. Para lograr este objetivo definimos los siguientes objetivos específicos:

- Revisar literatura de los modelos Testlet de uno y dos parámetros, con enlace logístico simétrico y asimétrico, y los modelos TRI de uno y dos parámetros con enlace logístico simétrico y asimétrico.
- Estudiar e implementar los modelos Testlet de uno y dos parámetros con enlace logístico simétrico y asimétrico desde la perspectiva bayesiana.
- Describir los modelos TRI de uno y dos parámetros con enlace logístico y logístico asimétrico desde la perspectiva bayesiana.
- Realizar estudios de simulación comparativos de los modelos Testlets y TRI desarrollados.
- Aplicar los modelos desarrollados a los resultados obtenidos de la prueba de Comprensión de Textos en la Evaluación Censal de Estudiantes 2010 en una muestra de estudiantes de 2do. grado de primaria de Lima Metropolitana.

1.3. Organización del Trabajo

En el Capítulo 2, se presenta la definición de la distribución logística y logística generalizada, éstas definiciones serán utilizadas como CCI en la definición del modelo TRI y Testlet. En el Capítulo 3, se define y se presentan propiedades de los modelos TRI y Testlet logístico y logístico de exponente positivo. Luego la estimación de los parámetros a partir del modelo más general, el modelo Testlet logístico de exponente positivo (2LPET) se presenta en el capítulo 4. Diversos estudios de simulación son presentados en el capítulo 5. En el Capítulo 6 se analizan los resultados de aplicar los modelos propuestos al estudio de la comprensión lectora usando datos del Ministerio de Educación. Finalmente se presentan conclusiones y sugerencias del presente trabajo en el capítulo 7.

En el anexo (Apéndice A) se presenta los métodos de estimación MCMC.

En el anexo (Apéndice **B**) se presenta los programas de simulación y los utilizados en las aplicaciones a conjuntos de datos reales.

En el anexo (Apéndice **C**) se presenta la prueba de Comprensión Lectora elaborada por el Ministerio de Educación para el 2do. grado de primaria y aplicada en la evaluación censal de estudiantes (ECE) del año 2010.



Capítulo 2

Distribución Logística y Logística Generalizada

En este capítulo se presenta la definición y características principales de la distribución logística y logística generalizada. Las referencias usadas son [Johnson y Kotz \(1994\)](#), [Johnson et al. \(1995\)](#), [Wikipedia \(2011\)](#) y [Balakrishnan y Leung \(1988\)](#).

2.1. Distribución Logística

Es una distribución de probabilidad continua simétrica que gráficamente es similar a la distribución normal pero con colas más pesadas. Su función de distribución acumulada es la función logística.

2.1.1. Función de densidad de probabilidad

Una variable aleatoria Y tiene distribución logística si su función de densidad de probabilidad (fdp) está dado por:

$$f(y | \mu, s) = \frac{e^{-(y-\mu)/s}}{s(1 + e^{-(y-\mu)/s})^2} \quad -\infty < y < \infty \quad (2.1)$$

donde la variable aleatoria y es una cantidad real, μ es el parámetro de localización y s es un parámetro de escala positivo proporcional a la desviación estándar.

En la figura 2.1 podemos observar la densidad de la distribución logística para $\mu = 0$ y diferentes valores de s .

2.1.2. Función de distribución acumulada

La función de distribución acumulada (fda), que pertenece a la familia de funciones logísticas, está dado por:

$$F(y | \mu, s) = \frac{1}{1 + e^{-(y-\mu)/s}} \quad (2.2)$$

2.1.3. Función cuantil

La función de distribución acumulada inversa o cuantil, es encontrada resolviendo $F(y) = p$, y viene dado por:

$$F^{-1}(p | \mu, s) = \mu + s \times \ln\left(\frac{p}{1-p}\right) \quad (2.3)$$

la cual es una generalización de la función logit, donde \ln es el logaritmo natural, definido como el logaritmo cuya base es $e \approx 2,718281828$.

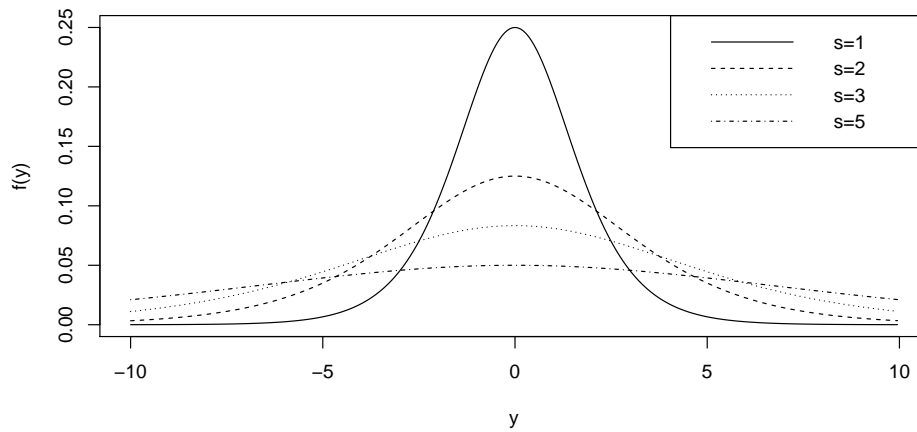


Figura 2.1: Densidades de la distribución logística para $\mu = 0$ y diferentes valores de s

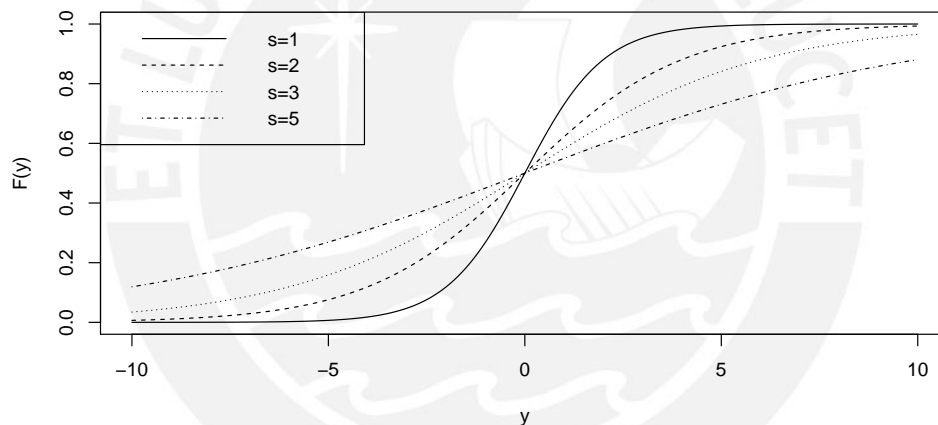


Figura 2.2: Distribución logística acumulada para $\mu = 0$ y diferentes valores de s

2.1.4. Momentos de orden superior

El n -ésimo momento de orden central, puede ser expresado en términos de la función cuantil:

$$E[(y - \mu)^n] = \int_{-\infty}^{\infty} (y - \mu)^n dF(y) = \int_0^1 (F^{-1}(p) - \mu)^n dp = s^n \int_0^1 \left[\ln\left(\frac{p}{1-p}\right) \right]^n dp \quad (2.4)$$

Esta integral luego se reduce a:

$$E[(y - \mu)^n] = s^n \pi^n (2^n - 2) \cdot |B_n| \quad (2.5)$$

con B_n que denota los números Bernoulli.

Los números de Bernoulli se puede definir mediante la siguiente forma recursiva:

$$B_0 = 1$$

$$B_m = - \sum_{k=0}^{m-1} \binom{m}{k} \frac{B_k}{m-k+1} \quad (2.6)$$

2.1.5. Propiedades

De la ecuación (2.5) obtenemos lo siguiente:

El valor esperado esta dado por:

$$E(y) = \mu$$

La varianza esta dado por:

$$Var(y) = \sigma^2 = \frac{\pi^2}{3} s^2 \quad (2.7)$$

de donde se deduce $s = \frac{\sqrt{3}}{\pi} \sigma$ que ha sido definido como el parámetro de escala de la distribución logística en la ecuación(2.1) y es proporcional a σ .

El coeficiente de asimetría esta dado por:

$$\begin{aligned} \gamma_1(y) &= E \left[\left(\frac{y - \mu}{\sigma} \right)^3 \right] = \frac{\mu_3}{\sigma^3} \\ &= \frac{s^3 \pi^3 (2^3 - 2) \cdot |B_3|}{\sigma^3} = 0 \end{aligned} \quad (2.8)$$

La kurtosis esta dada por:

$$\gamma_2(y) = \frac{\mu_4}{\sigma^4} - 3 \quad (2.9)$$

Reemplazando 2.7, tenemos:

$$\begin{aligned} &= \frac{s^4 \pi^4 (2^4 - 2) \cdot |B_4|}{\sigma^3} - 3 \\ &= 9 * (2^4 - 2) \cdot |B_4| - 3 = \frac{6}{5} \end{aligned}$$

Como se puede observar la distribución logística tiene un coeficiente de asimetría igual a cero $\gamma_1(y) = 0$ y la kurtosis es igual a $\gamma_2(y) = 6/5$.

2.1.6. Parametrización alternativa

Una parametrización alternativa esta dado por la siguiente función de densidad de probabilidad:

$$g(y | \mu, s) = f(y | \mu, s) = \frac{1}{4s} \operatorname{sech}^2 \left(\frac{y - \mu}{2s} \right) \quad (2.10)$$

donde, *sech* es la secante hiperbólica definido como $\operatorname{sech}(x) = \frac{2}{e^x + e^{-x}}$

2.2. Distribución Logística Generalizada

El término distribución logística generalizada hace referencia a varias familias de distribución de probabilidad. [Johnson y Kotz \(1994\)](#) listan cuatro formas. Nos centraremos en la

familia de Tipo I, también denominada distribución **skew-logística**.

La siguiente definición es para una variable x de una familia de versión estandarizada de la variable aleatoria y . La forma completa se define más adelante en 2.2.4.

2.2.1. Función de distribución acumulada

Dado la función

$$L(x) = \frac{1}{1 + e^{-x}} \quad (2.11)$$

se define,

$$F(x) = [L(x)]^d = \left[\frac{1}{1 + e^{-x}} \right]^d, \quad -\infty < x < \infty, \quad d > 0 \quad (2.12)$$

como la función de distribución acumulada.

En la figura 2.3, se puede observar el gráfico de las fda para diferentes valores de d .

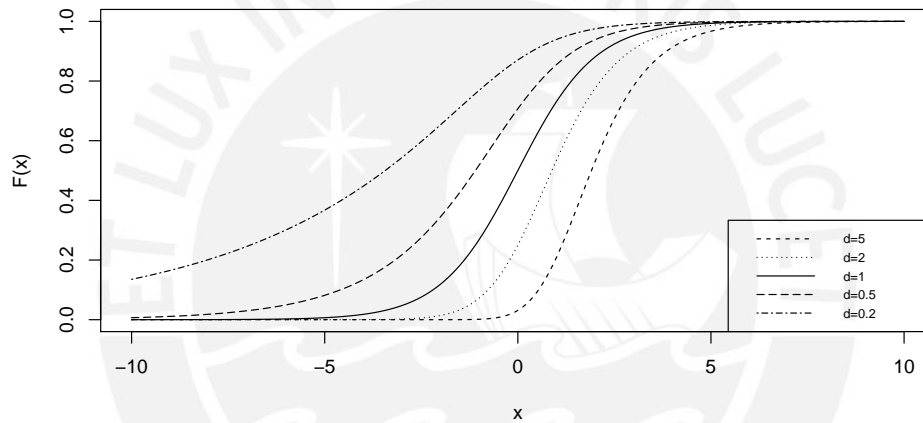


Figura 2.3: Función de distribución acumulada de la logística generalizada para diferentes valores de d

2.2.2. Función de densidad de probabilidad

Derivando 2.12 tenemos:

$$f(x) = d [L(x)]^{d-1} L'(x) \quad (2.13)$$

Reemplazando 2.11 en 2.13,

$$f(x) = d \left[\frac{1}{1 + e^{-x}} \right]^{d-1} \frac{e^{-x}}{(1 + e^{-x})^2} = \frac{de^{-x}}{(1 + e^{-x})^{d+1}} \quad (2.14)$$

Luego,

$$f(x) = \frac{de^{-x}}{(1 + e^{-x})^{d+1}} \quad -\infty < x < \infty, \quad d > 0 \quad (2.15)$$

es definido como la función de densidad de probabilidad.

Derivamos 2.15 e igualamos a cero, para encontrar el punto máximo que es definido como la moda (Mo).

$$f'(x) = \frac{de^{-x}(de^{-x} - 1)}{(1 + e^{-x})^{d+2}} \quad (2.16)$$

Igualando a cero, tenemos:

$$de^{-x} - 1 = 0 \Rightarrow Mo = \ln(d) \quad (2.17)$$

En la figura 2.4 se muestra algunas funciones de densidad de probabilidad para diferentes valores de d .

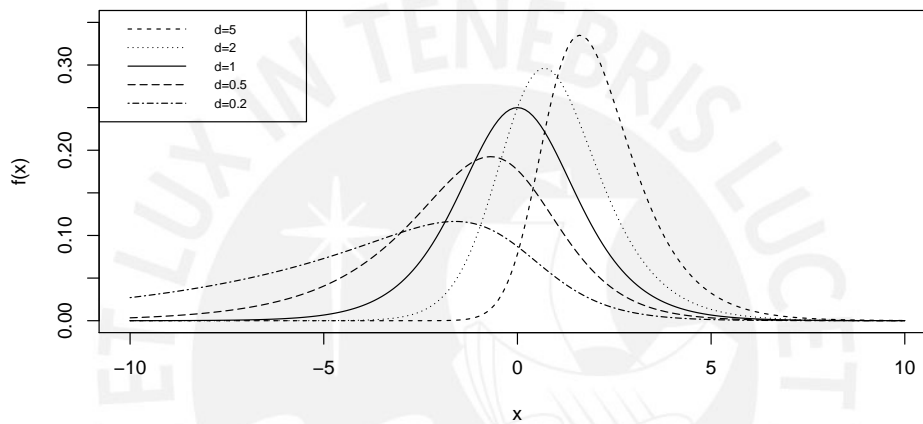


Figura 2.4: Densidades de la distribución logística generalizada para diferentes valores de d

2.2.3. Función generadora de momentos

Dado 2.15, la función generadora de momentos (fgm) esta dado por:

$$m(t) = E(e^{tx}) = dB(1 - t, d + t) \quad (2.18)$$

donde $B(.,.)$ es la función beta, luego la fgm es:

$$m(t) = d \frac{\Gamma(1 - t)\Gamma(d + t)}{\Gamma(1 + d)}, -d < t < 1 \quad (2.19)$$

$$m(t) = d \frac{\Gamma(1 - t)\Gamma(d + t)}{d\Gamma(d)}$$

$$m(t) = \frac{\Gamma(1 - t)\Gamma(d + t)}{\Gamma(d)}$$

De 2.19 podemos derivar:

Media:

$$E(x) = m'(0) = \psi(d) - \psi(1) \quad (2.20)$$

donde $\psi(\cdot)$ es la función psi ó digamma, definida por:

$$\psi(x) = \frac{d}{dx} \ln \Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)} \quad (2.21)$$

Varianza:

$$V(x) = E(x - E(x))^2 = \psi'(d) + \psi'(1) \quad (2.22)$$

Asimetría:

$$\gamma_1(x) = \frac{E(x - E(x))^3}{V(x)^{3/2}} = \frac{\psi''(d) - \psi''(1)}{(\psi'(d) + \psi'(1))^{3/2}} \quad (2.23)$$

Kurtosis:

$$\gamma_2(x) = \frac{E(x - E(x))^4}{V(x)^2} = \frac{\psi'''(d) + \psi'''(1)}{(\psi'(d) + \psi'(1))^2} \quad (2.24)$$

siendo $\psi'(\cdot)$, $\psi''(\cdot)$ y $\psi'''(\cdot)$ las derivadas sucesivas de la función digama definida en la ecuación (2.21).

En la figura 2.5, se muestra como cambia la asimetría para diferentes valores de d . Cuando $d < 0$ tenemos un coeficiente de asimetría negativo y cuando $d > 0$ obtenemos una asimetría positiva, notandose un mayor cambio en la asimetría para valores de $d < 0$. Podemos observar tambien que cuando $d = 1$, que corresponde a la distribución logística, el valor de la asimetría es cero.

En la figura 2.6 se muestra como cambia la kurtosis para diferentes valores de d . Note que cuando $d = 1$, la distribución logística, el valor de la kurtosis es igual a 1.2. La kurtosis representa un movimieto de masa que no es afectado por la varianza, sus componentes son el pico y las colas de la distribución (*peakedness* y *tailedness*) DeCarlo (1977). En consecuencia los valores positivos mayores 1.2 cuando $d < 1$ es debido a la presencia ua cola elevada al lado izquierdo de la distribución y cuando tenemos $d > 1$ es por la presencia de un pico más elevado en la distribución.

En ese sentido la distribución logística generalizada es una generalización de la distribución logística que puede ser asimétrica positiva o negativa y con kurtosis mayores que el valor de la distribución logística.

2.2.4. Función de distribución completa

Para una variable $y \in R$ no estandarizada, con media y varianza μ y σ^2 , respectivamente, su fda y fdp queda definida como en Saho (2002) por:

$$F(y; \mu, \sigma, d) = \left[\frac{1}{1 + e^{-(y-\mu)/\sigma}} \right]^d, \quad \sigma > 0, d > 0 \quad (2.25)$$

$$f(y; \mu, \sigma, d) = \frac{d}{\sigma} \frac{e^{-(y-\mu)/\sigma}}{[1 + e^{-(y-\mu)/\sigma}]^{d+1}}, \quad \sigma > 0, d > 0 \quad (2.26)$$

Si, estandarizamos la variable y , es decir calculamos $x = \frac{y-\mu}{\sigma}$, obtenemos la distribución logística generalizada estándar cuya fda y fdp esta dado por las ecuaciones 2.12 y 2.15 respectivamente. Note que μ y σ pueden ser estimados por máxima verosimilitud a partir de una muestra de datos para y , mayores detalles en (Saho, 2002).

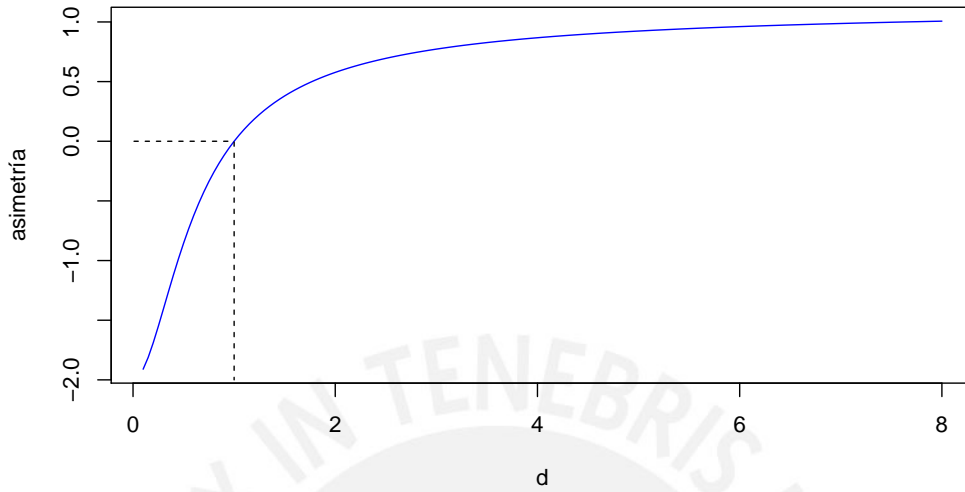


Figura 2.5: Coeficiente de asimetría de la distribución logística generalizada tipo I para d

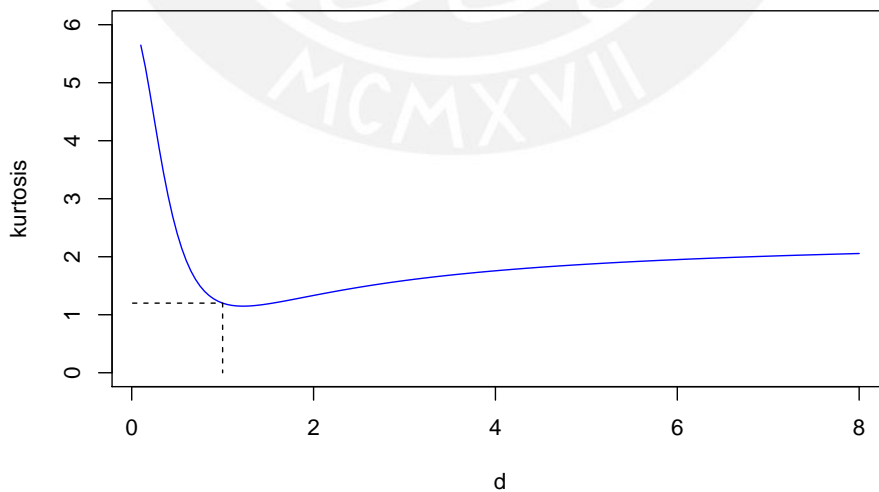


Figura 2.6: Kurtosis de la distribución logística generalizada tipo I para d

Capítulo 3

Teoría de Respuesta al Ítem y Testlet

Bock y Moustaki (2007) señalan que antes del desarrollo de la Teoría de Respuesta al Ítem (TRI) el tratamiento estadístico del logro educativo estuvo basado en lo que ahora se conoce como Teoría Clásica de los Test. Sin embargo los resultados de la teoría clásica son limitados en cuanto a su aplicación. Estos no son precisos en pruebas cortas e ignoran fuentes de error tales como la variación debido al efecto cuando las respuestas a los ítems deben ser juzgados subjetivamente. Otra gran limitación de la teoría clásica es su dependencia con respuestas dicotómicas.

La TRI supera esta limitación asumiendo una variable latente continua que representa la competencia o habilidad del sujeto en responder a los ítems de una prueba. Asume que la probabilidad de responder a un ítem con una, dos o más categorías mutuamente exclusivas esta en función de la localización del sujeto en el latente continuo y de ciertos parámetros característicos del ítem.

En este capítulo se desarrolla una definición general del modelo TRI para datos binarios o dicotómicos, luego se presenta de manera específica cada uno de los modelos a estudiar: El modelo logístico con enlace simétrico y el modelo logístico con enlace asimétrico, denominado modelo logístico de exponente positivo (LPE) que ha sido desarrollado por Samejima (1999), Bazan y Bolfarine (2010). Posteriormente agregaremos un parámetro adicional en la función de la variable latente, como lo realiza Bradlow, Wainer y Wang (1999), obteniendo con ello el modelo Testlet. Este nos permitirá evaluar el grado de dependencia condicional del ítem, utilizando un enlace simétrico y además utilizando también el enlace asimétrico LPE.

3.1. Definiciones generales

La teoría de respuesta al ítem (TRI) permite modelar de forma particular los modelos de rasgos latentes como lo mencionan Bazan et al. (2010); en ese sentido modela la relación entre respuestas multivariadas observables y variables latentes (tratamiento subyacente o habilidad no observada) de un conjunto de n individuos que han sido sometidos a una prueba con k ítems.

En la TRI, los modelos mas usados para datos dicotómicos son el modelo de ojiva normal (Lord, 1952) y tres formas del modelo logístico. Estas formas estan referidas al modelo logístico de un parámetro (Rash), dos parámetros y tres parámetros (Birnbbaum).

3.1.1. Definición general de un modelo TRI

Dado una matriz de datos binarios o dicotómicos Y_{ij} , que representan los scores o respuestas de n sujetos a k ítems de una prueba, se define el modelo TRI binario unidimensional como el sistema en el que para cada sujeto i se tiene un modelo de variable latente monótono unidimensional ¹ Bazan et al. (2009) (Y, Θ_i) , definido por:

$$Y_{ij} | \theta_i, \eta_j \sim \text{Bernoulli}(p_{ij}) \quad (3.1)$$

$$p_{ij} = P(Y_{ij} = 1 | \theta_i, \eta_j) = F(m_{ij}) \quad (3.2)$$

$$m_{ij} = a_j(\theta_i - b_j), \quad (3.3)$$

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

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 a la discriminación y dificultad del ítem respectivamente.
- θ_i es el valor de la variable latente Θ_i para el sujeto i .
- p_{ij} es la probabilidad condicionada de las respuestas a que $\Theta_i = \theta_i$
- F es la curva característica del ítem (CCI), definida por alguna función de distribución acumulada.
- m_{ij} es un predictor latente lineal relacionado con la variable latente y los parámetros del ítem.

3.1.2. Supuestos de un modelo TRI

Un modelo TRI, satisface las propiedades de un modelo de variable latente (Y, Θ) Holland y Rosenbaum (1986), a citar:

- **La independencia condicional latente**, Una condición básica en (Y, Θ) es la independencia condicional latente. La función de distribución condicional Y dado Θ es dada por

$$F(y_1, \dots, y_k | \theta) = P(Y_1 \leq y_1, \dots, Y_k \leq y_k | \Theta = \theta) \quad (3.4)$$

y la independencia condicional latente plantea que Y_1, \dots, Y_k son condicionalmente independientes dado Θ , o que:

$$F(y_1, \dots, y_k | \theta) = \prod_{j=1}^k F_j(y_j | \theta) \quad (3.5)$$

¹Cumple los supuestos de independencia condicional, monotonicidad latente y tiene unidimensionalidad latente

para todo y_1, \dots, y_k y θ , siendo $F_j(y_j|\theta)$ la función de distribución condicional de Y_j dado $\Theta = \theta$:

$$F_j(y_j|\theta) = P(Y_j \leq y_j | \Theta = \theta)$$

- **Monotonidad Latente**, un modelo de variable latente satisface la condición de monotonidad latente si las funciones:

$$1 - F_j(y|\theta) = P(Y_j > y | \Theta = \theta)$$

son funciones no decrecientes de θ para todos los valores de y y para $j = 1, \dots, k$. En caso de que Θ sea un vector, la monotonidad latente requiere que $1 - F_j(y|\theta)$ sea no decreciente en cada coordenada de θ .

La idea de la monotonidad latente es que valores grandes de θ implican distribuciones estocásticas grandes de Y_j , para cada j ; es decir, $Y_j|\Theta = \theta$ es estocásticamente mas grande que $Y_j|\Theta = \theta'$, si $\theta > \theta'$. Cuando Y_j es una variable dicotómica 0/1, la monotonidad latente es equivalente a los requerimientos de la curva característica del ítem.

- **Unidimensionalidad Latente**. Si, en un modelo de variable latente monótono, Θ es un escalar, entonces el modelo es denominado modelo de variable latente unidimensional monótono. Modelos unidimensionales son de interés especial porque son los modelos de variables latentes mas parsimoniosos, usualmente involucran pocos parámetros y tienen descripciones simples, permiten llevar formas de dependencia positiva, y un escalar θ permite una interpretación de una variable principal verdadera, que es medida a través de las respuestas observables en Y .

3.1.3. Características del modelo TRI

Siguiendo a [Bazan et al. \(2010\)](#), algunas características de los modelos TRI son:

- **Sobrep parametrización de parámetros**. Es una de las características mas resaltantes en el modelo TRI, Por ejemplo el modelo de dos parámetros tiene $n + 2 * k$ parámetros desconocidos. La cantidad de parámetros aumenta a mayor tamaño de muestra y mayor número de ítems.
- **Presencia de parámetros incidentales**. En la 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. Los parámetros estructurales no son necesariamente consistentes cuando son estimados de manera conjunta con parámetros incidentales.
- **Identificabilidad**. Un parámetro se dice identificado si distintos valores del parámetro implican diferentes distribuciones de probabilidad para los datos. Los parámetros del

modelo en TRI 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. Sin embargo, 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. 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.

3.1.4. Modelos Testlet

Según [Wainer et al. \(2007\)](#) los testlets han estado presentes, es decir fueron reconocidos, por un largo periodo de tiempo antes de 1987, aunque sin una denominación. Han sido utilizados para estimular pruebas de eficiencia en situaciones en que se examina la habilidad de un individuo para entender algún orden de estímulo; por ejemplo, un párrafo de lectura, una información gráfica, un pasaje musical o una tabla de números. Un testlet se refiere a un conjunto de ítems que están relacionados a un mismo estímulo. En tales situaciones, una cantidad importante del tiempo del evaluado es gastado en procesar el estímulo, y parece ser poco económico el esfuerzo de preguntar una sola vez sobre el estímulo presentado. Consecuentemente, estímulos largos fueron apareados con un conjunto de preguntas. Estas experiencias ayudaron a guiar el número de preguntas que fueron usadas para formar los testlet. Es fácil entender que si por ejemplo, nosotros hicieramos preguntas de párrafos compuestos por cerca de 250 palabras, encontraríamos gradualmente difícil escribir preguntas sobre algo nuevo.

Así, para el primer siglo, los testlets fueron usados simplemente como un método para elaborar pruebas que contengan un largo estímulo como una forma de eficiencia. Esto cambió en los 80 cuando técnica y económicamente fueron posibles aplicar pruebas adaptativas computarizadas (CAT). Una CAT es una prueba a medida para demostrar la habilidad del evaluado. El término Testlet fue definido explícitamente por [Wainer y Kiely \(1987\)](#) como:

Un grupo de ítems relacionados solo al contenido de un área, que es desarrollado como una unidad y contiene un número fijo de rutas predeterminadas que un evaluado puede seguir

quienes además propusieron un nuevo modelo con testlets para resolver los problemas que presentaba las CAT.

