

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

ESCUELA DE POSGRADO



**El Modelo de Respuesta Nominal:
Aplicación a datos educacionales**

**TESIS PARA OPTAR POR EL GRADO DE MAGÍSTER EN
ESTADÍSTICA**

Presentado por:

**José Manuel Rivera Espejo
jriveraespejo@gmail.com**

**Asesor: Enver Gerald Tarazona Vargas
enver.tarazona@pucp.edu.pe**

Miembros del jurado:

**Dr. Valdivieso Serrano, Luis Hilmar
Dr. Bayes Rodriguez, Cristian Luis
Mg. Tarazona Vargas, Enver Gerald**

Lima, 27 de junio de 2019

Dedicatoria

A Manuel, por la sabiduría impartida.

A Margarita y Susan, por el apoyo incondicional y paciencia.

A Andrea, Ana y Carla, por ser la razón de mis ganas.

A Gloria, Karim, Alvaro, Luz, Antonio, Gaby, Eleine, Vanessa, Fabiola, Tilsa, Juli, Daniel, Carlos y a los que me han apoyado en estos recientes años, gracias por compartir mi entusiasmo.

Espero llenarlos de orgullo.



Agradecimientos

A James A. Wollack, por su invaluable ayuda,

A Enver Tarazona, por la buena guía.



Resumen

La presente tesis centra sus esfuerzos en presentar y estudiar el Modelo de Respuesta Nominal o NRM por sus siglas en inglés (Bock, 1972, 1997), en el contexto de la Teoría de Respuesta al Ítem (IRT, por sus siglas en inglés). Se realizaron estudios de simulación para determinar la calidad de la recuperación de los parámetros del modelo, bajo la metodología clásica (MML) y bayesiana (MCMC) y finalmente, se aplicó el modelo estudiado en una muestra anónima, aleatoria y representativa de 1641 docentes de la modalidad de Educación Básica Regular de la especialidad de inglés, que fueron expuestos a la sub-prueba de Comprensión Lectora del Concurso de Nombramiento 2015.

En relación a la simulación, encontramos que el método bayesiano es un buen sustituto de su contraparte clásica, debido a que el mismo recupera de manera similarmente satisfactoria los parámetros de los ítems; sin embargo, la principal desventaja es que fue entre 620 a 14100 veces más lento que los métodos clásicos, pese a que se puso especial énfasis en hacer paralelos los procesos MCMC.

En relación a los resultados de la aplicación se tiene que el NRM: (i) facilita la recuperación de una mayor proporción información disponible en los ítems, frente a los modelos de respuestas dicotómicas (Bock, 1972; Thissen, 1976; Levine y Drasgow, 1983; Thissen y Steinberg, 1984), (ii) permite hallar el ordenamiento implícito en datos categóricos inicialmente no ordenados (Samejima, 1988; Bock, 1997) y (iii) brinda información relevante para la valoración de la calidad de un ítem (Thissen et al., 1989), especialmente en dos puntos: (a) les permitía identificar alternativas inservibles o forzadas y (b) les permitía identificar alternativas que se podían colapsar, dado que estas alternativas registraban similar temática.

Palabras-clave: TRI, Modelo de Respuesta Nominal, NRM, clásico, MML, bayesiano, MCMC.

Abstract

This thesis focuses its efforts on presenting and studying the Nominal Response Model or NRM (Bock, 1972, 1997), in the context of the Item Response Theory (IRT). Simulation studies are carried out to determine the quality of the recovery of the parameters of the model, under the Classic (MML) and Bayesian (MCMC) approach and finally, the studied model was applied to an random, representative and anonymous sample of 1641 teachers from the Basic Regular Education modality of the english specialty, who were exposed to the Reading-Comprehension sub-test of the “Concurso de Nombramiento 2015”.

Related to the simulation, we found the bayesian method is a good substitute for the classic counterpart, because it recovers in a similarly satisfactory fashion the parameters of the items; however, the main disadvantage was that the process was between 620 to 14,100 times slower than the classical approach, despite the special emphasis on making the MCMC processes parrallel.

Related to the results of the implementation of the model on real data, the NRM: (i) it facilitates the recovery of a greater proportion of information available in the items, compared to dichotomous response models (Bock, 1972; Thissen, 1976; Levine y Drasgow, 1983; Thissen y Steinberg, 1984), (ii) it allows to find the implicit order in initially not ordered categorical data (Samejima, 1988; Bock, 1997) and (iii) it provided relevant information for the examination of the quality of an item (Thissen et al., 1989), specially in two fronts: (a) it allowed the identification of useless or forced alternatives and (b) it allowed the identification of alternatives that could be collapsed, given that these alternatives registered a similar topics.

Keywords: IRT, Nominal Response Model, NRM, classic, MML, bayesian, MCMC.

Índice general

Lista de Abreviaturas	IX
Lista de Símbolos	X
Índice de figuras	XI
Índice de cuadros	XV
1. Introducción	1
1.1. Consideraciones preliminares	1
1.2. Objetivos	3
1.3. Organización del trabajo	3
2. El Modelo de Respuesta Nominal	5
2.1. Definición del Modelo	5
2.1.1. Observaciones	6
2.2. Supuestos del Modelo	7
2.3. Características del Modelo	8
2.3.1. Funciones de Respuestas a las Categorías del Ítem	8
2.3.2. Función de Información del Modelo	11
2.4. Relación con otros modelos de la Teoría de Respuesta al Ítem	15
2.5. Extensiones al Modelo	16
2.5.1. Modelo de Respuesta Nominal con “adivinación”	16
2.5.2. Nueva Parametrización del Modelo de Respuesta Nominal	17
3. Estimación Bayesiana	20
3.1. Estimación Bayesiana del NRM	20
3.1.1. Función de Verosimilitud	21
3.1.2. Especificación de Prioris	21
3.1.3. Especificación de Valores Iniciales	22
3.1.4. Distribución a posteriori de los parámetros	23
3.2. Implementación Computacional	23
4. Estudio de Simulación	24
4.1. Objetivos	24
4.2. Simulación	24
4.2.1. Condiciones de la simulación	24

4.2.2.	Algoritmo de simulación	25
4.2.3.	Criterios para Evaluar la Simulación	26
4.3.	Estimación de parámetros	27
4.3.1.	Métodos	27
4.3.2.	Replicas, Cadena, Burn-In y saltos en MCMC	27
4.4.	Resultados	28
4.4.1.	Diseño 1:	
	Distribución Normal Estándar, $p = 10$, $m_j = 4$, $n/p = 2.5 : 1$ y $n = 200$	28
4.4.2.	Diseño 2:	
	Distribución Normal Estándar, $p = 10$, $m_j = 4$, $n/p = 20 : 1$ y $n = 1600$	32
4.5.	Conclusiones	37
5.	Aplicación	38
5.1.	Instrumento de Evaluación	38
5.1.1.	Detalles del instrumento	38
5.1.2.	Recolección de Datos	39
5.1.3.	Diseño Muestral	39
5.2.	Resultados	40
5.2.1.	Análisis	40
5.2.2.	Convergencia de la distribución a posteriori	54
6.	Conclusión	61
6.1.	Conclusiones	61
6.2.	Sugerencias para Investigaciones Futuras	62
A.	Resultados teóricos	64
A.1.	Desarrollo conceptual del NRM	64
A.2.	Función de Información del NRM	66
B.	Códigos	70
B.1.	Código en R para simular datos aleatorios del NRM	70
B.1.1.	Función para la simulación de datos	70
B.1.2.	Estudio de simulación	71
B.2.	Código en R para realizar la estimación del modelo	73
B.2.1.	Función de las Categorías del Ítem	74
B.2.2.	Función de Información de las Alternativas e Ítem	75
B.2.3.	Estimación Clásica	77
B.2.4.	Estimación Bayesiana	80
B.2.5.	Código en R para la aplicación en el análisis de datos educacionales	83
B.3.	Código BUGS para la estimación del Modelo	86
C.	Diagnósticos de convergencia	88
C.1.	Simulación: Diagnósticos de Convergencia	88

C.1.1. Diseño 1:	
Distribución Normal Estándar, $p = 10$, $m_j = 4$, $n/p = 2.5 : 1$ y $n = 200$	88
C.1.2. Diseño 2:	
Distribución Normal Estándar, $p = 10$, $m_j = 4$, $n/p = 20 : 1$ y $n = 1600$	98
C.2. Aplicación: Diagnósticos de Convergencia	108
Bibliografía	118



Lista de Abreviaturas

La mayoría definidas por sus siglas en inglés.

MCQ	Ítem de opción múltiple.
IRT	Teoría de Respuesta al ítem.
CTT	Teoría Clásica de los Test.
MCMC	Montecarlo vía Cadenas de Markov.
ML	Máxima Verosimilitud.
MML	Máxima Verosimilitud Marginal.
JML	Máxima Verosimilitud Conjunta.
CML	Máxima Verosimilitud Condicional.
NRM	Modelo de Respuesta Nominal.
GRM	Modelo de Respuesta Graduada.
PCM	Modelo de Crédito Parcial.
GPCM	Modelo Generalizado de Crédito Parcial.
ICRF	Funciones de Respuestas a las Categorías del ítem.
ICIF	Función de Información de la Categoría del ítem.
IIF	Función de Información del ítem.
IF	Función de Información del Test.
3-PL	Modelo Dicotómico de Tres parámetros.
N()	Distribución Normal.
NM()	Distribución Normal Multivariada
LM()	Distribución Log-Normal.
RMSE	Raíz del Error Cuadrático Medio.
MAE	Error Absoluto Medio.
TCC	Curvas Características del Test.

Lista de Símbolos

n	número de individuos que enfrentan un test.
p	número de ítems disponibles en el test.
m_j	número de alternativas disponibles al interior del ítem j .
a_{jk}	parámetro de discriminación de la alternativa k del ítem j .
a_j	vector del parámetro de discriminación para todas las categorías del ítem j .
c_{jk}	parámetro de intercepto de la alternativa k del ítem j .
c_j	vector del parámetro de intercepto para todas las categorías del ítem j .
P_{jk}	Función de respuesta de la categorías k del ítem j (ICRF).
z_{jk}	vector logit multivariado de la categoría k del ítem j .
X_{ij}	variable aleatoria de elección de alguna alternativa en el ítem j , por parte del individuo i .
X_i	vector de elección en los p ítems por parte del individuo i .
$X_{.j}$	vector de elección de los n individuos en el ítem j .
$X_{..}$	vector aleatorio de las respuestas a p ítems por parte de los n individuos (conjunto integral).
θ_i	“habilidad” del individuo i .
θ	vector de “habilidad” de los n individuos
σ^2	varianza.
σ	desviación estándar.
I_{jk}	Información de la categoría k del ítem j .
I_j	Información del ítem j .
I	Información del test.

Índice de figuras

1.1. Pregunta de opción múltiple proveniente de Illinois Standards Achievement Test (ISAT) del 2006	1
2.1. Funciones de Respuestas a las Categorías de un ítem (ICRF) de cuatro ítems con distinta parametrización	9
2.2. Funciones de Respuestas de las Categorías de un Ítem (ICRF), Función de Información del ítem (IIF) y Contribución de Información de las Categorías de un Ítem (ICIF) con interceptos $\mathbf{c}_j = (-0.75; 0.25; 0.5)$ y pendientes $\mathbf{a}_j = (-0.2; -0.8; 1)$. Basado en un ejemplo de Baker y Kim (1992).	14
4.1. RMSE y MAE para todos los parámetros (interceptos y pendientes), según método de estimación.	30
4.2. RMSE y MAE para todos los parámetros (interceptos y pendientes), según método de estimación.	35
5.1. Funciones de respuestas a las categorías de los ítems (ICRF) de los primeros seis ítems de la competencia de Comprensión Lectora de la evaluación de Nombramiento 2015.	44
5.2. Funciones de respuestas a las categorías de los ítems (ICRF) de los ítems 7-12 de la competencia de Comprensión Lectora de la evaluación de Nombramiento 2015.	45
5.3. Funciones de respuestas a las categorías de los ítems (ICRF) de los ítems 13-18 de la competencia de Comprensión Lectora de la evaluación de Nombramiento 2015.	46
5.4. Funciones de respuestas a las categorías de los ítems (ICRF) de los ítems 19-25 de la competencia de Comprensión Lectora de la evaluación de Nombramiento 2015.	47
5.5. Funciones de información de las categorías de los ítems (ICIF) de los primeros seis ítems de la competencia de Comprensión Lectora de la evaluación de Nombramiento 2015.	48
5.6. Funciones de información de las categorías de los ítems (ICIF) de los ítems 7-12 de la competencia de Comprensión Lectora de la evaluación de Nombramiento 2015.	49

5.7. Funciones de información de las categorías de los ítems (ICIF) de los ítems 13-18 de la competencia de Comprensión Lectora de la evaluación de Nombramiento 2015.	50
5.8. Funciones de información de las categorías de los ítems (ICIF) de los ítems 19-25 de la competencia de Comprensión Lectora de la evaluación de Nombramiento 2015.	51
5.9. Distribución del estimado de las competencias de los individuos en Comprensión Lectora	52
5.10. Estimador puntual e Intervalo de confianza al 95 % de las competencias de los individuos en Comprensión Lectora	53
A.1. Líneas logit de regresión para un ítem de tres categorías con $\mathbf{a}_j = (-0.75; -0.25; 1.0)$ y $\mathbf{c}_j = (-1.5; -0.25; 1.75)$. Tomado de De Ayala (1993).	65
C.1. Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la habilidad del individuo 48, réplica 7, estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0, 1)$, $n = 200$ y $p=10$	88
C.2. Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la habilidad del individuo 194, réplica 8, estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0, 1)$, $n = 200$ y $p=10$	89
C.3. Iteraciones efectivas, gráfica de densidad y auto-correlaciones para el intercepto del ítem 6, alternativa 1, réplica 1, estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0, 1)$, $n = 200$ y $p=10$	90
C.4. Iteraciones efectivas, gráfica de densidad y auto-correlaciones para el intercepto del ítem 7, alternativa 2, réplica 4, estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0, 1)$, $n = 200$ y $p=10$	91
C.5. Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la pendiente del ítem 4, alternativa 1, réplica 2, estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0, 1)$, $n = 200$ y $p=10$	92
C.6. Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la pendiente del ítem 5, alternativa 2, réplica 3 estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0, 1)$, $n = 200$ y $p=10$	93
C.7. Iteraciones efectivas, gráfica de densidad y auto-correlaciones para el intercepto del ítem 3, alternativa 1, réplica 9, estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0, 1)$, $n = 200$ y $p=10$	94

C.8. Iteraciones efectivas, gráfica de densidad y auto correlaciones para el intercepto del ítem 2, alternativa 3, réplica 10, estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0, 1)$, $n = 200$ y $p=10$ 95

C.9. Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la pendiente del ítem 3, alternativa 4, réplica 5, estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0, 1)$, $n = 200$ y $p=10$ 96

C.10. Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la pendiente del ítem 10, alternativa 1, réplica 6 estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0, 1)$, $n = 200$ y $p=10$ 97

C.11. Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la habilidad del individuo 83, réplica 13, estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0, 1)$, $n = 1600$ y $p=10$ 98

C.12. Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la habilidad del individuo 342, réplica 2, estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0, 1)$, $n = 1600$ y $p=10$ 99

C.13. Iteraciones efectivas, gráfica de densidad y auto-correlaciones para el intercepto del ítem 1, alternativa 4, réplica 9, estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0, 1)$, $n = 1600$ y $p=10$ 100

C.14. Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la pendiente del ítem 2, alternativa 1, réplica 19, estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0, 1)$, $n = 1600$ y $p=10$ 101

C.15. Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la pendiente del ítem 5, alternativa 2, réplica 12, estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0, 1)$, $n = 1600$ y $p=10$ 102

C.16. Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la pendiente del ítem 3, alternativa 3, réplica 14, estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0, 1)$, $n = 1600$ y $p=10$ 103

C.17. Iteraciones efectivas, gráfica de densidad y auto-correlaciones para el intercepto del ítem 2, alternativa 3, réplica 10, estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0, 1)$, $n = 1600$ y $p=10$ 104

C.18. Iteraciones efectivas, gráfica de densidad y auto-correlaciones para el intercepto del ítem 10, alternativa 1, réplica 15, estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0, 1)$, $n = 1600$ y $p=10$ 105

C.19. Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la pendiente del ítem 2, alternativa 4, réplica 16, estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0, 1)$, $n = 1600$ y $p=10$ 106

C.20. Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la pendiente del ítem 2, alternativa 4, réplica 18, estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0, 1)$, $n = 1600$ y $p=10$ 107

C.21. Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la habilidad del individuo 406, estimado con el Modelo NRM, para los datos de aplicación 108

C.22. Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la habilidad del individuo 1255, estimado con el Modelo NRM, para los datos de aplicación 109

C.23. Iteraciones efectivas, gráfica de densidad y auto-correlaciones para el intercepto del ítem 1, alternativa 1, estimado con el Modelo NRM, para los datos de aplicación 110

C.24. Iteraciones efectivas, gráfica de densidad y auto-correlaciones para el intercepto del ítem 3, alternativa 3, estimado con el Modelo NRM, para los datos de aplicación 111

C.25. Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la pendiente del ítem 21, alternativa 1, estimado con el Modelo NRM, para los datos de aplicación 112

C.26. Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la pendiente del ítem 20, alternativa 3, estimado con el Modelo NRM, para los datos de aplicación 113

C.27. Iteraciones efectivas, gráfica de densidad y auto-correlaciones para el intercepto del ítem 2, alternativa 2, estimado con el Modelo NRM, para los datos de aplicación 114

C.28. Iteraciones efectivas, gráfica de densidad y auto-correlaciones para el intercepto del ítem 4, alternativa 5, estimado con el Modelo NRM, para los datos de aplicación 115

C.29. Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la pendiente del ítem 14, alternativa 3, estimado con el Modelo NRM, para los datos de aplicación 116

C.30. Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la pendiente del ítem 20, alternativa 2, estimado con el Modelo NRM, para los datos de aplicación 117

Índice de cuadros

3.1. Prioris establecidas para la estimación de ítems politómicos. $N(.,.)$ se refiere a la distribución normal, $NM(.,.)$ a la distribución Normal Multivariada y $LN(.,.)$ a la distribución log-normal. Para mayor detalle ver listado de abreviaturas.	22
4.1. Parámetros de los ítems usados en la simulación	25
4.2. Diseños de simulaciones: distribuciones, tamaños de muestras y longitud de evaluaciones	25
4.3. RMSE para los parámetros de interceptos y pendientes, en las cuatro alternativas de los diez ítems elegidos en el estudio, según método de estimación. . .	29
4.4. MAE para los parámetros de interceptos y pendientes, en las cuatro alternativas de los diez ítems elegidos en el estudio, según método de estimación. . .	29
4.5. RMSE para las probabilidades, en las cuatro alternativas de los diez ítems elegidos en el estudio, según método de estimación.	31
4.6. MAE para las probabilidades, en las cuatro alternativas de los diez ítems elegidos en el estudio, según método de estimación.	31
4.7. RMSE para los parámetros de interceptos y pendientes, en las cuatro alternativas de los diez ítems elegidos en el estudio, según método de estimación. . .	33
4.8. MAE para los parámetros de interceptos y pendientes, en las cuatro alternativas de los diez ítems elegidos en el estudio, según método de estimación. . .	34
4.9. RMSE para las probabilidades, en las cuatro alternativas de los diez ítems elegidos en el estudio, según método de estimación.	35
4.10. MAE para las probabilidades, en las cuatro alternativas de los diez ítems elegidos en el estudio, según método de estimación.	36
4.11. Duración del proceso de estimación por diseño y método de estimación	37
5.7. Características del estimador puntual de las competencias de los individuos en Comprensión Lectora	52
5.1. Estadísticos de los estimados de los parámetros de Interceptos de los ítems 1 al 10 en el contexto del modelo NRM, bajo el método MCMC de inferencia bayesiana	55
5.2. Estadísticos de los estimados de parámetros de Interceptos de los ítems 11 al 20 en el contexto del modelo NRM, bajo el método MCMC de inferencia bayesiana	56

5.3. Estadísticos de los estimados de parámetros de Interceptos de los ítems 21 al 25 en el contexto del modelo NRM, bajo el método MCMC de inferencia bayesiana 57

5.4. Estadísticos de los estimados de parámetros de Pendientes de los ítems 1 al 10 en el contexto del modelo NRM, bajo el método MCMC de inferencia bayesiana 58

5.5. Estadísticos de los estimados de parámetros de Pendientes de los ítems 11 al 20 en el contexto del modelo NRM, bajo el método MCMC de inferencia bayesiana 59

5.6. Estadísticos de los estimados de parámetros de Pendientes de los ítems 21 al 25 en el contexto del modelo NRM, bajo el método MCMC de inferencia bayesiana 60



Capítulo 1

Introducción

1.1. Consideraciones preliminares

En múltiples situaciones un individuo se ve enfrentado a una prueba “estandarizada”; es decir, una evaluación diseñada de tal manera que (a) las preguntas, (b) las condiciones para ser administrada, (c) los procedimientos de calificación y (d) las interpretaciones, son consistentes con una manera predeterminada o tipificada. En este escenario, el individuo es expuesto a un instrumento de evaluación cuyas preguntas usualmente cumplen las siguientes tres características: (i) son politómicas o de opción múltiple, (ii) parecen exhibir alternativas sin un ordenamiento específico, es decir, son de medida nominal y (iii) poseen, con frecuencia, una respuesta “correcta”, tal y como se observa en la Figura 1.1.