Para [Bradlow et al. \(1999\)](#), muchas pruebas estandarizadas comunes en educación tienen secciones que comprenden un grupo coherente de ítems de elección múltiple basados en un estímulo común. Estos grupos denominados testlets, son usados por muchas razones, principalmente por limitaciones de costo y tiempo.

Este tipo de estructuras de testlet no cumplen con el supuesto de independencia condicional de las respuestas a los ítems, en el sentido de que los ítems están relacionados. El efecto de testlets en la asociación de respuestas puede ser investigado por un análisis factorial [Bock y Moustaki \(2007\)](#). Si la asociación dentro de testlets es más grande que la atribuible a un factor general de comprensión de textos, entonces se encontrará un factor adicional específico de un grupo de ítems. La TRI basada en el análisis factorial provee pruebas estadísticas

para la existencia de estos factores "nuisance" o de incomodidad. El problema más serio de asumir independencia es la subestimación del error estándar de la escala del score. Esto ocurre porque el ignorar asociación incrementada dentro del testlet infla la varianza del score de test relativo a la medida de varianza del error.

Bradlow et al. (1999) desarrollaron el modelo testlet como una variante del modelo TRI para estimar este tipo de estructuras.

3.2. El modelo TRI Logístico

El modelo logístico fue introducido por Fisher (1938), luego Birnbaum (1957, 1958a, 1958b) obtuvo numerosos resultados para ordenar y clasificar sujetos reemplazando la distribución normal con la distribución logística, éste trabajo no fue muy difundido, hasta que Lord y Novick (1968) desarrollaron una versión actualizada con un análisis profundo del modelo logístico de uno, dos y tres parámetros.

3.2.1. Definición

Dado una matriz de datos binarios o dicotómicos Y_{ij} , que representan respuestas de n sujetos a k ítems de una prueba, donde $Y_{ij} = 1$ si la respuesta del i -ésimo individuo al j -ésimo ítem es correcta e $Y_{ij} = 0$ en caso contrario. Se define el modelo TRI logístico binario unidimensional como el sistema en el que para cada sujeto i se tiene un modelo de variable latente monótono unidimensional (Y, Θ_i) , definido por:

$$Y_{ij} | \theta_i, a_j, b_j \sim \text{Bernoulli}(p_{ij}) \quad (3.6)$$

$$p_{ij} = P(Y_{ij} = 1 | \theta_i, a_j, b_j) = \Psi(m_{ij}) \quad (3.7)$$

$$m_{ij} = a_j(\theta_i - b_j), \quad (3.8)$$

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

donde, $\Psi(\cdot)$ es la CCI, definida como la fda de la distribución logística; es decir:

$$\Psi(m_{ij}) = \frac{1}{1 + e^{-m_{ij}}} \quad -\infty < m_{ij} < +\infty \quad (3.9)$$

Aquí m_{ij} es el predictor lineal latente relacionado con la variable latente $\theta_i \in R$ y los parámetros del ítem, donde $a_j > 0$ es el parámetro de discriminación, $b_j \in R$ el parámetro de dificultad y p_{ij} es la probabilidad de que el i -ésimo sujeto responda correctamente el ítem j . El gráfico de la función de respuesta, ó curva característica del ítem (CCI), $\Psi(\theta)$ con respecto a θ es la curva ascendente de la fda de la distribución logística, con punto de inflexión en $\theta = b_j$.

En la figura 3.1 podemos observar la probabilidad de respuesta fijando $a = 1$ y diferentes valores de b y en la figura 3.2 la probabilidad de respuesta para diferentes valores de a y fijando $b = 0$.

En el gráfico de la probabilidad de responder correctamente a un ítem, CCI o función de respuesta, se puede observar que la curva es una imagen reflejada en el punto $(b_j, 0,5)$. Con

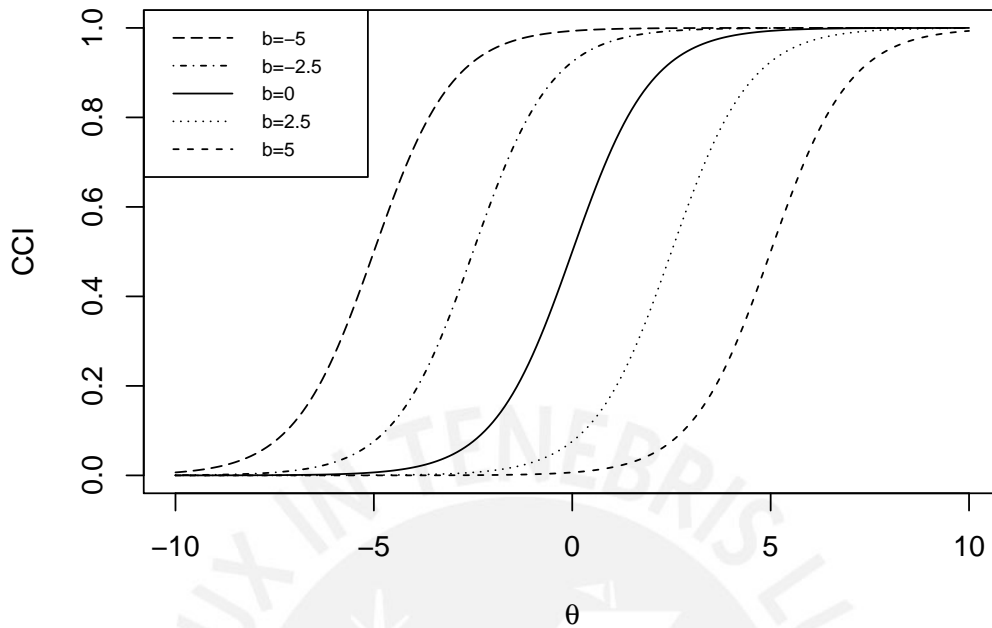


Figura 3.1: Probabilidad de responder correctamente a un ítem para $a = 1$ y diferentes valores de b

esto notamos que la CCI es simétrica. En este contexto, el parámetro a_j es denominado como el poder de discriminación del ítem y b_j la localización del ítem.

3.2.2. Información de Fisher

La información de Fisher es una medida de la cantidad de información que una variable aleatoria observable Y lleva sobre un parámetro desconocido θ que depende de la probabilidad de la variable observada.

La función de probabilidad de Y , que es también la función de verosimilitud de θ , es una función $f(Y, \theta)$; que es la densidad de probabilidad de la variable Y condicional al valor de θ . La derivada parcial con respecto a θ del logaritmo natural de la función de verosimilitud es denominado "score". Bajo ciertas condiciones de regularidad, se puede mostrar que el primer momento del score es cero. El segundo momento es denominado información de Fisher:

$$I(\theta) = E \left[\left(\frac{\partial}{\partial \theta} \ln f(Y; \theta) \right)^2 \middle| \theta \right] \quad (3.10)$$

donde, para un valor dado de θ , la expresión $E[\dots|\theta]$ denota la esperanza condicional de Y con respecto a la función de probabilidad $f(Y, \theta)$ dado θ . Note que $I(\theta) \geq 0$.

Como el valor esperado del "score" es cero, la función de Fisher es también la varianza del "score".

Para un sujeto con habilidad θ , $Y_j \sim \text{Bernoulli}(P_j)$ son las respuestas del sujeto, definimos

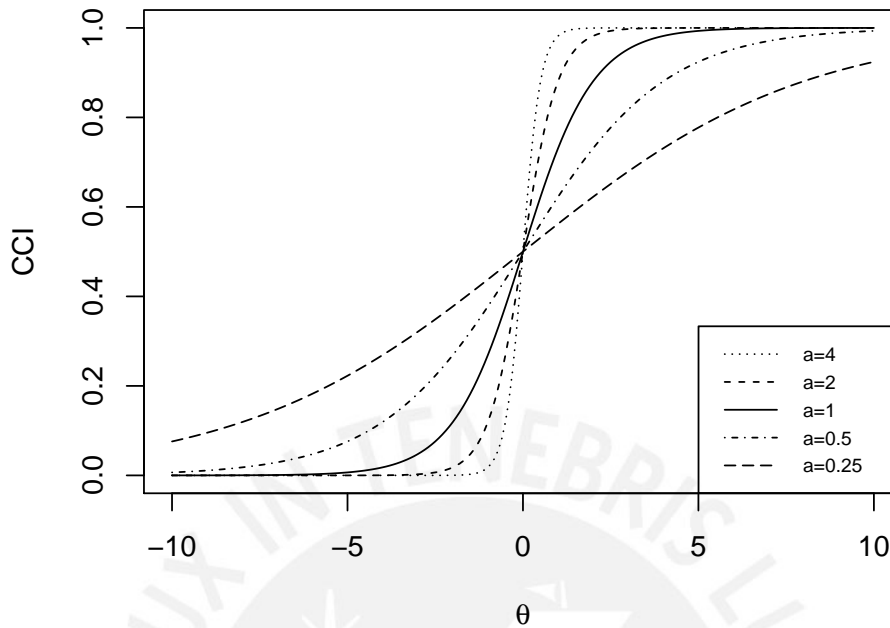


Figura 3.2: Probabilidad de responder correctamente a un ítem para $b = 0$ y diferentes valores de a

la función de verosimilitud de la variable latente θ como:

$$f(Y_j; \theta) = \prod_{j=1}^k p_j^{y_j} (1 - p_j)^{1-y_j} \quad (3.11)$$

donde p_j es la probabilidad de respuesta del sujeto al j -ésimo ítem, y k es la cantidad de ítems.

el logaritmo natural de la función verosimilitud esta dado por:

$$\ln f(Y_j; \theta) = \sum_{j=1}^k y_j \ln(p_j) + (1 - y_j) \ln(1 - p_j)$$

luego,

$$\frac{\partial}{\partial \theta} \ln f(Y_j; \theta) = \sum_{j=1}^k \frac{y_j}{p_j} \frac{\partial}{\partial \theta} p_j + \frac{(1 - y_j)}{1 - p_j} \frac{\partial}{\partial \theta} (1 - p_j)$$

reemplazando $\frac{\partial}{\partial \theta} p_j = p'_j(\theta)$

$$\frac{\partial}{\partial \theta} \ln f(Y_j; \theta) = \sum_{j=1}^k \frac{y_j}{p_j} p'_j(\theta) + \frac{(1 - y_j)}{1 - p_j} (-p'_j(\theta))$$

$$\begin{aligned}
 &= \sum_{j=1}^k p'_j(\theta) \left(\frac{y_j}{p_j} - \frac{(1-y_j)}{1-p_j} \right) \\
 &= \sum_{j=1}^k p'_j(\theta) \left(\frac{y_j(1-p_j) - p_j(1-y_j)}{p_j(1-p_j)} \right) \\
 &= \sum_{j=1}^k p'_j(\theta) \left(\frac{y_j - p_j}{p_j(1-p_j)} \right)
 \end{aligned}$$

finalmente,

$$\begin{aligned}
 I_j(\theta) &= E \left[\left(\sum_{j=1}^k p'_j(\theta) \frac{y_j - p_j}{p_j(1-p_j)} \right)^2 \right] \\
 &= \sum_{j=1}^k \left[\frac{p'_j(\theta)}{p_j(1-p_j)} \right]^2 E(y_j - p_j)^2 \\
 &= \sum_{j=1}^k \left[\frac{p'_j(\theta)}{p_j(1-p_j)} \right]^2 p_j(1-p_j)
 \end{aligned}$$

con lo cual la función de información del j -ésimo ítem [Lord y Novick \(1968\)](#), está dado por:

$$I_j(\theta) = \frac{p_j'^2(\theta)}{p_j(1-p_j)} \quad (3.12)$$

donde,

$$\begin{aligned}
 p'_j(\theta) &= \frac{\partial}{\partial \theta} p_j = \frac{\partial}{\partial \theta} \Psi(m_j) \\
 &= a_j p_j(1-p_j)
 \end{aligned}$$

La función de información de la prueba es la suma de la información de los ítems, dado por:

$$I(\theta) = \sum_j^k I_j(\theta) \quad (3.13)$$

Según [Chincaro \(2010\)](#), la estimación de la variable latente θ para cada sujeto, es asintóticamente insesgado cuando se tiene muchos ítems en la prueba, por lo que:

$$Var(\theta) = \frac{1}{I(\theta)} + s(k^{-1}) \quad (3.14)$$

donde, s es una función que depende de los k ítems y representa el sesgo del estimador. Si se considera k grande, se tiene:

$$Var(\theta) \approx \frac{1}{I(\theta)} \quad (3.15)$$

Con lo cual la función de información está inversamente relacionado con la varianza del estimador; es decir, a mayor función de información menor varianza del estimador y viceversa.

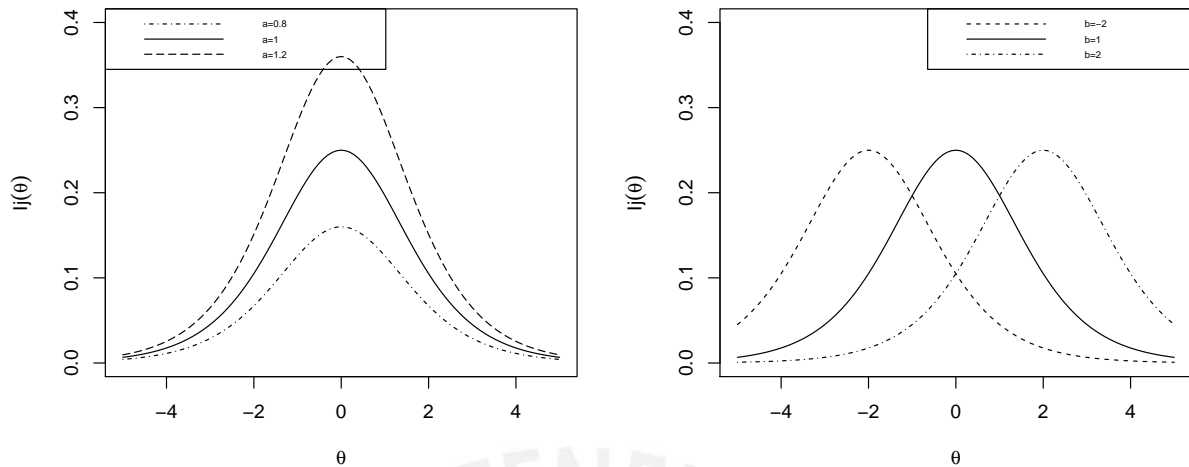


Figura 3.3: Función de información del ítem para diferentes valores de a (izquierda) y diferentes valores de b (derecha)

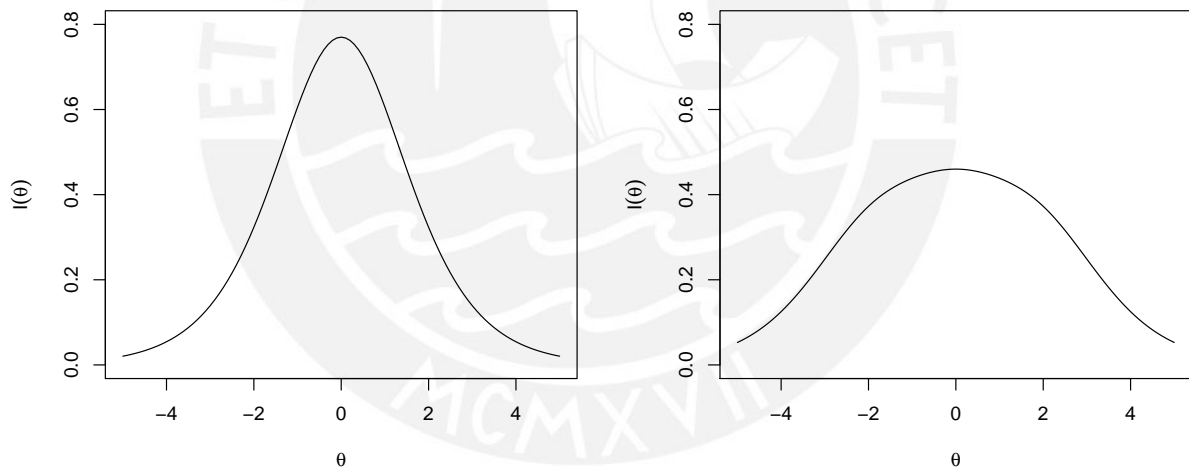


Figura 3.4: Función de información de test para los 3 ítems presentados en la figura 3.3

3.2.3. El modelo de un parámetro

Rasch (1960) introdujo una versión del modelo logístico en el cual el logit es definido como $m_j = \theta - b_j$. Este es un caso particular cuando $a_j = 1$ en la ecuación (3.8). Este modelo no produce discrepancias entre los rankings de la escala de los sujetos y el score de respuestas correctas. Esta característica es muchas veces requerida en programas de evaluación del rendimiento. Se debe resaltar que en aplicaciones que requieren solo un ordenamiento o ranking de los sujetos, este modelo no agrega información contenida en los scores del número de respuestas correctas.

En nuestro país [Chincaro \(2010\)](#) desarrolló el modelo Rash y su estimación desde el punto de vista clásico y bayesiano. Este modelo es ampliamente utilizado en el Ministerio de Educación para reportar las medidas de logro de los estudiantes y también es utilizado para reportar el score de los alumnos en el examen de admisión de la Pontificia Universidad Católica.

3.3. El modelo TRI Logístico de exponente positivo

Los modelos de la familia logística de exponente positivo para respuestas dicotómicas, ha sido propuesto por Samejima (1998). Samejima comenta que muchos modelos matemáticos en la teoría de respuesta al ítem unidimensional son representados por una CCI simétrica. Aunque esta familia de modelos ha sido propuesta, sus implicaciones y usos no son muy obvios para investigadores de medidas en educación.

Una característica de una curva CCI simétrica es que esta trata ambas las respuestas correctas e incorrectas simétricamente. Esto da una contradicción lógica en el ordenamiento de los examinados en el tratamiento latente o la escala de habilidad.

[Samejima \(1999\)](#) resalta por ejemplo la siguiente contradicción en el resultado del ordenamiento a través del estimador de máxima verosimilitud de θ para un modelo con CCI simétrico, el modelo de ojiva normal:

- Para cinco patrones de respuesta, cada uno de los cuales contiene solo una respuesta correcta, son ordenados en orden de dificultad del ítem que es respondido correctamente, y
- Para cinco patrones de respuesta, cada uno de los cuales contiene cuatro respuestas correctas son ordenados en orden de dificultad del ítem que no es respondido correctamente.

Estos dos principios son contradictorios el uno del otro. Si el primer principio es aceptado, luego debemos esperar que para los cinco patrones de respuesta que tienen cuatro respuestas correctas, el patrón de respuesta con los cuatro ítems más difíciles respondidos correctamente reciben la habilidad más alta. Sin embargo, si el segundo principio es aplicado, el patrón de respuesta que contesta correctamente el ítem más fácil recibe el valor más alto. Estas contradicciones son intrínsecas en todas las CCI simétricas, con excepción del modelo logístico, en el que la estimación de máxima verosimilitud no es afectada por el parámetro de dificultad. [Bazan y Bolfarine \(2010\)](#) señalan que en el contexto de regresión binaria, CCIs simétricas no siempre proveen buenos estimadores para algunos conjuntos de datos. En especial cuando la probabilidad de tener una respuesta binaria se aproxima a cero.

En ese sentido, se presenta la definición del modelo TRI logístico de exponente positivo cuya particularidad es que posee CCIs asimétricas.

3.3.1. Definición

Dado una matriz de datos binarios o dicotómicos Y_{ij} , que representan respuestas de n sujetos a k ítems de una prueba, donde $Y_{ij} = 1$ si la respuesta del i -ésimo individuo al j -ésimo ítem es correcta e $Y_{ij} = 0$ en caso contrario. Se define el modelo TRI logístico de exponente positivo, como una modificación de la CCI en el modelo TRI logístico (3.7), definido por:

$$Y_{ij}|\theta_i, a_j, b_j, d_j \sim \text{Bernoulli}(p_{ij})$$

$$p_{ij} = P(Y_{ij} = 1|\theta_i, a_j, b_j, d_j) = [\Psi(m_{ij})]^{d_j} \tag{3.16}$$

$$m_{ij} = a_j(\theta_i - b_j),$$

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

donde $d_j > 0$ es un nuevo parámetro de ítem, denominado parámetro de penalidad, p_{ij} es la probabilidad de que el i -ésimo sujeto responda correctamente el ítem j , m_{ij} es el predictor lineal latente relacionado con la variable latente $\theta_i \in R$ y los parámetros del ítem, siendo $a_j > 0$ el parámetro de discriminación y $b_j \in R$ el parámetro de dificultad.

Cuando $d_j = 1$, la CCI en el modelo LPE es la del modelo logístico. En este caso, la CCI es representada por una curva simétrica, tal como lo mencionamos en la sección (3.2).

La fortaleza del modelo LPE es que provee curvas asimétricas cuando $d_j \neq 1$ y permite el orden de individuos en la dimensión del tratamiento latente con una filosofía consistente. Una CCI asimétrica tiene su punto de inflexión en $(\theta_0 = b_j + \log(d_j)/a_j, [\Psi(a_j(\theta_0 - b_j))]^{d_j})$.

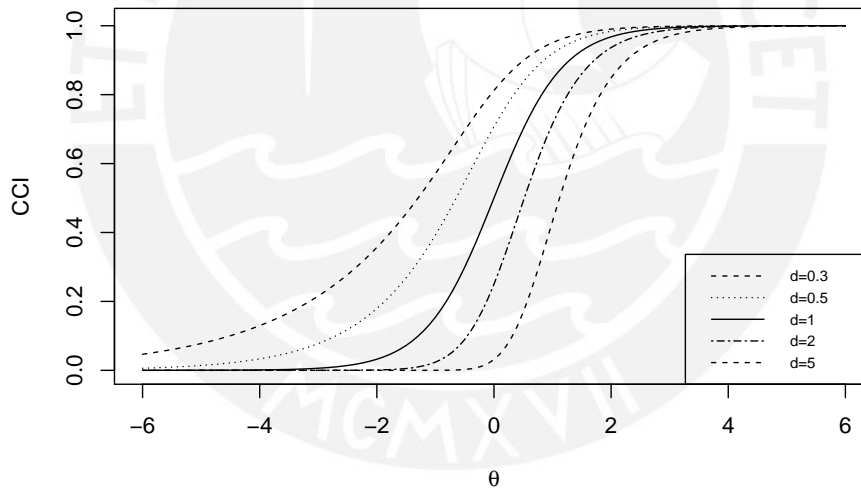


Figura 3.5: CCI en la familia de modelos logísticos de exponente positivo con parámetro de discriminación único $a_j = 1$ y con parámetro de dificultad $b_j = 0$ y diferente parámetro de penalidad $d = 0.3, 0.5, 1, 2$ y 5

En la figura 3.5 se grafica diferentes curvas características de ítems con un parámetro de discriminación $a_j = 1$, un mismo parámetro de dificultad $b_j = 0$ y diferentes valores para el parámetro d .

3.3.2. Función de Información

Para Samejima (1999) la función de información del ítem, esta dado por:

$$I_j(\theta) = \frac{[P_j'(\theta)]^2}{P_j(\theta)(1 - P_j(\theta))} \quad (3.17)$$

donde $P_j(\theta) = [\Psi(\theta)]^{d_j}$ y $P_j'(\theta) = \frac{\partial}{\partial \theta} \Psi_j(\theta)^{d_j}$.

$$\begin{aligned} \frac{\partial P_j(\theta)}{\partial \theta} &= \frac{\partial \Psi_j(\theta)^{d_j}}{\partial \theta} \quad (3.18) \\ &= d_j \left[\frac{1}{1 + e^{-m_j}} \right]^{d_j-1} (-a_j) \frac{e^{-m_j}}{(1 + e^{-m_j})^2} \\ &= -a_j d_j \left[\frac{1}{1 + e^{-m_j}} \right]^{d_j-1} \left(\frac{1}{1 + e^{-m_j}} \right) \left(\frac{e^{-m_j}}{1 + e^{-m_j}} \right) \\ &= -a_j d_j \left[\frac{1}{1 + e^{-m_j}} \right]^{d_j} \left(\frac{e^{-m_j}}{1 + e^{-m_j}} \right) \\ &= -a_j d_j \left[\frac{1}{1 + e^{-m_j}} \right]^{d_j} \left(1 - \frac{1}{1 + e^{-m_j}} \right) \\ &= -a_j d_j \Psi_j(\theta)^{d_j} (1 - \Psi_j(\theta)) \\ &= -a_j d_j P_j(\theta) (1 - \Psi_j(\theta)) \end{aligned}$$

Luego,

$$\begin{aligned} I_j(\theta) &= (a_j d_j)^2 \frac{P_j(\theta) [1 - \Psi_j(\theta)]^2}{(1 - P_j(\theta))} \\ I_j(\theta) &= (a_j d_j)^2 \frac{\Psi_j(\theta)^{d_j} [1 - \Psi_j(\theta)]^2}{(1 - \Psi_j(\theta)^{d_j})} \end{aligned}$$

En el gráfico 3.6 se muestra la función de información del ítem de la familia de modelos logísticos de exponente positivo con parámetro de discriminación único $a_j = 1$ y con parámetro de dificultad $b_j = 0$ y diferentes valores para el parámetro d .

Cuando el parámetro de penalización es mayor, $d > 1$, la probabilidad de acertar una respuesta dado la habilidad $P(\theta)$ es menor (ver 3.5) y la función de información del ítem tiene la curva mas alta (en 3.6), lo que significa que concentra mayor cantidad de información para valores cercanos de θ .

3.3.3. El modelo LPE de un parámetro

Al igual que en el modelo logístico, en el caso de la familia de modelos logísticos de exponente positivo, si $a_j = 1$, es decir, que no se considera la discriminación para los ítems, se tiene el modelo LPE de un parámetro cuyo score latente esta dado por: $m_j = \theta_i - b_j$.

3.4. El modelo Testlet Logístico

Considerando la modificación del predictor lineal latente m_{ij} en el modelo TRI de ojiva normal desarrollado por Bradlow et al. (1999), se realiza una modificación al predictor lineal latente m_{ij} en el modelo TRI Logístico definido en la sección (3.2) para definir el modelo Testlet Logístico.

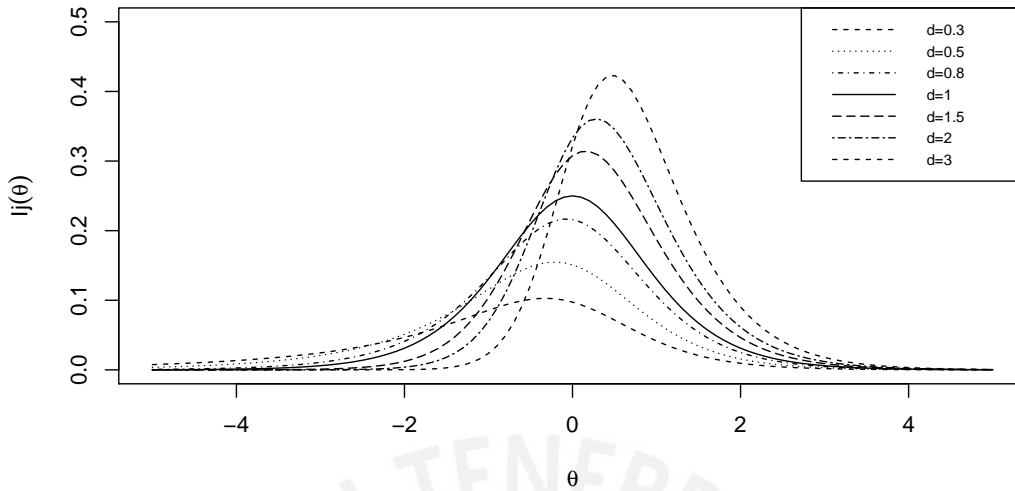


Figura 3.6: Función de información del ítem de la familia de modelos logísticos de exponente positivo con parámetro de discriminación único $a_j = 1$, parámetro de dificultad único $b_j = 0$ y siete valores para el parámetro de penalidad $d = 0.3, 0.5, 0.8, 1, 1.5, 2$ y 3

3.4.1. Definición

Dado una matriz de datos binarios o dicotómicos Y_{ij} , que representan respuestas de n sujetos a k ítems de una prueba, donde $Y_{ij} = 1$ si la respuesta del i -ésimo individuo al j -ésimo ítem es correcta e $Y_{ij} = 0$ en caso contrario. Se define el modelo Testlet logístico binario unidimensional como una modificación del predictor lineal latente m_{ij} en el modelo TRI logístico definido por la ecuación (3.8):

$$Y_{ij} | \theta_i, a_j, b_j, g_{ih(j)} \sim \text{Bernoulli}(p_{ij}) \quad (3.19)$$

$$p_{ij} = P(Y_{ij} = 1 | \theta_i, a_j, b_j, g_{ih(j)}) = \Psi(m_{ij}) \quad (3.20)$$

$$m_{ij} = a_j(\theta_i - b_j - g_{ih(j)}) \quad (3.21)$$

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

donde $g_{ih(j)}$ es el efecto testlet específico que corresponde a la i -ésima persona y el j -ésimo ítem anidado en el testlet $h(j)$, definido como:

$$g_{ih(j)} \sim N(0, \sigma_{g_{ih(j)}}^2) \quad (3.22)$$

éste efecto es independiente de la habilidad y los parámetros de ítem, $h(j)$ es un vector que nos indica la estructura de los testlets, donde:

- El número de testlets es $t = \text{máx}(h(j)) - 1$.
- Los testlets pueden tomar un sub-índice nuevo $l = h(j) = 1, \dots, t$.

- Un testlet esta formado por varios ítems con lo cual $l = h(j)$ identifica los ítems j que conforman el testlet.
- El tamaño de un testlet l es $n_{l=h(j)}$, j identifica los ítems que pertenecen al testlet.
- Los ítems que no estan asociados en un grupo testlet toman el valor de $t + 1$ y son considerados como un testlet de tamaño $n_{h(j)} = 1$

Con lo cual $h(j)$ representa el testlet al que pertenece cada ítem j , el tamaño de cada uno de los testlets y la cantidad de testlets. Por ejemplo:

$$h(j) = (1, 1, 1, 1, 2, 2, 2, 3, 3, 3, 5, 4, 4, 4, 4)$$

nos muestra una estructura de testlets que nos indica lo siguiente:

- el número total de ítems es $k = 15$ ($j = 1, 2, \dots, k$).
- el número de testlets $t = 4$ ($t = \text{máx}(h(j)) - 1$).
- el primer testlet es de tamaño 4, $n_{h(j)} = 4$, y esta formado por los ítems $j = 1, 2, 3, 4$; lo cual significa que $h(1) = h(2) = h(3) = h(4) = 1$ representa el efecto del primer testlet.
- el segundo testlet es de tamaño $n_{h(j)} = 3$ conformado por los ítems $j = 5, 6, 7$ y $h(5) = h(6) = h(7) = 2$ representa el efecto del segundo testlet.
- el tercer testlet es de tamaño $n_{h(j)} = 3$ conformado por los ítems $j = 8, 9, 10$ y $h(8) = h(9) = h(10) = 3$ representa el efecto del tercer testlet.
- el cuarto testlet es de tamaño $n_{h(j)} = 4$ conformado por los ítems $j = 12, 13, 14, 15$ y $h(12) = h(13) = h(14) = h(15) = 4$ representa el efecto del cuarto testlet.
- el ítem $j = 11$ no esta asociado con otro ítem por lo que éste ítem forma el quinto testlet de tamaño $n_{h(11)} = 1$.

Si el tamaño de un testlet es igual a 1 ($n_{h(j)} = 1$), por definición $g_{ih(j)} = 0$ para todos los sujetos.

El efecto testlet g esta centrado en cero para hacer énfasis en su estado de desviación en el score latente m_{ij} y para identificar el modelo; es decir,

$$\sum_{i=1}^n g_{ih(j)} = 0$$

Note que la CCI en este modelo es simétrica.

m_{ij} es el predictor lineal latente relacionado con la variable latente θ_i y los parámetros del ítem, donde $a_j > 0$ es el parámetro de discriminación y $b_j \in R$ el parámetro de dificultad. p_{ij} es la probabilidad de que el i -ésimo sujeto responda correctamente el ítem j .

Como caso particular cuando no existe el efecto testlet, es decir $g_{h(j)} = 0$ para todo $h(j)$, se tiene el modelo TRI logístico.

3.4.2. Función de Información

La presencia del efecto testlet afecta la función de información del testlet y la información de la prueba $I(\theta)$ así lo señalan [Bradlow et al. \(2002\)](#), por lo que se presenta la definición de la función de información del ítem y luego la función de información de la prueba.

La función de información del ítem, respecto a la variable latente θ , está dado por una modificación del predictor lineal latente m_{ij} en la ecuación (3.12), de la siguiente forma:

$$I_j(\theta) = a_j^2 P_j(\theta)(1 - P_j(\theta)) \quad (3.23)$$

donde $P_j(\theta) = \Psi(m_j)$ y $m_j = a_j(\theta - b_j - g_{h(j)})$. Por otro lado la función de información de la prueba definida en la ecuación (3.13) está dado por:

$$I(\theta) = \sum_j^k I_j(\theta)$$

3.4.3. El modelo Testlet logístico de un parámetro

Como un caso particular cuando $a_j = 1$ en la ecuación (3.26); es decir, no se considera el parámetro de discriminación, obtenemos el modelo Testlet logístico de un parámetro o conocido también como el modelo Rash testlet para datos dicotómicos [Wang y Wilson \(2005\)](#).

3.5. El modelo Testlet Logístico de Exponente Positivo

El modelo más general de todos los presentados hasta el momento es el modelo Testlet Logístico de Exponente Positivo (2LPET), que considera una modificación al predictor lineal latente m_{ij} en el modelo TRI Logístico de exponente positivo definido en (3.3).

3.5.1. Definición

Dado una matriz de datos binarios o dicotómicos Y_{ij} , que representan respuestas de n sujetos a k ítems de una prueba, donde $Y_{ij} = 1$ si la respuesta del i -ésimo individuo al j -ésimo ítem es correcta e $Y_{ij} = 0$ en caso contrario. Definimos el modelo Testlet logístico de exponente positivo, como una modificación del predictor lineal latente m_{ij} en el modelo TRI logístico de exponente positivo definido en (3.16)

$$Y_{ij} | \theta_i, a_j, b_j, d_j, g_{ih(j)} \sim \text{Bernoulli}(p_{ij})$$

$$p_{ij} = P(Y_{ij} = 1 | \theta_i, a_j, b_j, d_j, g_{ih(j)}) = [\Psi(m_{ij})]^{d_j} \quad (3.24)$$

$$m_{ij} = a_j(\theta_i - b_j - g_{ih(j)}), \quad (3.25)$$

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

donde $g_{ih(j)}$ recoge el efecto testlet específico que corresponde a la i -ésima persona y el j -ésimo ítem anidado en el testlet $h(j)$, definido como:

$$g_{ih(j)} \sim N(0, \sigma_{g_{ih(j)}}^2) \quad (3.26)$$

éste efecto es independiente de la habilidad y los parámetros de ítem, $h(j)$ es un vector que nos indica la estructura de los testlets, donde:

- El número de testlets es $t = \max(h(j))$.
- Los testlets pueden tomar un sub-índice nuevo $l = h(j) = 1, \dots, t$.
- Un testlet está formado por varios ítems con lo cual $l = h(j)$ identifica los ítems j que conforman el testlet.
- El tamaño de un testlet l es $n_{l=h(j)}$, y j identifica los ítems que pertenecen al testlet.
- Los ítems que no están asociados en un grupo testlet toman el valor de $t + 1$ y son considerados como un testlet de tamaño $n_{h(j)} = 1$

Si el tamaño de un testlet es igual a 1 ($n_{l=h(j)} = 1$), por definición $g_{ih(j)} = 0$ para todos los sujetos.

El efecto testlet g está centrado en cero para hacer énfasis en su estado de desviación en el score latente m_{ij} y para identificar el modelo; es decir,

$$\sum_{i=1}^n g_{ih(j)} = 0$$

p_{ij} es la probabilidad de que el i -ésimo sujeto responda correctamente el ítem j .

m_{ij} es el predictor lineal latente relacionado con la variable latente θ_i y los parámetros del ítem, donde $a_j > 0$ es el parámetro de discriminación y $b_j \in R$ el parámetro de dificultad.

$d_j > 0$ es el parámetro de ítem, denominado parámetro de penalidad que genera CCIs asimétricas.

3.5.2. Función de Información

La presencia del efecto testlet afecta la función de información del testlet $I_l(\theta)$ y la información de la prueba $I(\theta)$ así lo señalan [Bradlow et al. \(2002\)](#), por lo que se presenta la definición de la función de información del ítem y luego la función de información de la prueba.

La función de información del ítem, está dado por la ecuación (3.17):

$$I_j(\theta) = \frac{[P'_j(\theta)]^2}{P_j(\theta)(1 - P_j(\theta))}$$

donde $P'_j(\theta)$ es la derivada de $P_j(\theta) = \Psi_j(\theta)^{d_j}$ con respecto a θ y está dado por:

$$P'_j(\theta) = -a_j d_j P_j(\theta)(1 - \Psi_j(\theta))$$

luego,

$$I_j(\theta) = (a_j d_j)^2 \frac{P_j(\theta)[1 - \Psi_j(\theta)]^2}{(1 - P_j(\theta))}$$

$$\Psi_j(\theta) = \frac{1}{1 + e^{-m_j}}$$

Definimos la función de información de testlet como:

$$I_l(\theta) = \sum_{j=1}^{n_l} I_j(\theta) \quad (3.27)$$

La función de información de la prueba esta dado por 3.12.

$$I(\theta) = \sum_j^k I_j(\theta)$$

3.5.3. Casos particulares del Modelo Testlet logístico de exponente positivo

Como casos particulares del modelo 2LPET, se tiene: Cuando fijamos el valor de $a_j = 1$ en 3.25, se obtiene el modelo Testlet logístico de de exponente positivo de un parámetro (1LPET).

Si no se tiene efecto testlet; es decir, $n_{h(j)} = 1$ y $g_{ih(j)} = 0$ para todos los testlets, entonces se tiene como caso particular el modelo logístico de exponente positivo (2LPE).

Cuando el valor del parámetro de penalidad $d_j = 1$, se tiene el modelo Testlet logístico (2LT).

Si $g_{ih(j)} = 0$ para todos los testlets, y $d_j = 1$ entonces se tiene como caso particular el modelo TRI logístico (2L).

Capítulo 4

Estimación Bayesiana del Modelo Testlet LPE

Para la estimación de los parámetros se utiliza el análisis bayesiano debido a que varios autores consideran que es el método más adecuado en la estimación de parámetros en los modelos TRI. [Chincaro \(2010\)](#) muestra por ejemplo que el análisis bayesiano es el mejor método para estimar los parámetros en el modelo Rash.

Según [Bazan et al. \(2010\)](#) desde el punto de vista bayesiano el problema de sobreparametrización en los modelos TRI implica un tiempo computacional generalmente largo, pero la estimación de los parámetros de manera conjunta resuelve el problema de la distinción entre parámetros incidentales y estructurales, además resuelve también el problema de identificabilidad considerando la especificación de distribuciones a priori.

De acuerdo a [Fox \(2010\)](#) el modelamiento desde la perspectiva bayesiana permite realizar extensiones naturales en los modelos de respuesta al ítem y permite también una descripción inicial de los parámetros mediante la especificación de la priori, en niveles separados, para calcular diferentes fuentes de incertidumbre, dependencias complejas y otras fuentes de información. Estas flexibilidades en la definición de prioris para modelos de respuesta al ítem es una de las fortalezas del modelamiento bayesiano que hace posible manejar por ejemplo diseños de muestras complejas que recogen estructuras de dependencias complejas.

En este capítulo se presenta una definición general de la estimación bayesiana, vemos algunas características de las prioris usadas en los modelos TRI, luego se presenta la inferencia a partir del modelo más general, el modelo 2LPET, definido en el capítulo anterior (sec:testletlpe) y también se describen los criterios usados para comparar los modelos.

4.1. Definiciones Generales

Para definir una estimación mediante inferencia bayesiana, consideremos el vector aleatorio $y|\theta$ donde θ es un parámetro que caracteriza a la distribución de y . En la inferencia bayesiana:

- θ es un vector aleatorio no determinístico y por lo tanto tiene una distribución de probabilidad asociada.
- Es posible usar información preliminar acerca de θ , la cual se puede sintetizar proponiendo una distribución a priori para θ .
- Los datos se organizan en la función de verosimilitud $L(y|\theta)$.

- Usando el teorema de Bayes es posible obtener la distribución a posteriori de θ dados los datos:

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

La distribución a posteriori es proporcional a la verosimilitud y a la priori; es decir:

$$f(\theta|y) \propto L(y|\theta)f(\theta), \quad (4.2)$$

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

Observación:

Tomando logaritmo natural en 4.2 tenemos:

$$\ln f(\theta|y) \approx l(y|\theta) + \ln f(\theta). \quad (4.3)$$

Por tanto, si se considera una priori no informativa para θ , es decir, $f(\theta) = c$, donde c es una constante, maximizar $\ln f(\theta|y)$ para θ resulta equivalente a maximizar $l(y|\theta)$.

En otras palabras, la estimación de máxima verosimilitud (que maximiza $l(y|\theta)$) puede considerarse como un caso particular de inferencia Bayesiana sin información a priori.

Una vez identificada la distribución a posteriori para θ , $f(\theta|y)$, es posible realizar la inferencia y podemos obtener para esta distribución las medidas que se deseen, como por ejemplo: $E(\theta|y)$, $V(\theta|y)$, $Med(\theta|y)$, $Q_\lambda(\theta|y)$, que corresponden a la media, varianza, mediana y cuantil λ de la distribución a posteriori de θ dado $Y = y$. También podemos obtener un intervalo de credibilidad para $\theta|y$ donde por ejemplo con una probabilidad de 0.9, los límites del intervalo están definidos por el percentil 5 y 95.

4.2. Prioris en TRI

Usualmente la distribución a posteriori objetivo no es analíticamente tratable. En el pasado este problema se podía resolver usando una distribución apriori conjugada. Esta distribución apriori tiene la agradable propiedad de que el resultado de la distribución a posteriori es de la misma familia de distribución [Zoufras \(2009\)](#).

La especificación de la priori es un paso importante en el análisis bayesiano. Es más importante para muestras de tamaño pequeño donde la distribución a posteriori representan más de un compromiso entre la data observada y la opinion personal. Para muestras grandes, esto tiene menos importancia porque la data normalmente domina a la distribución a priori. [Bazan y Bolfarine \(2010\)](#)

No se conoce prioris conjugadas para los modelos TRI, En [Bazan et al. \(2010\)](#) se presentan varios teoremas en los que se tiene como resultado una distribución a posteriori impropia, por ejemplo:

- *En un modelo TRI de dos parámetro. Si la distribución a priori para (θ, a, b) es no informativa y proporcional a una constante, entonces la distribución a posteriori de (θ, a, b) es impropia.*

- En un modelo TRI de dos parámetros. Si la distribución a priori para al menos un (a_j) es impropia, entonces la distribución posterior de (θ, a, b) es impropia.
- En un modelo TRI de dos parámetros. Si la distribución a priori $f(\theta, a) \propto g(a)$, donde g es una función positiva arbitraria de a . Entonces la distribución posterior es siempre impropia.

En general, 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 prioris propias aseguran la identificabilidad del modelo y evitan el problema de presencia de parámetros incidentales.

Según [Bazan y Bolfarine \(2010\)](#) existe evidencia empírica que parece indicar presencia de correlación entre los parámetros de los ítems; sin embargo, es difícil asignar prioris dependientes para aquellos parámetros.

4.3. Función de Verosimilitud para el modelo 2LPET

Definimos a continuación la verosimilitud del modelo 2LPET, que es la probabilidad de encontrar el valor observado dado el modelo.

Sea la matriz de datos Y_{ij} asociada a los parámetros $\theta_i, a_j, b_j, d_j, g_{ih(j)}$ Definimos su función de verosimilitud L para un vector de datos observados y como:

$$L(y|\theta, a, b, d, g) = f(y|\theta, a, b, d, g) = \prod_{i=1}^n \prod_{j=1}^k [p_{ij}]^{y_{ij}} [1 - p_{ij}]^{1-y_{ij}} \quad (4.4)$$

$$= \prod_{i=1}^n \prod_{j=1}^k [\Psi_{d_j}(m_{ij})]^{y_{ij}} [1 - \Psi_{d_j}(m_{ij})]^{1-y_{ij}} \quad (4.5)$$

$$= \prod_{i=1}^n \prod_{j=1}^k \left[\frac{1}{(1 + e^{-a_j(\theta_i - b_j - g_{ih(j)})})d_j} \right]^{y_{ij}} \left[1 - \frac{1}{(1 + e^{-a_j(\theta_i - b_j - g_{ih(j)})})d_j} \right]^{1-y_{ij}} \quad (4.6)$$

4.4. Prioris en el modelo 2LPET

4.4.1. Especificación de la priori

Para el presente trabajo asumimos una estructura de prioris independientes y utilizamos la siguiente notación para referirnos a la especificación de las prioris.

$$\pi(\theta, a, b, d, g) = \prod_i^n \phi(\theta_i) \prod_j^n \pi_1(a_j) \pi_2(b_j) \pi_3(d_j) \pi_4(g_{ih(j)}) \quad (4.7)$$

donde $\phi(\theta_i)$, es la priori para la medida latente estimada θ_i definida como la distribución normal $N(0, 1)$.

Se definen prioris informativas porque la existencia de la distribución conjunta a posteriori no esta garantizada cuando una priori impropia es usada.

Específicamente se propone:

$$\begin{aligned}\theta_i &\sim N(0, 1) \\ a_j &\sim LN(\mu_a, \sigma_a^2) \\ b_j &\sim N(\mu_b, \sigma_b^2) \\ d_j &\sim G(m, n) \\ \theta_i &\sim N(0, 1) \\ g_{ih(j)} &\sim N(0, \sigma_{h(j)}^2)\end{aligned}\tag{4.8}$$

donde $LN(\mu, \sigma^2)$ denota la distribución log-normal con parámetros μ y σ^2 , $N(\mu, \sigma^2)$ denota la distribución normal con media μ y varianza σ^2 y $G(m, n)$ es la distribución Gamma con m y n como parametros de forma y escala respectivamente. Nótese que θ_i tiene como priori la distribución normal estandar y la varianza del efecto testlet $\sigma_{h(j)}^2$ es específica para cada testlet, permitiendo que la dependencia dentro de un testlet varíe a través de los testlets.

Los parámetros de las prioris son fijos, así en la priori de a_j ($\mu_a = 0$, $\sigma_a^2 = 1$), con lo cual se espera obtener $E(a_j) = e^{\mu_a + \sigma_a^2/2} = 1,649$ y $V(a_j) = (e^{\sigma_a^2} - 1)e^{(2\mu_a + \sigma_a^2)} = 4,671$.

En el caso de b_j : $\mu_b = 0$ y $\sigma_b^2 = 1$ y corresponden al esperado y varianza. En cuanto a la priori de d_j : $m = 0,25$ y $n = 4$, con lo que $E(d_j) = m * n = 1$ y $V(d_j) = m * n^2 = 4$. Para $g_{ih(j)}$ se considera incertidumbre en la varianza $\sigma_{h(j)}^2$ por lo que definimos una hiperpriori dada por:

$$\sigma_{h(j)}^2 \sim IG(a_{\sigma_g^2}, b_{\sigma_g^2})$$

donde $IG(a,b)$ es la distribución inversa gamma con parámetros $a_{\sigma_g^2} = 3$ y $b_{\sigma_g^2} = 2$, con lo que $E(\sigma_{h(j)}^2) = 0,5$ y $V(\sigma_{h(j)}^2) = 1$

4.5. Inferencia bayesiana

Considerando la funcion de verosimilitud dada en (4.6) y la priori definida en (4.7) definimos la distribución a posteriori de los parámetros dados los datos por:

$$f(\theta, a, b, d, g|y) \propto \pi(\theta, a, b, d, g) \times f(y|\theta, a, b, d, g)\tag{4.9}$$

$$\propto \pi(\theta, a, b, d, g) \times \prod_{i=1}^n \prod_{j=1}^k [\Psi_{d_j}(m_{ij})]^{y_{ij}} [1 - \Psi_{d_j}(m_{ij})]^{1-y_{ij}}\tag{4.10}$$

$$\propto \pi(\theta, a, b, d, g) \times \prod_{i=1}^n \prod_{j=1}^k \left[\frac{1}{(1 + e^{-a_j(\theta_i - b_j - g_{ih(j)})})d_j} \right]^{y_{ij}} \left[1 - \frac{1}{(1 + e^{-a_j(\theta_i - b_j - g_{ih(j)})})d_j} \right]^{1-y_{ij}}$$

Esta distribución no es identificable. Para la estimación bayesiana se consideran por tanto métodos de generación de distribuciones a partir de Cadenas de Markov de Monte Carlo (MCMC), específicamente Adaptive Rejection Sampling, que es implementado por defecto en el software WinBugs, para lo cual se escribe el código del modelo, teniendo la especificación de la verosimilitud jerárquica del modelo general:

$$y|\theta_i, a_j, b_j, d_j, g_{ih(j)} \sim \text{Bern}(F_{d_j}(a_j(\theta_i - b_j - g_{ih(j)}))) \quad (4.11)$$

$$\theta_i \sim N(0, 1) \quad (4.12)$$

$$a_j \sim LN(0, 1) \quad (4.13)$$

$$b_j \sim N(0, 1) \quad (4.14)$$

$$d_j \sim G(0, 25, 4) \quad (4.15)$$

$$g_{ih(j)} \sim N(0, \sigma_{h(j)}^2) \quad (4.16)$$

$$\sigma_{h(j)}^2 \sim IG(a_{\sigma_g^2}, b_{\sigma_g^2}) \quad (4.17)$$

4.6. Criterios de comparación de modelos

El principal criterio usado en este trabajo es el criterio de información de desvío (ó DIC, deviance information criterion) propuesto por Spiegelhalter (2002). El DIC es útil porque es fácil de calcular, interpretar y es satisfactorio para modelos jerárquicos complicados [Bazan y Bolfarine \(2006\)](#). Existen también otras medidas como el criterio de información bayesiana, el criterio de Akaike y el criterio de información bayesiana de Schwarz.

El desvío bayesiano es definido de manera general, usando la verosimilitud del modelo:

$$D(a, b, d, g, \theta) = -2\ln(f(y|a, b, d, g, \theta)) = -2 \sum_i^n \sum_j^k \ln P(Y_{ij} = y_{ij}|a, b, d, g, \theta) \quad (4.18)$$

Recordemos que y denota las respuestas observadas (0 ó 1) de los n sujetos para los k ítems, g son los parámetros de los ítems, θ las habilidades de los sujetos y g recoge el efecto testlet. También puede ser aproximada usando los resultados MCMC:

$$Dbar = \frac{1}{R} \sum_i^R D(a^r, b^r, d^r, \theta^r, g^r)$$

donde r representa la r -ésima realización del total de R realizaciones.

EAIC y EBIC pueden ser estimados mediante:

$$E\hat{AIC} = Dbar + 2p$$

$$E\hat{BIC} = Dbar + p\log(N)$$

$$D\hat{I}C = D\bar{bar} + \rho\hat{D}$$

donde p es el número de parámetros en el modelo, N es el número total de observaciones y ρD es el número efectivo de parámetros. [Bazan y Bolfarine \(2006\)](#).



Capítulo 5

Estudio de Simulación

En el presente capítulo, se desarrolla varios estudios de simulación, el primero esta en relación a la independencia condicional de los ítems, que es uno de los supuestos importantes en la teoría de respuesta al ítem. En este estudio se simulan datos con algunos ítems en grupos de testlets y se calculan los estimadores con un modelo Testlet logístico y TRI logístico, luego se comparan los estimadores. El segundo estudio introduce un componente asimétrico en la CCI, es decir, se simulan datos cuyas CCI son asimétricas y luego se calculan los estimadores con un modelo TRI logístico de exponente positivo y TRI logístico y se analizan los resultados. En el tercer estudio se simulan datos con algunas CCI asimétricas y algunos ítems en testlets y luego se calculan los estimadores con un modelo Testlet logístico de exponente positivo y TRI logístico de exponente positivo y se analizan los resultados.

5.1. Estudio 1

5.1.1. Objetivo

Simular un conjunto de datos binarios con grupos testlet y comparar los estimadores calculados a partir de un modelo Testlet logístico y de un modelo TRI logístico.

5.1.2. Algoritmo para simular los datos

Para simular una matriz de datos binarios con grupos de testlets, seguimos el siguiente procedimiento:

- Definir la cantidad de sujetos n , cantidad de ítems k y número de testlets t .
- Definir los ítems que conforman cada uno de los grupos testlet; es decir, definir $h(j)$.
- Generar los parámetros del modelo, el parámetro de discriminación para los k -ítems son generados mediante una distribución normal truncada a valores positivos, así: $a \sim N(1, 0,5)I(0, \infty)$, donde $I(0, \infty)$ especifica el rango de los valores generados; $b \sim N(0, 1)$, $\theta \sim N(0, 1)$ y el efecto testlet para cada sujeto como $g \sim N(0, \sigma_{h(j)}^2 = 0,5)$.
- Establecer restricciones de los parámetros: $\sum_i^n g_{ih(j)} = 0, \forall j$ (Bradlow et al., 1999), $\prod_j^k a_j = 1$ y $\sum_j^k b_j = 0$. (Bock y Aitkin, 1981) y (Spiegelhalter et al., 1996)
- Simular la probabilidad (p_{ij}) de que el sujeto i responda correctamente al ítem j a partir del modelo Testlet logístico.

- Simular las respuestas de cada sujeto i a cada pregunta j , para esto simular de una distribución uniforme $u_{ij} \sim U(0, 1)$ que nos dá una matriz de valores continuos entre 0 y 1. Luego comparar con la probabilidad generada anteriormente (p_{ij}) y si el valor generado es menor a la probabilidad ($u < p$) entonces y_{ij} tomará el valor 1, en caso contrario tomará valor 0.

El código para simular respuestas binarias con grupos testlet se muestra en el anexo B.5.

5.1.3. Método para estimar los parámetros

La estimación de los parámetros se realiza a través de inferencia bayesiana, mediante la estimación MCMC implementada en el software Openbugs. Se utilizará el esquema Adaptive Rejection Sampling (ARS) Gilks y Wild (1992) y se calculará la media estimada de la distribución a posteriori de los parámetros a estimar. Tenga en cuenta que se estiman los parámetros para cada uno de los ítems y los sujetos, por lo que se estimarán al menos $2k + n$ parámetros en el modelo 2L y $2k + n(k + 1)$ en el modelo 2LT.

5.1.4. Criterios para evaluar la simulación

Se compara los resultados a través de la raíz del error cuadrático medio (RMSE) y el promedio del error absoluto (MAE) Willmott y Matsuura (2005). Los estimadores de a, b son calculados para cada uno de los k ítems, en el caso de θ para cada uno de los n sujetos y en el caso de g para los n sujetos en los t testlets, siendo m es el número de simulaciones.

$$RMSE = \sqrt{\sum_{s=1}^m \frac{(\hat{\Theta}_s - \Theta)^2}{m}} \quad (5.1)$$

$$MAE = \sum_{s=1}^m \frac{|\hat{\Theta}_s - \Theta|}{m} \quad (5.2)$$

5.1.5. Escenario

Para la simulación se consideran $k = 20$ ítems, $t = 4$ testlets, con dos testlets de tamaño 3 y los otros dos testlets de tamaño 4, por lo que h queda definido como:

$$h = c(5, 1, 1, 1, 2, 2, 2, 3, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 5, 5)$$

ó usando letras:

$$h = c(E, A, A, A, B, B, B, C, C, C, C, D, D, D, D, E, E, E, E, E)$$

La cantidad de replicas es $m = 20$ con diferentes conjuntos de datos en cada réplica.

5.1.6. Resultados

En el cuadro (5.1) se muestran los resultados resumidos mediante el promedio de RMSE y MAE para los $k = 20$ ítems en los modelos analizados.

	Parámetro	RMSE	MAE
2LT	a	0.292	0.247
	b	0.265	0.225
	θ	0.468	0.387
	g	0.613	0.532
	σ_g^2	0.307	0.292
2L	a	0.308	0.267
	b	0.254	0.216
	θ	0.471	0.390

Cuadro 5.1: Resultados de medidas promedio de la simulación para comparar los modelos Testlet logístico (2LT) y TRI logístico (2L), con $m = 20$ réplicas

Los valores de las medidas RMSE y MAE para cada ítem se muestran en el cuadro (5.2).

Item	No.	t	2LT				2L			
			RMSE a	RMSE b	MAE a	MAE b	RMSE a	RMSE b	MAE a	MAE b
1	E		0.244	0.273	0.183	0.198	0.230	0.265	0.178	0.193
2	A		0.149	0.293	0.107	0.247	0.176	0.27	0.131	0.224
3	A		0.255	0.179	0.224	0.154	0.221	0.167	0.19	0.136
4	A		0.186	0.194	0.139	0.149	0.178	0.165	0.142	0.127
5	B		0.402	0.456	0.356	0.417	0.538	0.24	0.455	0.182
6	B		0.201	0.407	0.192	0.356	0.234	0.399	0.225	0.351
7	B		0.185	0.201	0.136	0.158	0.145	0.206	0.117	0.162
8	C		0.120	0.198	0.100	0.166	0.133	0.190	0.110	0.166
9	C		0.217	0.279	0.133	0.244	0.229	0.317	0.157	0.271
10	C		0.315	0.314	0.260	0.240	0.307	0.287	0.239	0.253
11	C		0.267	0.145	0.232	0.110	0.257	0.140	0.231	0.109
12	D		0.108	0.184	0.084	0.145	0.129	0.179	0.098	0.142
13	D		0.242	0.276	0.200	0.226	0.219	0.167	0.182	0.128
14	D		1.095	0.103	1.035	0.086	1.282	0.104	1.242	0.083
15	D		0.249	0.907	0.232	0.874	0.325	1.106	0.300	1.081
16	E		0.403	0.207	0.355	0.171	0.382	0.198	0.346	0.165
17	E		0.332	0.225	0.265	0.185	0.311	0.223	0.254	0.185
18	E		0.265	0.126	0.207	0.108	0.260	0.124	0.218	0.106
19	E		0.219	0.212	0.174	0.174	0.202	0.210	0.160	0.173
20	E		0.392	0.115	0.338	0.090	0.402	0.117	0.360	0.090

Cuadro 5.2: Raíz del error cuadrático medio y Media absoluta del error de los estimadores en la simulación - Estudio 1, considerando $m = 20$ réplicas

Según los resultados mostrados en el cuadro (5.3) en el estudio de simulación 1 el modelo Testlet logístico (2LT) ajusta mejor los datos, considerando el criterio DIC. Note que los criterios EAIC y EBIC presentan valores altos para el modelo 2LT porque cuentan con una cantidad mayor de parámetros, siendo de esta manera inconveniente para la comparación.