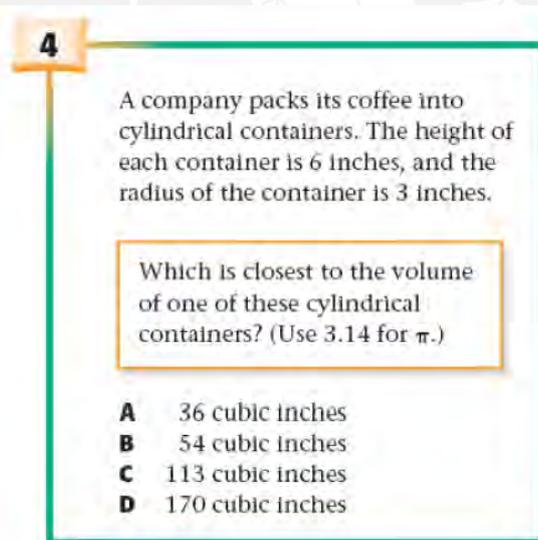


Figura 1.1: Pregunta de opción múltiple proveniente de Illinois Standards Achievement Test (ISAT) del 2006

Es conocido que cuando un individuo es enfrentado a este tipo de preguntas, este rara vez distribuye de manera uniforme su respuesta entre todas las alternativas disponibles; es decir, la alternativa seleccionada no tiene la misma probabilidad de ser elegida frente a sus pares. Thissen et al. (1989) señalan que para muchos examinados y constructores de ítems, las alternativas son la parte más importante del mismo. De hecho, algunos examinados primero observan las alternativas antes de leer el enunciado de la pregunta, pues este último solo sirve de guía para elegir “la más correcta”. Del mismo modo, cuando un constructor de

ítems intenta juzgar la calidad de una pregunta de opción múltiple, el análisis profundo de las alternativas juega un papel importante en el análisis global del ítem.

Estos resultados nos llevan a inferir que la elección de las alternativas contiene información que podría ser aprovechada para estimar la competencia latente de los individuos (Bock, 1972); información útil que se pierde al dicotomizar las respuestas (De Ayala, 1993; Thissen y Steinberg, 1984). Autores como De Ayala (1993) también plantean, que incluir la información de las alternativas abre no solo la posibilidad de “localizar mejor” a los individuos (en el continuo de la competencia de interés) sino también proveen información útil que, a grandes rasgos, permitiría generar perfiles de individuos (acordes con las alternativas elegidas).

De este modo, el estudio del Modelo de Respuesta Nominal (NRM) se sustenta, en primer lugar, en que permite estimar simultáneamente los parámetros de las categorías de respuesta de un ítem de opción múltiple, aún cuando dichas alternativas no parecen exhibir un ordenamiento específico. Múltiples argumentos se podrían realizar a favor del uso de modelos para categorías ordenadas en categorías nominales (p.e. PCM, GPCM y GRM); sin embargo, debido a que las alternativas no elegidas no representan respuestas parcialmente correctas, su uso no sería apropiado (De Ayala, 1993).

Aunado a lo anterior, la investigación del modelo también se sustenta en que: (i) el modelo facilita la recuperación de una mayor proporción información disponible en los ítems, frente a los modelos de respuestas dicotómicas (Bock, 1972; Thissen, 1976; Levine y Drasgow, 1983; Thissen y Steinberg, 1984); (ii) permite hallar el ordenamiento implícito en datos categóricos inicialmente no ordenados (Samejima, 1988; Bock, 1997); (iii) brinda información relevante para la valoración de la calidad de un ítem (Thissen et al., 1989); e incluso según lo señalado por De Ayala (1993) (iv) podría proveer de información útil para caracterizar a los individuos expuestos al instrumento de evaluación.

Así, el NRM modela la probabilidad de elegir una alternativa de un ítem mediante una distribución categórica, de la siguiente manera:

$$P_{jk}(\theta_i) = f [z_{jk}(\theta_i)] = \frac{e^{c_{jk} + a_{jk} \cdot \theta_i}}{\sum_{h=1}^{m_j} e^{c_{jh} + a_{jh} \cdot \theta_i}}$$

donde para la alternativa k en un ítem j , la probabilidad de “adhesión” a la categoría $P_{jk}(\theta_i)$ se encuentra explicada por una relación no lineal, entre el nivel de habilidad latente del individuo θ_i y un conjunto de características de la alternativa: (a) cuanto cambia el logaritmo del ratio de oportunidades (ratio de Odds) cuando la competencia del individuo aumenta o disminuye, relación explicada por el parámetro de pendiente a_{jk} , y (b) la tendencia global de selección o “popularidad” de la alternativa, definida por el parámetro del intercepto c_{jk} .

Se considera que el aporte del presente trabajo es teórico, metodológico y práctico. El aporte teórico se centra en ofrecer un análisis y descripción exhaustiva del modelo, el aporte metodológico concentra su esfuerzo en profundizar en los métodos de estimación asociados

con la inferencia bayesiana y el aporte práctico se centra en aplicar el modelo en un conjunto de datos reales del campo educacional.

1.2. Objetivos

El principal objetivo de la tesis consistirá en estudiar y presentar el modelo de respuesta para datos nominales propuesto por Bock (1972, 1997), en el contexto de la IRT. Así mismo, se realizará un estudio de simulación para determinar la calidad de los estimados de los parámetros de interés del modelo. Finalmente, la tesis pretende aplicar el modelo estudiado a un conjunto de datos reales desde el enfoque de la inferencia bayesiana.

De manera específica, el presente trabajo:

- Realiza una extensiva revisión de la literatura acerca del modelo de interés, con el propósito de estudiar y presentar el modelo de respuesta para datos nominales, sus características, supuestos y propiedades;
- Plantea e implementa métodos MCMC para la estimación de los parámetros de interés del modelo;
- Realiza estudios de simulación para comprobar la calidad de los estimados de los parámetros de interés de los métodos clásicos (MML), frente a los métodos bayesianos (MCMC);
- Aplica el mencionado modelo, desde el enfoque de la inferencia bayesiana, a un conjunto de datos reales pertenecientes al campo de la evaluación educacional.

Con el propósito en mente, la implementación computacional del modelo estudiado será desarrollada usando el programa R (R Core Team, 2015) y WinBUGS (Lunn et al., 2000).

1.3. Organización del trabajo

En el **Capítulo 2: El Modelo de Respuesta Nominal**, presentamos los conceptos previos al desarrollo del NRM. Así mismo, presentamos la definición del modelo, características, supuestos y propiedades del NRM. Finalmente se plantea la relación del NRM con otros modelos tanto dicotómicos como politómicos, y también algunas extensiones al NRM original.

En el **Capítulo 3: Estimación Bayesiana**, se plantean las razones de la elección de los métodos bayesianos, beneficios y principales problemas, en el contexto de los modelos IRT. Se presentan la función de verosimilitud de los datos, la propuesta de prioris para los parámetros de interés y la distribución a posteriori resultante. Finalmente se presenta la implementación computacional del modelo.

En el **Capítulo 4: Estudio de Simulación**, se realizará estudios de simulación para comparar la calidad de los estimados de los parámetros de interés. El estudio establecerá el algoritmo de simulación de los datos y comparará los resultados obtenidos provenientes del método clásico frente al bayesiano.

En el **Capítulo 5: Aplicación**, se hace una breve reseña del constructo bajo análisis, así como del instrumento utilizado en la evaluación y la “dimensión” medida. Del mismo modo,

se detallará el diseño muestral, la recolección de datos y la población evaluada; así como, un análisis de los estimados de los parámetros de los ítems y habilidades de las personas.

Finalmente, en el **Capítulo 6: Conclusiones** discutimos algunas conclusiones obtenidas en la presente investigación. Resumimos lo realizado en el documento, los resultados obtenidos de la implementación del modelo estudiado, tanto a nivel de simulación como aplicación en datos reales y finalmente damos luces acerca de las investigaciones futuras que se podrían derivar de la presente labor.



Capítulo 2

El Modelo de Respuesta Nominal

El Modelo de Respuesta Nominal o NRM, introducido por Bock (1972, 1997), comprende a una amplia familia de funciones útiles para describir el comportamiento cualitativo de un individuo ante un estímulo. El principal propósito del NRM es maximizar la precisión de los estimados de competencia, utilizando la información contenida en todas y cada una de las respuestas que los individuos dan a un ítem (Hambleton y Swaminathan, 1991).

El modelo fue originalmente diseñado para ser aplicado a ítems de opción múltiple, cuyas alternativas eran de categorías nominales; es decir, alternativas que no presentan un ordenamiento específico, en términos del rasgo latente a ser medido. Sin embargo, investigaciones posteriores (Samejima, 1988; Bock, 1997), señalaron que el modelo también era útil para hallar el orden implícito de categorías inicialmente no ordenadas.

El presente capítulo estudiará y presentará el NRM, sus características, supuestos y propiedades.

2.1. Definición del Modelo

En el contexto de una evaluación suponemos que n examinados responden p ítems de opción múltiple eligiendo una de las m_j alternativas disponibles. Las alternativas son mutuamente excluyentes, exhaustivas y no presentan un ordenamiento específico. De este modo, el NRM es definido como:

$$X_{ij} = x_{ij} \mid \theta_i, \mathbf{c}_j, \mathbf{a}_j \sim \text{Categórica} \left[P_{j1}(\theta_i), \dots, P_{jk}(\theta_i), \dots, P_{j(m_j)}(\theta_i) \right] \quad (2.1)$$

$$P_{jk}(\theta_i) = f[z_{jk}(\theta_i)] = \frac{e^{z_{jk}(\theta_i)}}{\sum_{h=1}^{m_j} e^{z_{jh}(\theta_i)}} \quad (2.2)$$

$$z_{jh}(\theta_i) = c_{jh} + a_{jh} \cdot \theta_i \quad (2.3)$$

donde $i = 1, \dots, n$, $j = 1, \dots, p$ y $h = 1, \dots, k, \dots, m_j$. Además,

- X_{ij} es la variable aleatoria manifiesta cuya realización x_{ij} puede tomar un valor entre 1 y m_j , para la persona i en el ítem j .
- θ_i representa la competencia latente de examinado i . La variable es continua y puede

tomar cualquier valor dentro del rango de los números reales ($\theta_i \in \mathbb{R}$). Se define a $\boldsymbol{\theta} = [\theta_1, \dots, \theta_n]^T$ como el vector de competencias latentes de los n examinados.

- $z_{jh}(\theta_i)$ es el valor del logit multivariado del examinado i para la alternativa h en el ítem j .
- c_{jh} es el parámetro del intercepto. Indica el valor del vector logit multivariado de la alternativa h en el ítem j , cuando $\theta_i = 0$. Puede ser interpretado como el parámetro de “popularidad global” de la alternativa h en el ítem j ; es decir, el que refleja la probabilidad de elegir la alternativa de interés en el ítem cuando la competencia del individuo es igual a cero (Thissen et al., 1989; Stone y Zhu, 2006). El vector de los interceptos se encuentra definido por $\mathbf{c}_j = [c_{j1}, \dots, c_{jk}, \dots, c_{j(m_j)}]$.
- a_{jh} es el parámetro de pendiente. Indica como cambia el valor del vector logit multivariado de la alternativa h en el ítem j , a medida que aumenta/disminuye θ_i . Puede ser interpretado como la discriminación de la alternativa h en el ítem j ; es decir, el que refleja la fuerza y dirección de la relación de la alternativa de interés en el ítem y la competencia del individuo (De Ayala, 1993; Ostini y Nering, 2006; Stone y Zhu, 2006). El vector de pendientes se encuentra definido por $\mathbf{a}_j = [a_{j1}, \dots, a_{jk}, \dots, a_{j(m_j)}]$.
- $P_{jk}(\theta_i)$ refleja la probabilidad de que el individuo i elija la alternativa k en el ítem j . Considerando la distribución categórica, las probabilidades de la elección de las alternativas del ítem deben cumplir $\sum_{h=1}^{m_j} P_{jh}(\boldsymbol{\theta}) = \mathbf{1}$, donde $\mathbf{1}$ es un vector de unos de dimensión $(n \times 1)$.

Para mayor detalle acerca del desarrollo conceptual del NRM referirse a la sección [A.1](#).

2.1.1. Observaciones

1. El NRM corresponde al grupo de modelos de **trazo latente unidimensional**.
2. Pese a que en términos prácticos se suele asumir que el número de alternativas disponibles es la misma en todos los ítems, la parametrización del modelo no requiere la imposición de tal restricción; por el contrario, el número de alternativas disponibles m_j puede ser distinto de ítem a ítem. Bock (1972) señaló que incluso los valores perdidos u omitidos pueden ser utilizados como una categoría más, sin que el modelo necesite modificación alguna.
3. Intentos de estimar los parámetros de la ecuación (2.3) encontrarán problemas de identificación (Baker y Kim, 1992). Thissen y Steinberg (1997) señalaron que el modelo original de Bock era muy flexible, por lo que el mismo requería que sus parámetros sean restringidos de tal manera que le permita funcionar apropiadamente; es decir, los parámetros debían ser anclados de alguna manera en algún punto. El presente trabajo seguirá la propuesta de Bock (1972), al restringir a cero la suma del logit multivariado (ecuación 2.3) al interior de cada ítem. Para mayor detalle del método seguido referirse a la sub-sección [2.3.1](#).

2.2. Supuestos del Modelo

Para determinar si el ajuste a un modelo particular será limitado o si el modelo será de cuestionable aplicación, considerando las características de los datos, será de especial ayuda reconocer los supuestos asociados a un modelo específico y la adecuación de los datos a estas asunciones. Acorde con Hambleton y Swaminathan (1991) y Azevedo (2003) el NRM debe cumplir:

(S1) Dimensionalidad del Espacio Latente, el espacio latente se considera completo si todos los rasgos o competencias que influyen los resultados de los examinados han sido especificados (Hambleton y Swaminathan, 1991). Se considera que el espacio latente es *unidimensional* sí y solo sí, una competencia es necesaria para explicar el desempeño de los examinados. En la práctica, para que el supuesto se cumpla adecuadamente solo es requerido que un factor o rasgo “dominante” explique los resultados de los examinados.

(S2) Adecuación al Modelo Multivariado de Bernoulli, la probabilidad de que los individuos elijan la opción h del ítem j puede ser representada por el modelo Multivariado de Bernoulli, también conocido como Distribución Generalizada de Bernoulli o Distribución Categórica, del siguiente modo:

$$P(X_{ij} = x_{ij} | \theta_i, \mathbf{c}_j, \mathbf{a}_j) = \prod_{h=1}^{m_j} [P_{jh}(\theta_i)]^{1(x_{ij}=h)} \quad (2.4)$$

Donde $P_{jh}(\theta_i)$ es definida acorde con la ecuación (2.2), mientras que $1(x_{ij} = h)$ es una función indicadora que toma el valor de 1 cuando $x_{ij} = h$.

(S3) Independencia Local, dada la competencia del individuo, las alternativas elegidas en diferentes ítem son estadísticamente independientes entre sí; incluso si el orden de la presentación de los ítems es distinto de evaluación a evaluación, de este modo:

$$P(\mathbf{X}_i = \mathbf{x}_i | \theta_i, \mathbf{c}_j, \mathbf{a}_j) = \prod_{j=1}^p P(X_{ij} = x_{ij} | \theta_i, \mathbf{c}_j, \mathbf{a}_j) = \prod_{j=1}^p \prod_{h=1}^{m_j} [P_{jh}(\theta_i)]^{1(x_{ij}=h)} \quad (2.5)$$

donde $\mathbf{X}_i = [X_{i1}, \dots, X_{ip}]$ es el vector de selecciones del examinado i para los p ítems y $\mathbf{x}_i = [x_{i1}, \dots, x_{ip}]$ son sus realizaciones. Se define además el vector de selecciones de los n individuos en el ítem j como $\mathbf{X}_j = [X_{1j}, \dots, X_{nj}]^T$ y $\mathbf{x}_j = [x_{1j}, \dots, x_{nj}]^T$ como sus realizaciones.

(S4) Independencia de Individuos, Las respuestas provenientes de diferentes individuos son estadísticamente independientes entre sí, de este modo:

$$P(\mathbf{X}_{..} = \mathbf{x}_{..} | \theta, \mathbf{c}_j, \mathbf{a}_j) = \prod_{i=1}^n P(\mathbf{X}_i = \mathbf{x}_i | \theta_i, \mathbf{c}_j, \mathbf{a}_j) = \prod_{i=1}^n \prod_{j=1}^p \prod_{h=1}^{m_j} [P_{jh}(\theta_i)]^{1(x_{ij}=h)} \quad (2.6)$$

donde $\mathbf{X}_{..} = [\mathbf{X}_{.1}, \dots, \mathbf{X}_{.p}] = [\mathbf{X}_{1.}, \dots, \mathbf{X}_{n.}]^T$ es la matriz o conjunto integral de respuestas de los n individuos a los p ítems y $\mathbf{x}_{..} = [\mathbf{x}_{.1}, \dots, \mathbf{x}_{.p}] = [\mathbf{x}_{1.}, \dots, \mathbf{x}_{n.}]^T$ son sus realizaciones.