En las figuras 5.1, 5.2 y 5.3 se muestran las medidas de RMSE, MAE y el sesgo en la

Modelo	DIC	pD	Tiempo (seg.)	Dbar	EAIC	EBIC
2LT	4261	385	3232	3876	5956	7623
2L	4302	189	1382	4113	4593	4977

Cuadro 5.3: Resultados de medidas de ajuste de la simulación para comparar los modelos Testlet logístico (2LT) y TRI logístico (2L)

estimación de los parámetros de ítem de los modelos 2LT y 2L.

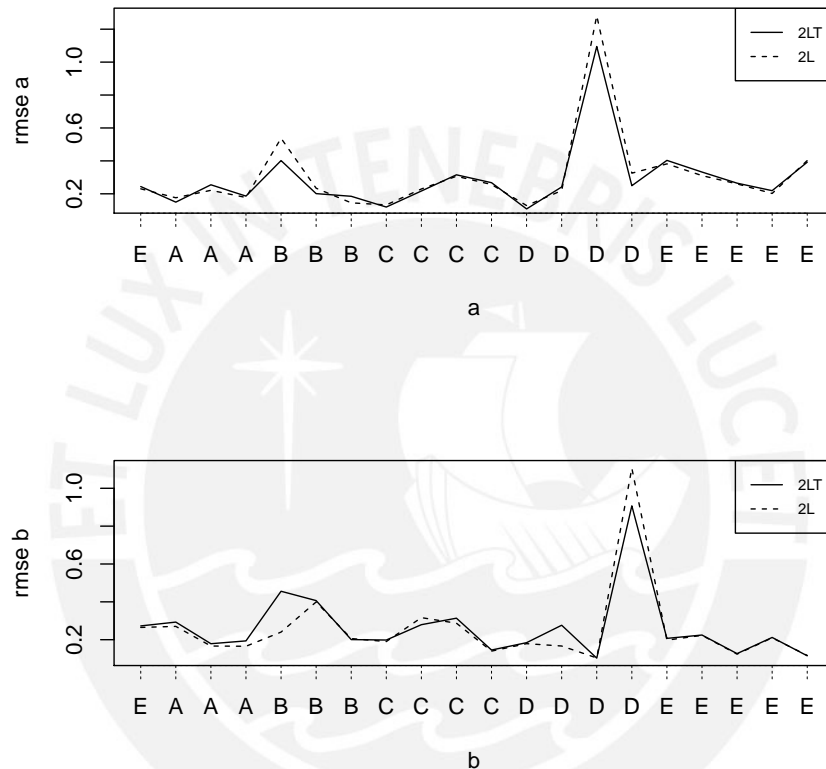


Figura 5.1: Raíz del error cuadrático medio (RMSE) de los estimadores simulados para la discriminación (a) y dificultad (b) de cada ítem para un ajuste Testlet logístico (2LT) y un ajuste TRI logístico (2L), considerando $m = 20$ réplicas

La figura 5.4 muestra la estimación de habilidades bajo ambos modelos, en comparación con los valores simulados. Note que las estimaciones bajo el modelo 2L se encuentran más alejados de los verdaderos valores. Finalmente el cuadro 5.4 muestra las correlaciones entre las estimaciones de los parámetros de ítem y de habilidades bajo los dos modelos con sus correspondientes valores verdaderos. Note que el modelo 2LT recupera mejor estos valores, especialmente para la estimación de la discriminación.

La conclusión del primer estudio de simulación es que:

- Cuando los datos están en grupos de testlets, el mejor modelo es el modelo Testlet Logístico (2LT) antes que el modelo TRI Logístico (2L), esto es debido a que el modelo

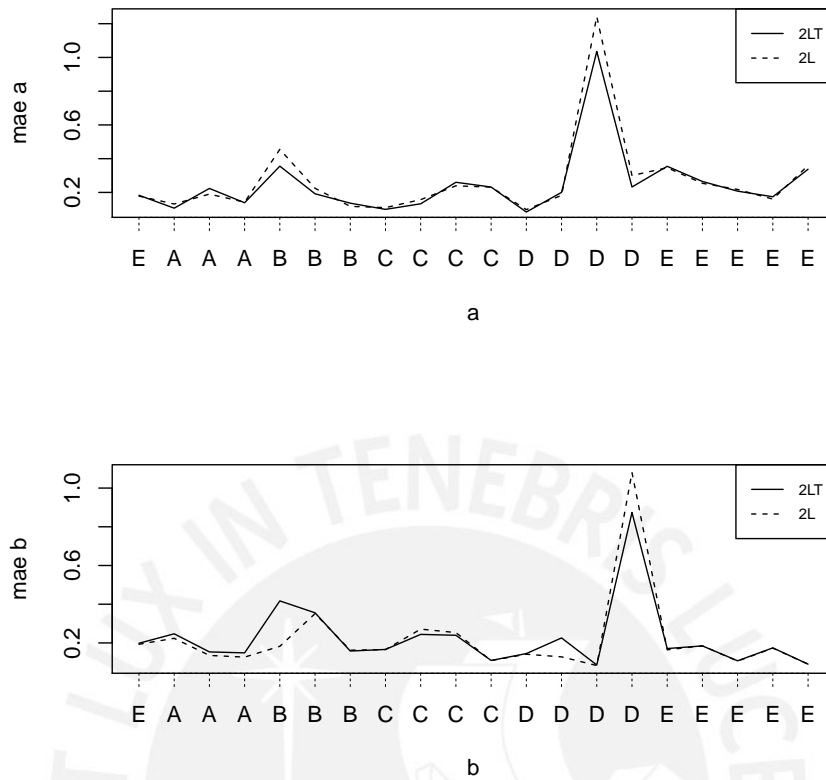
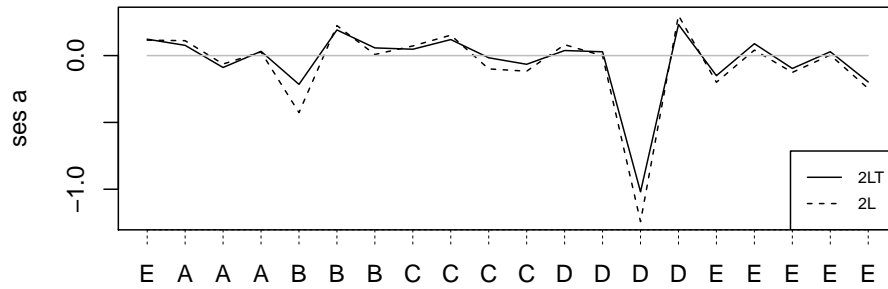


Figura 5.2: Media absoluta del error (MAE) de los estimadores simulados para la discriminación (a) y dificultad (b) de cada ítem para un ajuste Testlet logístico (2LT) y un ajuste TRI logístico (2L), considerando $m = 20$ réplicas

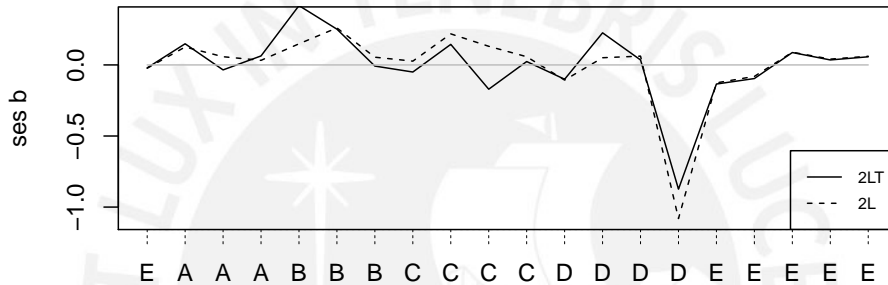
Correlación	2LT	2L
r_a	0.949	0.922
r_b	0.980	0.978
r_θ	0.977	0.970

Cuadro 5.4: Medidas de correlación entre los valores simulados y los estimados promedio, de los modelos Testlet logístico (2LT) y TRI logístico (2L), considerando $m = 20$ réplicas

presenta menor sesgo, menor error cuadrático medio y mayor correlación con los valores verdaderos.



a



b

Figura 5.3: Sesgo en la estimación de los parámetros a y b obtenidos por los modelos Testlet logístico (2LT) y TRI logístico (2L), considerando $m = 20$ réplicas

5.2. Estudio 2

5.2.1. Objetivo

Simular un conjunto de datos binarios con algunas CCI asimétricas, calcular los estimadores mediante un modelo TRI logístico de exponente positivo (2LPE) y el modelo TRI logístico (2L), y comparar los resultados obtenidos.

5.2.2. Algoritmo para simular los datos

Para simular una matriz de datos binarios con CCIs asimétricas, seguimos el siguiente procedimiento:

- Establecer la cantidad de sujetos n , cantidad de ítems k y los ítems j que tendrán una CCI asimétrica.
- Generar los parámetros del modelo: Los parámetros de discriminación (a) son generados por una normal truncada a valores positivos, así: $a \sim N(1, 0,5)I(0, \infty)$, donde $I(0, \infty)$ especifica el rango de los valores generados; $b \sim N(0, 1)$ y $\theta \sim N(0, 1)$.

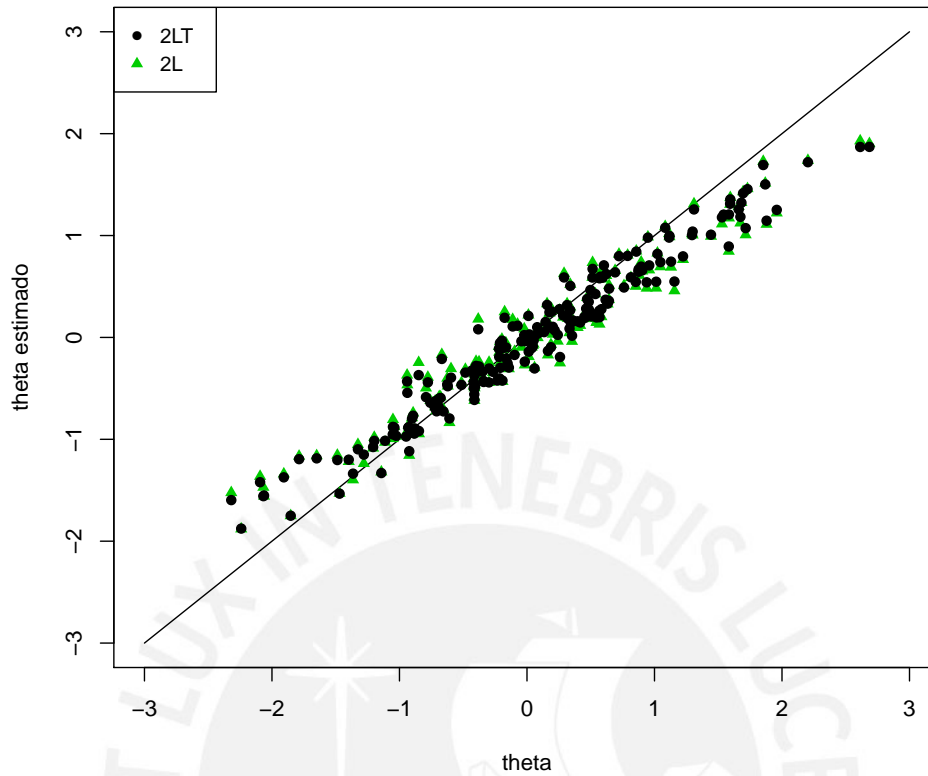


Figura 5.4: Estimación de la habilidad para un ajuste Testlet logístico (2LT) y un ajuste TRI logístico (2L)

- Calcular las restricciones de los parámetros: $\prod_j^k a_j = 1$ y $\sum_j^k b_j$ (Bock y Aitkin, 1981) y (Spiegelhalter et al., 1996).
- Definir los ítems cuyo parámetro de penalidad d_j generan una CCI asimétrica.
- Simular la probabilidad (p_{ij}) de que el sujeto i responda correctamente al ítem j a partir del modelo TRI logístico de exponente positivo.
- Simular las respuestas de cada sujeto i a cada pregunta j , para esto simular $u_{ij} \sim U(0, 1)$ que nos dá una matriz de valores entre 0 y 1. Luego comparar con la probabilidad generada anteriormente (p_{ij}) y si el valor generado es menor a la probabilidad la respuesta y_{ij} tomará el valor 1, en caso contrario tomará valor 0.

El código para simular respuestas binarias con CCI asimétricas se muestra en el anexo (B.6).

5.2.3. Método para estimar los parámetros

La estimación de los parámetros se realiza a través de inferencia bayesiana. Se utiliza la estimación MCMC implementada en el software Openbugs, bajo el esquema ARS Gilks y Wild (1992) y se calcula la media estimada de la distribución a posteriori de los parámetros

a estimar. Considere que se estima los parámetros para cada uno de los ítems y los sujetos, por lo que se estimarán al menos $3k + n$ parámetros.

5.2.4. Criterios para evaluar la simulación

Se compara los resultados a través de la raíz del error cuadrático medio (RMSE) y el promedio del error absoluto (MAE) Willmott y Matsuura (2005), el estimador de a, b, d es calculado para cada uno de los k ítems, en el caso de θ para cada uno de los n sujetos y m es el número de simulaciones.

$$RMSE = \sqrt{\sum_s^m \frac{(\hat{\Theta}_s - \Theta)^2}{m}} \quad (5.3)$$

$$MAE = \sum_s^m \frac{|\hat{\Theta}_s - \Theta|}{m} \quad (5.4)$$

5.2.5. Escenario

Para la simulación se consideran evaluar las respuestas de $n = 200$ sujetos, $k = 20$ ítems, de los cuales 13 ítems tienen CCIs asimétricas cuyo parámetro de penalización en 8 casos toma valores $d < 1$ y en 5 casos toma valores $d > 1$. Así d queda definido como:

$$d = c(0.1, 0.2, 0.3, 0.3, 0.3, 0.4, 0.7, 0.8, 1, 1, 1, 1, 1, 1, 1, 1, 1.1, 1.4, 1.5, 1.8, 3)$$

La cantidad de réplicas simuladas es $m = 20$ con diferentes conjuntos de datos en cada réplica.

5.2.6. Resultados

Los resultados del estudio de simulación 2 se presentan a continuación. En el cuadro 5.5 se muestran los resultados resumidos mediante el promedio de las medidas RMSE y MAE para los modelos 2LPE y 2L.

	Parámetro	RMSE	MAE
2LPE	a	0.311	0.269
	b	0.590	0.564
	d	0.421	0.397
	θ	0.468	0.388
2L	a	0.308	0.270
	b	0.920	0.896
	θ	0.476	0.398

Cuadro 5.5: Resultados de medidas promedio de la simulación para comparar el modelo TRI logístico de exponente positivo (2LPE) y TRI logístico (2L), considerando $m = 20$ réplicas

Las medidas de RMSE y MAE para cada ítem, de los modelos 2LPE y 2L se presentan en el cuadro 5.6.

En el cuadro 5.7 se muestran los resultados del estudio de simulación 2, en el cual el modelo 2LPE ajusta mejor los datos, considerando el criterio DIC. Nuevamente los criterios EAIC y EBIC presentan valores altos para el modelo 2LPE porque cuentan con una cantidad mayor de parámetros, siendo de esta manera inconveniente para la comparación.

Item		2LPE				2L			
No.	d	RMSE a	RMSE b	MAE a	MAE b	RMSE a	RMSE b	MAE a	MAE b
1	0.1	0.316	0.856	0.296	0.832	0.535	2.434	0.522	2.428
2	0.2	0.481	0.622	0.441	0.59	0.493	1.619	0.467	1.604
3	0.3	0.223	0.376	0.204	0.341	0.151	1.283	0.136	1.252
4	0.3	0.491	1.046	0.46	1.028	0.625	1.065	0.602	1.046
5	0.3	0.33	0.31	0.307	0.259	0.37	1.485	0.355	1.47
6	0.4	0.198	0.592	0.164	0.57	0.254	1.134	0.211	1.118
7	0.7	0.402	0.861	0.364	0.847	0.367	0.254	0.331	0.191
8	0.8	0.383	0.531	0.329	0.493	0.366	0.296	0.299	0.267
9	1.0	0.247	0.959	0.196	0.936	0.175	0.566	0.142	0.533
10	1.0	0.333	0.178	0.28	0.137	0.328	0.485	0.283	0.469
11	1.0	0.238	0.369	0.184	0.335	0.223	0.507	0.174	0.479
12	1.0	0.66	0.742	0.64	0.73	0.326	0.518	0.286	0.483
13	1.0	0.251	0.177	0.178	0.146	0.181	0.427	0.136	0.383
14	1.0	0.231	1.529	0.201	1.523	0.146	0.768	0.119	0.737
15	1.0	0.251	0.509	0.188	0.495	0.181	0.535	0.156	0.515
16	1.1	0.22	0.273	0.185	0.233	0.17	0.599	0.132	0.569
17	1.4	0.223	0.743	0.159	0.717	0.17	0.761	0.144	0.745
18	1.5	0.189	0.311	0.157	0.282	0.201	0.958	0.171	0.951
19	1.8	0.213	0.163	0.179	0.136	0.387	1.03	0.309	1.016
20	3.0	0.338	0.659	0.261	0.646	0.503	1.681	0.422	1.677

Cuadro 5.6: Raíz del error cuadrático medio y Media absoluta del error de los estimadores en la simulación - Estudio 2, considerando $m = 20$ réplicas

Modelo	DIC	pD	Tiempo (seg.)	Dbar	EAIC	EbIC
2LPE	4127	194	3171	3933	4453	4869
2LT	4133	193	1560	3940	4420	4805

Cuadro 5.7: Resultados de medidas de ajuste de la simulación para comparar los modelos TRI Logístico de Exponente Positivo (2LPE) y TRI Logístico (2L), considerando $m = 20$ réplicas

En las figuras 5.5 y 5.6 se muestran las medidas de RMSE y MAE para la estimación de los parámetros de ítem de los modelos 2LPE y 2L.

La figura 5.7 muestra la estimación de las habilidades, con ambos modelos, en comparación con los valores simulados, note que la estimación en el modelo 2L se encuentra más alejada de los verdaderos valores, especialmente para las habilidades más bajas .

Finalmente el cuadro 5.8 muestra las correlaciones entre las estimaciones de los parámetros de ítem y las habilidades, bajo ambos modelos, con sus correspondientes valores verdaderos. Note que cuando los datos presentan asimetría, el modelo 2LPE recupera mejor los valores, especialmete en el parámetro de dificultad b .

Correlación	2LPE	2L
r_a	0.770	0.744
r_b	0.616	0.537
r_θ	0.989	0.988

Cuadro 5.8: Medidas de correlación entre los valores verdaderos y los estimados promedio de los modelos TRI logístico de exponente positivo (2LPE) y TRI logístico (2L)

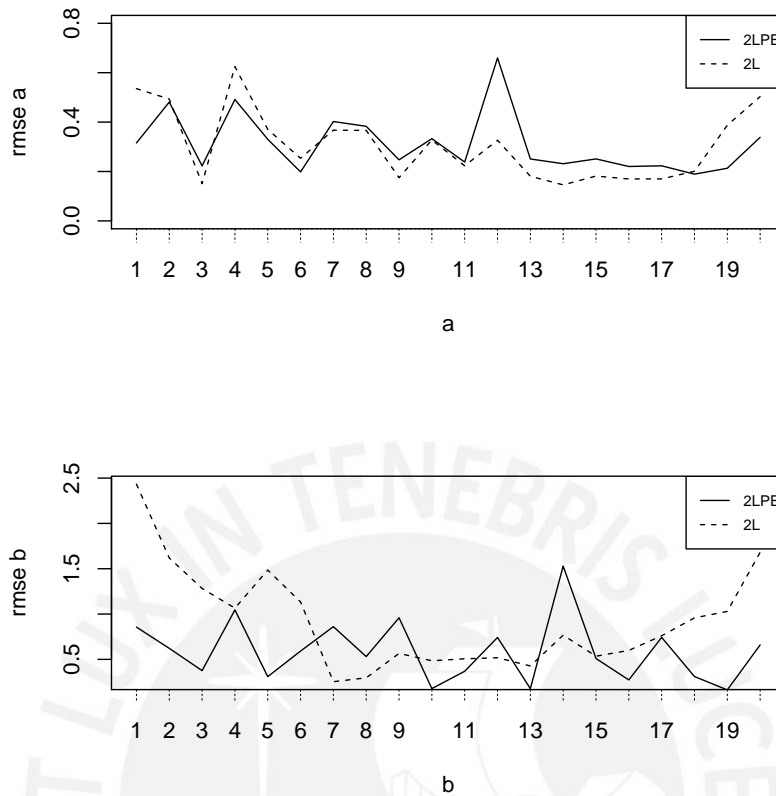
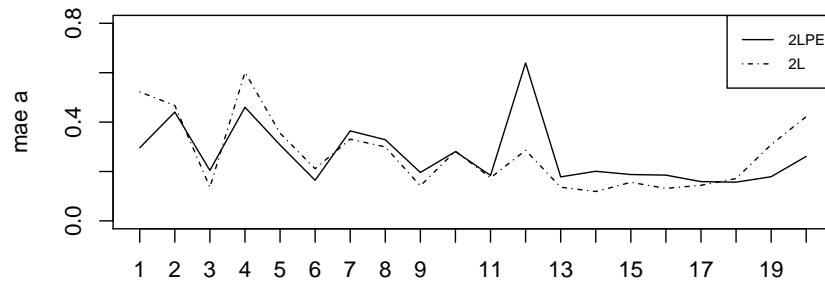


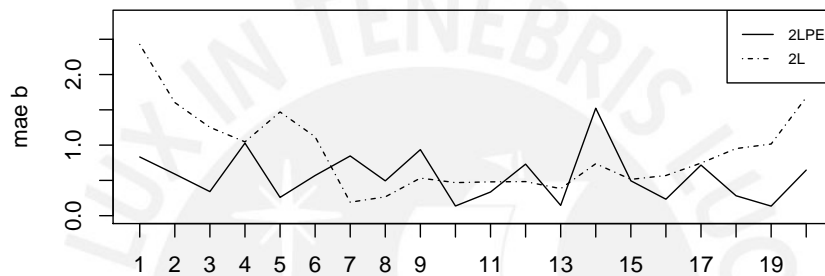
Figura 5.5: Raíz del Error Cuadrático Medio (RMSE) de los estimadores simulados para la discriminación (a) y dificultad (b) de cada ítem con un modelo TRI logístico de exponente positivo (2LPE) y TRI logístico (2L), considerando $m = 20$ réplicas

Las conclusiones del estudio 2 son:

- Cuando los datos presentan asimetría, el mejor modelo es el modelo 2LPE, esto es verdad especialmente en los ítems asimétricos, como por ejemplo el ítem 20 ó el ítem 1.
- El principal efecto de ignorar la asimetría se produce en la estimación de a y b .



a



b

Figura 5.6: Media Absoluta del Error de los estimadores simulados para la discriminación (a) y dificultad (b) de cada ítem con un modelo TRI logístico de exponente positivo (2LPE) y TRI logístico (2L), considerando $m = 20$ réplicas

5.3. Estudio 3

5.3.1. Objetivo

Simular un conjunto de datos binarios, con CCIs asimétricas y grupos de testlet, y comparar los estimadores calculados mediante un modelo Testlet Logístico de Exponente Positivo (2LPET) con los estimadores de un modelo TRI Logístico de Exponente Positivo (2LPE).

5.3.2. Algoritmo para simular los datos

Para simular una matriz de datos binarios con grupos de testlets y con CCIs asimétricas, seguimos el siguiente procedimiento:

- Establecer la cantidad de sujetos n , cantidad de ítems k , los ítems j que tendrán una CCI asimétrica y el número de testlets t .
- Definir los ítems que conforman cada uno de los grupos testlet, es decir definir $h(j)$, si un ítem j no está en ningún grupo testlet $h(j) = t + 1$.

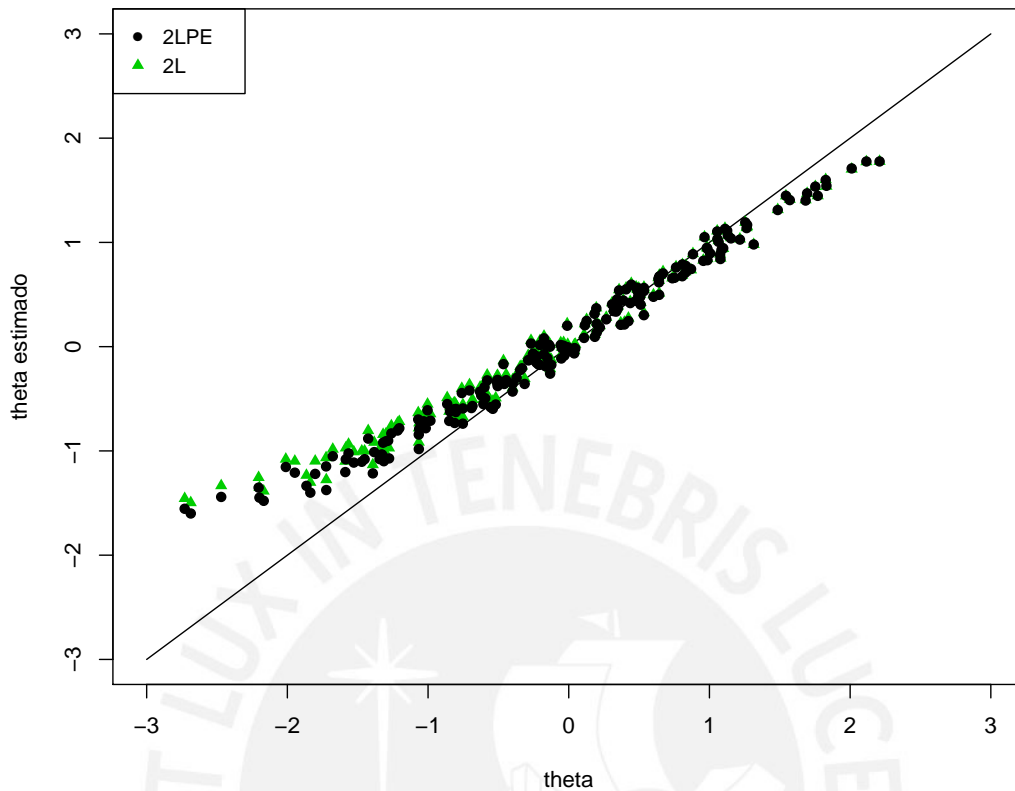


Figura 5.7: Estimación de la habilidad (θ) por el modelo TRI logístico de exponente positivo (2LPE) y TRI logístico (2L), considerando $m = 20$ réplicas

- Generar los parámetros del modelo, para los k -ítems $a \sim N(1, 0,5)I(0, \infty)$, donde $I(0, \infty)$ especifica el rango de los valores generados; $b \sim N(0, 1)$, $\theta \sim N(0, 1)$ y el efecto testlet para cada sujeto $g \sim N(0, \sigma_{h(j)}^2 = 0,5)$.
- Establecer restricciones de los parámetros: $\sum_i^n g_{ih(j)} = 0, \forall j$ (Bradlow et al., 1999), $\prod_j^k a_j = 1$ y $\sum_j^k b_j = 0$. (Bock y Aitkin, 1981) y (Spiegelhalter et al., 1996)
- Definir los ítems cuyo parámetro de penalidad d_j genera una CCI asimétrica.
- Simular la probabilidad (p_{ij}) de que el sujeto i responda correctamente al ítem j a partir del modelo 2LPET.
- Simular las respuestas de cada sujeto i a cada pregunta j , para esto simular $u_{ij} \sim U(0, 1)$ que nos dá una matriz de valores continuos entre 0 y 1. Luego comparar con la probabilidad generada anteriormente (p_{ij}) y si el valor generado es menor a la probabilidad ($u < p$) entonces y_{ij} tomará el valor 1, en caso contrario tomará valor 0.

En el proceso de simulación se generan los parámetros a, b, d, θ, g y quedan fijos, luego las respuestas son generadas para cada replica. El código para simular respuestas binarias con grupos testlet se muestra en el anexo (B.7).

5.3.3. Método para estimar los parámetros

La estimación de los parámetros se realiza mediante inferencia bayesiana, se utiliza la estimación MCMC implementada en el Openbugs, bajo el esquema ARS Gilks y Wild (1992) y se calcula la media estimada de la distribución a posteriori de los parámetros a estimar, considere que se estiman parámetros para cada uno de los ítems y los sujetos, por lo que se estimarán $3k + n(t + 1)$ parámetros en el modelo 2LPET y $3k + n$ parámetros en el modelo 2LPE.

5.3.4. Criterios para evaluar la simulación

Se compara los resultados a través de la raíz del error cuadrático medio (RMSE) y el promedio del error absoluto (MAE) Willmott y Matsuura (2005), los estimadores de a, b, d son calculados para cada uno de los k ítems, en el caso de θ para cada uno de los n sujetos y g para los n sujetos en t testlets. m es el número de simulaciones.

$$RMSE = \sqrt{\sum_s^m \frac{(\hat{\Theta}_s - \Theta)^2}{m}} \quad (5.5)$$

$$MAE = \sum_s^m \frac{|\hat{\Theta}_s - \Theta|}{m} \quad (5.6)$$

5.3.5. Escenario

Para la simulación se consideran $k = 20$ ítems y $t = 4$ testlets, 2 testlets de tamaño 3 y los otros dos testlets de tamaño 4, por lo que h queda definido como

$$h = c(5, 1, 1, 1, 2, 2, 2, 3, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 5, 5)$$

ó usando letras:

$$h = c(E, A, A, A, B, B, B, C, C, C, C, D, D, D, D, E, E, E, E, E)$$

Luego 13 ítems tienen CCIs asimétricas cuyo parámetro de penalización en 8 casos toma valores $d < 1$ y en 5 casos toma valores $d > 1$. Así d queda definido como:

$$d = c(0.1, 0.2, 1, 0.3, 0.3, 0.4, 1.5, 0.8, 1, 1, 1.1, 1.4, 0.7, 1.8, 3, 1, 1, 0.3, 1, 1)$$

La cantidad de replicas es $m = 20$ con diferentes conjuntos de datos en cada réplica.

5.3.6. Resultados

Los resultados del estudio de simulación se presentan a continuación. En el cuadro 5.9 se muestran las medidas promedio del RMSE y MAE.

Los resultados de las medidas RMSE y MAE para cada ítem se presentan en el cuadro 5.10.

	Parámetro	$RMSE$	MAE
2LPET	a	0.332	0.282
	b	0.625	0.597
	d	0.434	0.403
	θ	0.508	0.424
	g	0.592	0.514
	σ_g^2	0.383	0.374
2LPE	a	0.317	0.275
	b	0.641	0.613
	d	0.431	0.395
	θ	0.507	0.425

Cuadro 5.9: Resultados del promedio de RMSE y MAE en el estudio de simulación 3, para comparar los modelos Testlet logístico de exponente positivo (2LPET) y TRI logístico de exponente positivo (2LPE), considerando $m = 20$ réplicas

Item		2LPET						2LPE						
No.	t	d	RMSE a	RMSE b	RMSE d	MAE a	MAE b	MAE d	RMSE a	RMSE b	RMSE d	MAE a	MAE b	MAE d
1	E	0.1	0.413	0.45	0.165	0.318	0.368	0.145	0.362	0.413	0.162	0.276	0.332	0.143
2	A	0.2	0.463	1.026	0.313	0.436	1.005	0.306	0.487	1	0.338	0.458	0.981	0.331
3	A	1.0	0.397	0.149	0.285	0.289	0.128	0.233	0.223	0.174	0.285	0.159	0.139	0.21
4	A	0.3	0.378	1.24	0.442	0.338	1.223	0.43	0.392	1.271	0.477	0.361	1.254	0.46
5	B	0.3	0.239	0.368	0.063	0.197	0.326	0.046	0.245	0.289	0.114	0.198	0.261	0.092
6	B	0.4	0.372	1.361	0.137	0.304	1.343	0.127	0.375	1.366	0.123	0.303	1.35	0.111
7	B	1.5	0.579	0.144	0.298	0.526	0.121	0.244	0.616	0.155	0.299	0.592	0.133	0.274
8	C	0.8	0.398	0.626	0.876	0.329	0.608	0.849	0.318	0.691	0.851	0.245	0.665	0.814
9	C	1.0	0.239	0.596	0.864	0.202	0.573	0.82	0.234	0.684	0.855	0.202	0.655	0.796
10	C	1.0	0.26	0.252	0.584	0.219	0.2	0.509	0.222	0.281	0.593	0.182	0.231	0.49
11	C	1.1	0.495	0.339	0.933	0.422	0.311	0.88	0.481	0.418	0.943	0.413	0.399	0.861
12	D	1.4	0.177	0.457	0.219	0.157	0.441	0.204	0.266	0.481	0.245	0.243	0.457	0.232
13	D	0.7	0.219	0.171	0.188	0.176	0.138	0.155	0.234	0.179	0.274	0.186	0.138	0.217
14	D	1.8	0.268	0.191	0.482	0.212	0.156	0.429	0.137	0.224	0.251	0.104	0.187	0.193
15	D	3.0	0.291	1.607	0.938	0.282	1.605	0.918	0.283	1.583	1.011	0.27	1.58	0.994
16	E	1.0	0.23	1.19	0.283	0.203	1.181	0.272	0.223	1.203	0.281	0.203	1.198	0.271
17	E	1.0	0.357	0.871	0.622	0.288	0.831	0.613	0.341	1.021	0.58	0.274	1.008	0.568
18	E	0.3	0.232	0.604	0.186	0.214	0.569	0.167	0.264	0.559	0.171	0.241	0.522	0.157
19	E	1.0	0.39	0.365	0.697	0.296	0.332	0.625	0.399	0.356	0.655	0.367	0.308	0.59
20	E	1.0	0.254	0.485	0.112	0.236	0.477	0.083	0.238	0.479	0.119	0.219	0.469	0.093

Cuadro 5.10: Raíz del error cuadrático medio y Media absoluta del error de los estimadores en la simulación - Estudio 3, considerando $m = 20$ réplicas.

En el cuadro 5.11 se muestran las medidas de ajuste del estudio de simulación 3, para los modelos 2LPET y 2LPE, considerando el criterio DIC, se puede observar que el modelo 2LPE tiene un mejor ajuste. Note que al igual que en los estudios anteriores los criterios EAIC y EBIC presentan valores altos para el modelo con mayor cantidad de parámetros, por lo que estos criterios no son adecuados para una buena comparación.

Modelo	DIC	pD	Tiempo (seg.)	Dbar	EAIC	EBIC
2LPET	3861	384	5603	3477	5597	7295
2LPE	3851	196	3226	3655	4175	4592

Cuadro 5.11: Resultados de medidas promedio de ajuste de la simulación para comparar los modelos Testlet logístico de exponente positivo (2LPET) y TRI logístico de exponente positivo (2LPE)

En las figuras 5.8, 5.9 y 5.10 se muestran el RMSE, MAE y el sesgo para la estimación de los parámetros de ítem de los modelos 2LPET y 2LPE.

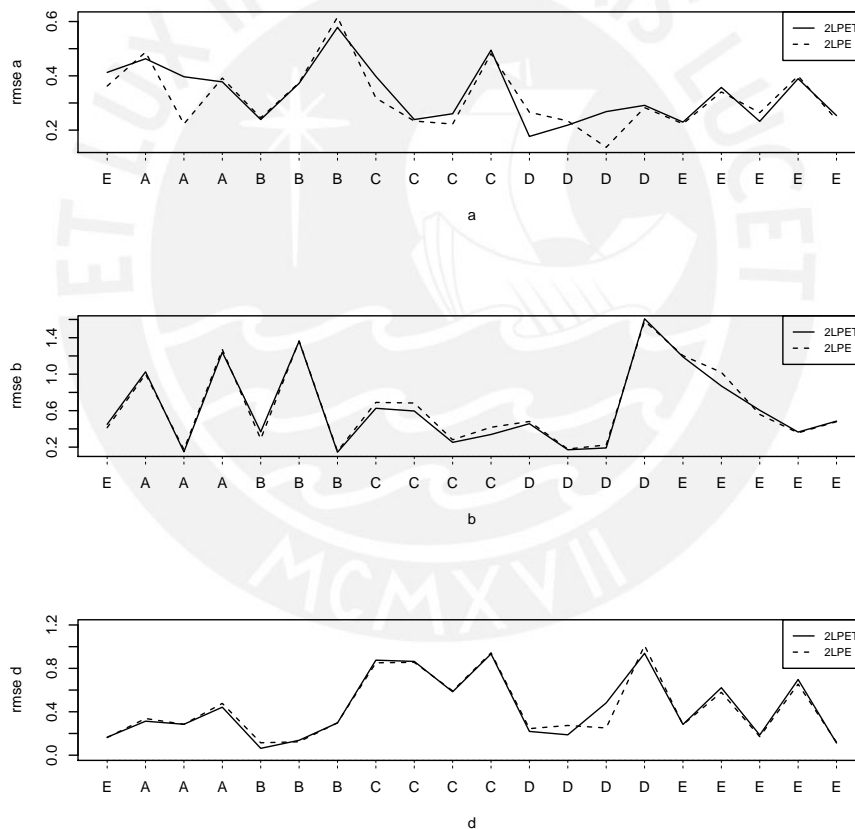


Figura 5.8: Raíz del Error Cuadrático Medio (RMSE) de los estimadores simulados, para la discriminación (a), dificultad (b) y el parámetro de penalidad (d) de cada ítem, con el modelo Testlet logístico de exponente positivo (2LPE) y TRI logístico de exponente positivo (2LPE), considerando $m = 20$ réplicas

La figura 5.11 muestra la estimación de las habilidades bajo ambos modelos, en comparación con los valores simulados.

Finalmente en el cuadro 5.12 se muestra las correlaciones entre las estimaciones de los paráme-

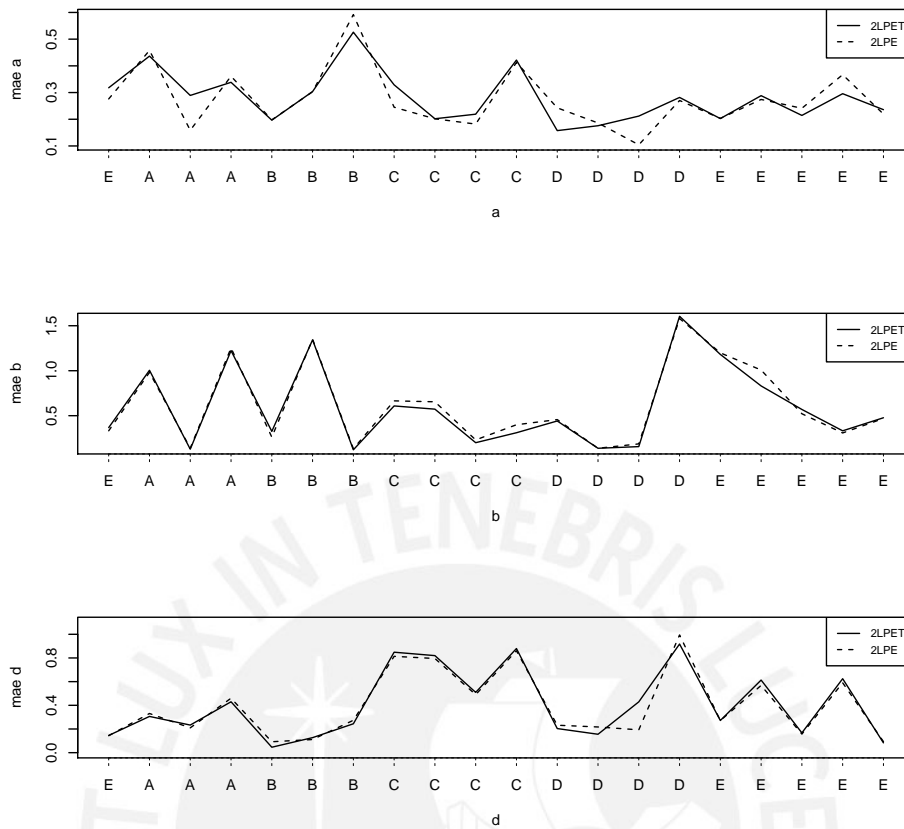


Figura 5.9: Media Absoluta del Error de los estimadores simulados, para la discriminación (a), dificultad (b) y penalidad (d) de cada ítem, con un modelo Testlet logístico de exponente positivo (2LPET) y TRI logístico de exponente positivo (2LPE), considerando $m = 20$ réplicas

tros de ítem y de las habilidades, bajo los dos modelos, con sus correspondientes valores verdaderos. Note que la estimación del parámetro de dificultad y de la variable latente para el modelo 2LPET tiene mayor correlación respecto al modelo 2LPE.

Correlación	2LPET	2LPE
r_a	0.908	0.912
r_b	0.730	0.718
r_θ	0.957	0.950

Cuadro 5.12: Medidas de correlación entre los valores verdaderos y los estimados promedio de los modelos Testlet logístico de exponente positivo (2LPET) y TRI logístico de exponente positivo (2LPE), considerando $m = 20$ réplicas

Las conclusiones del estudio 3 son:

- Cuando existe efecto conjunto de asimetría y testlets, el modelo 2LPE tiene un mejor ajuste. Note que se consideró 12 ítems asimétricos, 8 con $d < 1$ y 3 ítems con penalidad $d > 1$
- El modelo 2LPET recupera mejor que el 2LPE, el parámetro de dificultad b y la medida de habilidad, como se muestra en las correlaciones con los valores verdaderos.

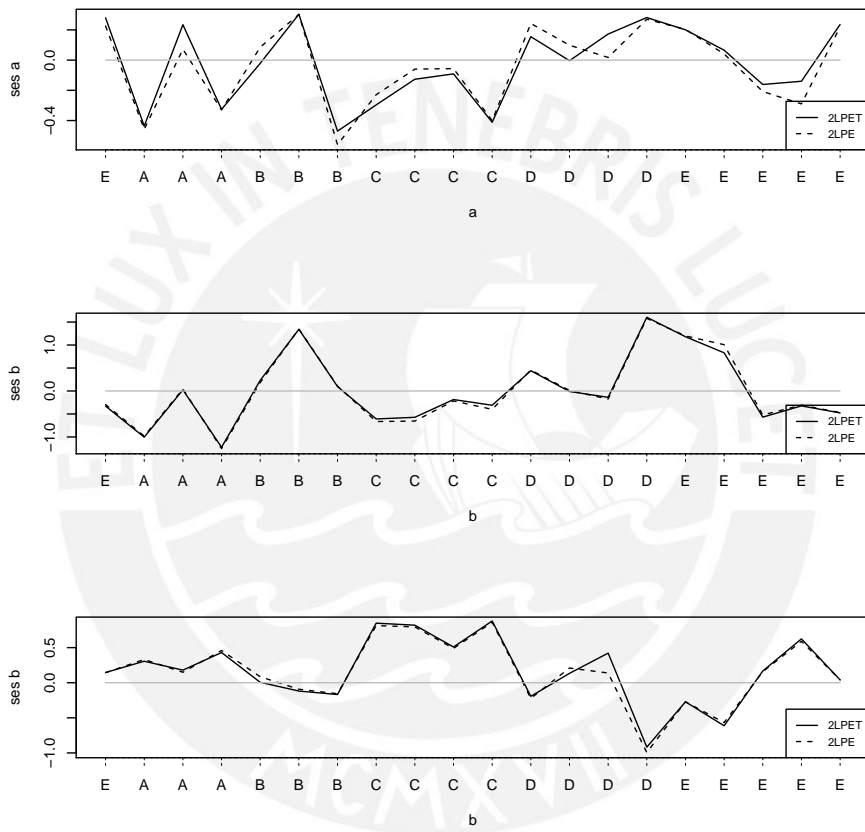


Figura 5.10: Sesgo de los estimadores simulados, para la discriminación (a), dificultad (b) y penalidad (d) de cada ítem, con un modelo Testlet logístico de exponente positivo (2LPET) y TRI logístico de exponente positivo (2LPE), considerando $m = 20$ réplicas

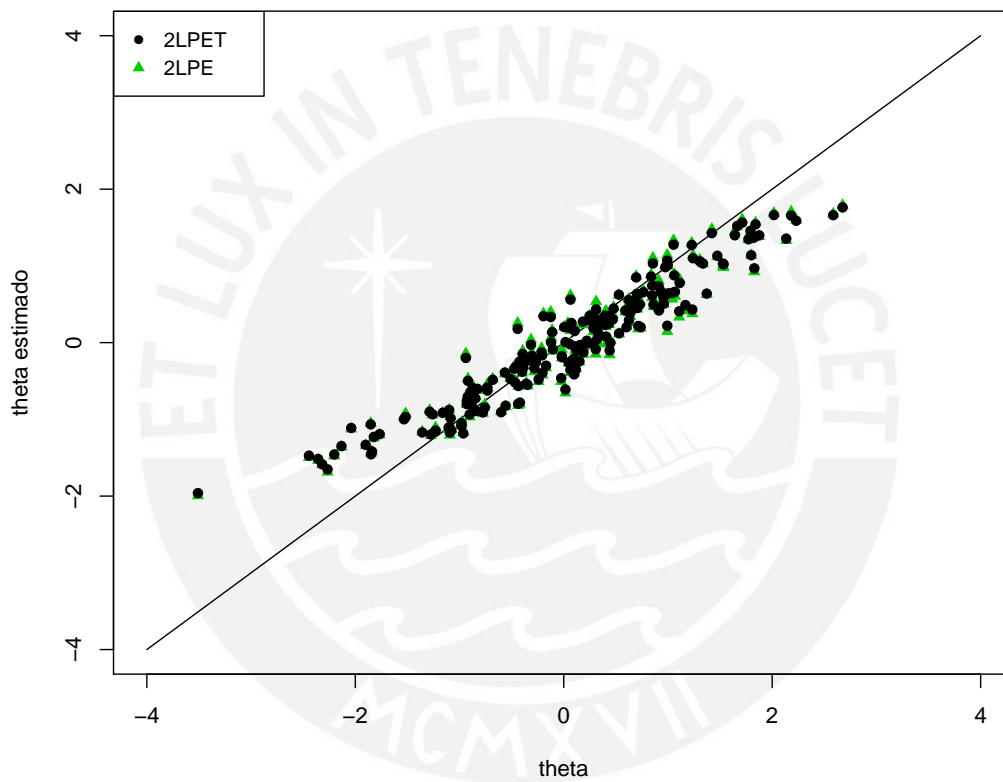


Figura 5.11: Estimación promedio de la habilidad (θ) por el modelo Testlet logístico de exponente positivo (2LPET) y TRI logístico de exponente positivo (2LPE), considerando $m = 20$ réplicas

Capítulo 6

Aplicación

La aplicación desarrollada en el presente trabajo esta orientada al área de educación. Una medida importante en esta área para la toma de decisiones de políticas de estado, a nivel nacional, incluso a nivel distrital, es el rendimiento de los estudiantes. El ministerio de educación a través de la Unidad de Medición de la Calidad educativa (UMC) a partir del año 2007 realiza la Evaluación Censal de Estudiantes (ECE) [MINEDU \(2009\)](#). La ECE evalúa anualmente, en comprensión de lectora y matemática, a todos los estudiantes de segundo grado de primaria de las Instituciones Educativas a nivel nacional que tienen 5 o más estudiantes matriculados en el segundo grado de primaria, y reporta el rendimiento de los estudiantes a partir de una medida de habilidad latente estimada desde el punto de vista clásico a través del modelo TRI logístico de un parámetro (modelo de Rash).

Las pruebas de comprensión de textos en la ECE, esta conformado por grupos de testlets. Esta estructura genera dependencia en los ítems de un mismo testlet y en este caso cuando se asume independencia condicional en un modelo TRI, los estimadores estan sesgados, tal como se muestra en el proceso de simulación [5.3.6](#), por lo que realizaremos la estimación de la habilidad latente y los parámetros de los ítems con los modelos presentados en el capítulo 3, en una muestra delimitada de estudiantes que fueron evaluados en la ECE 2010 y se evalúa el efecto de la dependencia generada por los testlet en los estimadores.

6.1. Instrumento de Medición

Se utiliza como instrumento de medición una de las pruebas de Comprensión Lectora, diseñado para la Evaluación Censal de Estudiantes 2010. En la ECE se evalúan a los estudiantes de segundo grado mediante dos cuadernillos de comprensión de textos y dos de matemática. La ECE busca explorar los procesos que el estudiante realiza al interpretar un texto en una situación concreta de comunicación y al enfrentarse con situaciones cotidianas que demandan el uso de su conocimiento matemático [MINEDU \(2009\)](#). Algunas características de la ECE son:

Capacidades Evaluadas. Las capacidades evaluadas en la ECE corresponden a aquellas que son las más relevantes por ser la base del resto de aprendizajes; además, por ser evaluables a gran escala. La selección de dichas capacidades se realiza a partir de lo que se propone en el Diseño Curricular Nacional (DCN) vigente al momento de la aplicación de las evaluaciones. Estas capacidades son:

- Lee oraciones: Establece correspondencia entre una oración y un dibujo.

- Localiza información literal: Identifica datos explícitos. Reconoce sucesiones de hechos o acciones.
- Infiere información: Deduce relaciones de causa-efecto. Deduce el significado de palabras o expresiones a partir del contexto, deduce el tema central del texto. Deduce la idea principal de un párrafo del texto. Deduce las cualidades o defectos de los personajes de una narración. Deduce la enseñanza del texto. Deduce el propósito del texto.

Contenidos evaluados. Son los tipos de texto propuestos en las pruebas

- Narrativo: anécdota, cuento
- Informativo: descripción, artículo enciclopédico, nota, carta, noticia breve
- Instructivo: receta, reglas de juego, texto de recomendaciones
- Argumentativo: aviso

Debido a que se devuelven resultados por alumno, todos los niños rinden las mismas pruebas (4 cuadernillos), estos son evaluados en dos días de aplicación de la siguiente forma:

	Nro. Cuadernillo	Cantidad de ítems	Día de aplicación
Comprensión de Textos	C1	23	1er. día
	C2	23	2do. día
Matemática	M1	21	1er. día
	M2	21	2do. día

Cuadro 6.1: Cuadernillos ECE 2010

Para la presente aplicación, se utiliza el Cuadernillo 1 (C1) de Comprensión de Textos que consta de 23 preguntas o ítems. Esta prueba esta conformada por 5 testlets. A continuación el detalle de la estructura de la prueba identificando los ítems y la cantidad de ítems por testlets:

Nro. de testlet	ítem	Cantidad de ítems
6	1	1
6	2	1
1	3, 4, 5, 6	4
2	7, 8, 9, 10	4
6	11	1
3	12, 13	2
4	14, 15, 16, 17, 18	5
5	19, 20, 21, 22, 23	5

Cuadro 6.2: Estructura de los testlet en el instrumento de medición

Como se puede observar en el cuadro (6.2), el cuadernillo C1 esta formado por 5 testlet, de diferentes tamaños y existen solo 3 ítems (1,2 y 11) que no estan dentro de un grupo testlet. Para identificar estos ítems en el modelo se les asigna el código testlet igual a la cantidad de teslets mas uno, por ello en la columna Nro. testlet estos ítems tienen el valor 6.

6.2. Muestra

El objetivo de esta aplicación es comparar los estimadores a partir de los diferentes modelos planteados, mas no hacer inferencias en alguna población específica. En tal sentido, delimitaremos nuestro análisis a un grupo de estudiantes de segundo grado de primaria de Lima metropolitana, específicamente del distrito de Santiago de Surco y analizaremos las respuestas de los estudiantes a la prueba de comprensión de textos, Cuadernillo 1, de las Instituciones Educativas seleccionadas para la muestra control del año 2010.

La muestra control esta formada por un grupo de escuelas seleccionadas del marco muestral de la ECE con la finalidad de reportar resultados con inferencia a nivel regional y tienen características operativas rigurosas que permiten controlar los errores de aplicación.

6.3. Recolección de datos

La recolección de información se realiza durante dos días de evaluación. Existe un aplicador asignado para cada sección de la Institución Educativa (IE) evaluada. En el caso de que una IE sea parte de la muestra control adicionalmente se le asigna un aplicador lider para que garantice un estandar adecuado en el momento de la evaluación. Una vez finalizada la evaluación los aplicadores son los encargados de trasladar la información de los cuadernillos a fichas ópticas que son revisadas por el director de la IE, mediante un proceso de control de calidad a un grupo de cuadernillos seleccionados por el mismo director. Luego las fichas ópticas y los cuadernillos en el caso de la muestra control, son digitalizados mediante un proceso de lectura de fichas ópticas y lectura de información de marcas de los cuadernillos. La base de datos obtenida del proceso de lectura tiene un control de calidad en cada etapa que involucra este proceso, y luego pasa por un proceso de consistencia y depuración para luego obtener la información de las respuestas de los estudiantes. Estas respuestas son finalmente codificadas como aciertos ($y=1$) o desaciertos ($y=0$); es decir, en una matriz de datos dicotómicos donde las filas nos indican el i -ésimo estudiante y cada pregunta el j -ésimo ítem.

6.4. Análisis de los datos

En la teoría de los scores mentales ó variables latentes, la estimción del score se realiza a partir de información de otras variables observadas. En la presente aplicación nuestro interés es la habilidad del estudiante en la comprensión de textos que es definida como una medida que será estimada a partir de las respuestas de los estudiantes a un conjunto de ítems de un test. El test es un instrumento de medición que se elabora siguiendo ciertos procesos de definición, validación y consistencia. De tal forma que se obtiene un test que recoja la información precisa que nos permita estimar el score adecuado.

La validación es un proceso en el cual se verifica que la medida obtenida se acerca a la definición planteada.

La confiabilidad se puede definir como la estabilidad de los resultados y que tan bien respresentan los puntajes observados al puntaje verdadero. Puede ser medido con la consistencia interna (alpha-cronbach).

6.4.1. Análisis clásico de los datos

Desde el punto de vista de la teoría clásica de los test, en el análisis clásico del ítem, la primera preocupación es obtener una descripción de las características estadísticas de los parámetros de un ítem de la prueba, sin embargo el ítem individual solo es de interés a través del efecto que tiene sobre el score total. Así es importante el análisis del efecto individual de las características del ítem en la característica de la medida total del instrumento [Lord y Novick \(1968\)](#). Los datos que forman parte de la aplicación son las respuestas de 344 estudiantes al cuadernillo C1, que fueron evaluados en 12 escuelas del distrito de Santiago de Surgo. Las puntuaciones totales ó "scores" se distribuyen como se muestra en la figura 6.1.

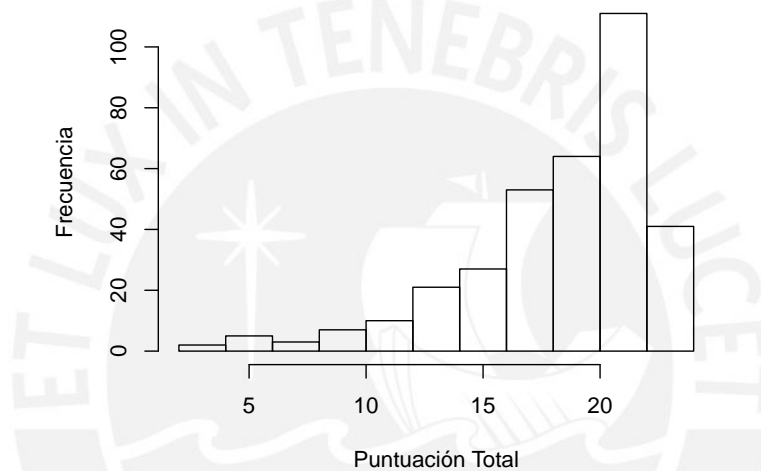


Figura 6.1: Distribución de las puntuaciones totales

Algunas estadísticas de los scores se presentan en el cuadro (6.3).

Medida	valor
media	18.8
mediana	20
des. est.	3.99
asimetría	-1.4
kurtosis	2.0

Cuadro 6.3: Estadísticos del score

Note que los scores tienen un coeficiente de asimetría negativo, con lo cual los datos se encuentran concentrados en la cola derecha.

Dificultad del ítem. La proporción de aciertos $pa_j = \frac{\sum_i^n y_{ij}}{n}$ es un estimador insesgado, para la muestra de examinados, de la dificultad del ítem poblacional.

Poder de discriminación del ítem. Dado por la varianza de la puntuación del ítem j $s_j^2 = p_j - p_j^2$. Items con varianza grande son elegidos de manera preferente frente a otros en

la construcción de una prueba.

Confiabilidad interna de los datos. Medido por el alpha de cronbach,

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_j^k \sigma_j^2}{\sigma_t^2}\right) = 0,84$$

6.5. Estimación de parámetros mediante los modelos desarrollados

Para la implementación de los modelos se utiliza el software Openbugs. Existe para el efecto un código escrito en Openbugs para el modelo LPE realizado por Bazan y Bolfarine (2010). Se utilizó esta fuente así como también el código escrito para modelos testlet en el reciente artículo publicado por McKay (2010).

1[ex] Para comparar la estimación realizada por los diferentes modelos, se corre la aplicación para 15 mil iteraciones con una cadena para cada modelo planteado y se utiliza la medida de ajuste DIC (ver sección 6.6).