2.3. Características del Modelo

2.3.1. Funciones de Respuestas a las Categorías del Ítem

El NRM (Bock, 1972, 1997) expresa la probabilidad de elegir alguna de las m_j alternativas de un ítem, mediante m_j Funciones de Respuestas a las Categorías del Ítem o ICRF (Ostini y Nering, 2006), donde las ICRF se encuentran definidas como en las ecuaciones (2.2) y (2.3):

$$P_{jk}(\theta_i) = f[z_{jk}(\theta_i)] = \frac{e^{z_{jk}(\theta_i)}}{\sum_{h=1}^{m_j} e^{z_{jh}(\theta_i)}}$$

$$z_{jh}(\theta_i) = c_{jh} + a_{jh} \cdot \theta_i$$

donde $i = 1, \dots, n$, $j = 1, \dots, p$ y $h = 1, \dots, k, \dots, m_j$.

La localización de las ICRF, en el continuo de competencia, y sus formas depende de la combinación de los parámetros del set completo de respuestas del ítem. Baker y Kim (1992) y Hambleton y Swaminathan (1991) señalaron que dado las m_j categorías de respuestas del ítem, se esperaría que una de las ICRF sea monotónicamente creciente, una monotónicamente decreciente y las $(m_j - 2)$ ICRF restantes sean curvas unimodales con colas asintóticamente cercanas a cero.

La Figura 2.1 presenta las ICRF para cuatro ítems con distinta parametrización. Siguiendo la interpretación de los parámetros se tiene que:

- El panel superior izquierdo denota a un ítem en el que a medida que la competencia del individuo aumenta, la probabilidad de que el mismo se decante por la alternativa 4 aumenta rápidamente. Dicha probabilidad es alta ($P_{j4} = 0.7$) incluso entre los individuos que se encuentran en el medio del continuo de competencia ($\theta = 0$). Así, la alternativa 4 se hace rápidamente la más popular entre individuos de competencia media y sigue haciéndose más popular a medida que la competencia del individuo aumenta, en contraste con el resto de alternativas del ítem, en especial la 3, la cual pierde popularidad casi igual de rápido con la que la gana la alternativa 4.
- Por su parte, el panel superior derecho, denota a un ítem en el que a medida que la competencia del individuo aumenta, mayor es la probabilidad de que el mismo elija la alternativa 2; sin embargo, la probabilidad de elegir dicha alternativa es baja ($P_{j2} = 0.2$) entre los individuos que se encuentran en el medio del continuo de competencia ($\theta = 0$), los cuales se decantan por la alternativa 4. De este modo, la elección de una alternativa depende del nivel competencia del individuo, donde un evaluado de habilidad baja se decantaría por la alternativa 1, uno de habilidad media elegiría más la alternativa 4 y uno de habilidad alta elegiría cada vez más la alternativa 2 frente a la 3.

- El panel inferior izquierdo, por su parte, denota a un ítem en el que a medida que la competencia del individuo aumenta, mayor es la probabilidad de que el mismo elija la alternativa 4. Sin embargo, dicha probabilidad es casi tan alta como la probabilidad de elegir las alternativas 2 y 3 ($P_{j4} = 0.39$, $P_{j2} = 0.24$ y $P_{j3} = 0.24$) entre los individuos que se encuentran en el medio del continuo de competencia ($\theta = 0$). Así, un individuo de habilidad baja se decanta por la alternativa 1 frente a cualquier otra. A medida que los individuos se mueven al medio del continuo de competencia, estos consideran que las alternativas 1, 2 y 3 presentan similar atractivo, distribuyendo su elección entre ellas, casi equitativamente. Poco antes de superar el corte de competencia media ($\theta = 0$), es claro que un individuo con mayor habilidad dejará de elegir cualquier otra alternativa y se decantará por la 4.
- Finalmente, el panel inferior derecho muestra un ítem cuya alternativa 3 se hace rápidamente popular ($P_{j3} = 0.61$) para individuos en el rango medio del continuo de competencia ($\theta = 0$). De este modo, los individuos con competencia baja eligen principalmente la alternativa 4 (con casos eligiendo la alternativa 1). Este comportamiento se observa hasta que $\theta = -1$, donde cada una de las alternativas tiene igual popularidad o probabilidad de ser elegida. A partir de $\theta = 0$, los individuos eligen la opción 2 o 3, siendo solo la última, la que se hace más popular a medida que la competencia del evaluado aumenta.

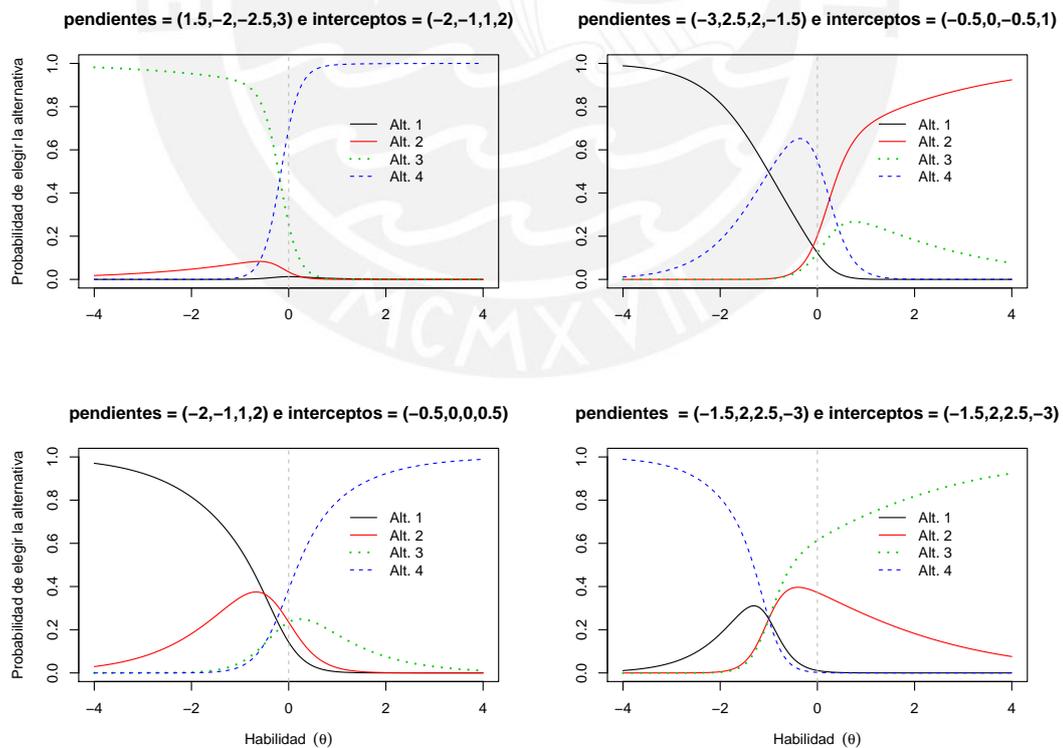


Figura 2.1: Funciones de Respuestas a las Categorías de un ítem (ICRF) de cuatro ítems con distinta parametrización

Como ya se había adelantado en párrafos anteriores, intentos de estimar los parámetros

de la ecuación (2.3) encontrarán problemas de identificación (Baker y Kim, 1992). Esto sucede debido a que las ICRFs presentadas en la ecuación (2.2) son invariantes respecto a la traslación del logit multivariado (Baker y Kim, 1992; Bock, 1972); es decir, la adición de una constante a cualquiera de los parámetros de pendiente e intercepto, a_{jh} y c_{jh} respectivamente, resulta en un mismo valor para la probabilidad $P_{jh}(\theta_i)$, por lo que estos deben ser fijados en un punto de alguna forma.

Bock (1972) propone superar el problema forzando a que la suma del logit multivariado $z_{jh}(\boldsymbol{\theta}) = [z_{jh}(\theta_i), \dots, z_{jh}(\theta_n)]^T$ sea cero al interior de cada ítem; es decir,

$$\sum_{h=1}^{m_j} z_{jh}(\boldsymbol{\theta}) = \mathbf{0} \quad (2.7)$$

donde $\mathbf{0}$ es un vector de ceros de dimensión $(n \times 1)$. Lo establecido en la ecuación (2.7) implica a su vez,

$$\sum_{h=1}^{m_j} a_{jh} = 0 \quad (2.8)$$

$$\sum_{h=1}^{m_j} c_{jh} = 0 \quad (2.9)$$

Un método alternativo, propuesto por Thissen (1991), consiste en restringir a cero los parámetros de intercepto y pendiente de la primera categoría de cada ítem; es decir, $a_{j1} = c_{j1} = 0$ para todo $j = 1, \dots, p$.

Ostini y Nering (2010) definieron la matriz *logit multinomial* compuesta por m_j logits multivariados de la siguiente forma: $\mathbf{z}_j = [z_{j1}(\boldsymbol{\theta}), \dots, z_{jk}(\boldsymbol{\theta}), \dots, z_{j(m_j)}(\boldsymbol{\theta})]$. Expresando al *logit multinomial* de manera matricial se tendría:

$$\mathbf{z}_j = \mathbf{K} \cdot \mathbf{B}_j$$

$(n \times m_j)$

donde

$$\mathbf{K} = [\mathbf{1} \quad \boldsymbol{\theta}]$$

$(n \times 2)$

$$\mathbf{B}_j = \begin{bmatrix} \mathbf{c}_j \\ \mathbf{a}_j \end{bmatrix}$$

$(2 \times m_j)$

De manera general, Bock (1972, 1997) planteó que el *logit multinomial*, expresado en función de $2 \times (m_j)$ parámetros, puede ser expresado en términos de $2 \times (m_j - 1)$ parámetros “libres”, si se reemplaza los parámetros originales por funciones paramétricas lineales de la siguiente manera:

$$\mathbf{B}_j = \mathbf{A}_j \cdot \mathbf{T}_j$$

donde

$$\mathbf{A}_j = \begin{bmatrix} \boldsymbol{\gamma}_j \\ \boldsymbol{\alpha}_j \end{bmatrix}$$

$$\begin{aligned} \boldsymbol{\gamma}_j &= [\gamma_{j1}, \dots, \gamma_{jk}, \dots, \gamma_{j(m_j-1)}] \\ \boldsymbol{\alpha}_j &= [\alpha_{j1}, \dots, \alpha_{jk}, \dots, \alpha_{j(m_j-1)}] \end{aligned}$$

$$\mathbf{T}_j = \begin{pmatrix} \frac{1}{m_j} & \frac{1}{m_j} - 1 & \frac{1}{m_j} & \dots & \frac{1}{m_j} \\ \frac{1}{m_j} & \frac{1}{m_j} & \frac{1}{m_j} - 1 & \dots & \frac{1}{m_j} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{1}{m_j} & \frac{1}{m_j} & \frac{1}{m_j} & \dots & \frac{1}{m_j} - 1 \end{pmatrix}$$

siendo la *matriz de transformación* \mathbf{T}_j la que toma la forma requerida para que las restricciones impuestas al vector logit se cumplan (Bock, 1972). De esta forma, la matriz *logit multinomial* es transformada de la siguiente manera:

$$\mathbf{z}_j = \mathbf{K} \cdot \mathbf{A}_j \cdot \mathbf{T}_j$$

Esta transformación es más conveniente pues reduce el número de parámetros a estimar por $2n$; es decir, usa procedimientos computacionales más económicos que los requeridos bajo la parametrización original. Recordar que la optimización bajo la parametrización original se llevaría a cabo mediante una maximización sujeta a restricciones, las que serían operacionales mediante multiplicadores de Lagrange.

2.3.2. Función de Información del Modelo

En términos generales, cuando uno se refiere a *poseer información*, uno quiere decir que *conoce algo* acerca de un objeto o tema en particular. En Estadística y Psicometría, el concepto de *información* contempla una interpretación similar, pero más técnica (Baker, 2001).

Se atribuye a Fisher la definición de *información*, y esta se entiende como el recíproco de la precisión con el que los parámetros de una función pueden ser estimados. De este modo, si uno estima con mayor precisión, implicaría que conoce más acerca del valor real del parámetro, que si uno lo estimara con menor precisión. Estadísticamente, la precisión con la que un parámetro es estimado es medido por el inverso de la variabilidad del estimador alrededor del valor real del parámetro; es decir, como el inverso de la varianza (σ^2) del estimador; así, la *información* se puede definir de la siguiente manera:

$$I = \frac{1}{\sigma^2} \quad (2.10)$$

De la ecuación (2.10) se puede deducir que si la cantidad de información es grande, la precisión

con la que se estimará el parámetro será mayor, asumiendo que los estimados se encuentran razonablemente cerca de los valores reales de los parámetros de interés; es decir, sean insesgados

En Psicometría, el analista tiene un particular interés en la *información* y específicamente en lo que se le conoce como la *Función de Información del Test* (IF, por sus siglas en inglés), pues es el indicador de la precisión con el que las competencias de los individuos serán estimadas (Baker y Kim, 1992). Sin embargo, la noción del concepto de *información* no solo está asociado a la precisión de la estimación, sino que da paso a varias aplicaciones que son centrales al campo de la medición psicológica y/o educacional, entre ellas la construcción de evaluaciones, selección de ítems, comparación entre evaluaciones, entre otras (Hambleton et al., 1991).

Baker y Kim (1992) notaron que la IF bajo el NRM empleaba un enfoque similar al utilizado por los modelos politómicos ordenados, el cual a su vez se encontraba basado en el caso dicotómico. Mellenberg (1995) señaló que un ítem dicotómico puede ser pensado como un caso especial de un ítem politómico con dos categorías.

Cuando un ítem es calificado de forma dicotómica, la *Función de Información del Ítem* (IIF) toma la siguiente forma (Birnbbaum, 1968):

$$I_j(\boldsymbol{\theta}) = -E \left[\frac{\partial^2 \log P_j(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2} \right] = \frac{[P'_j(\boldsymbol{\theta})]^2}{P_j(\boldsymbol{\theta})Q_j(\boldsymbol{\theta})} \quad (2.11)$$

donde

$$P'_j(\boldsymbol{\theta}) = \partial P_j(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}$$

Si se considera a $P_{j1}(\boldsymbol{\theta}) = P_j(\boldsymbol{\theta})$ como el vector de probabilidad que corresponde a la elección de la alternativa correcta en el ítem j y a $P_{j2}(\boldsymbol{\theta}) = Q_j(\boldsymbol{\theta})$ como el vector de probabilidad que corresponde a la elección de la alternativa incorrecta en el ítem j , o en otras palabras, la probabilidad de elegir la segunda categoría disponible en el ítem j (que en este caso resulta ser la “incorrecta”), estaríamos transformando el típico caso dicotómico a un caso politómico de dos categorías.

Siguiendo el hilo de razonamiento anterior, podríamos generalizar al caso de m_j categorías disponibles. Siendo $P_{jh}(\boldsymbol{\theta}) = [P_{jh}(\theta_1), \dots, P_{jh}(\theta_n)]^T$ el vector de probabilidades de la elección de la alternativa h en el ítem j para los n examinados, podemos definir la **Función de Información de la Categoría del Ítem (ICIF)**; función que determina la cantidad de información utilizada para estimar la localización del individuo en el continuo de la competencia, provista por la elección particular de la categoría en el ítem (De Ayala, 1993):

$$I_{jk}(\boldsymbol{\theta}) = -E \left[\frac{\partial^2 \log P_{jk}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2} \right] = -\frac{\partial}{\partial \boldsymbol{\theta}} \left[\frac{P'_{jk}(\boldsymbol{\theta})}{P_{jk}(\boldsymbol{\theta})} \right] \quad (2.12)$$

donde $P'_{jk}(\boldsymbol{\theta}) = \partial P_{jk}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$. Continuando con el despeje de la ecuación tenemos que,

$$I_{jk}(\boldsymbol{\theta}) = \frac{[P'_{jk}(\boldsymbol{\theta})]^2 - P_{jk}(\boldsymbol{\theta})P''_{jk}(\boldsymbol{\theta})}{P_{jk}(\boldsymbol{\theta})^2} = \left[\frac{P'_{jk}(\boldsymbol{\theta})}{P_{jk}(\boldsymbol{\theta})} \right]^2 - \left[\frac{P''_{jk}(\boldsymbol{\theta})}{P_{jk}(\boldsymbol{\theta})} \right] \quad (2.13)$$

donde $P''_{jk}(\boldsymbol{\theta}) = \partial^2 P_{jk}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}^2$. Después de algunos despejes la ecuación (2.13) quedaría expresada de la siguiente forma:

$$I_{jk}(\boldsymbol{\theta}) = \left[\frac{\sum_{h=1}^{m_j} e^{z_{jh}(\boldsymbol{\theta})} \cdot (a_{jk} - a_{jh})}{D} \right]^2 - \frac{\sum_{h=1}^{m_j} e^{z_{jh}(\boldsymbol{\theta})} \cdot (a_{jk}^2 - a_{jh}^2)}{D} + \frac{2 \cdot \sum_{h=1}^{m_j} a_{jh} \cdot e^{z_{jh}(\boldsymbol{\theta})} \cdot \left[\sum_{h=1}^{m_j} e^{z_{jh}(\boldsymbol{\theta})} \cdot (a_{jk} - a_{jh}) \right]}{D^2} \quad (2.14)$$

donde $D = \sum_{h=1}^{m_j} e^{z_{jh}(\boldsymbol{\theta})}$. Para mayor detalle a las derivaciones asociadas con las funciones de información referirse al apéndice A.2.

Sin embargo, los examinados pueden elegir cualquiera de las alternativas disponibles en el ítem con probabilidad $P_{jh}(\boldsymbol{\theta})$, por lo que la cantidad de información con la que la categoría k contribuye al ítem j , estaría dada por $I_{jk}(\boldsymbol{\theta}) \cdot P_{jk}(\boldsymbol{\theta})$. De este modo en el NRM, la **Función de Información del Ítem (IIF)** se define de la siguiente manera:

$$I_j(\boldsymbol{\theta}) = \sum_{h=1}^{m_j} I_{jh}(\boldsymbol{\theta})P_{jh}(\boldsymbol{\theta}) \quad (2.15)$$

Finalmente, gracias a la propiedad aditiva de las funciones de información, el total de información provista por el instrumento de evaluación, la **Función de Información del Test (IF)**, estaría definida por la suma de todas las IIF individuales, de la siguiente manera:

$$I(\boldsymbol{\theta}) = \sum_{j=1}^p I_j(\boldsymbol{\theta}) = \sum_{j=1}^p \sum_{h=1}^{m_j} I_{jh}(\boldsymbol{\theta})P_{jh}(\boldsymbol{\theta}) \quad (2.16)$$

Un ejemplo de IIF y contribuciones de información de las categorías del ítem se observa en la Figura 2.2.

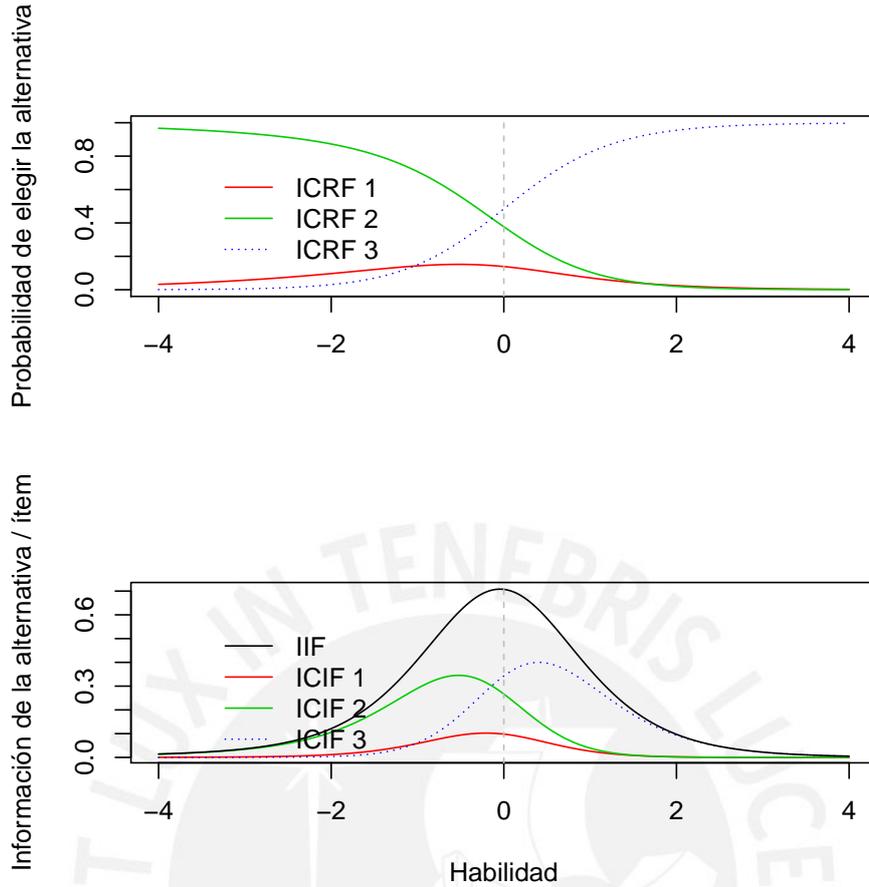


Figura 2.2: Funciones de Respuestas de las Categorías de un Ítem (ICRF), Función de Información del ítem (IIF) y Contribución de Información de las Categorías de un Ítem (ICIF) con interceptos $\mathbf{c}_j = (-0.75; 0.25; 0.5)$ y pendientes $\mathbf{a}_j = (-0.2; -0.8; 1)$. Basado en un ejemplo de Baker y Kim (1992).

Bock (1972) presentó las funciones de información en términos de álgebra matricial, donde define a la **ICIF** de la categoría k del ítem j para el individuo i como:

$$I_{jk}(\theta_i) = \begin{matrix} (1 \times 1) \\ \left[\begin{matrix} \mathbf{a}_j & \cdot & \mathbf{W}_{ij} & \cdot & \mathbf{a}_j^T \\ \hline (1 \times m_j) & & (m_j \times m_j) & & (m_j \times 1) \end{matrix} \right] & \cdot & P_{jk}(\theta_i) \\ (1 \times 1) \end{matrix} \quad (2.17)$$

donde

$$\mathbf{W}_{ij} = \begin{pmatrix} P_{j1}(\theta_i)^T \cdot (1 - P_{j1}(\theta_i)) & \cdots & -P_{j1}(\theta_i)^T \cdot P_{jk}(\theta_i) & \cdots & -P_{j1}(\theta_i)^T \cdot P_{jm_j}(\theta_i) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ -P_{jk}(\theta_i)^T \cdot P_{j1}(\theta_i) & \cdots & P_{jk}(\theta_i)^T \cdot (1 - P_{jk}(\theta_i)) & \cdots & -P_{jk}(\theta_i)^T \cdot P_{jm_j}(\theta_i) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ -P_{jm_j}(\theta_i)^T \cdot P_{j1}(\theta_i) & \cdots & -P_{jm_j}(\theta_i)^T \cdot P_{jk}(\theta_i) & \cdots & P_{jm_j}(\theta_i)^T \cdot (1 - P_{jm_j}(\theta_i)) \end{pmatrix}$$

Finalmente, dada la propiedad aditiva de las funciones de información, se tiene que la **IIF** del ítem j para cada individuo i , donde $i = 1, \dots, n$, toma la siguiente forma:

$$\begin{aligned}
I_j(\theta_i) &= \sum_{h=1}^{m_j} I_{jh}(\theta_i) \\
&= \mathbf{a}_j \cdot \mathbf{W}_{ij} \cdot \mathbf{a}_j \cdot \sum_{h=1}^{m_j} P_{jh}(\theta_i) \\
&= \mathbf{a}_j \cdot \mathbf{W}_{ij} \cdot \mathbf{a}_j
\end{aligned} \tag{2.18}$$

Es importante mencionar, que la ecuación (2.18) es igual a la ecuación (2.15), que tras el correcto despeje algebraico.

2.4. Relación con otros modelos de la Teoría de Respuesta al Ítem

A diferencia de los modelos para respuestas dicotómicas, como el Rasch (1980), de uno, dos, tres parámetros (Lord y Novik, 2008) y cuatro parámetros (McDonald, 1967), que expresan la probabilidad de elegir la alternativa “correcta” en función de la competencia del individuo; el NRM expresa la probabilidad de elegir una de las alternativas del ítem en función de la misma competencia (ecuación 2.2). Se espera que al igual que los modelos para respuestas dicotómicas, la curva asociada con la respuesta “correcta” sea monótonicamente creciente respecto a la competencia, mientras que la forma de las ICRFs restantes dependerá de como el evaluado perciba la alternativa (Baker y Kim, 1992; Hambleton y Swaminathan, 1991), tal como se observa en la Figura 2.2.

Es importante recalcar, que no se puede afirmar que un modelo NRM con dos alternativas es igual a un modelo de respuesta dicotómica, ni mucho menos que los modelos dicotómicos son un caso especial del NRM. Esto se debe a que la función logística multivariada definida mediante la ecuación (2.3) no se puede reducir a una función logística univariada, difieren entre ellas por un factor de 2; es decir, tanto el parámetro del intercepto como el de discriminación son la mitad de lo que representan los mismos parámetros en una función logística univariada, utilizada en los modelos para ítems dicotómicos (para mayor detalle respecto a la afirmación referirse a Baker y Kim, 1992, Capítulo 9, p. 256).

En relación con los modelos politómicos, Thissen y Steinberg (1986) los clasifican de manera general en tres tipos: (i) la familia de los modelos en diferencia, (ii) la familia de los modelos *divide-by-total* y (iii) la familia de modelos *left-side added divide-by-total*. Por su parte, Mellenberg (1995) clasifica a los ítems, sobre los cuales se pueden ajustar un modelo, según el número de categorías de respuestas y si estas poseen o no un ordenamiento intrínseco. El autor señala que los ítems pueden presentar dos categorías de respuestas (dicotómicos) o más de dos (politómicos), y que estas a su vez, pueden ser del tipo nominales (sin orden) u ordinales. De este modo, se pueden tener cuatro tipos de ítems: Nominales-Dicotómicos (ND), Ordinales-Dicotómicos (OD), Nominales-Politómicos (NP) y Ordinales-Politómicos (OP).

En relación a la división establecida por Thissen y Steinberg (1986), el NRM no se sustenta

sobre el concepto de la dicotomización de las alternativas, que derivan en los umbrales por categorías característicos de la familia de modelos en diferencia, a los que pertenecen los modelos de respuesta graduada (GRM; Samejima, 1970, 1972, 1991); por el contrario, la probabilidad correspondiente a cada alternativa es modelada directamente, implementando una generalización multivariada del modelo de rasgos latentes logístico (Bock 1972, Ostini y Nering 2006). La forma de la ecuación general, planteada en (2.2), es familiar debido a sus similitudes con la ecuación general de los modelos Rasch politómicos (Masters y Wright, 1984), dándonos indicios de que el NRM es uno más de los modelos clasificados como *divide-by-total*. Los autores señalaron incluso que los modelos *divide-by-total* son, de hecho, versiones restringidas del NRM presentado por Bock.

Mellenberg (1995) por su parte, desarrollando el modelo bajo un enfoque unificado, usando métodos y restricciones que preserven el orden de las categorías cuando sea requerido, mostró que el NRM se puede adaptar no solo a ítems del tipo Nominal-Politómico (NP), acorde con la propuesta original de Bock, sino a cualquier otro tipo (OD, ND y OP). De este modo, es posible expresar los modelos politómicos más comunes (PC, GPC, entre otros) como casos especiales del modelo nominal.

2.5. Extensiones al Modelo

En esta sección se presentan dos extensiones al NRM original propuesto por Bock (1972, 1997).

2.5.1. Modelo de Respuesta Nominal con “adivinación”

Propuesto por Samejima (1979) incorpora el factor de “adivinación”, al igual que el modelo de tres parámetros (3-PL) propuesto por Lord y Novik (2008). Bajo esta propuesta una porción d_k (de individuos) elige una de las alternativas del ítem mediante la adivinación, de este modo la autora define las ICRFs como:

$$P_{jk}(\theta_i) = \frac{e^{z_{jk}(\theta_i)} + d_k e^{z_{j0}(\theta_i)}}{\sum_{h=0}^{m_j} e^{z_{jh}(\theta_i)}} \quad (2.19)$$

donde $d_k = 1/m_j$ (que es un parámetro no estimado). La nueva categoría “No sé” (DK, por su siglas en inglés), denotada por el subíndice cero, es latente y dependiente de la competencia del individuo; y $z_{j0}(\theta_i) = c_0 + a_0 \cdot \theta_i$ representa el logit multivariado asociado al comportamiento “adivinator”. La principal asunción de la autora es que los examinados con competencias bajas eligen de manera aleatoria entre las m_j alternativas disponibles con igual probabilidad.

Thissen y Steinberg (1984, 1986) encontraron que la asunción de aleatoriedad e igualdad de probabilidades de elección de alternativas no se encontraba soportada empíricamente, por lo que desarrollaron una extensión al modelo de Samejima, donde d_k es una función de otros parámetros d_k^* que serán estimados. Debido a que d_k representa un set de m_j proporciones, la suma de los mismos debe ser restringida dentro del intervalo $[0, 1]$, mediante una función

logística multivariada,

$$d_k = \frac{e^{d_k^*}}{\sum_{h=0}^{m_j} e^{d_h^*}} \quad (2.20)$$

y los parámetros d_k^* , al igual que a_{jk} y c_{jk} , deben satisfacer $\sum_{h=1}^{m_j} d_h^* = 0$.

Los autores señalaron que el modelo planteado por las ecuaciones (2.19) y (2.20), incluye de manera jerárquica a los modelos previos. Si $d_k = 1/m_j$ la ecuación (2.19) se convierte en el modelo de Samejima; mientras que si $a_0 = c_0 = 0$, el parámetro d_k en la ecuación (2.19) se elimina mecánicamente, convirtiéndose en el NRM propuesto por Bock.

Si bien los autores encontraron atractivo al modelo, señalaron que la interpretación semántica de los parámetros era más compleja, y en mucho de los casos era mejor representada gráficamente. Por su parte, comentaron también que el modelo requiere la estimación de un mayor número de parámetros, llevándolo incluso a convertirse en un posible caso de sobreparametrización. Encontraron dificultoso aplicar el modelo a grandes programas de evaluación, debido a que el proceso de eliminar las redundancias en la parametrización era tedioso y propenso al error. Finalmente, resaltaron que el proceso de estimación presentaba múltiples retos y problemas.

2.5.2. Nueva Parametrización del Modelo de Respuesta Nominal

Propuesto por Thissen et al. (2010), su desarrollo se encuentra guiado principalmente por tres objetivos:

- **Facilidad de interpretación**, la re-parametrización planteada debería facilitar no solo la interpretación de los parámetros, sino que también haga que algunos de ellos, por ser únicos (el parámetro de discriminación a_h^* por ejemplo), sean comparables a símiles de otros modelos IRT.
- **Uso de modelos intermedios**, mediante la implementación de una Matriz de Transformación de Base Fourier la nueva parametrización permite acceder a modelos que se encuentren entre el modelo nominal de rango completo (Bock, 1972, 1997) y un modelo de Crédito Parciales Generalizados (GPC, por sus siglas en inglés) altamente restringido.
- **Extensión a un modelo multidimensional**, mediante una re-parametrización del vector logit multivariado los autores plantean que el modelo es extensible al caso multidimensional.

De esta forma, los autores representan las curvas ICRFs por:

$$P_{jk}(\theta_i) = \frac{e^{z_{jk}(\theta_i)}}{\sum_{h=1}^{m_j} e^{z_{jh}(\theta_i)}} \quad (2.21)$$

donde $z_{jh}(\theta_i) = c_{jh} + a_j^* \cdot a_{jh}^s \cdot \theta_i$. Por su parte, a_j^* es el parámetro de discriminación general del ítem j , a_{jh}^s denota la función de puntuación de la alternativa h en el ítem j y c_{jh} denota

al parámetro de intercepto, tal como lo denota en el modelo original.

Esta parametrización ayuda a identificar el ordenamiento empírico de las categorías mediante solo la inspección de los valores a_{jh}^s . Valores altos para estos parámetros indican que la categoría del ítem se encuentra altamente relacionada con el/los rasgos latentes medidos. Por ejemplo, si un ítem fuera realmente ordinal (como en un escala de Likert), y tuviera cuatro categorías de respuesta, se esperaría ver un ordenamiento del tipo $a_{j1}^s < a_{j2}^s < a_{j3}^s < a_{j4}^s$ posterior a la estimación.

Con el propósito de obtener los estimados para los parámetros, los autores deben asegurar la identificación del vector logit multivariado anclando algunos de los valores que componen el vector. Siguiendo una propuesta similar a la de Thissen (1991), los autores establecen $a_{j1}^s = c_{j1} = 0$, añadiendo a la misma que $a_{j(m_j)}^s = m - 1$. Por su parte, siguiendo la propuesta de Bock (1972), replantean los parámetros originales en función de parámetros “libres” γ_j y α_j , y la matriz de transformación T_j de la siguiente manera:

$$\begin{bmatrix} c_j \\ a_j^s \end{bmatrix} = \begin{bmatrix} \gamma_j \\ \alpha_j \end{bmatrix} \mathbf{T}_j \quad (2.22)$$

donde la matriz de transformación T_j se encuentra conformada por un vector columna lineal y una matriz de base Fourier de la siguiente forma:

$$\mathbf{T}_j = \mathbf{T}_F = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 1 & f_{22} & \cdots & f_{2(m_j-1)} \\ 2 & f_{32} & \cdots & f_{3(m_j-1)} \\ \vdots & \vdots & \ddots & \vdots \\ m-1 & 0 & \cdots & 0 \end{pmatrix}$$

en la cual $f_{kj} = \sin[\pi(j-1)(k-1)/(m_j-1)]$.

De esta forma, la nueva parametrización de la matriz de transformación \mathbf{T}_j permite obtener numerosas variantes útiles del modelo nominal:

- (a) Cuando a_j^* , $\alpha_1, \dots, \alpha_{m_j-1}$ y γ_j son estimados, se obtiene el modelo nominal de rango completo;
- (b) Si $\alpha_1, \dots, \alpha_{m_j-1} = 0$, el modelo se convierte en una versión reparametrizada del modelo GPC, y
- (c) las bases de Fourier proveen un modo para crear modelos que se encuentren entre el GPC y el modelo nominal, tal como fue planteado por Thissen et al. (1989).

Finalmente, siguiendo la parametrización descrita, los autores indicaron que el modelo se puede extender al caso multidimensional, al representar al vector logit multivariado de la siguiente forma: $z_{ij} = c_j + \mathbf{a}'_j^* \cdot \mathbf{a}_{jh}^s \cdot \boldsymbol{\theta}^*$. El valor \mathbf{a}'_j^* denota el vector de parámetros de discriminación, valores que indican la pendiente en cada dirección del espacio $\boldsymbol{\theta}^*$ -dimensional.

Por su parte, \mathbf{a}_{jh}^s representan las funciones de puntuación de las categorías y se asumen las mismas en todas las direcciones de espacio θ^* -dimensional. Por último, el valor \mathbf{c}_j representa el vector de parámetros de interceptos.

Es importante mencionar que la presente parametrización, desarrollada para un modelo de dos dimensiones, se encuentra implementada en el paquete MIRT (Chalmers, 2012).



Capítulo 3

Estimación Bayesiana

El uso práctico del NRM implica la estimación de todos los parámetros asociados a los ítems y sus alternativas, y al menos un atributo cuantitativo de competencia por cada examinado. Con este propósito en mente, la estimación del modelo puede ser realizada mediante:

- Métodos clásicos o frecuentistas, entre los que se encuentran los métodos de Máxima Verosimilitud Conjunta (JML), Máxima Verosimilitud Condicional (CML) y Máxima Verosimilitud Marginal (MML).
- Inferencia bayesiana, usando los métodos MCMC.

Para mayor detalle acerca del desarrollo teórico y práctico de los métodos de estimación clásica del NRM referirse a Bock (1972, 1997), Baker y Kim (1992) o Azevedo (2003).

En este trabajo desarrollaremos la metodología de inferencia Bayesiana usando métodos MCMC. Así mismo, es descrita también la implementación computacional necesaria para la estimación de los parámetros de interés, usando R (R Core Team, 2015) y WinBUGS (Lunn et al., 2000).

3.1. Estimación Bayesiana del NRM

De manera preliminar, se creyó conveniente resaltar algunas razones por las cuales es atractivo estimar los parámetros del modelo para datos politómicos nominales mediante los métodos bayesianos:

1. Acorde con Hambleton et al. (1991), elimina los problemas de no convergencia y estimación impropia de los parámetros encontrados en los procedimientos de Máxima Verosimilitud (ML), especialmente en los métodos de Máxima Verosimilitud Conjunta (JML, por sus siglas en inglés) y Máxima Verosimilitud Marginal (MML, por sus siglas en inglés) ;
 - a) Los métodos ML presentan problemas de estimación cuando los individuos tienen *scores* nulos o perfectos, o cuando presentan patrones de respuestas aberrantes (Azevedo, 2003),
 - b) En general los métodos bayesianos presentan igual, y en algunos casos mejor, capacidad de recuperación de los parámetros de interés (Baker, 1998; Wollack

et al., 2002; Hsieh et al., 2010); dependiendo de las características de los datos e ítems.

2. Baker (1998) y Kim y Cohen (1999) señalaron que los métodos MCMC podrían ser más efectivos frente a modelos donde la complejidad se incrementa o ante modelos altamente parametrizados, pues usa simulaciones en vez de métodos numéricos. Fox (2010) secunda la afirmación, al resaltar que el método también puede manejar todo tipo de prioris y modelos generadores de datos;
3. Los modelos MCMC se vuelven particularmente útiles cuando los datos son dispersos o cuando es poco probable que la teoría asintótica se mantenga (Fox, 2010); y
4. Presenta la posibilidad de incorporar, en el análisis, información no presente en los datos, más allá de las respuestas observadas (Fox, 2010).