6.5.1. Modelo TRI Logístico (2L)

En este modelo la probabilidad de responder correctamente a un ítem depende de la habilidad del estudiante, de la dificultad del ítem y de la discriminación de cada ítem. Esta probabilidad es modelada como se define en la sección 3.2 por:

$$P(y_{ij} = 1 | \theta_i, a_j, b_j) = \Psi(m_{ij})$$

donde $m_{ij} = a_j(\theta_i - b_j)$ y $\Psi(\cdot)$ es la acumulada de la distribución logística. Los valores estimados de los parámetros de ítem, se presentan en el cuadro 6.4.

Los estimadores de los parámetros de ítem nos dan una idea de caracterización del ítem; es decir, b_j mide que tan complejo o difícil es la capacidad que mide el ítem j y a_j mide el poder de discriminación del ítem j . De los resultados podemos observar en el cuadro 6.4 que el ítem $j = 18$ es el más difícil y el ítem $j = 1$ es el que discrimina más que los demás ítems. Estas estimaciones ayudan a describir lo que el estudiante logra desarrollar y cuales son las capacidades mas complicadas en las que se debe incidir desde la parte pedagógica.

6.5.2. Modelo Testlet Logístico (2LT)

En el modelo anterior (2L) para el cálculo de la verosimilitud se considera que los ítems son independientes, de tal forma que podemos escribir:

$$f(y; \theta, a, b) = \prod \prod p_{ij}^{y_{ij}} (1 - p_{ij})^{(1-y_{ij})}$$

Sin embargo, la estructura de nuestro instrumento de medición (ver cuadro 6.2) nos muestra que existen ítems que están relacionados con un mismo estímulo. Por ejemplo, los ítems $j = 3, 4, 5$ y 6 son preguntas referidas al texto *a volar nuestras cometas*, estas estructuras las denominamos como testlets y tal como lo definimos en la sección 3.4 modificamos el predictor lineal latente como: $m_{ij} = a_j(\theta_i - b_j - g_{ih(j)})$, de tal forma que se modela la dependencia de estos ítems. En nuestro instrumento de medición tenemos 5 testlets, por lo cual el parámetro $g_{ih(j)}$ recoge el efecto de interacción de cada estudiante i con el testlet $h(j) = 1, 2, 3, 4, 5$.

Item no.	a_j	ee_{a_j}	b_j	ee_{b_j}
1	2	0.6	-1	0.2
2	1.9	0.5	-0.7	0.2
3	0.9	0.1	0.2	0.2
4	1	0.1	0.8	0.1
5	1	0.2	0	0.2
6	0.8	0.1	0.3	0.2
7	0.8	0.1	-1	0.3
8	1.1	0.2	0.5	0.1
9	1.1	0.2	0	0.2
10	1.2	0.2	-0.1	0.2
11	0.7	0.1	0.7	0.2
12	0.6	0.1	-0.8	0.3
13	1.5	0.3	-0.5	0.2
14	1	0.2	0.4	0.1
15	1.1	0.2	-0.4	0.2
16	0.6	0.1	0.3	0.2
17	0.8	0.1	-0.6	0.3
18	0.5	0.1	0.9	0.2
19	1.7	0.3	-0.4	0.2
20	0.6	0.1	0.6	0.2
21	1.6	0.3	0.4	0.1
22	1.6	0.3	-0.1	0.1
23	1.1	0.2	0.5	0.1

Cuadro 6.4: Estimación de los parámetros del modelo 2L

Si este efecto es muy variable la varianza de $g_{ih(j)}$ será grande por lo que nuestro interés es conocer $\sigma_{g_{h(j)}}^2$ para los 5 testlets.

En este modelo es importante definir la estructura de los testlets, es decir cuantos y qué ítems pertenecen a cada uno de los testlet. Esta información se ingresa como parte de los datos, y según el cuadro 6.2, el vector que recoge la información de ítems por testlet esta dado por:

$$h = c(6, 6, 1, 1, 1, 1, 2, 2, 2, 2, 6, 3, 3, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5)$$

ó considerando letras:

$$h = c(F, F, A, A, A, A, B, B, B, B, F, C, C, D, D, D, D, D, E, E, E, E, E)$$

Note que los ítems independientes tienen el valor del número total de testlets más uno.

Al ajustar los datos mediante este modelo (2LT) se obtuvo los siguientes resultados mostrados en el cuadro 6.5. La varianza generada por los testlets se muestran en el cuadro 6.7.

Podemos observar que el testlet 3 genera mayor varianza y en consecuencia tiene un mayor efecto en el score de los estudiantes y esto genera nuevos estimadores de los parámetros de ítem.

Item no.	t	a_j	ee_{a_j}	b_j	ee_{b_j}
1	F	2.3	0.9	-0.8	0.2
2	F	2.3	0.8	-0.6	0.2
3	A	0.8	0.1	0.2	0.2
4	A	0.9	0.2	0.8	0.1
5	A	1	0.2	0	0.2
6	A	0.8	0.1	0.3	0.2
7	B	0.8	0.1	-1	0.3
8	B	1.1	0.2	0.5	0.1
9	B	1.1	0.2	0.1	0.2
10	B	1.2	0.2	0	0.2
11	F	0.7	0.1	0.8	0.2
12	C	0.6	0.1	-0.9	0.4
13	C	2.7	1.7	-0.7	0.3
14	D	1	0.2	0.4	0.2
15	D	1.1	0.2	-0.5	0.2
16	D	0.5	0.1	0.3	0.2
17	D	0.8	0.2	-0.7	0.3
18	D	0.5	0.1	1	0.2
19	E	1.6	0.3	-0.5	0.2
20	E	0.6	0.1	0.7	0.2
21	E	1.5	0.3	0.4	0.1
22	E	1.6	0.3	-0.2	0.2
23	E	1.1	0.2	0.5	0.1

Cuadro 6.5: Estimación de los parámetros de ítem del modelo 2LT

Testlet	$s_{g_{ih(j)}}^2$
1	0.27
2	0.13
3	0.84
4	0.39
5	0.21

Cuadro 6.6: Varianza estimada de los testlets en el modelo 2LT

6.5.3. Modelo TRI Logístico de Exponente Positivo (2LPE)

En los dos modelos anteriores 2L y 2LT la probabilidad de una respuesta es modelada con una CCI simétrica. Sin embargo, Samejima (1999) afirma que esto es contradictorio en el ordenamiento de los estudiantes; es decir, la probabilidad de respuesta correcta para los estudiantes más hábiles no necesariamente es distribuido de manera simétrica a la de los estudiantes menos hábiles. Por ello se adiciona el parámetro d_j , denominado de penalidad en el modelo 2L y así obtenemos el modelo 2LPE definido en (3.3) con lo cual se generan CCIs asimétricas, de la siguiente forma:

$$p_{ij} = P(y_{ij} = 1 | \theta_i, a_j, b_j, d_j) = \Psi(m_{ij})^{d_j}$$

donde $\Psi(m_{ij})^{d_j}$ es la acumulada de la distribución logística generalizada definido en 2.2.

Los resultados de la estimación mediante el modelo 2LPE se presenta en el cuadro 6.7. Note que los valores estimados del parámetro $d_j < 1$ en 19 ítems, lo que nos indica que éstos ítems tienen una probabilidad mayor de responder correctamente el ítem que considerando CCI's simétricas y en un solo caso $d_{23} > 1$, para éste ítem la probabilidad de responder correctamente es menor a la de una CCI simétrica.

Item no.	a_j	ee_{a_j}	b_j	ee_{b_j}	d_j	ee_{d_j}
1	1.7	0.6	-1.1	0.5	0.3	0.4
2	2.1	0.9	-0.8	0.4	0.3	0.3
3	0.9	0.2	0.2	0.6	0.6	0.3
4	1.1	0.5	0.6	0.7	0.7	0.6
5	1.2	0.5	0.1	0.6	0.4	0.4
6	0.9	0.3	0.3	0.7	0.6	0.4
7	0.8	0.3	0	0.8	0.2	0.2
8	1.7	0.7	0.6	0.4	0.4	0.3
9	1	0.2	-0.4	0.6	0.8	0.6
10	1.1	0.3	-0.4	0.6	0.7	0.6
11	1	0.6	1.2	0.5	0.4	0.2
12	0.5	0.2	0.1	0.9	0.3	0.2
13	1.3	0.3	-0.9	0.5	0.7	0.6
14	1	0.3	-0.1	0.5	0.9	0.6
15	1.1	0.3	-0.3	0.6	0.4	0.3
16	0.6	0.3	0.7	0.9	0.6	0.4
17	1	0.3	0.4	0.5	0.2	0.1
18	0.6	0.3	0.8	0.8	0.8	0.4
19	1.5	0.4	-0.9	0.4	0.8	0.6
20	0.6	0.2	0.5	0.8	0.8	0.4
21	1.8	0.7	0	0.5	0.7	0.5
22	1.6	0.6	-0.5	0.5	0.7	0.6
23	1	0.2	-0.3	0.5	1.3	0.6

Cuadro 6.7: Estimación de los parámetros de ítem del modelo 2LPE

En la figura 6.2, se muestra la probabilidad de respuesta al ítem 1, en el modelo 2L la CCI es simétrica, en el modelo 2LPE la CCI tiene asimetría negativa porque la mayor cantidad de estudiantes tienen una probabilidad alta de acertar el ítem.

El código del modelo LPE se encuentra en el anexo A3. Los estimadores tienen un error de estimación mayor, esto podría ser debido a la alta correlación con la iteración anterior, se han quitado más iteraciones que el caso de los modelos simétricos, el ajuste del modelo LPE no mejora mucho, sin embargo es importante notar como se corrige la información de la prueba respecto a la medida de habilidad, que estaba concentrada alrededor de la media con los modelos simétricos, ver figura 6.5.

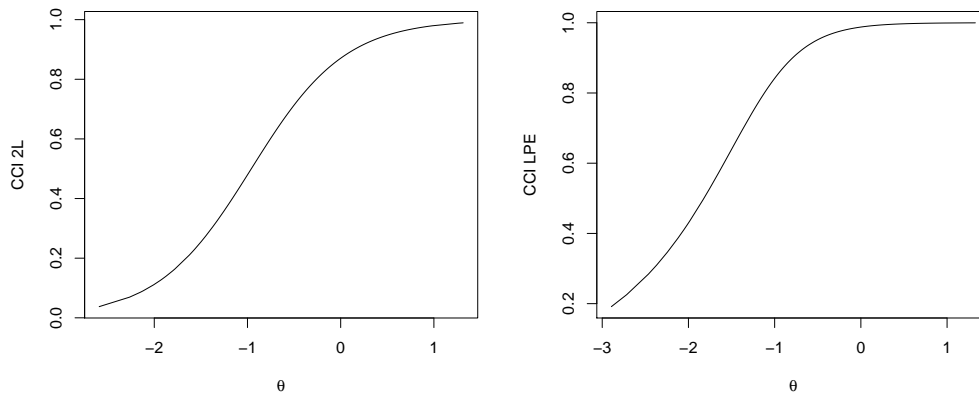


Figura 6.2: Curvas características del ítem para los modelos 2L y 2LPE

6.5.4. Modelo Testlet Logístico de Exponente Positivo (2LPET)

Si en el modelo 2LPE se tiene ítems que están relacionados mediante un mismo estímulo, como es el caso de nuestro instrumento de medición, se puede modelar el efecto de esta dependencia para CCIs asimétricas. Mediante el modelo que proponemos, el modelo 2LPET definido en la sección 3.4, en el cual se modifica el predictor lineal latente

$$m_{ij} = a_j(\theta_i - b_j - g_{ih(j)})$$

donde $g_{ih(j)}$ recoge el efecto testlet. El instrumento de medición utilizado tiene 5 testlets por lo que $h(j) = 1, 2, 3, 4, 5$. g está centrado en cero para cada testlet $\sum_i^n g_{ih(j)} \forall h(j)$. De la misma forma que el modelo 2LT nuestro interés es conocer la varianza de cada testlet $\sigma_{g_{h(j)}}^2$ en los ítems con CCIs asimétricas.

Los resultados de la estimación de los parámetros se muestran en el cuadro 6.8. Podemos observar que la CCI para el ítem 13 es asimétrica, tiene dependencia con el ítem 12 y cuando se recoge el efecto de esta dependencia el parámetro de penalización es $d_{13} = 0,3$. Esto resulta interesante porque el ítem 12 y 13, ambos, miden la capacidad de localizar información literal en un texto y esta es una capacidad sencilla para los estudiantes evaluados.

En el cuadro 6.9 se muestra la estimación de la varianza de cada testlet. El testlet 3 es el que presenta mayor varianza.

6.6. Comparación de modelos

Si nos basamos en el DIC, el modelo que mejor estima los resultados es el modelo 2LT, sin embargo el modelo 2LPET distribuye mejor la información del score θ , ver 6.5.

En la figura 6.3 se muestra la dispersión de las habilidades estimadas con el modelo Logístico (2L) y el modelo Logístico Testlet (2LT), podemos observar que los puntos que están arriba de la línea son valores que tienen mayor puntaje en el eje y, es decir tienen un valor mayor para la estimación mediante 2L.

En la figura 6.4 se compara las habilidades estimadas con el modelo 2L y 2LPET, y existen

Item no.	t	a_j	ee_{a_j}	b_j	ee_{b_j}	d_j	ee_{d_j}
1	F	2	1	-1.2	0.5	0.4	0.5
2	F	2.4	1.1	-0.9	0.4	0.3	0.4
3	A	0.9	0.3	0.3	0.6	0.5	0.3
4	A	1.2	0.6	0.9	0.5	0.5	0.3
5	A	1.1	0.4	-0.2	0.6	0.6	0.4
6	A	0.9	0.4	0.4	0.7	0.6	0.4
7	B	0.8	0.3	0.1	0.7	0.2	0.1
8	B	1.6	0.8	0.5	0.6	0.4	0.4
9	B	1.1	0.3	-0.2	0.6	0.6	0.4
10	B	1.2	0.4	-0.6	0.6	0.8	0.6
11	F	0.9	0.4	0.9	0.7	0.5	0.4
12	C	0.8	0.4	0.0	0.6	0.3	0.2
13	C	0.9	0.4	0.1	0.8	0.1	0.1
14	D	1.1	0.4	0.1	0.6	0.7	0.5
15	D	1.1	0.4	-0.5	0.6	0.4	0.4
16	D	0.5	0.2	0.5	0.9	0.6	0.3
17	D	0.9	0.4	0.1	0.7	0.3	0.2
18	D	0.5	0.3	0.8	0.9	0.8	0.4
19	E	1.6	0.6	-0.8	0.5	0.6	0.5
20	E	0.6	0.2	0.3	0.8	0.8	0.4
21	E	2	1	0.1	0.4	0.5	0.5
22	E	1.7	0.6	-0.5	0.4	0.5	0.4
23	E	1	0.2	-0.4	0.5	1.2	0.7

Cuadro 6.8: Estimación de los parámetros de ítem del modelo 2LPET

Testlet	$s_{g_{ih(j)}}^2$
1	0.25
2	0.18
3	0.39
4	0.31
5	0.21

Cuadro 6.9: Varianza estimada de los testlets en el modelo 2LPET

Modelo	Dbar	Dhat	DIC	pD
2L	5428.0	5140.0	5717.0	288.5
2LT	5129.0	4649.0	5609.0	480.3
2LPE	5418.0	5129.0	5707.0	288.9
2LPET	5124.0	4605.0	5642.0	518.8

Cuadro 6.10: Medidas de comparación de los modelos

mayor cantidad de puntos que están alejados de la línea en comparación con la figura 6.3, lo cual nos permite afirmar que el modelo 2L sobreestima la habilidad de los estudiantes que tienen puntuaciones muy bajas, respecto al modelo 2LPET.

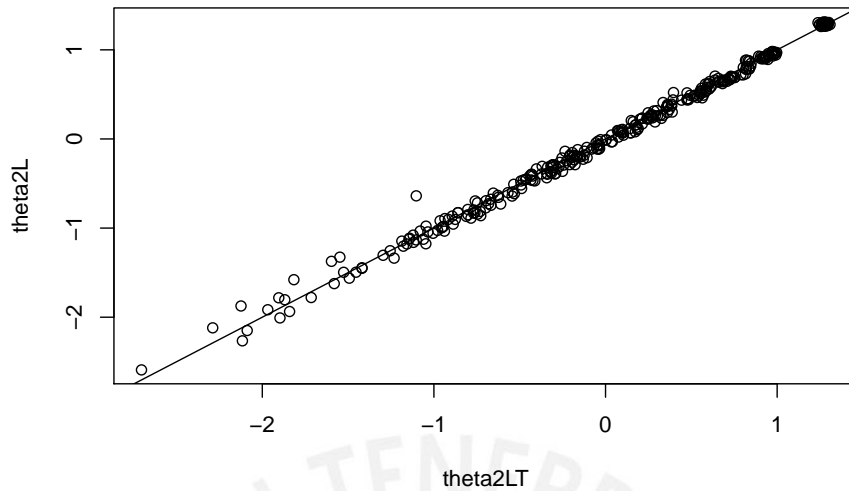


Figura 6.3: Dispersión de las habilidades estimadas con el modelo 2L y 2LT

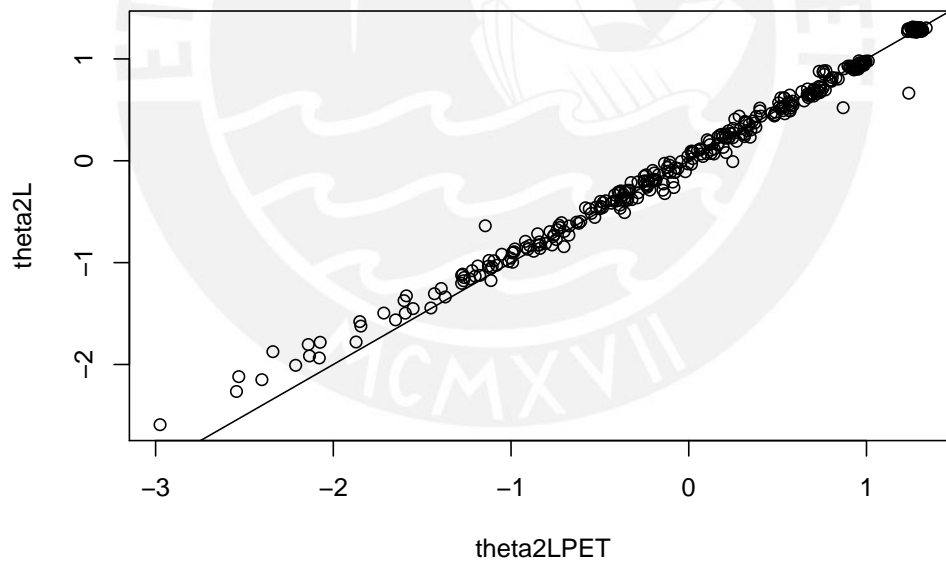


Figura 6.4: Dispersión de las habilidades estimadas con el modelo 2L y 2LPE Testlet

6.7. Función de Información

La información de fisher ha sido calculada con la estimación promedio de cada parámetro en los modelos desarrollados, el código en R se encuentra en el anexo (B.8). Los resultados de la función de información de θ indican que si se ignora el efecto testlet la información de θ es menor, en el caso simétrico; es decir, mayores errores estándar. Sin embargo, en el

caso asimétrico el considerar el efecto testlet podría equilibrar mejor la información para los valores extremos.

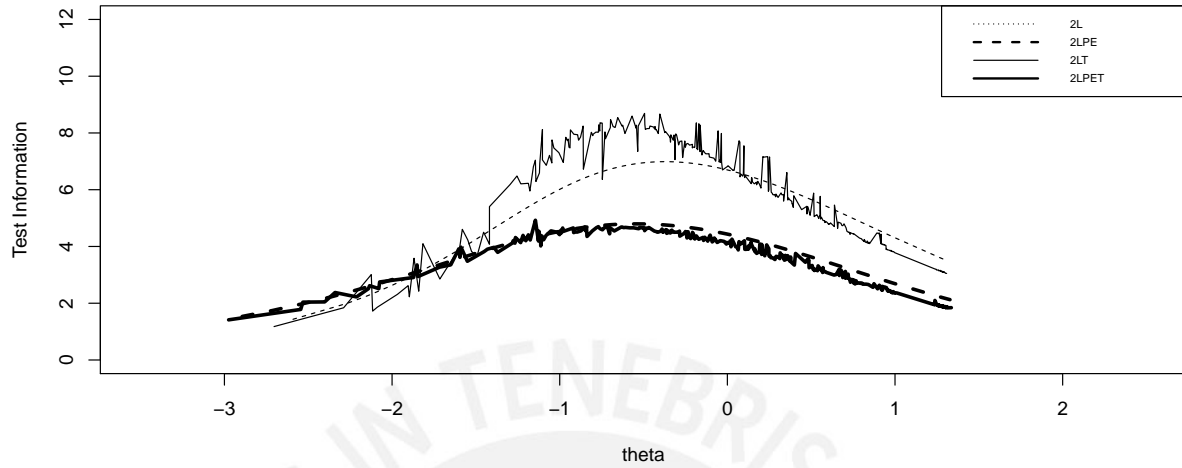
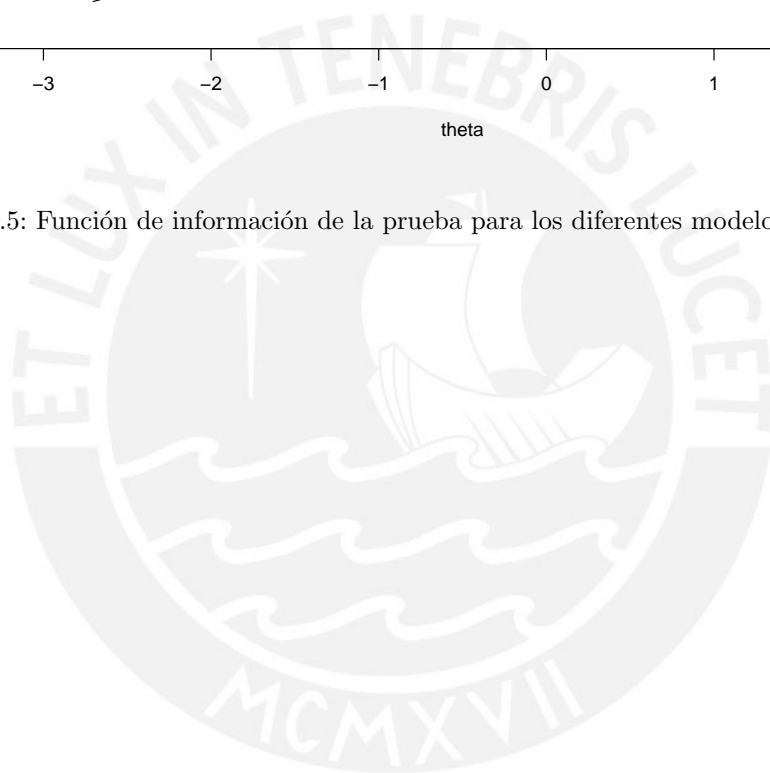


Figura 6.5: Función de información de la prueba para los diferentes modelos planteados



6.8. Resultados del modelo 2LPET

En la figura 6.6 y el cuadro se muestra estadísticas descriptivas de la medida latente estimada.

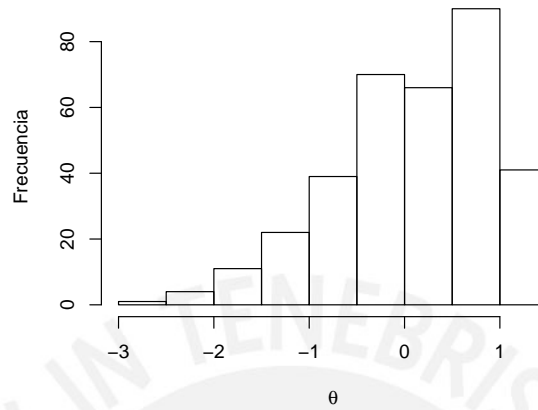


Figura 6.6: Medida de habilidad (θ) estimada por el modelo 2LPET

Medida	valor
media	0.10
mediana	0.19
des. est.	0.85
asimetría	-0.57
kurtosis	-0.19

Cuadro 6.11: Estadísticos de la medida de habilidad (θ) estimada por el modelo 2LPET

En la figura (6.7) los ítems 18, 11 y 4 son los ítems más difíciles en contraste con los ítems 10, 22, 19, 2 y 1 que son los más fáciles. Además los ítems con mayor discriminación son el 8, 19, 21, 22 y 2, y los menos discriminativos son el ítem 16, 18 y 20.

Encontramos de acuerdo a la figura 6.8 y 6.9, que la mayoría de los ítems de la prueba presentan parámetro de penalidad $d < 1$, lo que de acuerdo a Samejima la CCI para estos ítems asume valores más altos que la CCI simétrica para todo el rango de θ , es decir alumnos con habilidades muy bajas tienen alta probabilidad de pasar el ítem.

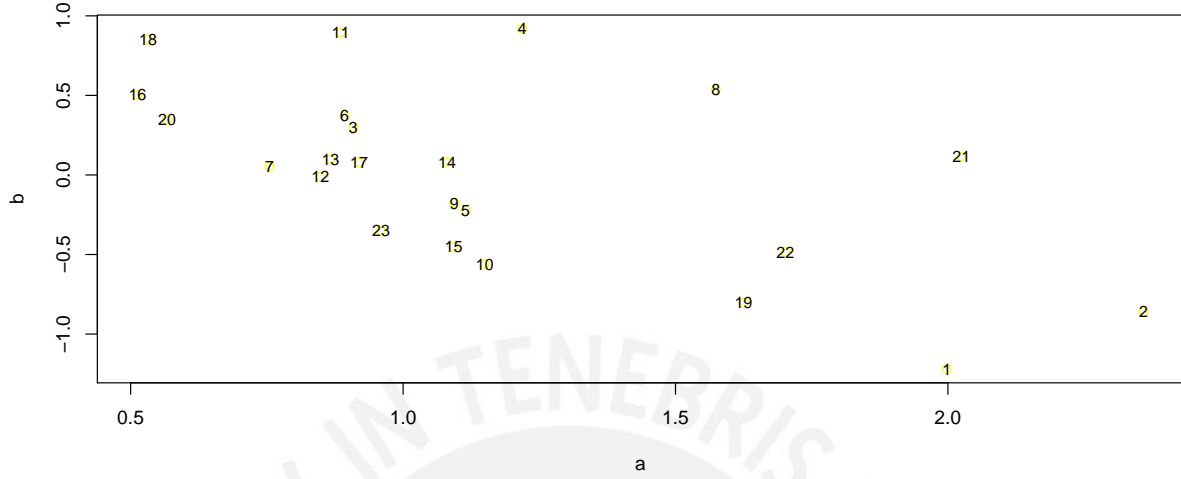


Figura 6.7: Discriminación y dificultad de los ítems, en el modelo 2LPET

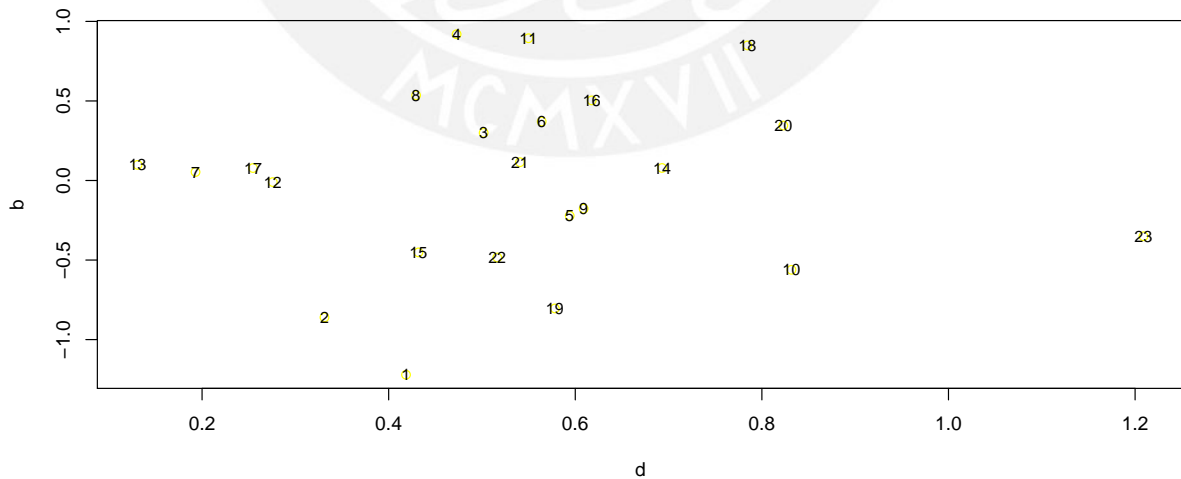


Figura 6.8: Dificultad y penalidad de los ítems en el modelo 2LPET

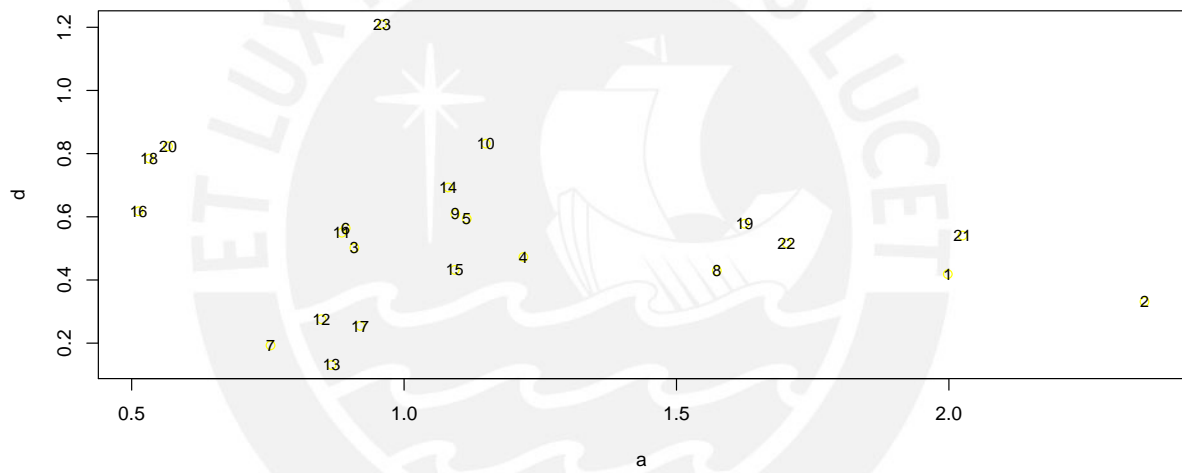


Figura 6.9: Penalidad y dificultad de los ítems en el modelo 2LPET

Capítulo 7

Conclusiones

7.1. Conclusiones

- En el presente trabajo de tesis se desarrolla el modelo TRI logístico unidimensional de respuesta binaria que considera enlaces simétricos y el modelo TRI logístico de exponente positivo unidimensional de respuesta binaria que considera enlaces asimétricos.
- Se presenta el modelo Testlet logístico que consideran un enlace simétrico y tienen una parametrización adicional en el predictor lineal latente para recoger el efecto testlet.
- Se propone e implementa el modelo Testlet logístico de exponente positivo que es una extensión del modelo LPE y considera enlaces asimétricos.
- La estimación de los modelos se realizó mediante la inferencia bayesiana, usando el software Openbugs, debido a la facilidad en su implementación, no obstante el procedimiento de estimación es relativamente demorado.
- Se realizaron tres estudios de simulación: (a) para medir el efecto de ignorar testlet en presencia de éste, (b) para medir el efecto de ignorar asimetría cuando se presentan datos asimétricos y (c) para medir el efecto conjunto de tener testlets y asimetría lo que permite concluir:
 - Cuando los datos presentan testlets el mejor modelo es el modelo Testlet logístico antes que el modelo TRI logístico.
 - Cuando los datos presentan asimetría el mejor modelo es el modelo TRI logístico de exponente positivo, especialmente en los ítems asimétricos.
 - Cuando los datos presentan testlets y asimetría es mejor el modelo Testlet logístico de exponente positivo antes que el modelo TRI logístico de exponente positivo, pero no para los ítems en testlets.
- La aplicación fue realizada con datos del Ministerio de Educación, específicamente los resultados de la ECE 2010 dirigido a estudiantes de segundo grado de primaria en un conjunto de escuelas de Lima metropolitana. De los resultados obtenidos se concluye que:
 - Los datos consiguen un mejor ajuste con los modelos Testlets que los modelos sin considerar dicho efecto y

- Los modelos TRI sobreestiman la medida de habilidad respecto a los modelos Testlets
- La información de la prueba es mejor distribuida por el modelo propuesto 2LPET.
- La estimación de los parámetros nos permite concluir que los ítems más difíciles son 18, 11 y 4; los más fáciles son el 10, 22, 19, 2 y 1. Los que más discriminan son 8, 19, 21, 22 y 2 y los menos discriminativos son el ítem 16, 18 y 20. La mayoría de los ítems presentan asimetría con parámetro de penalidad $d < 1$.

7.2. Sugerencias para investigaciones futuras

- Los datos de la aplicación corresponden a una distribución de puntajes con asimetría negativa (mayor presencia de puntajes altos), se recomienda explorar el uso del modelo Testlet logístico de exponente positivo para este tipo de poblaciones.
- Explorar el uso de los modelos presentados con datos educacionales para promover el reconocimiento de la importancia del efecto testlet tanto en la elaboración de pruebas pero sobre todo en la estimación de los niveles de habilidad de los estudiantes.
- A partir de la metodología presentada en este estudio es posible proponer una extensión testlets del modelo de exponente positivo recíproco (RLPE en inglés) propuesto por [Bazan y Bolfarine \(2010\)](#) sustituyendo en la expresión 3.23 $p_{ij} = \Psi(m_{ij})^{d_j}$ por $p_{ij} = 1 - \Psi(-m_{ij})^{d_j}$ en la sección 3.5 o equivalentemente ajustando el modelo 2LPET a los datos donde los aciertos son categorizados como 0 y los desaciertos como 1.
- Muchas respuestas en evaluaciones contienen datos politómicos, nominales u ordinales por lo que podrían estudiarse modelos para este tipo de respuestas, CCI's asimétricas.
- Buscar alternativas computacionales para superar la limitación del tiempo de estimación.

Apéndice A

Markov Chain Monte Carlo Methods (MCMC)

Los métodos MCMC fueron introducidos en 1953 en el área de la ciencia física, en una versión simplificada por Metropolis y sus asociados. Posteriores publicaciones incluyen la generalización del algoritmo de Metropolis por Hastings (1970) y el desarrollo de Gibbs sampler por Geman and Geman (1984). Luego fue redescubierto 35 años más tarde, por Tanner and Wong, (1987) y es una de las herramientas principales en la inferencia estadística moderna. Permite a los investigadores cuantitativos estimar la distribución posterior, de modelos muy complicados, con precisión. Zoufras (2009).

Las técnicas MCMC están basadas en la construcción de cadenas de Markov que eventualmente convergen a la distribución objetivo, denominada estacionaria o de equilibrio, que en el análisis bayesiano es la distribución a posteriori $f(\theta|y)$.

Los métodos MCMC más populares son:

- Metropolis-Hastings (Metropolis et al., 1953; Hastings, 1970)
- Gibbs sampling (Geman and Geman 1984)

A.1. Metropolis-Hastings Sampling

El algoritmo Metropolis-Hastings fue formulado inicialmente por Metropolis (1953), introduciendo los métodos de simulación basados en cadenas de Markov. Luego Hastings (1970) generalizó la idea original en lo que se conoce ahora como algoritmo Metropolis-Hastings (M-H).

Dado una distribución objetivo $f(x)$ para la que deseamos generar una muestra de tamaño T , el algoritmo M-H se resume en los siguientes pasos, donde $x^{(t)}$ es el vector de valores generados en la iteración t .

1. Dar valores iniciales a $x^{(0)}$.
2. Para $t=1, \dots, T$ repetir los siguientes pasos:
 - a. Generar el valor de un nuevo candidato x' con distribución propuesta $q(x \rightarrow x') = q(\frac{x'}{x})$
 - b. Calcular $\alpha = \min(1, \frac{f(x')q(x|x')}{f(x)q(x'|x)})$
 - c. Actualizar $x^{(t)} = x'$ con probabilidad α y $x^{(t)} = x = x^{(t-1)}$ con probabilidad $1-\alpha$

El algoritmo M-H converge a su distribución de equilibrio al margen de la distribución q elegida. Sin embargo, en la práctica la elección es importante dado que elecciones pobres demoran la convergencia a la distribución de equilibrio.

Casos especiales del algoritmo M-H son Metropolis de camino aleatorio, muestreo independiente, de componente única y Muestreador de Gibbs.

A.2. Muestreador de Gibbs (MGibbs)

El muestreador de Gibbs (Gibbs sampler) fue introducido por Geman and Geman (1984). Es un caso especial del algoritmo single-component M-H usando como densidad propuesta $q(\theta'|\theta^{(t)})$ a la distribución a posteriori condicional completa $f(\theta_j|\theta_j, y)$, donde $\theta_j = (\theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_d)^T$. Tales distribuciones propuestas tienen una probabilidad de aceptación $\alpha = 1$, y por tanto la propuesta es aceptada en todas las iteraciones. A pesar de que MGibbs es un caso especial del algoritmo M-H, es usualmente citado como una técnica de simulación separada por su popularidad y conveniencia. Una ventaja de MGibbs es que en cada paso, se deben generar valores aleatorios de distribuciones unidimensionales, frecuentemente estas distribuciones condicionales tienen una forma conocida y así se pueden generar fácilmente números aleatorios usando funciones estándar en softwares estadísticos y computacionales. MGibbs toma un nuevo valor en cada iteración y lo más importante es que no requiere la especificación de una distribución propuesta. Sin embargo, MGibbs puede ser inefectivo cuando el espacio paramétrico es complicado o los parámetros están altamente correlacionados. El algoritmo puede ser resumido de la siguiente forma:

1. Dar valores iniciales a $x^{(0)}$.
2. Para $t=1, \dots, T$ repetir los siguientes pasos:
 - a. Fijar $\theta = \theta^{(t-1)}$
 - b. Para $j = 1, \dots, d$ actualizar θ_j de $\theta_j \sim f(\theta_j|\theta_j, y)$
 - c. $\theta^{(t)} = \theta$ y guardar esto como un conjunto de valores generados en la iteración $t + 1$.

Apéndice B

Programas en OpenBUGS y R

Este primer anexo contiene los códigos de los modelos desarrollados en Openbugs.

B.1. Modelo TRI Logístico de dos parámetros (2L)

$$\begin{array}{l}
 a_j \sim \text{dlnorm}(0, 1) \\
 aa_j = a_j / (\text{prod}(a_{\dots}))^{1/I} \\
 b_j \sim \text{dnorm}(0, 1) \\
 bb_j = b_j - \text{mean}(b_{\dots})
 \end{array}
 \left. \vphantom{\begin{array}{l} a_j \\ aa_j \\ b_j \\ bb_j \end{array}} \right\} 1 \leq j \leq I$$

$$\begin{array}{l}
 m_{i,j} = a_j \cdot (\theta_i - b_j) \\
 p_{i,j} = e^{m_{i,j}} / (1 + e^{m_{i,j}}) \\
 y_{i,j} \sim \text{dbern}(p_{i,j})
 \end{array}
 \left. \vphantom{\begin{array}{l} m_{i,j} \\ p_{i,j} \\ y_{i,j} \end{array}} \right\} 1 \leq j \leq I$$

$$\theta_i \sim \text{dnorm}(0, 1) \left. \vphantom{\theta_i} \right\} 1 \leq i \leq N$$

B.2. Modelo Testlet Logístico de dos parámetros (2LT)

$$\begin{array}{l}
 a_j \sim \text{dlnorm}(0, 1) \\
 aa_j = a_j / (\text{prod}(a_{\dots}))^{1/I} \\
 b_j \sim \text{dnorm}(0, 1) \\
 bb_j = b_j - \text{mean}(b_{\dots})
 \end{array}
 \left. \vphantom{\begin{array}{l} a_j \\ aa_j \\ b_j \\ bb_j \end{array}} \right\} 1 \leq j \leq I$$

$$\begin{array}{l}
 y_{i,j} \sim \text{dbern}(p_{i,j}) \\
 m_{i,j} = a_j \cdot (\theta_i - b_j - g_{i,h_j}) \\
 p_{i,j} = e^{m_{i,j}} / (1 + e^{m_{i,j}})
 \end{array}
 \left. \vphantom{\begin{array}{l} y_{i,j} \\ m_{i,j} \\ p_{i,j} \end{array}} \right\} 1 \leq j \leq I$$

$$\theta_i \sim \text{dnorm}(0, 1) \left. \vphantom{\theta_i} \right\} 1 \leq i \leq N$$

$$\begin{array}{l}
 g_{i,k} \sim \text{dnorm}(0, p_{ig,k}) \\
 gg_{i,k} = g_{i,k} - \text{mean}(g_{\dots,k}) \\
 g_{i,T+1} = 0
 \end{array}
 \left. \vphantom{\begin{array}{l} g_{i,k} \\ gg_{i,k} \\ g_{i,T+1} \end{array}} \right\} 1 \leq k \leq T$$

$$\begin{array}{l}
 p_{g,k} \sim \text{dgamma}(3, 1) \\
 p_{ig,k} = 1/p_{g,k}
 \end{array}
 \left. \vphantom{\begin{array}{l} p_{g,k} \\ p_{ig,k} \end{array}} \right\} 1 \leq k \leq T$$

B.3. Modelo TRI Logístico de Exponente Positivo (LPE)

$$\left. \begin{aligned}
 y_{i,j} &\sim \text{dbern}(p_{i,j}) \\
 m_{i,j} &= a_j \cdot (\theta_i - b_j) \\
 \text{logist}_{i,j} &= e^{m_{i,j}} / (1 + e^{m_{i,j}}) \\
 p_{i,j} &= \text{logist}_{i,j}^{\text{dd}_j}
 \end{aligned} \right\} 1 \leq j \leq I \left. \vphantom{\begin{aligned} y_{i,j} \\ m_{i,j} \\ \text{logist}_{i,j} \\ p_{i,j} \end{aligned}} \right\} 1 \leq i \leq N$$

$$\left. \begin{aligned}
 \theta_i &\sim \text{dnorm}(0, 1) \\
 a_j &\sim \text{dlnorm}(0, 1) \\
 \text{aa}_j &= a_j / (\text{prod}(a_{\dots}))^{1/I} \\
 b_j &\sim \text{dnorm}(0, 1) \\
 \text{bb}_j &= b_j - \text{mean}(b_{\dots}) \\
 d_j &\sim \text{dnorm}(0, 2) \\
 \text{dd}_j &= e^{d_j}
 \end{aligned} \right\} 1 \leq j \leq I$$

B.4. Modelo Testlet Logístico de Exponente Positivo (LPET)

$$\left. \begin{aligned}
 a_j &\sim \text{dlnorm}(0, 1) \\
 \text{aa}_j &= a_j / (\text{prod}(a_{\dots}))^{1/I} \\
 b_j &\sim \text{dnorm}(0, 1) \\
 \text{bb}_j &= b_j - \text{mean}(b_{\dots}) \\
 d_j &\sim \text{dnorm}(0, 2) \\
 \text{dd}_j &= e^{d_j}
 \end{aligned} \right\} 1 \leq j \leq I$$

$$\left. \begin{aligned}
 y_{i,j} &\sim \text{dbern}(p_{i,j}) \\
 m_{i,j} &= a_j \cdot (\theta_i - b_j - g_{i,h_j}) \\
 \text{logist}_{i,j} &= e^{m_{i,j}} / (1 + e^{m_{i,j}}) \\
 p_{i,j} &= \text{logist}_{i,j}^{\text{dd}_j}
 \end{aligned} \right\} 1 \leq j \leq I \left. \vphantom{\begin{aligned} y_{i,j} \\ m_{i,j} \\ \text{logist}_{i,j} \\ p_{i,j} \end{aligned}} \right\} 1 \leq i \leq N$$

$$\left. \begin{aligned}
 \theta_i &\sim \text{dnorm}(0, 1) \\
 g_{i,k} &\sim \text{dnorm}(0, p_{ig,k}) \\
 \text{gg}_{i,k} &= g_{i,k} - \text{mean}(g_{\dots,k}) \\
 g_{i,T+1} &= 0
 \end{aligned} \right\} 1 \leq k \leq T$$

$$\left. \begin{aligned}
 p_{g,k} &\sim \text{dgamma}(3, 1) \\
 p_{ig,k} &= 1/p_{g,k}
 \end{aligned} \right\} 1 \leq k \leq T$$

B.5. Código en R para el proceso de simulación - Estudio 1

```

# Programa que simula datos Testlet y ajusta modelos 2LT y 2L
#####

# Librerías básicas

library(ltm)
    
```

```

library(msm)
require(coda)
require(lattice)
require(BRugs)
library(R2WinBUGS)

m<-20 # numero de simulaciones

# GENERACION DEL CONJUNTO DE DATOS
#####

genera2Ltest<- function(a,b,theta,g,h)
{
n<-length(theta)
k<-ncol(a)
m<-matrix(,n,k)
p<-matrix(,n,k)
data2ltest<-matrix(,n,k)

for(i in 1:n){
for(j in 1:k){
m[i,j]<-a[,j]*(theta[i]-b[,j]-g[i,h[j]])
p[i,j]<-plogis(m[i,j])
}
}
u<-matrix(runif(n*k),n)
data2Ltest<-ifelse(u<p,1,0)
return(data2Ltest)
}

n<-200 # Número de sujetos
k<-20 # Número de ítems
t<-4 # Número de testlets
s.a<-sqrt(0.5)
s.g<-sqrt(0.5)

a<-matrix(rtnorm(k,mean=1,sd=s.a,lower=0,upper=Inf),,k)
b<-matrix(rnorm(k,0,1),,k)
theta<-matrix(rnorm(n,0,1),n,)

#Restricciones.
b<-b-mean(b)

```



```

den<-1
for(de in 1:k)
{den<-den*a[1,de]}
a<-a/(den^(1/k))

a<-round(a,3)
b<-round(b,3)

g<-matrix(0,n,t+1)
hh<-matrix(c(5,1,1,1,2,2,2,3,3,3,3,4,4,4,4,5,5,5,5,5),k,)

for (j in 1:t)
{g[,j]<-rnorm(n,0,s.g)
ma<-mean(g[,j])
  for (i in 1:n)
  {g[i,j]<-g[i,j] - ma}
}
colMeans(g)
res<-matrix(0,m,6)
colnames(res)<-c(
"DIC2LT",
"pD2LT",
"time2LT",
"DIC2L",
"ph2L",
"time2L")
a_fit<-matrix(0,k,m)
b_fit<-matrix(0,k,m)
theta_fit<-matrix(0,n,m)
g1_fit<-matrix(0,n,m)
g2_fit<-matrix(0,n,m)
g3_fit<-matrix(0,n,m)
g4_fit<-matrix(0,n,m)
a2_fit<-matrix(0,k,m)
b2_fit<-matrix(0,k,m)
theta2_fit<-matrix(0,n,m)

for( l in 1:m){
data2LT<-genera2Ltest(a,b,theta,g,hh)
y=dput(data2LT)
d=c(5,1,1,1,2,2,2,3,3,3,3,4,4,4,4,5,5,5,5,5)

```

```

data2LT <- list(d=d,I=k,N=n,T=t,y=y)
data.inits <- function(){
  list(a=rep(1,k), b=rep(0,k), theta=rep(0,n), g=structure(.Data=rep(0,n*4),.Dim=c(200,
bugs.data(data2LT,dir="C:/SIMULACION/Testlet")
tiempo2LT<-system.time(
model.sim<-bugs(
  data="C:/SIMULACION/Testlet/data.txt",
  inits=data.inits,
  model.file="C:/SIMULACION/Testlet/2LTmodel.bug",
  parameters=c("aa","bb","theta","gg"),
  debug=F,
  n.chains=2,
  digits=5,
  n.iter=15000,
  n.burnin=1000,
  n.thin=5,
  program="openbugs"
))
a_fit[,1]<-model.sim$mean$aa
b_fit[,1]<-model.sim$mean$bb
theta_fit[,1]<-model.sim$mean$theta
g1_fit[,1]<-model.sim$mean$gg[,1]
g2_fit[,1]<-model.sim$mean$gg[,2]
g3_fit[,1]<-model.sim$mean$gg[,3]
g4_fit[,1]<-model.sim$mean$gg[,4]
tlt<-tiempo2LT[3]
DIC<-model.sim$DIC
pD<-model.sim$pD

data_2L <- list(I=k,N=n,y=y)
bugs.data(data_2L,dir="C:/SIMULACION/Testlet")
data.inits <- function(){
  list(a=rep(1,k), b=rep(0,k), theta=rep(0,n))}
tiempo2L<-system.time(
model.sim2<-bugs(
  data="C:/SIMULACION/Testlet/data.txt",
  inits=data.inits,
  model.file="C:/SIMULACION/Testlet/2Lmodel19.bug",
  parameters=c("aa","bb","theta"),
  debug=F,
  n.chains=2,
  digits=5,

```

```

      n.iter=15000,
      n.burnin=1000,
      n.thin=5,
      program="openbugs"
    ))
a2_fit[,1]<-model.sim2$mean$aa
b2_fit[,1]<-model.sim2$mean$bb
theta2_fit[,1]<-model.sim2$mean$theta
DIC2<-model.sim2$DIC
pD2<-model.sim2$pD
res[1,]<-c(
DIC,
pD,
tlt,
DIC2,
pD2,
tiempo2L[3])
}
write.table(res, file = "C:/SIMULACION/Testlet/salida8.txt")

#calculando ecm para estimacion 2LT
#####
ecm_a<-matrix(0,k)
ecm_b<-matrix(0,k)
ecm_theta<-matrix(0,n)
ecm_g1<-matrix(0,n)
ecm_g2<-matrix(0,n)
ecm_g3<-matrix(0,n)
ecm_g4<-matrix(0,n)

for(j in 1:m)
{
ecm_a<-ecm_a+(a_fit[,j]-t(a))^2
ecm_b<-ecm_b+(b_fit[,j]-t(b))^2
ecm_theta<-ecm_theta+(theta_fit[,j]-theta)^2
ecm_g1<-ecm_g1+(g1_fit[,j]-g[,1])^2
ecm_g2<-ecm_g2+(g2_fit[,j]-g[,2])^2
ecm_g3<-ecm_g3+(g3_fit[,j]-g[,3])^2
ecm_g4<-ecm_g4+(g4_fit[,j]-g[,4])^2
}

ecm_a<-ecm_a/m

```

```

ecm_b<-ecm_b/m
ecm_theta<-ecm_theta/m
ecm_g1<-ecm_g1/m
ecm_g2<-ecm_g2/m
ecm_g3<-ecm_g3/m
ecm_g4<-ecm_g4/m

#Calculando RMSE 2LT
#####
rmse_a<-sqrt(ecm_a)
rmse_b<-sqrt(ecm_b)
rmse_theta<-sqrt(ecm_theta)
rmse_g1<-sqrt(ecm_g1)
rmse_g2<-sqrt(ecm_g2)
rmse_g3<-sqrt(ecm_g3)
rmse_g4<-sqrt(ecm_g4)

#Calculando MAE 2LT
#####
mae_a<-matrix(0,k)
mae_b<-matrix(0,k)
mae_theta<-matrix(0,n)
mae_g1<-matrix(0,n)
mae_g2<-matrix(0,n)
mae_g3<-matrix(0,n)
mae_g4<-matrix(0,n)

for(j in 1:m)
{
mae_a<-mae_a+abs(a_fit[,j]-t(a))
mae_b<-mae_b+abs(b_fit[,j]-t(b))
mae_theta<-mae_theta+abs(theta_fit[,j]-theta)
mae_g1<-mae_g1+abs(g1_fit[,j]-g[,1])
mae_g2<-mae_g2+abs(g2_fit[,j]-g[,2])
mae_g3<-mae_g3+abs(g3_fit[,j]-g[,3])
mae_g4<-mae_g4+abs(g4_fit[,j]-g[,4])
}
mae_a<-mae_a/m
mae_b<-mae_b/m
mae_theta<-mae_theta/m
mae_g1<-mae_g1/m
mae_g2<-mae_g2/m

```

```

mae_g3<-mae_g3/m
mae_g4<-mae_g4/m

#calculando ecm para estimacion 2L
#####
ecm2_a<-matrix(0,k)
ecm2_b<-matrix(0,k)
ecm2_theta<-matrix(0,n)
for(j in 1:m)
{
ecm2_a<-ecm2_a+(a2_fit[,j]-t(a))^2
ecm2_b<-ecm2_b+(b2_fit[,j]-t(b))^2
ecm2_theta<-ecm2_theta+(theta2_fit[,j]-theta)^2
}
ecm2_a<-ecm2_a/m
ecm2_b<-ecm2_b/m
ecm2_theta<-ecm2_theta/m

#Calculando RMSE 2L
#####
rmse2_a<-sqrt(ecm2_a)
rmse2_b<-sqrt(ecm2_b)
rmse2_theta<-sqrt(ecm2_theta)

#Calculando MAE 2L
#####
mae2_a<-matrix(0,k)
mae2_b<-matrix(0,k)
mae2_theta<-matrix(0,n)

for(j in 1:m)
{
mae2_a<-mae2_a+abs(a2_fit[,j]-t(a))
mae2_b<-mae2_b+abs(b2_fit[,j]-t(b))
mae2_theta<-mae2_theta+abs(theta2_fit[,j]-theta)
}
mae2_a<-mae2_a/m
mae2_b<-mae2_b/m
mae2_theta<-mae2_theta/m

```

B.6. Código en R para el proceso de simulación - Estudio 2

```
# Programa que simula datos Testlet LPE y ajusta modelos 2LPE Testlet y 2LPE
```

```
#####
```

```
library(msm)
require(coda)
require(lattice)
require(BRugs)
library(R2WinBUGS)

setwd("C:/SIMULACION/LPE")
m<-20 # numero de simulaciones

genera2LPE<- function(a,b,theta,d)
{
n<-length(theta)
k<-ncol(a)
m<-matrix(,n,k)
p<-matrix(,n,k)
datag2lpe<-matrix(,n,k)

for(i in 1:n){
for(j in 1:k){
m[i,j]<-a[,j]*(theta[i]-b[,j])
p[i,j]<-(plogis(m[i,j]))^d[,j]
}
}
u<-matrix(runif(n*k),n)
datag2lpe<-ifelse(u<p,1,0)
return(datag2lpe)
}

n<-200 # Número de sujetos
k<-20 # Número de ítems

a<-matrix(rtnorm(k,mean=1,sd=sqrt(0.5),lower=0,upper=Inf),,k)
b<-matrix(rnorm(k,0,1),,k)
theta<-matrix(rnorm(n,0,1),n,)
d<-matrix(c(0.1, 0.2, 0.3, 0.3, 0.3, 0.4, 0.7, 0.8, 1.0, 1.0, 1.0, 1.0,
1.0, 1.0, 1.0, 1.1, 1.4, 1.5, 1.8, 3.0),,k)

#Restricciones.
b<-b-mean(b)
den<-1
```

```

for(de in 1:k)
{den<-den*a[1,de]}
a<-a/(den^(1/k))

a<-round(a,3)
b<-round(b,2)

res<-matrix(0,m,6)
colnames(res)<-c(
"DIC2LT",
"pD2LT",
"time2LT",
"DIC2L",
"pD2L",
"time2L"
)

a_fit<-matrix(0,k,m)
b_fit<-matrix(0,k,m)
d_fit<-matrix(0,k,m)
theta_fit<-matrix(0,n,m)

a2_fit<-matrix(0,k,m)
b2_fit<-matrix(0,k,m)
theta2_fit<-matrix(0,n,m)

#Inicia iteraciones
for(l in 1:m){
data2LPEa<-genera2LPE(a,b,theta,d)
y=dput(data2LPEa)
data2lpe <- list(I=k,N=n,y=y)

data.inits <- function(){
  list(a=rep(1,k), b=rep(0,k), d=rep(1,k), theta=rep(0,n) )}

bugs.data(data2lpe,dir="C:/SIMULACION/LPE")

tiempo2lpe<-system.time(
model.sim<-bugs(
  data="C:/SIMULACION/LPE/data.txt",
  inits=data.inits,
  model.file="C:/SIMULACION/LPE/2LPEmodel19.txt",

```

```

parameters=c("aa","bb","d","theta"),
debug=F,
n.chains=2,
digits=5,
n.iter=15000,
n.burnin=1000,
n.thin=5,
program="openbugs"
))

a_fit[,1]<-model.sim$mean$aa
b_fit[,1]<-model.sim$mean$bb
d_fit[,1]<-model.sim$mean$d
theta_fit[,1]<-model.sim$mean$theta
tlt<-tiempo2lpe[3]
DIC<-model.sim$DIC
pD<-model.sim$pD

data2l <- list(I=k,N=n,y=y)
data.inits <- function(){
  list(a=rep(1,k), b=rep(0,k), theta=rep(0,n))}
bugs.data(data2l,dir="C:/SIMULACION/LPE")

tiempo2L<-system.time(
model.sim2<-bugs(
  data="C:/SIMULACION/LPE/data.txt",
  inits=data.inits,
  model.file="C:/SIMULACION/LPE/2Lmodel9.bug",
  parameters=c("aa","bb","theta"),
  debug=F,
  n.chains=2,
  digits=5,
  n.iter=15000,
  n.burnin=1000,
  n.thin=5,
  program="openbugs"
))
a2_fit[,1]<-model.sim2$mean$aa
b2_fit[,1]<-model.sim2$mean$bb
theta2_fit[,1]<-model.sim2$mean$theta

```



```

DIC2<-model.sim2$DIC
pD2<-model.sim2$pD

res[1,]<-c(
DIC,
pD,
tlt,
DIC2,
pD2,
tiempo2L[3])
} #Termina Iteraciones
write.table(res, file = "C:/SIMULACION/LPE/salida12.txt")

```

B.7. Código en R para el proceso de simulación - Estudio 3

```

# Programa que simula datos Testlet LPE y ajusta modelos 2LPE Testlet y 2LPE
#####