3.1.1. Función de Verosimilitud

Considerando las ecuaciones (2.2) y (2.3), y el cumplimiento de los supuestos (S1), (S2), (S3) y (S4) referidos en la sección 2.2, la función de verosimilitud puede ser escrita como:

$$L(x_{..}|\boldsymbol{\theta}, \mathbf{c}_j, \mathbf{a}_j) = f(x_{..}|\boldsymbol{\theta}, \mathbf{c}_j, \mathbf{a}_j) = \prod_{i=1}^n \prod_{j=1}^p \prod_{h=1}^{m_j} [P_{jh}(\theta_i)]^{1(x_{ij}=h)} = \prod_{i=1}^n \prod_{j=1}^p \prod_{h=1}^{m_j} \left[\frac{e^{z_{jh}(\theta_i)}}{\sum_{h=1}^{m_j} e^{z_{jh}(\theta_i)}} \right]^{1(x_{ij}=h)} \quad (3.1)$$

3.1.2. Especificación de Prioris

Similar a lo planteado por Patz y Junker (1999), Bazán (2005) y Tarazona (2013), para la especificación de las prioris conjuntas de los parámetros de las personas e ítems, se considera que existe una estructura de distribuciones independientes, de la siguiente forma:

$$f(\boldsymbol{\theta}, \mathbf{c}_j, \mathbf{a}_j|x_{..}) = f(\boldsymbol{\theta}) \cdot f(\mathbf{c}_j, \mathbf{a}_j|x_{..}) = \prod_{i=1}^n f(\theta_i) \cdot \prod_{j=1}^p \prod_{k=1}^{m_j} f(a_{jk}) \cdot f(c_{jk}) \quad (3.2)$$

Bazán (2005) y Tarazona (2013) recomiendan que se establezcan prioris informativas y no impropias, pues la existencia de una distribución conjunta a posteriori no se encuentra garantizada cuando se establecen prioris impropias. En el caso del NRM, se mantiene el requerimiento de que la prior no sea impropia; sin embargo, ya no es requerido que la prior sea informativa. Ejemplos donde se alcanza la convergencia de los parámetros con prioris no informativas son Wollack et al. (2002) y Hsieh et al. (2010).

Es importante señalar, que a diferencia de los modelos dicotómicos o algunos de los modelos politómicos ordenados (como el GRM), el NRM ya no necesita imponer restricciones a ninguno de sus parámetros. Muchos de los modelos dicotómicos requieren que su parámetro de discriminación sea positivo, mientras que en el caso del modelo GRM la restricciones debían asegurar que se mantenga el ordenamiento de las categorías. El NRM no necesita cumplir con ningún requerimiento específico en sus parámetros.

De este modo tenemos:

- Bazán (2005) considera que para el vector de “habilidades” de los examinados se puede utilizar una distribución normal con componentes independientes,

$$\theta_i \sim N(0, 1) \quad (3.3)$$

- Las prior para los parámetros a_{jk} y c_{jk} serán establecidas de manera similar a lo establecido por Wollack et al. (2002); debido a que los autores realizaron un estudio con similares propósitos al presente trabajo usando el mismo modelo. De esta forma, se utilizarán las siguientes prioris para los parámetros de pendiente e intercepto:

$$a_{jk} \sim N(0, 100) \quad (3.4)$$

$$c_{jk} \sim N(0, 100) \quad (3.5)$$

Es importante señalar que Wollack et al. (2002) también estableció prioris con media igual a cero, pero con una varianza mayor.

El cuadro 3.1 muestra una lista de las prioris establecidas en estudios relacionados con modelos politómicos del tipo NRM o GRM, tanto como modelos dicotómicos:

Autor	Parámetros		
	a_{ijk}	c_{ijk}	θ
Patz y Junker (1999)	$N(2, 0.5)$	n.d.	$N(0, 1)$
Wollack et al. (2002)	$N(0, 100^2)$	$N(0, 100^2)$	$N(0, 1)$
Azevedo (2003)	$NM(0, 1)$	$NM(0, 1)$	$N(0, 1)$
Bazán (2005)	$N(0, 1)$	n.d.	$N(0, 1)$
Hsieh et al. (2010)	$LN(0, 100^2)$	n.d.	$N(0, 1)$
Kieftenbeld y Natesan (2012)	$LN(0, \sqrt{2})$	n.d.	$N(0, 1)$

Cuadro 3.1: Prioris establecidas para la estimación de ítems politómicos. $N(., .)$ se refiere a la distribución normal, $NM(., .)$ a la distribución Normal Multivariada y $LN(., .)$ a la distribución log-normal. Para mayor detalle ver listado de abreviaturas.

3.1.3. Especificación de Valores Iniciales

Para poder iniciar el método MCMC, se necesita establecer los *valores de partida* de las iteraciones o cadenas; es decir, proveer los *valores iniciales* para cada parámetro desconocido en el modelo de interés. El presente trabajo optó porque el software proponga los valores iniciales de las prioris “no informativas” propuestas para los interceptos c_{jk} , pendientes a_{jk} y competencias θ_i . En el caso de la estimación clásica, la función `nrm` del paquete `mcIRT` (Reif, 2014) estableció un valor de partida de -0.1 para todos los parámetros;

Los resultados reportados en el Apéndice C.1 indican que la estimación de los parámetros presentaron convergencia en sus cadenas.

3.1.4. Distribución a posteriori de los parámetros

Considerando la función de verosimilitud dada en (3.1) y las prioris consideradas en (3.2), (3.3), (3.4) y (3.5), definimos la distribución a posteriori de los parámetros como:

$$\begin{aligned}
 f(\boldsymbol{\theta}, \mathbf{c}_j, \mathbf{a}_j | x_{..}) &\propto f(x_{..} | \boldsymbol{\theta}, \mathbf{c}_j, \mathbf{a}_j) \cdot f(\boldsymbol{\theta}, \mathbf{c}_j, \mathbf{a}_j) \\
 &\propto \left[\prod_{i=1}^n f(\theta_i) \cdot \prod_{j=1}^p \prod_{k=1}^{m_j} f(a_{jk}) \cdot f(c_{jk}) \right] \left[\prod_{i=1}^n \prod_{j=1}^p \prod_{h=1}^{m_j} [P_{jh}(\theta_i)]^{1(x_{ij}=h)} \right] \\
 &\propto \left[\prod_{i=1}^n f(\theta_i) \cdot \prod_{j=1}^p \prod_{k=1}^{m_j} f(a_{jk}) \cdot f(c_{jk}) \right] \left[\prod_{i=1}^n \prod_{j=1}^p \prod_{h=1}^{m_j} \left[\frac{e^{z_{jh}(\theta_i)}}{\sum_{h=1}^{m_j} e^{z_{jh}(\theta_i)}} \right]^{1(x_{ij}=h)} \right]
 \end{aligned} \tag{3.6}$$

La distribución a posteriori dada en la ecuación (3.6) 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 usando el algoritmo de Metrópolis-Hastings, que es implementado por defecto en el software WinBUGS.

Otros trabajos encontraron distribuciones a posteriori que se asemejan, pero que no son iguales, a distribuciones conocidas (ver Azevedo, 2003).

3.2. Implementación Computacional

El estudio mantuvo presente que existen múltiples programas de inferencia bayesiana, entre ellos `OpenBUGS` (Lunn et al., 2009), `Stan` (Carpenter et al., 2016), `WinBUGS` (Lunn et al., 2000), entre otros. En este trabajo, la implementación del método MCMC de inferencia bayesiana fue implementada por `WinBUGS` mientras que la gestión de la simulación fue implementada mediante el programa `R` (R Core Team, 2015), y su librería `R2WinBUGS` (Sturtz et al., 2005).

Las líneas de código requeridas, para implementar el método MCMC de la inferencia bayesiana del modelo NRM, se muestran en los apéndices B.3 y B.2.

Capítulo 4

Estudio de Simulación

4.1. Objetivos

- Simular múltiples conjuntos de datos politómicos nominales generados a partir del Modelo de Respuesta Nominal (NRM).
- Comparar la capacidad de recuperación de los parámetros de los ítems y competencias de los individuos del método clásico de estimación (MML) frente el bayesiano (MCMC), en el contexto del modelo de interés.
- Investigar el ratio *tamaño de muestra entre número de ítems* y el papel que juega en la recuperación de los parámetros, en el contexto del modelo de interés.

4.2. Simulación

4.2.1. Condiciones de la simulación

Siguiendo una propuesta cercana a la de Ayala y Sava-Bolesta (1999), se definieron los meta-parámetros de la simulación:

- los ratios de *tamaño de muestra : número de ítems* ($n : p$) utilizados en la simulación fueron de 2.5 : 1 y 20 : 1;
- el número de ítems (p) o longitud de la evaluación fue de 10 ítems; y
- el número de alternativas (m_j) disponibles por ítem fue de 4.

Con la información provista se procedió a calcular el tamaño de muestra, donde el número de observaciones o individuos (n) simulados se calculó mediante la siguiente fórmula:

$$n = \text{ratio} \times p \times m_j \times \text{número de parámetros}$$

Los parámetros de competencia (θ_i) asociados a los examinados fueron generados de una distribución Normal Estándar; mientras que los parámetros a_{jk} y c_{jk} (pendiente e intercepto, respectivamente) fueron extraídos del trabajo de Wollack et al. (2002).

Tanto los parámetros de competencias como los de pendientes e interceptos se mantuvieron fijos a lo largo de todas las réplicas de simulación. Es importante señalar, que el uso

de ítems calibrados (provenientes de un estudio previo), intentó asegurar que la simulación sea lo más real posible (Harwell et al., 1996); mientras que mantener fijo los parámetros de competencias, pendientes e interceptos, a través de todas réplicas de simulación, fue realizado con la intención de minimizar el error experimental (Kieftenbeld y Natesan, 2012).

El Cuadro 4.1 muestra los valores para cada uno de los parámetros de los 10 ítems elegidos.

Ítems	Pendientes (a_{jk})				Interceptos (c_{jk})			
	1	2	3	4	1	2	3	4
1	-1.14	-0.15	0.15	1.14	-1.65	0.16	0.92	0.57
2	1.25	0.07	0.42	-1.74	1.73	0.38	0.62	-2.73
3	1.09	0.69	-0.20	-1.58	1.20	-0.26	0.34	-1.28
4	0.30	-1.17	1.01	-0.14	0.20	-0.98	0.74	0.04
5	0.94	0.42	-0.50	-0.86	0.70	-0.73	0.11	-0.08
6	-0.21	-1.07	0.39	0.89	0.30	0.10	-0.84	0.44
7	-0.89	-0.29	-0.51	1.69	-0.10	0.34	0.33	-0.57
8	-0.32	-0.80	0.99	0.13	0.78	-0.87	-0.22	0.31
9	-0.42	-0.62	0.73	0.31	0.50	1.40	-0.97	-0.93
10	-0.53	1.62	-0.55	-0.54	0.02	-1.86	-0.67	2.51

Cuadro 4.1: Parámetros de los ítems usados en la simulación

Así, de la combinación de los distintos meta-parámetros (tamaño de muestra, longitud de evaluación, alternativas disponibles por ítem, parámetros de los ítems y parámetros de las competencias) se obtuvieron dos escenarios que se detallan en el Cuadro 4.2.

Diseño	Distribución (Competencia)	Longitud (Evaluación)	Alternativas (número)	Ratio ($n : p$)	Muestra (Tamaño)
1	$N(0, 1)$	10	4	2.5 : 1	200
2	$N(0, 1)$	10	4	20 : 1	1600

Cuadro 4.2: Diseños de simulaciones: distribuciones, tamaños de muestras y longitud de evaluaciones

4.2.2. Algoritmo de simulación

Para realizar la simulación de los datos se procedió a seguir los siguientes pasos:

1. **Simulación de habilidades:** se generaron los parámetros de competencia de una distribución normal estándar, $\theta_i \sim N(0, 1)$.
2. **Cálculo de probabilidades:** se calcularon las probabilidades para los parámetros originales, así como para cada réplica de cada uno de los diseños de simulación. La probabilidad $P_{jk}(\theta_i)$ es definida como en las ecuaciones (2.2) y (2.3).
3. **Elección de Alternativas:** Se simula la alternativa elegida por cada uno de los n examinados en cada uno de los p ítems. Con este propósito se toma una muestra alea-

toria simple de tamaño uno, para una población de m_j elementos, con probabilidades iguales a las calculadas en el paso anterior. Esto es repetido n veces.

Las líneas de código asociadas a la simulación de los datos politómicos nominales basados en el NRM, se encuentra en el apéndice B.1.

4.2.3. Criterios para Evaluar la Simulación

La precisión de los estimados fue evaluada usando la Raíz del Error Cuadrático Medio (RMSE) y el Error Absoluto Medio (MAE) para cada réplica (acorde con los diseños establecidos en 4.2.1).

El RMSE es definido como la raíz cuadrada del promedio de las diferencias entre los valores reales y los estimados al cuadrado, de la siguiente manera:

$$RMSE(a_{jk}) = \sqrt{\frac{1}{R} \sum_{r=1}^R (\hat{a}_{jkr} - a_{jk})^2} \quad (4.1)$$

$$RMSE(c_{jk}) = \sqrt{\frac{1}{R} \sum_{r=1}^R (\hat{c}_{jkr} - c_{jk})^2} \quad (4.2)$$

$$RMSE(\theta_i) = \sqrt{\frac{1}{R} \sum_{r=1}^R (\hat{\theta}_{ir} - \theta_i)^2} \quad (4.3)$$

Del mismo modo, se calculó el MAE de la siguiente manera:

$$MAE(a_{jk}) = \frac{1}{R} \sum_{r=1}^R |\hat{a}_{jkr} - a_{jk}| \quad (4.4)$$

$$MAE(c_{jk}) = \frac{1}{R} \sum_{r=1}^R |\hat{c}_{jkr} - c_{jk}| \quad (4.5)$$

$$MAE(\theta_i) = \frac{1}{R} \sum_{r=1}^R |\hat{\theta}_{ir} - \theta_i| \quad (4.6)$$

donde \hat{a}_{jkr} , \hat{c}_{jkr} y $\hat{\theta}_{ir}$ son los estimados de los parámetros de los ítems y competencias de los evaluados en la réplica $r = 1, \dots, R$, respectivamente.

Tal y como se observa en las fórmulas, los criterios serán aplicados a los parámetros de ítem e individuos. Sin embargo, dado que en el NRM, el impacto de cada categoría depende ampliamente de los valores y comportamiento de las otras, en determinados casos se podría observar diferencias entre el parámetro real y estimado que no necesariamente corresponderían con un error en la estimación de la elección de la categoría por parte del individuo (Wollack et al., 2002). Así, para corregir este inconveniente se comparará no solo los estimados de los parámetros de los ítems e individuos, sino también las probabilidades condicionales asociadas a la elección de cada alternativa específica (Yen, 1987; Wollack et al.,

2002). De este modo, dado que las probabilidades $P_{jk}(\theta_i)$ difieren entre sí, no solo por los parámetros de los ítems (a_{jk} y c_{jk}) sino también en función de las habilidades (θ_i), para el análisis de las diferencias entre las probabilidades estimadas y las reales, se planteó utilizar los dos criterios (RMSE y MAE) con algunos ajustes:

$$RMSE(P_{jk}) = \sqrt{\frac{1}{n} \sum_{i=1}^n \left[\frac{1}{R} \sum_{r=1}^R (\hat{P}_{jkr}(\theta_i) - P_{jk}(\theta_i))^2 \right]} \quad (4.7)$$

$$MAE(P_{jk}) = \frac{1}{n} \sum_{i=1}^n \left[\frac{1}{R} \sum_{r=1}^R |\hat{P}_{jkr}(\theta_i) - P_{jk}(\theta_i)| \right] \quad (4.8)$$

donde $\hat{P}_{jkr}(\theta_i)$ es el estimado de las probabilidades condicionales en la réplica $r = 1, \dots, R$.

Como se observa en las formulas (4.7) y (4.8), las diferencias entre probabilidades se agregan y promedian a nivel de replicas y posteriormente a nivel de individuos en cada una de las réplicas ($R = 20$). De esta manera, se tienen una medida global de ajuste a los datos.

4.3. Estimación de parámetros

4.3.1. Métodos

Los parámetros del NRM fueron estimados, en cada una de las réplicas, usando el método MML mediante el paquete `mcIRT` (Reif, 2014) y el método MCMC implementado mediante el programa `WinBUGS` (Lunn et al., 2000). En el caso del método MCMC, se usó como estimador puntual de los parámetros al promedio de los parámetros estimados provenientes de las distribuciones a posteriori. Por su parte, se usó el error estándar de las cadenas generadas de los parámetros como una aproximación del error estándar de la estimación.

La implementación computacional para la inferencia bayesiana del modelo en `WinBUGS` se encuentra en el apéndice B.3 y la función en R, de la librería `R2WinBUGS`, que permite ejecutarla en el apéndice B.2.3 y B.2.4.

4.3.2. Réplicas, Cadena, Burn-In y saltos en MCMC

Se decidió establecer 20 réplicas ($R = 20$) para los dos diseños de estudios de simulación detallados en la sección 4.2.1.

Se realizó un breve estudio preliminar que buscó determinar el largo de la cadena (número de iteraciones) que asegure la convergencia de las simulaciones MCMC, asegurando así mismo una estimación más precisa de los parámetros. El estudio determinó que una cadena de 37000 iteraciones eran suficientes para minimizar la auto-correlación en la serie. Las primeras 3000 iteraciones fueron descartadas (proceso conocido como *burn-in*), mientras que las siguientes 34000 fueron reducidas en saltos de 17 (proceso conocido como *thinning*). Este proceso generó una cadena de 2000 iteraciones efectivas, que sirvieron para estimar las distribuciones a posteriori de los parámetros.

Finalmente, es importante mencionar que el proceso de generación de cadenas MCMC

fue paralelizado mediante los paquetes `foreach` (Weston, 2015) y `doParallel` (Weston y Calaway, 2015), reduciendo de esta manera el tiempo requerido por el proceso.

4.4. Resultados

4.4.1. Diseño 1:

Distribución Normal Estándar, $p = 10$, $m_j = 4$, $n/p = 2.5 : 1$ y $n = 200$

Los criterios RMSE y MAE para las estimaciones de los parámetros de cada ítem son mostrados en los Cuadros 4.3 y 4.4.

Bajo el primer diseño de simulación, los métodos bayesianos tienen más dificultades en recuperar los interceptos del modelo. Se observa en los ítems 1 y 10, que el método bayesiano registra un RMSE y MAE que es entre 0.32 a 1.79 logits más grande que su contraparte clásica, mientras que lo inverso se observa en el ítem 2, en donde el método estima el parámetro con hasta 5.3 logits más de precisión que su contraparte clásica. En el resto de ítems los interceptos fueron estimados con menor precisión de manera consistente; sin embargo, dichas diferencias pueden considerarse marginalmente inferiores. En contraste, los métodos bayesianos recuperan las pendientes consistentemente de manera más precisa frente a su contraparte clásica (en promedio 0.36 de diferencia).

Los resultados obtenidos presentan evidencia de lo afirmado por Wollack et al. (2002), de que la estimación bajo los métodos bayesianos no presentan diferencias marcadas respecto a los clásicos, sugiriendo que el MCMC sería un buen sustituto para el MML.

Estimación	Ítems	Interceptos				Pendientes			
		1	2	3	4	1	2	3	4
Clásica	1	0.55	0.26	0.21	0.24	2.36	0.36	0.39	2.32
	2	2.99	2.83	2.93	8.73	3.44	1.67	2.04	6.8
	3	0.22	0.24	0.26	0.49	2.33	1.49	0.46	3.4
	4	0.2	0.3	0.11	0.17	0.69	2.42	2.09	0.4
	5	0.18	0.21	0.15	0.27	1.91	0.87	0.98	1.77
	6	0.17	0.22	0.38	0.21	0.57	2.26	0.98	1.88
	7	0.15	0.14	0.14	0.26	1.78	0.67	1.06	3.47
	8	0.15	0.25	0.15	0.14	0.69	1.69	2.06	0.34
	9	0.3	0.28	0.69	0.24	0.92	1.35	1.67	0.69
	10	0.6	1.35	0.55	0.51	1.37	3.68	1.14	1.33
Bayesiana	1	1.63	0.58	0.55	0.57	2.1	0.36	0.48	1.77
	2	1.3	1.08	1.15	3.48	2.31	0.86	1.23	4.06
	3	0.39	0.36	0.42	1.03	1.88	1.27	0.48	2.9
	4	0.23	0.43	0.14	0.2	0.57	1.84	1.49	0.34
	5	0.19	0.22	0.16	0.29	1.35	0.68	0.72	1.26
	6	0.18	0.23	0.42	0.22	0.49	1.64	0.83	1.35
	7	0.17	0.17	0.18	0.38	1.34	0.61	0.83	2.71
	8	0.15	0.28	0.16	0.15	0.5	1.26	1.49	0.29
	9	0.46	0.45	1.17	0.37	0.84	1.14	1.64	0.55
	10	1.37	3.14	0.95	1.33	1.63	3.89	1.05	1.56

Cuadro 4.3: RMSE para los parámetros de interceptos y pendientes, en las cuatro alternativas de los diez ítems elegidos en el estudio, según método de estimación.