```

```

library(msm)
library(ltm)
require(coda)
require(lattice)
require(BRugs)
library(R2WinBUGS)
setwd("C:/SIMULACION/TestletLPE")

m<-20 # numero de simulaciones
genera2LPEtest<- function(a,b,theta,g,h,d)
{
n<-length(theta)
k<-ncol(a)
m<-matrix(,n,k)
p<-matrix(,n,k)
data2ltest<-matrix(,n,k)

for(i in 1:n){
for(j in 1:k){
m[i,j]<-a[,j]*(theta[i]-b[,j]-g[i,h[j]])
p[i,j]<-(plogis(m[i,j]))^d[,j]
}
}
u<-matrix(runif(n*k),n)
data2Ltest<-ifelse(u<p,1,0)
return(data2Ltest)
}

```

```

}

n<-200 # Número de sujetos
k<-20 # Número de ítems
t<-4 # Número de testlets
s.a<-sqrt(0.5)
s.g<-sqrt(0.5)

a<-matrix(rtnorm(k,mean=1,sd=s.a,lower=0,upper=Inf),,k)
b<-matrix(rnorm(k,0,1),,k)
theta<-matrix(rnorm(n,0,1),n,)
d<-matrix(c(0.1, 0.2, 1, 0.3, 0.3, 0.4, 1.5, 0.8, 1.0, 1.0, 1.1, 1.4,
0.7, 1.8, 3.0, 1.0, 1.0, 0.3, 1.0, 1.0),,k)

#Restricciones.
b<-b-mean(b)
den<-1
for(de in 1:k)
{den<-den*a[1,de]}
a<-a/(den^(1/k))

a<-round(a,3)
b<-round(b,3)
g<-matrix(0,n,t+1)
hh<-matrix(c(5,1,1,1,2,2,2,3,3,3,3,4,4,4,4,5,5,5,5,5),k,)

for (j in 1:t){
g[,j]<-rnorm(n,0,s.g)
ma<-mean(g[,j])
for (i in 1:n){g[i,j]<-g[i,j] - ma}
}
colMeans(g)

res<-matrix(0,m,6)
colnames(res)<-c(
"DIC2LPET",
"ph2LPET",
"time2LPET",
"DIC2LPE",
"ph2LPE",
"time2LPE"
)

```

```

a_fit<-matrix(0,k,m)
b_fit<-matrix(0,k,m)
d_fit<-matrix(0,k,m)
theta_fit<-matrix(0,n,m)
g1_fit<-matrix(0,n,m)
g2_fit<-matrix(0,n,m)
g3_fit<-matrix(0,n,m)
g4_fit<-matrix(0,n,m)

a2_fit<-matrix(0,k,m)
b2_fit<-matrix(0,k,m)
d2_fit<-matrix(0,k,m)
theta2_fit<-matrix(0,n,m)

#Inicia iteraciones
for(l in 1:m){
data2LT<-genera2LPEtest(a,b,theta,g,hh,d)
y=dput(data2LT)
h=c(5,1,1,1,1,2,2,2,3,3,3,3,3,4,4,4,4,5,5,5)
data2lpetest <- list(h=h,I=k,N=n,T=t,y=y)

data.inits <- function(){
  list(a=rep(1,k), b=rep(0,k), d=rep(1,k), theta=rep(0,n),g=structure(.Data=rep(0,n*4),
bugs.data(data2lpetest,dir="C:/SIMULACION/TestletLPE") # No modificar

tiempo2lpet<-system.time(
model.sim<-bugs(
  data="C:/SIMULACION/TestletLPE/data.txt",
  inits=data.inits,
  model.file="C:/SIMULACION/TestletLPE/2LPEmodel1.bug",
  parameters=c("aa","bb","d","theta","gg"),
  debug=F,
  n.chains=2,
  digits=5,
  n.iter=15000,
  n.burnin=1000,
  n.thin=5,
  program="openbugs"
))

a_fit[,l]<-model.sim$mean$aa

```

```

b_fit[,1]<-model.sim$mean$bb
d_fit[,1]<-model.sim$mean$d
theta_fit[,1]<-model.sim$mean$theta
g1_fit[,1]<-model.sim$mean$gg[,1]
g2_fit[,1]<-model.sim$mean$gg[,2]
g3_fit[,1]<-model.sim$mean$gg[,3]
g4_fit[,1]<-model.sim$mean$gg[,4]
tlpet<-tiempo2lpet[3]
DIC<-model.sim$DIC
pD<-model.sim$pD

data2lpe <- list(I=k,N=n,y=y)
data.inits <- function(){
  list(a=rep(1,k), b=rep(0,k), d=rep(1,k), theta=rep(0,n) )}
bugs.data(data2lpe,dir="C:/SIMULACION/TestletLPE")

tiempo2LPE<-system.time(
model.sim2<-bugs(
  data="C:/SIMULACION/TestletLPE/data.txt",
  inits=data.inits,
  model.file="C:/SIMULACION/TestletLPE/2LPEmodel9.txt",
  parameters=c("aa","bb","d","theta"),
  debug=F,
  n.chains=2,
  digits=5,
  n.iter=15000,
  n.burnin=1000,
  n.thin=5,
  program="openbugs"
))
a2_fit[,1]<-model.sim2$mean$aa
b2_fit[,1]<-model.sim2$mean$bb
d2_fit[,1]<-model.sim2$mean$d
theta2_fit[,1]<-model.sim2$mean$theta
tlpe<-tiempo2LPE[3]
DIC2<-model.sim2$DIC
pD2<-model.sim2$pD

res[1,]<-c(
DIC,
pD,
tlpet,

```

```

DIC2,
pD2,
tlpe)
}# Termina
write.table(res, file = "C:/SIMULACION/TestletLPE/salida3.txt")

```

B.8. Código en R para calcular la Función de Información

```

#Calculos para Función de Información
#####

#Leyendo datos
setwd("D:/.../resultados2L")
theta.data <- read.table("theta.txt", header=F, sep="\t")
a.data <- read.table("a_fit.txt", header=T, sep="\t")
b.data <- read.table("b_fit.txt", header=T, sep="\t")

#1.Funcion de informacion del item para 2L
#####
fii <- function(x,a){
  a^2* (1/(1+exp(-x)))*(1-1/(1+exp(-x))))}
theta2L <- sort(theta.data$V1)
IIF<-matrix(0,344,23)
for(i in 1:23)
{
a<-a.data$mean[i]
b<-b.data$mean[i]
y<-a*(theta2L - b)
IIF[,i]<-fii(y,a)
}
I_2L<-IIF[,1]
for(j in 2:23){
I_2L<-I_2L+IIF[,j]}

# 2.Funcion de informacion del item para 2LPE
#####

#Leyendo datos
setwd("D:/.../resultados2LPE")
theta.data <- read.table("theta_2lpe.txt", header=F, sep="\t")
a.data <- read.table("a_2lpe.txt", header=T, sep="\t")
b.data <- read.table("b_2lpe.txt", header=T, sep="\t")
d.data <- read.table("d_2lpe.txt", header=T, sep="\t")

```

```

fda <- function(x,d){
  1/(1+exp(-x))^d}
iif <- function(x,d,a){
  (a*d)^2*fda(x,d)*(1-fda(x,1))^2/(1-fda(x,d))}

theta2LPE <- sort(theta.data$V1)
IIF_2lpe<-matrix(0,344,23)
for(i in 1:23)
{
a<-a.data$mean[i]
b<-b.data$mean[i]
d<-d.data$mean[i]
y<-a*(theta2LPE - b)
IIF_2lpe[,i]<-iif(y,d,a)
}
I_2LPE<-IIF_2lpe[,1]

for(j in 2:23){
I_2LPE<-I_2LPE+IIF_2lpe[,j]}

#####3.Funcion de informacion del item para 2L Testlet
#####
#Leyendo datos
setwd("D:/.../resultados2Ltestlet")
theta.data <- read.table("2ltest_theta.txt", header=F, sep="\t")
a.data <- read.table("a_2LT.txt", header=T, sep="\t")
b.data <- read.table("b_2LT.txt", header=T, sep="\t")
g.data <- read.table("g_2LT.txt", header=T, sep="\t")

fii <- function(x,a){
  a^2* (1/(1+exp(-x)))*(1-1/(1+exp(-x))))}
theta <- matrix(theta.data$V1,344,1)

g2lt<-matrix(g.data$mean,344,5,byrow=TRUE)
g2lt<-cbind(g2lt,rep(0,344))
g2lt<-cbind(g2lt,theta)
write.table(g2lt, file = "g_thet.txt") #Ordenar datos por la variable theta.
go.data <- read.table("g_thet_o.txt", header=T, sep="\t")
g2lt_o<-data.matrix(go.data[2:8])
g2lt_o[1,7]

```

```

h=c(6,6,1,1,1,1,2,2,2,2,6,3,3,4,4,4,4,4,5,5,5,5,5)
for(i in 1:23)
{
a<-a.data$mean[i]
b<-b.data$mean[i]
y<-a*(g2lt_o[,7] - b - g2lt_o[,h[i]])
IIF_2LT[,i]<-fii(y,a)
}
I_2LT<-IIF_2LT[,1]
for(j in 2:23){
I_2LT<-I_2LT+IIF_2LT[,j]}

##### 4.Funcion de informacion del item para 2LPE Testlet
#####

#Leyendo datos
setwd("D:/.../resultados2LPE_testlet")
theta.data <- read.table("theta_2lpet.txt", header=F, sep="\t")
a.data <- read.table("a_2lpet.txt", header=T, sep="\t")
b.data <- read.table("b_2lpet.txt", header=T, sep="\t")
d.data <- read.table("d_2lpet.txt", header=T, sep="\t")
g.data <- read.table("g_2lpet.txt", header=T, sep="\t")

fda <- function(x,d){
  1/(1+exp(-x))^d}
iif <- function(x,d,a){
  (a*d)^2*fda(x,d)*(1-fda(x,1))^2/(1-fda(x,d))}

theta <- theta.data$V3
g2lpet<-matrix(g.data$mean,344,5,byrow=TRUE)
g2lpet<-cbind(g2lpet,rep(0,344))
g2lpet<-cbind(g2lpet,theta)
write.table(g2lpet, file = "glpet_thet.txt")
go.data <- read.table("glpet_thet_o.txt", header=T, sep="\t")
g2lpet_o<-data.matrix(go.data[2:8])
g2lpet_o[1,1:7]
h=c(6,6,1,1,1,1,2,2,2,2,6,3,3,4,4,4,4,4,5,5,5,5,5)

IIF_2lpet<-matrix(0,344,23)
for(i in 1:23)
{
a<-a.data$mean[i]

```

```

b<-b.data$mean[i]
d<-d.data$mean[i]
y<-a*(g2lpet_o[,7] - b - g2lpet_o[,h[i]])
IIF_2lpet[,i]<-iif(y,d,a)
}
I_2LPET<-IIF_2lpet[,1]

for(j in 2:23){
I_2LPET<-I_2LPET+IIF_2lpet[,j]}

plot(theta2L,I_2L,type='l',lty=1, xlab='theta', ylab='Información de la prueba', ylim=c(0
lines(theta2LPE,I_2LPE,lty=1,col=1,lwd=3)
lines(g2lt_o[seq(1,344,by=1),7],I_2LT[seq(1,344,by=1)],lty=2, col=1)
lines(g2lpet_o[seq(1,344,by=1),7],I_2LPET[seq(1,344,by=1)],lty=2,col=1,lwd=3)
legend("topright",legend=c("2L","2LPE","2LT","2LPET"),lty=c(1,1,3,2),
lwd=c(1,2,1,2),cex=0.6)

```



Apéndice C

Prueba de Comprensión de Textos - ECE2010





PERÚ Ministerio de Educación



N° Folio

Comprensión lectora

SEGUNDO GRADO

Cuadernillo 1
• Primer día

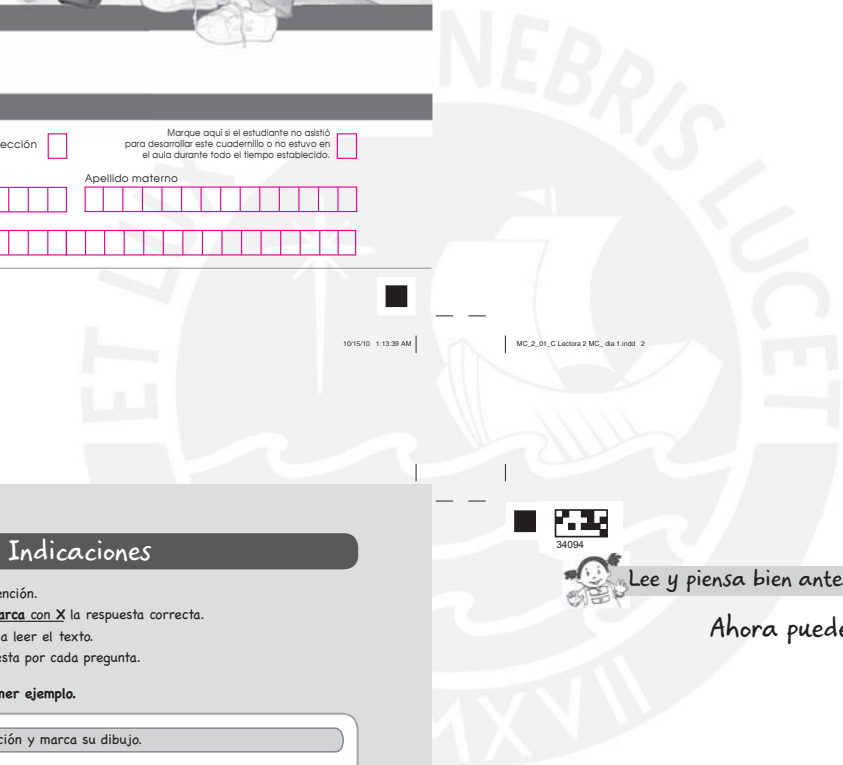
Datos de la I.E.

Datos del estudiante

Código del estudiante Sección Marque aquí si el estudiante no asistió para desarrollar este cuadernillo o no estuvo en el aula durante todo el tiempo establecido.

Apellido paterno Apellido materno

Nombres



MC_2_01_C_Lectura 2 MC_ da 1.indd 1

10/15/10 1:13:39 AM

MC_2_01_C_Lectura 2 MC_ da 1.indd 2

10/15/10 1:13:39 AM

Indicaciones

- Lee cada texto con mucha atención.
- Luego, lee cada pregunta y **marca con X** la respuesta correcta.
- Si lo necesitas, puedes volver a leer el texto.
- Solo debes marcar una respuesta por cada pregunta.

Vamos a resolver juntos el primer ejemplo.



Lee la siguiente oración y marca su dibujo.

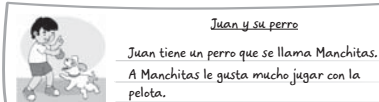
Juana lee un libro.



Ahora resuelve tú solo el segundo ejemplo.



Lee con atención el siguiente texto:



Ahora marca la respuesta correcta.

¿Cómo se llama el perro de Juan?

- a Peluchín.
b Negrito.
c Manchitas.

- ▶ Resuelve el resto del cuadernillo en silencio.
- ▶ Trabaja sin mirar los cuadernillos de tus compañeros.
- ▶ Solo podrás preguntar si tienes dudas de cómo **marcar** tus respuestas.

Espera.



34094



Lee y piensa bien antes de marcar tus respuestas.

Ahora puedes empezar.



Lee la siguiente oración y marca su dibujo.

1.

Horacio jala su oveja y Esteban carga su conejo.





Lee con atención la siguiente oración:

Mónica y Efraín irán a jugar al parque cuando Josefina les dé permiso.



Ahora marca la respuesta correcta.

2. ¿Qué harán Mónica y Efraín en el parque?

- a Comer.
- b Leer.
- c Jugar.



Pasa a la siguiente página.



Lee con atención el siguiente texto:

¡A volar nuestras cometas!

Seguramente alguna vez has volado una cometa. A continuación, te damos algunas ideas para cuando te animes a hacerlo.

Cuando salgas a volar tu cometa, asegúrate de que no haya demasiado viento, ya que podría ser difícil manejarla bien. Lo ideal es que el viento no sopla muy fuerte, para que puedas hacer piruetas en el cielo con tu cometa.



No debes volar tu cometa cerca de postes de luz, y menos si está lloviendo. Podría enredarse en el poste y atraer la electricidad.

Para volar tu cometa, debes pedirle a un amigo que la sostenga. Después, coge el hilo de la cometa y aléjate un poco, de espaldas al viento. Luego, pídele a tu amigo que suelte la cometa y corre para que se mantenga en el aire. Cuando la cometa esté volando, suelta un poco de hilo para que se eleve.

Si deseas, puedes colocarle una cola a tu cometa. La cola ayudará a que la cometa permanezca en el aire.



Ahora marca la respuesta correcta de cada pregunta.

3. ¿Para qué sirve la cola de una cometa?

- a Para permitir que la cometa vuele más alto.
- b Para que la cometa haga piruetas en el aire.
- c Para ayudar a que la cometa permanezca en el aire.

4. ¿Qué es lo primero que debes hacer para volar tu cometa?

- a Soltar un poco de hilo para que se eleve.
- b Pedirle a un amigo que sostenga tu cometa.
- c Alejarte un poco, de espaldas al viento.

5. ¿Por qué no debes volar tu cometa cerca de los postes de luz?

- a Porque podría atraer la electricidad de los postes.
- b Porque la cometa puede malograr los postes.
- c Porque la cometa puede quemarse con los focos.

6. ¿Para qué se escribió este texto?

- a Para darnos una opinión.
- b Para darnos recomendaciones.
- c Para contarnos una historia.



Pasa a la siguiente página.



Lee con atención el siguiente texto:



El algarrobo es un árbol que crece en la costa norte del Perú, especialmente en Lambayeque y Piura. Tiene un tronco grueso y pequeñas florecillas amarillas.

El algarrobo tiene unas semillas que son muy dulces. Estas semillas sirven para hacer galletas y caramelos. También sirven para preparar un jarabe muy conocido llamado algarrobina.

El algarrobo puede vivir en lugares donde no llueve. Sus profundas raíces le permiten tomar el agua subterránea.

Sin embargo, los algarrobos están en peligro de extinción. Muchos de estos árboles se talan con hachas y sierras, pero no se vuelven a sembrar. Luego son usados como leña y carbón en pollerías y supermercados.



Ahora marca la respuesta correcta de cada pregunta.

7. ¿Dónde crece el algarrobo?
- a) En el sur del Perú.
 - b) En la costa norte del Perú.
 - c) Solo en Lambayeque.
8. En el texto, ¿qué quiere decir que "muchos de estos árboles se talan"?
- a) Que están sembrando muchos árboles.
 - b) Que están quemando muchos árboles.
 - c) Que están cortando muchos árboles.
9. ¿Por qué el algarrobo puede vivir en lugares donde no llueve?
- a) Porque toma el agua subterránea con sus profundas raíces.
 - b) Porque hay personas que le echan agua todos los días.
 - c) Porque puede almacenar el agua en su grueso tronco.
10. ¿De qué trata principalmente este texto?
- a) Trata de cómo es la costa.
 - b) Trata de cómo es el algarrobo.
 - c) Trata de los usos de la algarrobina.



11. Lee la siguiente parte del texto:

El algarrobo tiene unas semillas que son muy dulces. Estas semillas sirven para hacer galletas y caramelos. También sirven para preparar un jarabe muy conocido llamado algarrobina.

¿Cuál es la idea más importante de esta parte del texto?

- a) Para qué sirven las semillas del algarrobo.
- b) Cómo se prepara el jarabe de algarrobina.
- c) Qué sabor tienen las galletas de algarrobina.

Lee con atención la siguiente nota:

Estimado profesor:
Mañana jueves, mi hija Mariana irá a la posta médica. Le toca su vacuna. El viernes, Mariana regresará a la escuela. Muchas gracias.
Antonio

Ahora marca la respuesta correcta de cada pregunta.

12. ¿Cuándo irá Mariana a la posta médica?
- a) El lunes.
 - b) El jueves.
 - c) El viernes.
13. ¿Por qué Mariana irá a la posta médica?
- a) Porque le toca su vacuna.
 - b) Porque está muy enferma.
 - c) Porque quiere visitar a Antonio.



Lee con atención el siguiente cuento:

Hace mucho tiempo, un tigrillo quería cruzar al otro lado del río, pero no podía. Entonces, le pidió ayuda a una tortuga:
—Tortuguita, por favor, déjame subir a tu espalda para que me lleves a la isla.

La tortuga cargó al tigrillo a cuestas y pasó el río.

Al poco rato, el tigrillo intentó morder a la tortuga y le dijo:

—Tortuguita, tengo muchísima hambre. Te voy a comer.

La tortuga se asustó mucho. Se metió en su duro caparazón y ya no quiso salir de él.

Pasaron los días y el tigrillo, al ver que no podía comerse a la tortuga, le dijo:

—Tortuguita, me muerdo de hambre. Por favor, regresa a la selva.

La tortuga creyó que el tigrillo ya no intentaría hacerle daño; sin embargo, el tigrillo pensaba comerse a la tortuga apenas llegaran a la selva.

La tortuga nadó llevando al tigrillo a cuestas nuevamente, pero se dio cuenta de sus intenciones.

Antes de llegar a la orilla, la tortuga se metió debajo del agua para escaparse del tigrillo, quien casi se ahoga. Cuando llegó a la orilla del río, el tigrillo comprendió que había actuado mal con la tortuga que lo había ayudado.





Ahora marca la respuesta correcta de cada pregunta.

14. ¿Cuál de estos hechos ocurrió primero en el cuento?
- a El tigrillo casi se ahoga en el río.
 - b La tortuga se metió debajo del agua.
 - c El tigrillo intentó morder a la tortuga.
15. ¿Por qué el tigrillo intentó morder a la tortuga?
- a Porque no quiso ayudarlo.
 - b Porque tenía mucha hambre.
 - c Porque intentó ahogarlo.
16. ¿Qué quiere decir que la tortuga nadó "llevando al tigrillo a cuestas"?
- a Que nadó llevándolo sobre su espalda.
 - b Que nadó llevándolo muy asustada.
 - c Que nadó llevándolo con facilidad.
17. ¿Cómo era la tortuga?
- a Era egoísta.
 - b Era abusiva.
 - c Era inteligente.
18. Este cuento nos enseña principalmente que:
- a no debemos pelearnos con nuestros amigos.
 - b nunca debemos hacer daño a los animales.
 - c debemos ser agradecidos con los que nos ayudan.



Pasa a la siguiente página. >>>



Lee con atención el siguiente texto:

Lunes 20 de abril de 2009

Un domingo para San Vicente

El domingo 19 de abril hubo una fiesta en nuestra comunidad. Era el día de San Vicente, nuestro patrono.

La fiesta patronal se inició por la mañana con una misa en la capilla, totalmente iluminada por las velas, y siguió con una procesión acompañada por la banda de músicos de San Vicente. Por la tarde, varios cantantes se presentaron en la plaza de armas de nuestra comunidad para animar el ambiente, mientras el sol brillaba en lo alto.

Por la noche, hubo fuegos artificiales. Mientras la gente disfrutaba del espectáculo, el cielo se iluminó de todos los colores. La fiesta terminó con un discurso de nuestro alcalde.

MC_2_01_C_Lectura 2 MC_1 da 1.indd 13

10/15/10 1:13:51 AM

MC_2_01_C_Lectura 2 MC_1 da 1.indd 14

10/15/10 1:13:51 AM



19192



Ahora marca la respuesta correcta de cada pregunta.

19. ¿Por qué hubo una fiesta en la comunidad?
- a Porque era el día de San Vicente.
 - b Porque era el cumpleaños del alcalde.
 - c Porque se construyó una capilla.
20. Según el texto, ¿qué sucedió por la tarde?
- a Los músicos acompañaron la procesión.
 - b Varios cantantes se presentaron en la plaza.
 - c Prendieron velas para iluminar la capilla.
21. ¿Cómo terminó la fiesta patronal de San Vicente?
- a Con un discurso del alcalde.
 - b Con una misa en la capilla.
 - c Con una procesión.
22. ¿De qué trata principalmente este texto?
- a Trata de cómo son los fuegos artificiales.
 - b Trata de la fiesta patronal de San Vicente.
 - c Trata de una misa en la capilla del pueblo.
23. ¿Para qué se escribió este texto?
- a Para darnos una opinión.
 - b Para darnos un consejo.
 - c Para darnos información.



Felicitaciones.
Has terminado.



PERÚ
Ministerio
de Educación

Secretaría
de Planificación
Estratégica

Unidad de
Medición de la Calidad
Educativa



Evaluación
Nacional de
Mejora de la
Calidad
2010

Bibliografía

- Balakrishnan, N. y Leung, M. Y. (1988). Order statistics from the type i generalized logistic distribution, *Commun. Statist. - Simula.* pp. 25–50.
- Bazan, J. y Bolfarine, H. (2010). Bayesian estimation of the logistic positive exponent irt model, *Journal of Educational and Behavioral Statistics* **35**, No. 6: 693–713.
- Bazan, J., Bolfarine, H. y Aparecida, R. (2006). Sensitivity analysis of prior specification for the probit-normal irt model: An empirical study, *Estadística* **58**, No. 170-171: 17–42.
- Bazan, J., Calderon, A. y Valdivieso, L. (2009). Modelos de teoría de respuesta al ítem bajo enfoque bayesiano. Texto del Minicurso desarrollado en el XXVII Coloquio de la Sociedad Matemática Peruana, 10-14 de agosto de 2009.
- Bazan, J., Calderon, A. y Valdivieso, L. (2010). Enfoque bayesiano en modelos de teoría de respuesta al ítem. Reporte del Minicurso desarrollado en la I Jornada Internacional de Probabilidad y Estadística, (Lima, 3-5 de febrero de 2010).
- Bazan, J. L. y Bolfarine, H. (2006). Skewed logit item response models.
- Bock, D. y Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an em algorithm, *Psychometrika* **46**, No. 4.
- Bock, D. R. y Moustaki, I. (2007). Item response theory in a general framework, *Handbook of Statistics* **26**: 469–513.
- Bradlow, E. T., Wainer, H. y Wang, X. (1999). A bayesian random effects model for testlets, *Psychometrika* **64**, No. 2: 153–168.
- Bradlow, E., Wang, X. y Wainer, H. (2002). A general bayesian model for testlets: Theory and applications. GRE Board Report No. 98-01P. This report presents the findings of a research project by and carried out under the auspices of the Graduate Record Examinations Board and Educational Testing Service. February 2002.
- Chincaro, O. A. (2010). Modelo rash dicotómico con aplicación a la educación. Tesis para optar el grado de Magister en Estadística - PUCP.
- DeCarlo, L. T. (1977). On the meaning and use of kurtosis, *Psychological Methods* **2**, No. 3: 292–307.
- Fox, J.-P. (2010). *Bayesian Item Response Modeling Theory and Applications*, Springer-Verlag New York, LLC.
- Gilks, W. R. y Wild, P. (1992). Adaptive rejection sampling for gibbs sampling, *Appl. Statist.* **41**, No. 2: 337–348.
- Holland, P. B. y Rosenbaum, P. R. (1986). Conditional association and unidimensionality in monotone latent variable models, *The Annals of Statistics* **14**, No. 4: 1523–1543.

- Johnson, N. L. y Kotz, S. (1994). *Continuous Univariate Distributions, Vol. 2*, John Wiley & Sons.
- Johnson, N. L., Kotz, S. y N., B. (1995). *Continuos Univariate Distributions* , John Wiley & Sons.
- Lord, F. M. (1952). A theory of test scores, *Psychometric Corporation* . Psychometric Monograph No. 7.
- Lord, F. M. y Novick, M. (1968). *Statistical Theories of Mental Tests Scores*, Addison-Wesley.
- McKay, C. S. (2010). Bugs code for item response theory, *Journal of Statistical Software* **36**.
- MINEDU, U. (2009). Marco de trabajo evaluacion censal de estudiantes.
- Saho, Q. (2002). Maximum likelihood estimation for generalised logistic distributions, *Communicationns in Statistics* **31**, No. 10: 1687–1700.
- Samejima, F. (1999). Usefulness of the logistic positive exponent family of models in educational measurement. Paper presented at the Annual Meeting of the American Educational Research Association (Montreal, Quebec, Canada, April 19-23. 1999).
- Spiegelhalter, D., Thomas, A., Best, N. y Gilks, W. (1996). *BUGS 0.5 Examples Volumen 1 (version i)*. MRC Biosttistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 2SR.
- Wainer, H., Bradlow, E. T. y Wang, X. (2007). *Testlet Response Theory and Its Applications*, Cambridge University Press.
- Wainer, H. y Kiely, G. L. (1987). Item clusters and computerized adaptative testing: A case for testlets, *Journal of Educational Measurement* **24**, No. 3: 185–201.
- Wang, W.-C. y Wilson, M. (2005). The rash testlet model, *Applied Psychological Measurement* **29**, No. 2: 126–149.
- Wikipedia (2011). Logistic distribution. http://en.wikipedia.org/wiki/Logistic_distribution.
- Willmott, C. J. y Matsuura, K. (2005). Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance, *Climate Research* **30**: 79–82.
- Zoufras, N. (2009). *Bayesian Modelling Using WinBugs*, John Wiley & Sons.