Estimación	Ítems	Interceptos				Pendientes			
		1	2	3	4	1	2	3	4
Clásica	1	0.32	0.21	0.16	0.17	2.33	0.32	0.33	2.3
	2	1.19	1.14	1.11	3.3	3.11	0.76	1.39	5.14
	3	0.17	0.17	0.22	0.35	2.31	1.43	0.4	3.36
	4	0.15	0.24	0.09	0.14	0.65	2.4	2.08	0.34
	5	0.14	0.16	0.12	0.23	1.9	0.8	0.95	1.75
	6	0.14	0.18	0.27	0.15	0.51	2.25	0.9	1.86
	7	0.12	0.12	0.11	0.21	1.76	0.63	1.04	3.43
	8	0.12	0.19	0.12	0.12	0.67	1.67	2.05	0.3
	9	0.2	0.18	0.42	0.18	0.88	1.33	1.57	0.63
	10	0.44	0.99	0.44	0.37	1.28	3.6	1.07	1.27
Bayesiana	1	0.61	0.31	0.26	0.25	1.5	0.26	0.29	1.27
	2	0.99	0.84	0.82	2.59	1.9	0.65	1	3.42
	3	0.25	0.25	0.31	0.63	1.39	1.02	0.38	2.23
	4	0.17	0.29	0.11	0.16	0.47	1.41	1.15	0.27
	5	0.15	0.17	0.13	0.24	1.03	0.57	0.58	0.97
	6	0.15	0.19	0.3	0.17	0.39	1.26	0.62	1.02
	7	0.14	0.14	0.14	0.31	0.99	0.47	0.64	2.04
	8	0.12	0.21	0.13	0.13	0.4	0.95	1.13	0.23
	9	0.26	0.23	0.55	0.24	0.56	0.81	1.05	0.42
	10	0.93	2.23	0.8	0.94	1.18	3.1	0.82	1.2

Cuadro 4.4: MAE para los parámetros de interceptos y pendientes, en las cuatro alternativas de los diez ítems elegidos en el estudio, según método de estimación.

Por su parte, en relación a la recuperación de la variable latente θ , tanto el MML como MCMC parecen registrar una similar capacidad de recuperación del parámetro en cuestión, con errores que superan los 0.5 logits llegando hasta los 2 logits. En la Figura 4.1 se puede notar que en este diseño, la estimación bayesiana genera resultados marginalmente mejores que los obtenidos por el método clásico. Se tiene que de el total de 200 parámetros θ_i , 108 (54%) tienen un RMSE o MAE cuya magnitud es menor que la registrada bajo el método clásico.

Por su parte, los resultados del RMSE y MAE, aplicados en las probabilidades de elección de las alternativas por ítem, son mostrados en los cuadros 4.5 y 4.6. Se puede observar que ambas metodologías tienen un RMSE que no supera el 0.14 y un MAE que no supera el 0.38, además de presentar diferencias marginales en los criterios, indicativo de que en agregado ambos métodos tienen buen ajuste a los datos.

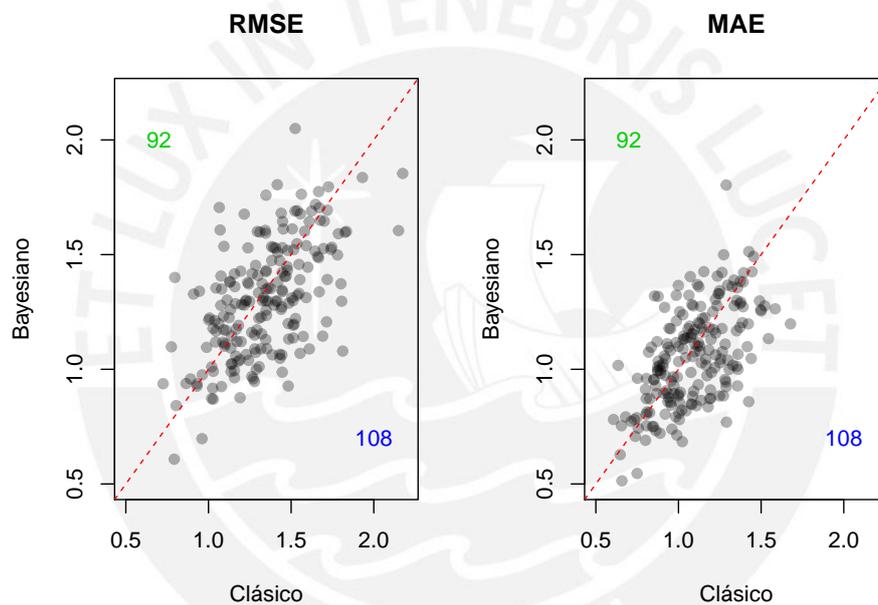


Figura 4.1: RMSE y MAE para todos los parámetros (interceptos y pendientes), según método de estimación.

Estimación	Ítems	Probabilidades			
		1	2	3	4
Clásica	1	0.12	0.11	0.13	0.28
	2	0.3	0.14	0.1	0.16
	3	0.33	0.05	0.16	0.25
	4	0.08	0.22	0.3	0.12
	5	0.34	0.04	0.15	0.22
	6	0.11	0.31	0.05	0.32
	7	0.16	0.12	0.14	0.37
	8	0.18	0.1	0.23	0.08
	9	0.05	0.18	0.14	0.06
	10	0.03	0.17	0.02	0.15
Bayesiana	1	0.13	0.11	0.13	0.29
	2	0.3	0.14	0.1	0.17
	3	0.33	0.05	0.17	0.26
	4	0.08	0.22	0.3	0.12
	5	0.34	0.04	0.15	0.23
	6	0.11	0.31	0.05	0.33
	7	0.16	0.12	0.15	0.38
	8	0.18	0.1	0.24	0.08
	9	0.05	0.19	0.15	0.07
	10	0.03	0.18	0.02	0.17

Cuadro 4.5: RMSE para las probabilidades, en las cuatro alternativas de los diez ítems elegidos en el estudio, según método de estimación.

Estimación	Ítems	Probabilidades			
		1	2	3	4
Clásica	1	0.08	0.08	0.08	0.12
	2	0.12	0.08	0.07	0.09
	3	0.13	0.05	0.09	0.11
	4	0.06	0.1	0.12	0.08
	5	0.13	0.05	0.09	0.11
	6	0.07	0.12	0.05	0.13
	7	0.09	0.08	0.08	0.14
	8	0.09	0.07	0.11	0.06
	9	0.05	0.1	0.08	0.06
	10	0.04	0.09	0.03	0.09
Bayesiana	1	0.08	0.08	0.08	0.12
	2	0.12	0.08	0.07	0.09
	3	0.13	0.05	0.09	0.11
	4	0.06	0.11	0.12	0.08
	5	0.13	0.05	0.09	0.11
	6	0.08	0.12	0.05	0.13
	7	0.09	0.08	0.09	0.14
	8	0.09	0.07	0.11	0.06
	9	0.05	0.1	0.09	0.06
	10	0.04	0.1	0.03	0.09

Cuadro 4.6: MAE para las probabilidades, en las cuatro alternativas de los diez ítems elegidos en el estudio, según método de estimación.

Finalmente, se procedió a evaluar la convergencia de la distribución a posteriori de los parámetros hacia una distribución estacionaria, en cada réplica. La convergencia se examinó de manera visual a través de diversos gráficos de diagnóstico obtenidos por la librería `coda` (Plummer et al., 2006).

Se observa que en todas las réplicas estimadas, cuatro de los ítems (y sus respectivos parámetros) y las competencias, presentan cadenas que convergen a una distribución estacionaria, con densidades definidas y auto correlaciones casi nulas. Por su parte, los seis ítems restantes (1, 2, 3, 4, 7 y 10) no registraron similar comportamiento. A través de todas las réplicas, de manera usual se observó que los ítems 3 y 7 presentaron cadenas con auto-correlaciones entre 0.1 a 0.3 para las primeras 100 a 500 iteraciones efectivas resultantes; sin embargo, esto no impidió que los parámetros alcanzaran estacionariedad en las restantes iteraciones efectivas. Similar resultado, pero con mayor magnitud, se observaron en los ítems 1, 2, 4 y 10. En estos, los niveles de auto-correlación de la cadena fue significativamente más alto (alrededor de 0.7), pero nuevamente, se observa que la cadena termina por alcanzar convergencia en las restantes iteraciones efectivas. En relación a los parámetros de habilidad (θ), en su mayoría alcanzaron distribuciones estacionarias con casi nulas auto-correlaciones, sin embargo hubieron algunos casos en los que se registró auto-correlaciones de alrededor de 0.1. No se identificaron observaciones en las que las cadenas de las habilidades registraran similares problemas a través de múltiples réplicas.

Parte de los resultados son reportados en el Anexo C.1.

4.4.2. Diseño 2:

Distribución Normal Estándar, $p = 10$, $m_j = 4$, $n/p = 20 : 1$ y $n = 1600$

Los criterio RMSE y MAE para las estimaciones de los parámetros de cada ítem son mostrados en los cuadros 4.7 y 4.8. Los resultados indican que los parámetros de los ítems también fueron recuperados de manera satisfactoria.

Bajo el segundo diseño, los métodos bayesianos estiman los interceptos con una precisión similar a su contraparte clásica. Es importante señalar que la afirmación no es aplicable al ítems 10, en donde se observa que el método bayesiano tiene mayor RMSE y MAE respecto al método clásico (diferencia mayor a 0.5). Por otro lado, los métodos bayesianos recuperan todos los parámetros de pendientes de manera más precisa respecto a su contraparte clásica.

Los resultados brindan soporte a lo afirmado por Ayala y Sava-Bolesta (1999), de que la recuperación de los parámetros de interés bajo los métodos bayesianos serian mejores con ratios *tamaño de muestra - número de ítems* de 20 : 1.

Estimación	Ítems	Interceptos				Pendientes			
		1	2	3	4	1	2	3	4
Clásica	1	0.15	0.07	0.06	0.06	2.29	0.3	0.3	2.29
	2	0.11	0.12	0.13	0.33	2.54	0.16	0.88	3.55
	3	0.07	0.09	0.07	0.19	2.2	1.36	0.4	3.16
	4	0.04	0.09	0.05	0.07	0.6	2.37	2.04	0.28
	5	0.05	0.07	0.03	0.06	1.91	0.82	1.01	1.72
	6	0.05	0.05	0.08	0.06	0.44	2.14	0.8	1.78
	7	0.08	0.05	0.06	0.12	1.79	0.55	1.01	3.35
	8	0.04	0.1	0.09	0.05	0.65	1.61	2	0.26
	9	0.06	0.06	0.1	0.1	0.85	1.25	1.47	0.64
	10	0.11	0.2	0.11	0.09	1.05	3.23	1.08	1.1
Bayesiana	1	0.15	0.07	0.06	0.06	1.71	0.24	0.23	1.72
	2	0.13	0.14	0.14	0.38	1.89	0.14	0.65	2.64
	3	0.07	0.09	0.07	0.19	1.66	1.01	0.3	2.37
	4	0.31	0.93	0.31	0.33	0.56	2.03	1.56	0.37
	5	0.05	0.07	0.03	0.06	1.41	0.64	0.76	1.28
	6	0.05	0.05	0.09	0.06	0.32	1.6	0.59	1.33
	7	0.08	0.05	0.06	0.12	1.36	0.4	0.75	2.51
	8	0.04	0.1	0.09	0.05	0.49	1.2	1.49	0.2
	9	0.06	0.07	0.11	0.1	0.64	0.94	1.12	0.47
	10	0.4	1.01	0.91	0.45	0.94	2.71	0.96	0.98

Cuadro 4.7: RMSE para los parámetros de interceptos y pendientes, en las cuatro alternativas de los diez ítems elegidos en el estudio, según método de estimación.

Estimación	Ítems	Interceptos				Pendientes			
		1	2	3	4	1	2	3	4
Clásica	1	0.12	0.06	0.05	0.04	2.29	0.29	0.29	2.29
	2	0.08	0.1	0.09	0.24	2.54	0.14	0.87	3.55
	3	0.05	0.08	0.06	0.14	2.2	1.36	0.4	3.16
	4	0.03	0.07	0.04	0.06	0.6	2.37	2.04	0.27
	5	0.04	0.06	0.03	0.05	1.91	0.81	1	1.72
	6	0.04	0.05	0.06	0.05	0.43	2.14	0.79	1.78
	7	0.06	0.04	0.05	0.1	1.79	0.55	1.01	3.34
	8	0.04	0.08	0.08	0.05	0.65	1.61	2	0.26
	9	0.05	0.05	0.08	0.08	0.85	1.25	1.47	0.63
	10	0.09	0.17	0.09	0.07	1.05	3.23	1.08	1.1
Bayesiana	1	0.13	0.06	0.05	0.04	1.32	0.2	0.18	1.3
	2	0.09	0.11	0.1	0.26	1.46	0.12	0.52	2.07
	3	0.05	0.08	0.06	0.14	1.26	0.79	0.24	1.83
	4	0.1	0.29	0.11	0.13	0.43	1.59	1.23	0.23
	5	0.04	0.06	0.03	0.05	1.08	0.51	0.59	0.97
	6	0.04	0.05	0.06	0.05	0.25	1.21	0.47	1.01
	7	0.06	0.04	0.05	0.1	1.04	0.32	0.58	1.92
	8	0.04	0.09	0.08	0.05	0.38	0.93	1.15	0.16
	9	0.05	0.06	0.08	0.09	0.49	0.73	0.85	0.38
	10	0.23	0.48	0.41	0.24	0.73	2.03	0.74	0.75

Cuadro 4.8: MAE para los parámetros de interceptos y pendientes, en las cuatro alternativas de los diez ítems elegidos en el estudio, según método de estimación.

En relación a la variable latente, tanto el MML como MCMC recuperan el parámetro con similar precisión, con errores que superan los 0.5 logits llegando hasta los 2 logits. Notamos también en la Figura 4.2, que al igual que el primer diseño, la estimación bayesiana genera resultados marginalmente mejores que la obtenida mediante los métodos de estimación clásica. Se tiene que de el total de 1600 parámetros θ_i , entre 849 y 858 (poco más de 53%) tienen un RMSE o MAE cuya magnitud es menor que la registrada bajo el método clásico.

Los criterios, aplicados a las probabilidades de la elección de alternativas por ítem, son mostrados en los cuadros 4.9 y 4.10. En agregado ambas metodologías presentan diferencias marginales, indicativo general del buen ajuste del modelo a los datos.

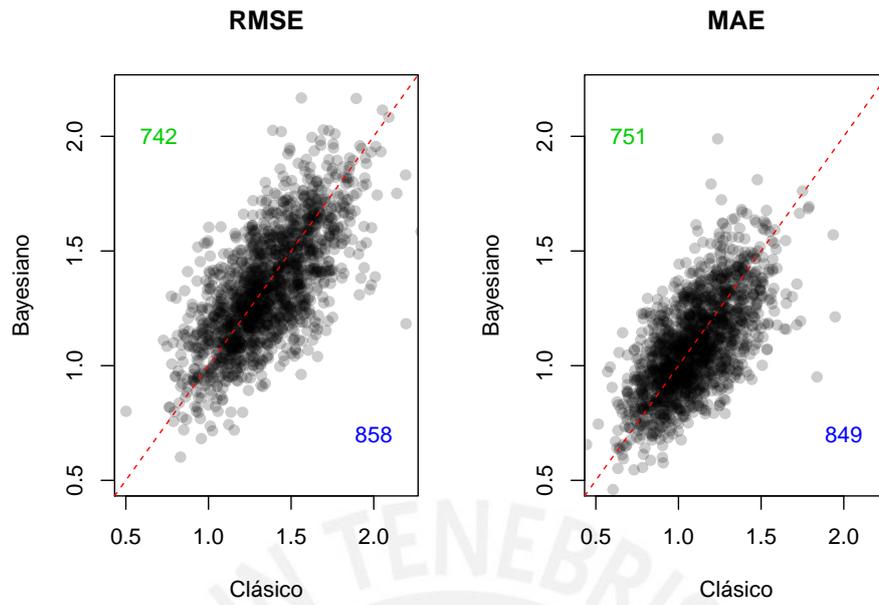


Figura 4.2: RMSE y MAE para todos los parámetros (interceptos y pendientes), según método de estimación.

Estimación	Ítems	Probabilidades			
		1	2	3	4
Clásica	1	0.11	0.11	0.12	0.28
	2	0.29	0.13	0.09	0.15
	3	0.32	0.04	0.15	0.24
	4	0.06	0.22	0.29	0.1
	5	0.34	0.03	0.15	0.22
	6	0.09	0.29	0.04	0.31
	7	0.16	0.1	0.13	0.36
	8	0.17	0.09	0.23	0.06
	9	0.03	0.17	0.13	0.06
	10	0.01	0.15	0.01	0.13
Bayesiana	1	0.11	0.11	0.12	0.28
	2	0.29	0.13	0.09	0.15
	3	0.32	0.04	0.15	0.24
	4	0.07	0.22	0.29	0.11
	5	0.34	0.03	0.15	0.22
	6	0.09	0.29	0.04	0.31
	7	0.16	0.1	0.13	0.36
	8	0.17	0.09	0.23	0.06
	9	0.03	0.17	0.13	0.06
	10	0.01	0.15	0.01	0.13

Cuadro 4.9: RMSE para las probabilidades, en las cuatro alternativas de los diez ítems elegidos en el estudio, según método de estimación.

Estimación	Ítems	Probabilidades			
		1	2	3	4
Clásica	1	0.08	0.07	0.08	0.12
	2	0.12	0.08	0.07	0.09
	3	0.13	0.05	0.09	0.11
	4	0.06	0.1	0.12	0.07
	5	0.13	0.04	0.09	0.11
	6	0.07	0.12	0.04	0.12
	7	0.09	0.07	0.08	0.13
	8	0.09	0.07	0.11	0.05
	9	0.04	0.09	0.08	0.05
	10	0.03	0.09	0.02	0.08
Bayesiana	1	0.08	0.07	0.08	0.12
	2	0.12	0.08	0.07	0.09
	3	0.13	0.05	0.09	0.11
	4	0.06	0.11	0.12	0.07
	5	0.13	0.04	0.09	0.11
	6	0.07	0.12	0.04	0.12
	7	0.09	0.07	0.08	0.13
	8	0.09	0.07	0.11	0.05
	9	0.04	0.09	0.08	0.05
	10	0.03	0.09	0.02	0.08

Cuadro 4.10: MAE para las probabilidades, en las cuatro alternativas de los diez ítems elegidos en el estudio, según método de estimación.

Finalmente, se evaluó la convergencia de la distribución a posteriori de los parámetros hacia una distribución estacionaria.

Al igual que en el primer diseño, se observa que en todas las réplicas estimadas, cuatro de los ítems (y sus respectivos parámetros) y las competencias, presentan cadenas que convergen a una distribución estacionaria, con densidades definidas y auto-correlaciones casi nulas. Por su parte, los seis ítems restantes (2, 3, 4, 5, 7 y 10) no registraron similar comportamiento. A través de todas las réplicas, de manera usual se observó que los ítems 3, 4, 5 y 7 presentaron cadenas con auto-correlaciones entre 0.1 a 0.3 para las primeras 100 a 500 iteraciones efectivas resultantes; sin embargo, esto no impidió que los parámetros alcanzaran estacionariedad en las restantes iteraciones efectivas. Similar resultado, pero con mayor magnitud, se observaron en los ítems 2 y 10. En estos, los niveles de auto-correlación de la cadena fue significativamente más alto (alrededor de 0.7), pero nuevamente, se observa que la cadena termina por alcanzar convergencia en las restantes iteraciones efectivas. En relación a los parámetros de habilidad (θ), en su mayoría alcanzaron distribuciones estacionarias con casi nulas auto-correlaciones, sin embargo existieron algunos casos en los que se registró auto-correlaciones cercanas a 0.1. Al igual que en el primer diseño, tampoco se identificaron observaciones en las que las cadenas de los parámetros de habilidades registraran similares problemas (para considerarlas

un patrón) a través de múltiples réplicas.

Parte de los resultados son reportados en el Anexo C.1.

4.5. Conclusiones

Del presente estudio de simulación se puede concluir:

1. Mediante los métodos de inferencia bayesiana (MCMC), los parámetros de los ítems e individuos en el contexto del NRM, fueron recuperados de manera adecuada.
2. Para lograr convergencia en las estimaciones fue necesario realizar un gran número de iteraciones. Se requirieron 37000 iteraciones; donde las primeras 3000 fueron descartadas, mientras que las siguientes 34000 fueron reducidas en saltos de 17, dejando así una cadena de 2000 iteraciones efectivas. En tres de los ítems (6, 8 y 9), el diseño de las cadenas de simulación (iteraciones totales, descartes y saltos) permitieron que los parámetros de las alternativas tuvieran cadenas de 2000 iteraciones efectivas que convergen a una distribución estacionaria, con densidades definidas y auto-correlaciones casi nulas. En los otros siete ítems también se logró alcanzar cadenas de iteraciones efectivas que convergen hacia densidades definidas, pero el número de estas iteraciones fue menor.
3. Los diseños de muestras permitieron corroborar que los métodos bayesianos recuperan con similar, e incluso mejor precisión, los parámetros de interés frente a los métodos clásicos. Este resultado se confirma bajo los supuestos de que las competencias provienen de una Distribución Normal Estándar, la longitud de la evaluación es de 10 ítems con 4 alternativas por ítem y un ratio tamaño de muestra : numero de ítems de 20 a 1 ($n/p = 20 : 1$). Este resultado nos lleva a afirmar que el método MCMC puede ser un buen sustituto del método clásico.
4. El proceso de estimación de los parámetros bajo el método bayesiano toma demasiado tiempo y suele requerir muchos recursos de la máquina. Esto es de particular importancia, puesto que la afirmación es cierta aún cuando se establecieron procesos que hicieron paralelas las tareas de simulación MCMC. Así, dependiendo del diseño, el método bayesiano puede llegar a ser entre 620 a 14100 veces más lento que el método clásico:

Diseño					Clásico	Bayesiano
Distribución	Longitud	Alternativas	Ratio n/p	Muestra	(min.)	(min.)
$N(0, 1)$	10	4	2.5 : 1	200	0.30	186
$N(0, 1)$	10	4	20 : 1	1600	0.20	2820

Cuadro 4.11: Duración del proceso de estimación por diseño y método de estimación

Capítulo 5

Aplicación

La presente aplicación se apoya en este marco teórico de medición y análisis, y tiene como principal propósito el aplicar el modelo investigado a un conjunto de datos reales; y como propósitos secundarios comprobar si el modelo permite hallar un ordenamiento en las categorías inicialmente no ordenadas, si brinda información relevante para valorar la calidad de un ítem y sus alternativas, y finalmente, corroborar si es útil para caracterizar a los individuos expuestos al instrumento de evaluación; es decir, encontrar perfiles de docentes. El modelo fue aplicado en un conjunto de datos que se caracterizó por:

1. Una muestra anónima, aleatoria y representativa de 1641 docentes del nivel secundario de la especialidad de Inglés, modalidad de Educación Básica Regular (EBR), evaluados en el primer concurso de Ingreso a la Carrera Pública Magisterial y Contratación Docente aplicado en el 2015 (en adelante Nomenclario 2015); y
2. Un conjunto de 25 ítems correspondientes a la sub-prueba de Comprensión de Textos del mismo concurso.

5.1. Instrumento de Evaluación

En la presente sección se detallan algunos de los aspectos relacionados con el instrumento utilizado en la aplicación.

5.1.1. Detalles del instrumento

El instrumento completo consistió en una evaluación estandarizada de 90 preguntas de opción múltiple, con cuatro alternativas por ítem, organizados en tres sub-prueba: (i) **Comprensión Lectora**, a la que le corresponden los primeros 25 ítems, (ii) **Razonamiento Lógico**, evaluados en los siguientes 25 ítems y (iii) **Conocimientos Pedagógicos y de la Especialidad**, evaluado en los últimos 40 ítems.

La sección elegida para ser analizada bajo el modelo estudiado fue la de **Comprensión Lectora**. La sub-prueba fue diseñada para evaluar la capacidad que un docente posee para reconstruir el significado de diversos tipos de textos en distintos formatos. La sub-prueba evaluó tres dominios:

- **Comprensión literal:** Capacidad para ubicar información explícita en textos complejos,
- **Comprensión inferencial:** Capacidad para integrar información de un texto con el objetivo de inferir su temática, propósito o relaciones lógicas implícitas en sus distintos componentes,
- **Reflexión sobre el texto:** Capacidad de reflexionar críticamente sobre el contenido y forma de textos diversos.

Es importante mencionar, que pese a que la prueba fue dividida en tres dominios, dicha división fue realizada con el propósito de realizar una mejor descripción de los resultados; por lo tanto, la división no necesariamente representa sub-constructos o sub-escalas que se deban analizar por separado. Otro aspecto que es importante recalcar, es que pese a que el instrumento fue diseñado con cuatro alternativas, los patrones de respuesta revelaron que en realidad existen hasta seis categorías de respuestas:

- Cuatro (4) alternativas previstas en el diseño del ítem (A,B,C,D);
- Una alternativa más (1), denotada por la posibilidad que el individuo tiene para marcar más de una opción; no considerada en el puntaje del mismo; y
- una última alternativa (1), que consiste en dejar en blanco el ítem, también no considerado en la calificación del individuo.

Para el presente análisis, se prefirió unir las respuestas múltiples y en blanco en una sola categoría: (i) por la facilidad de estimación e interpretación de un mismo número de categorías para todos los ítems; (ii) para alcanzar la masa crítica que dicha categoría podría necesitar para tener una estimación precisa de sus parámetros; y (iii) porque se busca entender que tipos de personas y cuales ítems tiene patrones particulares para esta nueva categoría. Apalancados en la versatilidad del NRM (Bock, 1972, 1997), el modelo no requirió modificación alguna para realizar la tarea.

5.1.2. Recolección de Datos

Para tener acceso a los datos referidos se siguió el trámite de acceso a la información en la sede principal del Ministerio de Educación (MINEDU).

5.1.3. Diseño Muestral

Con el fin de comparar los estimados que se obtengan del modelo NRM, proveniente de los métodos MML y MCMC, se accedió a una muestra anónima, aleatoria y representativa de 1641 docentes que fueron expuestos a la evaluación de interés.

5.2. Resultados

Como ya se había descrito en el Capítulo 3 la implementación del modelo se realizó mediante el programa R (R Core Team, 2015), y la librería R2WinBUGS (Sturtz et al., 2005). Por su parte, se utilizó la librería coda (Plummer et al., 2006) para mostrar los estadísticos de interés para cada parámetro.

Debido a los resultados presentados en el capítulo de simulación se creyó conveniente aumentar el número de iteraciones, con el propósito de que las simulaciones pudieran alcanzar cadenas convergentes a una distribución específica con bajos niveles de auto-correlación. Para comparar la estimación realizada por el modelo, se ejecutó la aplicación con una sola cadena de 84000 iteraciones, de las cuales las primeras 28000 iteraciones fueron descartadas (*burn-in*) y las siguientes 56000 fueron reducidas en saltos de 28 (*thinning*), generando un total de 2000 iteraciones efectivas, que sirvieron para estimar las distribuciones a posteriori de los parámetros (ver sección 4.3.2).

Para mayor detalle de los códigos utilizados en las respectivas estimaciones, referirse al apéndice y B.2.5.

5.2.1. Análisis

Ítems

Los valores estimados de los parámetros de intercepto (Media), el error estándar (EE) y los principales estadísticos de orden se presentan en los Cuadros 5.1, 5.2, 5.3.

A grandes rasgos, se puede notar que gran parte de los interceptos de las alternativas se estimaron con una elevada precisión (menor a 0.01); sin embargo, se nota también que algunos de estos parámetros fueron estimados con hasta diez veces menos precisión que sus contra-partes en otros ítems. En términos específicos, se observa que los parámetros de interceptos que se estimaron con mayor precisión fueron los relacionados con los ítems 3, 10, 13, 21, 22, 23 y 24; sin embargo es importante resaltar, que en ninguno de estos ítems, los interceptos asociados con la alternativa “en blanco” o multimarca fueron estimados con similar precisión, esto debido a que la proporción de observaciones en dicha alternativa es significativamente menor que la observada en sus pares. Por otro lado, el método tiene mayor dificultad en estimar los interceptos de los ítems 9 y 18, esto se puede observar en el error estándar de la estimación, el cual registra entre dos a diez veces menos precisión que sus contra-partes en otros ítems.

Por su parte, los estimados de los parámetros de pendientes (Media), el error estándar (EE) y los principales estadísticos de orden se presentan en los Cuadros 5.4, 5.5, 5.6.

A grandes rasgos, se puede notar que los parámetros de pendientes se estimaron con mayor precisión que los parámetros de interceptos, donde el error máximo alcanzado fue de 0.053. En este sentido, se puede afirmar que el método no tiene mayor dificultad en la estimación de las pendientes de las alternativas de los ítems. Específicamente, se estimaron

con mayor precisión los ítems 3, 4, 8, del 10 al 15 y del 19 al 25; sin embargo, es importante resaltar, que en ninguno de estos ítems, las pendientes asociadas con la alternativa que representa multimarca o “en blanco” fueron estimados con similar precisión, esto debido a que la proporción de observaciones en dicha alternativa es significativamente menor que la observada en sus pares.

Con estos resultados se halla soporte a lo previsto por (Baker y Kim, 1992), de que el modelo permite estimar simultáneamente todos los parámetros de las alternativas de respuesta de un ítem de opción múltiple. Este resultado es de especial relevancia debido a que las alternativas de los ítems son de medida nominal.

Finalmente, mediante el análisis visual de las ICRFs (Figuras 5.1, 5.2, 5.3 y 5.4), las ICIFs (Figuras 5.5, 5.6, 5.7 y 5.8) y la ayuda de los especialistas que elaboraron el instrumento de evaluación, se procedió a investigar los beneficios adicionales de utilizar el modelo, entre ellos:

1. Si el modelo facilita la recuperación de una mayor proporción información disponible en los ítems, frente a los modelos de respuestas dicotómicas (Bock, 1972; Thissen, 1976; Levine y Drasgow, 1983; Thissen y Steinberg, 1984)
2. Si permite hallar el ordenamiento implícito en datos categóricos inicialmente no ordenados (Samejima, 1988; Bock, 1997),
3. Si brinda información relevante para la valoración de la calidad de un ítem (Thissen et al., 1989), y
4. Si el modelo podría proveer de información útil para caracterizar a los individuos expuestos al instrumento de evaluación (De Ayala, 1993) .

En relación al primer punto, el presente trabajo encuentra resultados similares a los encontrados por Bock (1972), Thissen (1976) y Thissen y Steinberg (1984). Se observa en las ICIFs, que la incorporación de las alternativas “en blanco” y multimarca incrementa el nivel de información provisto por el ítem, especialmente en los niveles bajos de habilidad. La implicancia práctica de este resultado es, que el NRM nos permite estimar con mayor precisión las habilidades de individuos que se encuentran en los niveles bajos de habilidad, sin sacrificar la precisión en los niveles altos. Es importante mencionar que, pese a que no se ha hecho una comparación directa entre los modelos dicotómicos y el NRM, los autores Bock, Thissen y Steinberg han, de hecho, realizado dichas comparaciones múltiples veces, encontrando siempre la misma conclusión.

Respecto al segundo punto, los especialistas señalaron que en 23 de los 25 ítem existe un ordenamiento creciente de las opciones de respuesta, en relación a la habilidad de “comprensión lectora” requerida. Por ejemplo, en el primer ítem, los individuos con bajos niveles de competencias tienden a distribuir sus respuestas entre las alternativas 1, 2 y 4, todas alternativas incorrectas. A medida que los evaluados se acercan a un nivel “promedio” de competencia

(alrededor de cero), tienden a favorecer de las alternativas 3 y 4, siendo finalmente la número 3 la que prevalece entre individuos con habilidad superior a la competencia “promedio”, puesto que es la opción correcta. Similares resultados se pueden extender para los ítems 4 al 25:

- Los evaluados con menor habilidad suelen dejar las respuestas en blanco, realizar multimarca o elegir las alternativas de respuestas “más erradas”. Llama la atención que el modelo asocia las alternativas multimarca y en blanco con curvas que usualmente presentan marcadas pendientes negativas; es decir, representa la noción de que un docente con baja competencia suele apoyarse en estos recursos para intentar lograr un resultado positivo (especialmente cuando se elige marcar más de una opción) y que a medida que la competencia del docente es más alta, este rápidamente deja de utilizar dichos recursos.
- Los docentes con habilidades promedio, o superiores a las competencias mínimas, aún favorecieron alternativas incorrectas, pero dichas alternativas usualmente presentaron diferencias sutiles o matices distintos, pero errados, en relación a la alternativa correcta. Este último detalle nos lleva afirmar que la elección de estas alternativas requiere un mayor nivel de competencia que las competencias “mínimas”. Por su parte, también se puede afirmar que la no elección de estas alternativas usualmente demanda un nivel aún mayor que el de las competencias “promedio”.
- Finalmente, los evaluados que registraron los más altos niveles de competencia usualmente eligieron la alternativa correcta, la que registró una relación monotónicamente creciente con la habilidad del evaluado; es decir, que a mayor habilidad mayor la probabilidad de elegir la alternativa correcta. Es importante señalar, que esta relación también es observada en los modelos del tipo dicotómico.

Algo interesante de señalar, es que en todos los ítems, el modelo asignó una pendiente monotónicamente creciente a las alternativas correctas. Sin embargo, se notó que en los ítems 8, 10, 14, 17, 18 y 20 el modelo registró también algunas ICRFs de alternativas incorrectas con pendientes monotónicamente crecientes, en gran parte del rango de la competencia (p.e alternativa 4 del ítem 8 o la alternativa 2 del ítem 10). Esto fue de gran utilidad para los especialistas, que indicaron que estas alternativas permitían diferenciar entre los docentes que tenían la competencia requerida y los docentes que estaban por alcanzarla, debido a que las alternativas incorrectas que competían presentaban diferencias sutiles respecto a las correctas, y por lo tanto requerían un conocimiento que sepa discriminar entre esas sutilezas.

De esta manera, los resultados dan soporte a lo planteado por Samejima (1988) y Bock (1997), puesto que el modelo logra encontrar el ordenamiento implícito de las alternativas del ítem respecto a la competencia requerida (excepto en los ítems 2 y 3), lo cual es especialmente importante debido a que el instrumento de evaluación fue diseñado con solo una alternativa correcta y un conjunto de alternativas incorrectas sin un ordenamiento definido. Del mismo modo, brinda soporte a lo planteado por Baker y Kim (1992) y Hambleton y Swaminathan

(1991) en el sentido de que dadas las m_j categorías de respuestas del ítem, la ICRF asociada con la respuesta correcta siempre tiene una relación monótonicamente creciente con la competencia requerida, mientras que existe una alternativa monótonicamente decreciente (usualmente asociada a las respuestas en blanco y multimarcas) y las $m_j - 2$ ICRFs restantes usualmente son curvas unimodales con colas asintóticamente cercanas a cero, que reflejan como el evaluado percibe la alternativa, en relación a la competencia requerida.

En relación al tercer punto, los especialistas resaltaron dos aspectos en los que el modelo les permitía valorar la calidad del ítem: (a) les permitía identificar alternativas inservibles o forzadas y (b) les permitía identificar alternativas que se podían colapsar, debido a que estas alternativas registraban similar temática. De este modo, los especialistas señalaron que muchos ítems pudieron estar definidos solo por tres (y hasta dos) alternativas. Por ejemplo en los ítems 5 al 6, 9 al 15 y 19 al 21, se encontraron alternativas que se consideraron forzadas, debido a que su temática no se relacionaba con el estímulo principal del ítem. Estas alternativas se caracterizaron por distribuciones con bajas probabilidades de selección. Por otro lado, en los ítems 1, 7, 9, 11, 12, 15, 16 y del 18 al 25, los especialistas observaron que existían alternativas que podían ser colapsadas. Estas alternativas se caracterizaron por tener ICIFs similares, en términos del rango en el que proveen mayor información. De este modo, muchos de los mencionados ítems pudieron haber sido evaluados solo con tres alternativas seleccionables (sin considerar las alternativas “artificiales” de multimarca y respuestas en blanco) y en el caso del ítem 23 solo se hubiesen necesitado dos.

De esta forma, los resultados brindan soporte a lo afirmado por Thissen et al. (1989). El modelo brinda información relevante para la valoración de la calidad de un ítem; que en este caso se traduce en la labor de encontrar alternativas forzadas y la posibilidad de colapsar otras.

Finalmente, pese a que en términos prácticos, eliminar o colapsar alternativas también significa que los diseñadores de instrumentos podrían clasificar a los individuos, en grupos de competencias según la elección de las alternativas; posterior al análisis, no se encontró medios para brindar soporte a la afirmación de De Ayala (1993). Se encontró por el contrario que la capacidad del modelo, para develar perfiles de individuos asociados a las alternativas elegidas, depende ampliamente del diseño del instrumento. Es decir, la capacidad que el NRM tiene de proveer información útil para caracterizar a los individuos en perfiles, depende ampliamente de la similitud temática entre las distintas alternativas de los diversos ítems, establecidas durante el diseño del instrumento.

Para mayor detalle acerca de la función que genera las funciones de probabilidad (ICRFs) e información de las categorías de los ítems (ICIFs) y del ítem (IIF), referirse a los apéndices [B.2.1](#) y [B.2.2](#).

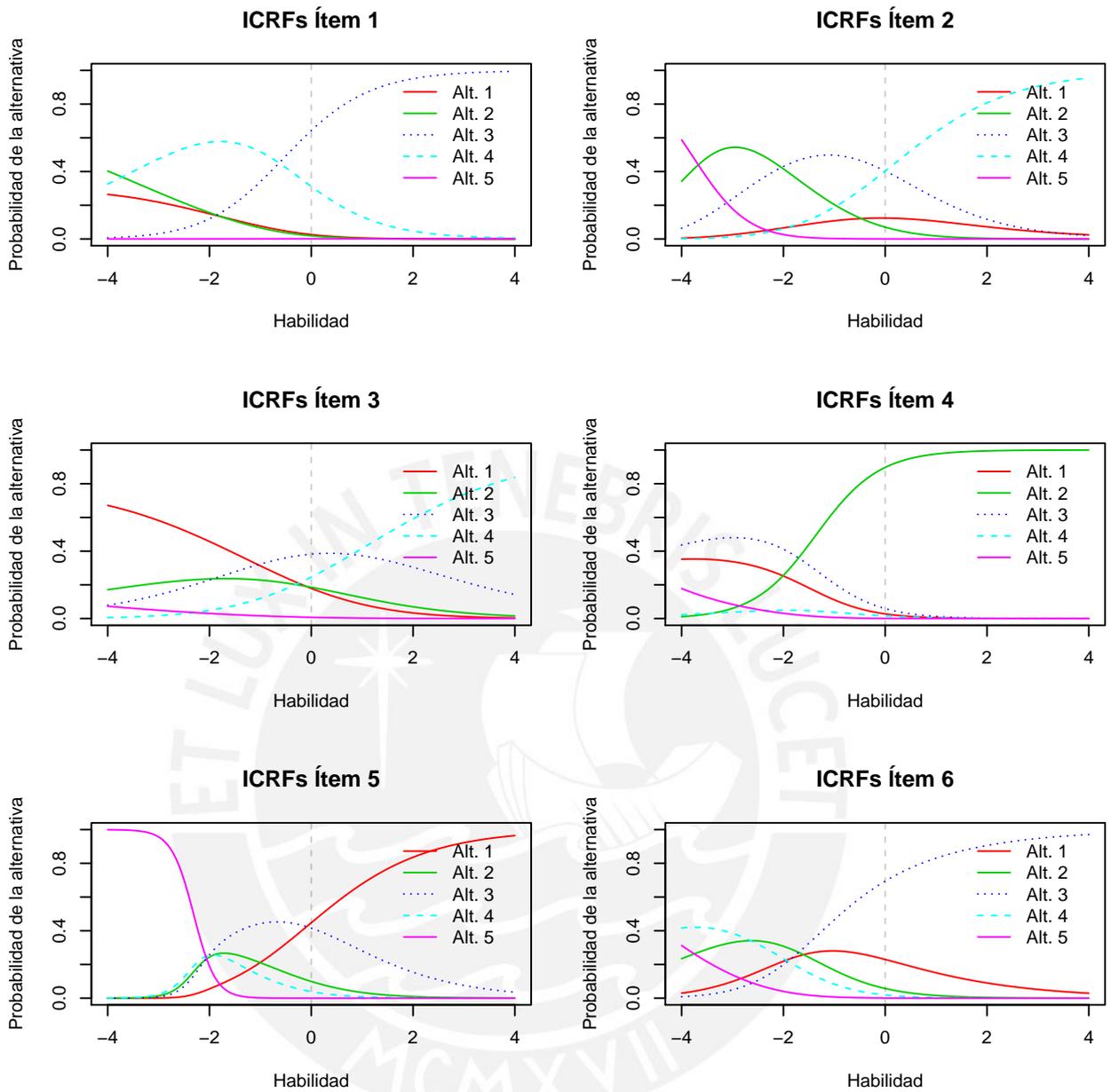


Figura 5.1: Funciones de respuestas a las categorías de los ítems (ICRF) de los primeros seis ítems de la competencia de Comprensión Lectora de la evaluación de Nombramiento 2015.

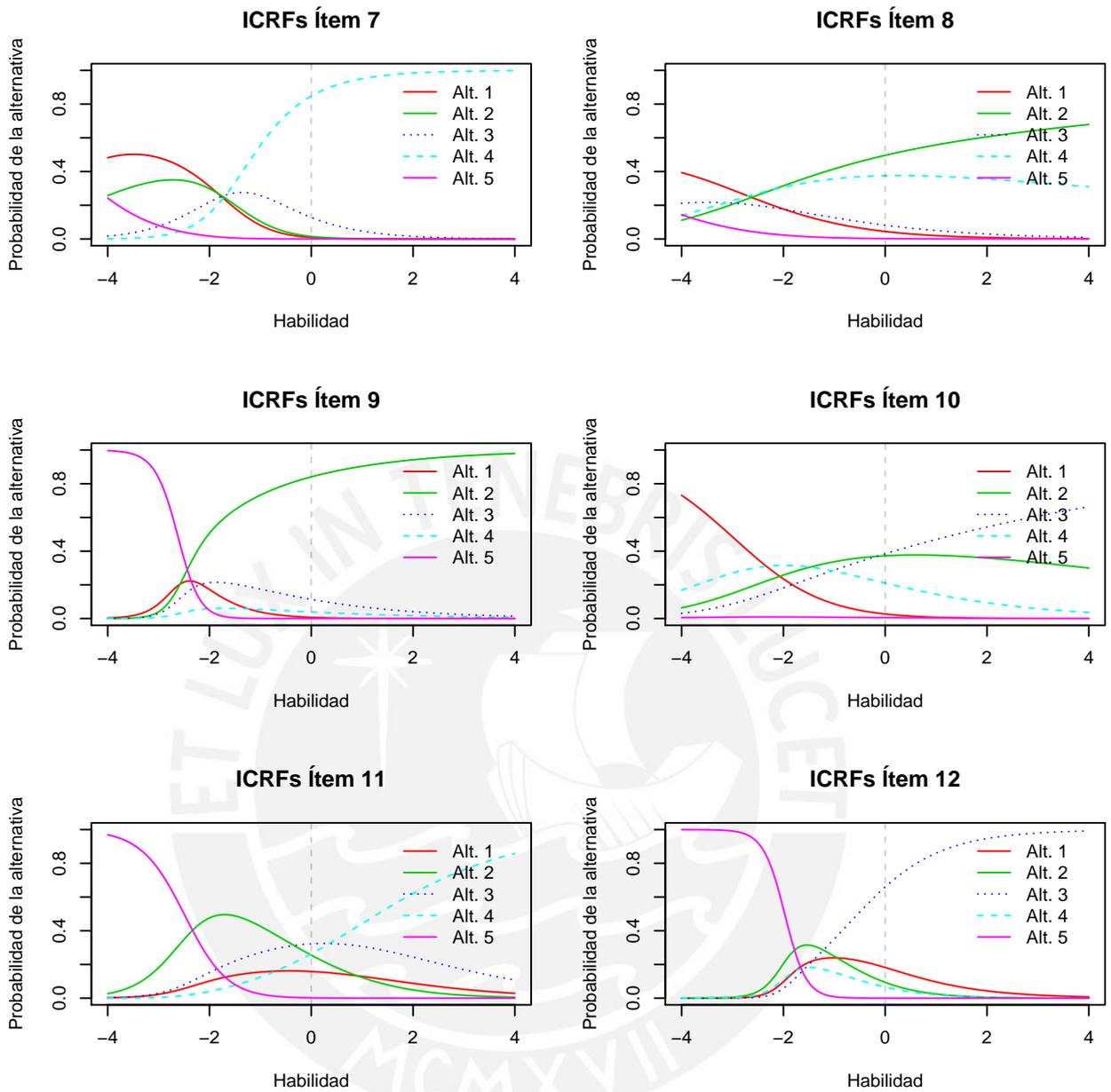


Figura 5.2: Funciones de respuestas a las categorías de los ítems (ICRF) de los ítems 7-12 de la competencia de Comprensión Lectora de la evaluación de Nombramiento 2015.

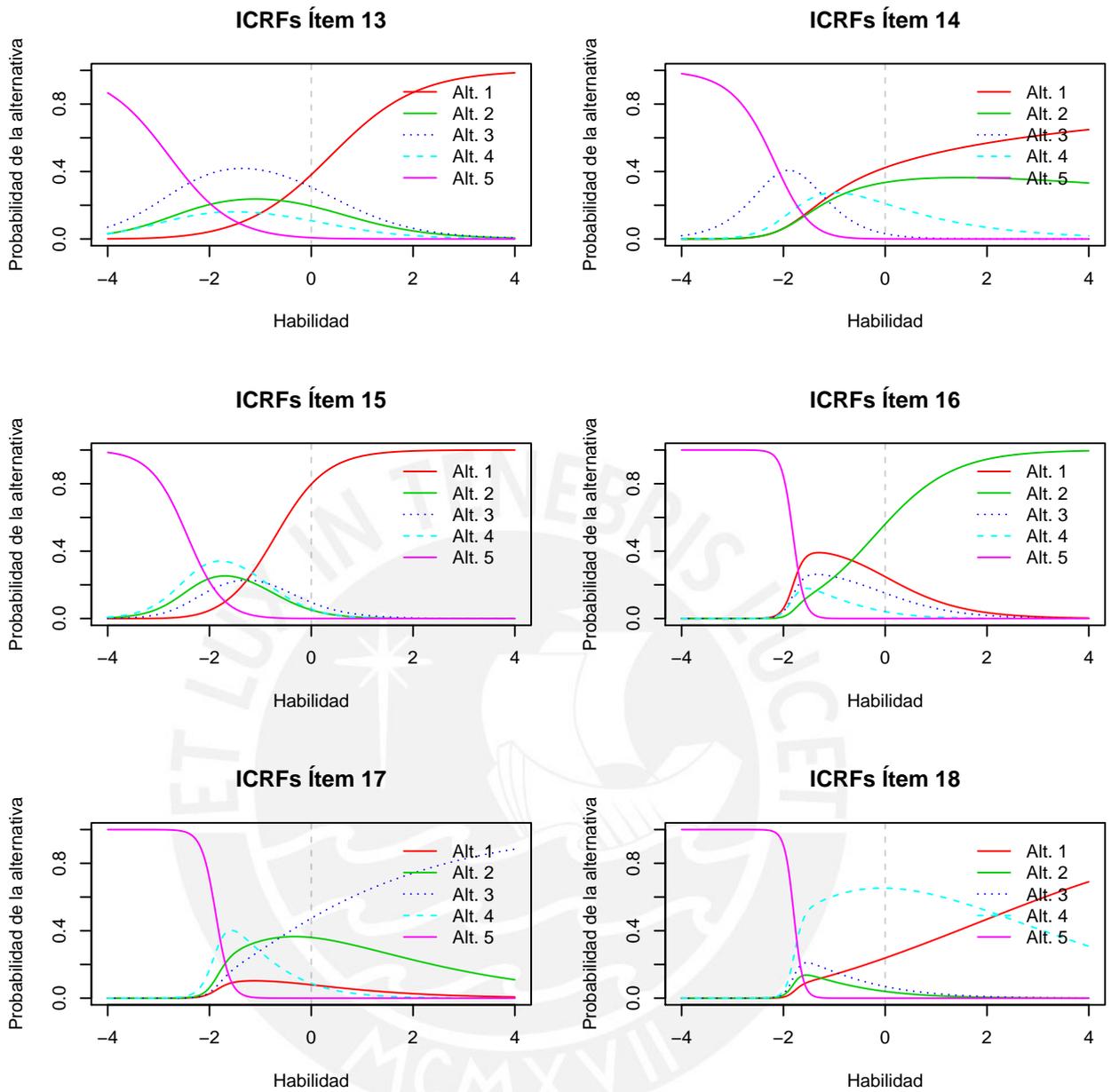


Figura 5.3: Funciones de respuestas a las categorías de los ítems (ICRF) de los ítems 13-18 de la competencia de Comprensión Lectora de la evaluación de Nombramiento 2015.

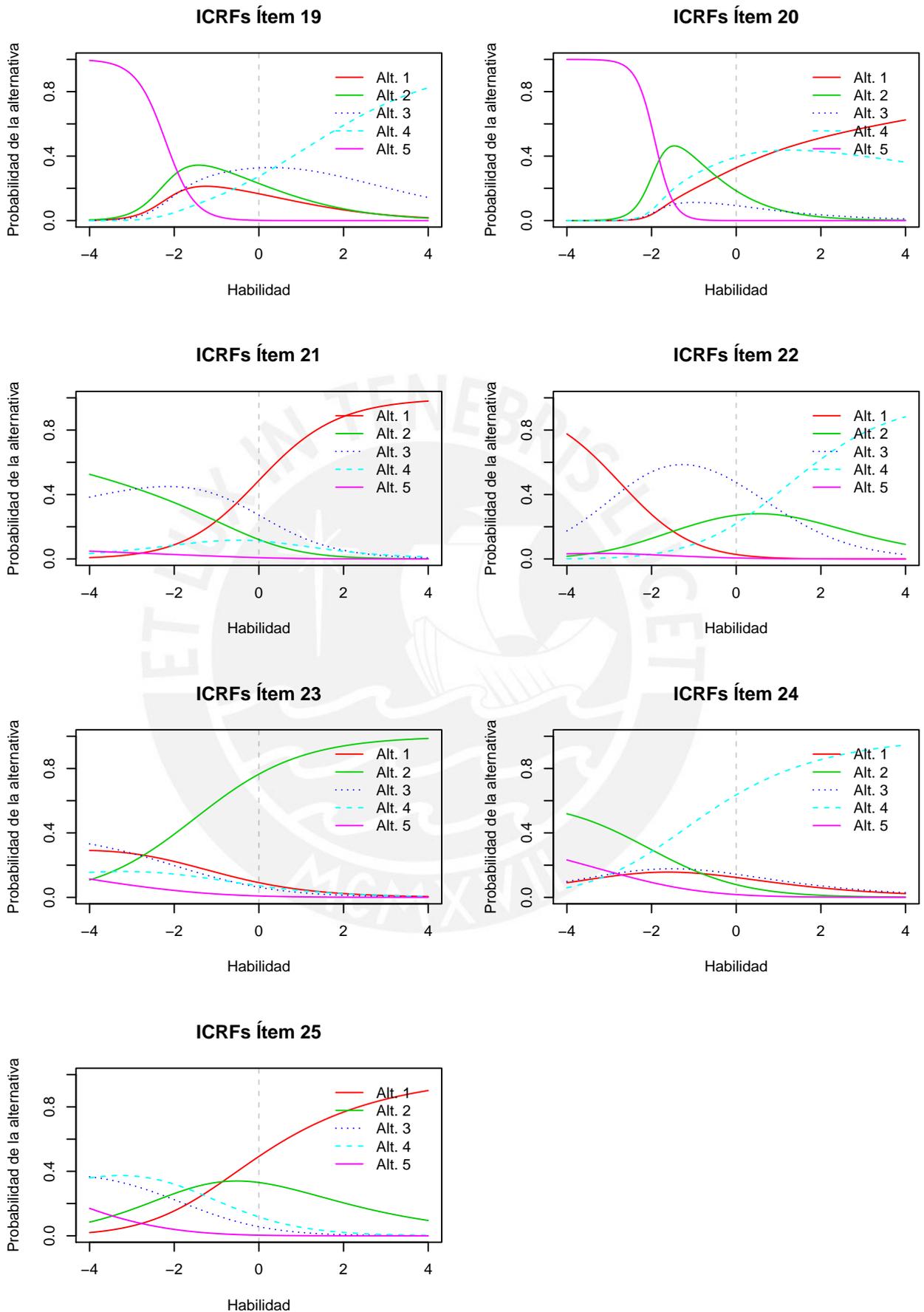


Figura 5.4: Funciones de respuestas a las categorías de los ítems (ICRF) de los ítems 19-25 de la competencia de Comprensión Lectora de la evaluación de Nombramiento 2015.

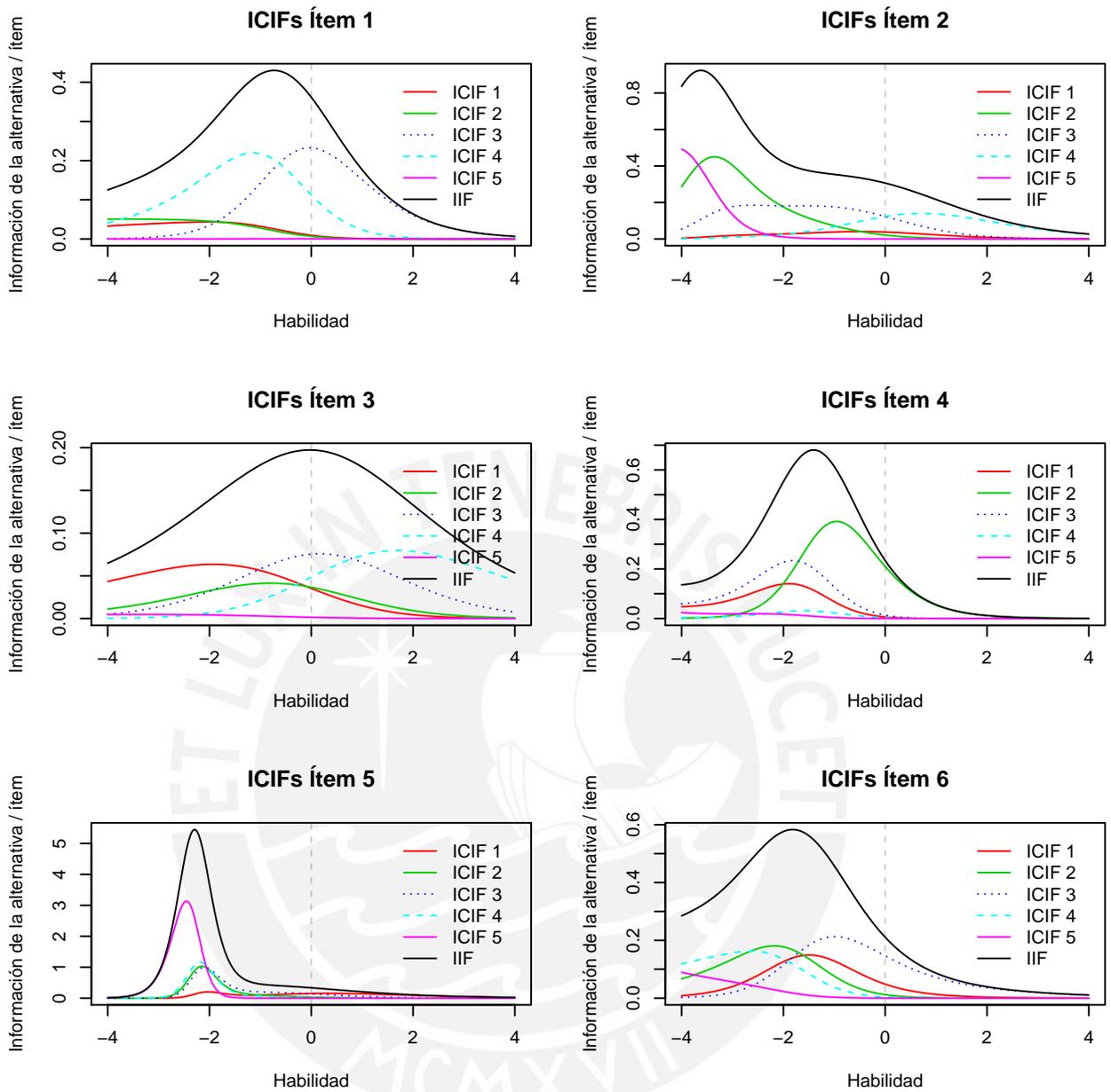


Figura 5.5: Funciones de información de las categorías de los ítems (ICIF) de los primeros seis ítems de la competencia de Comprensión Lectora de la evaluación de Nombramiento 2015.

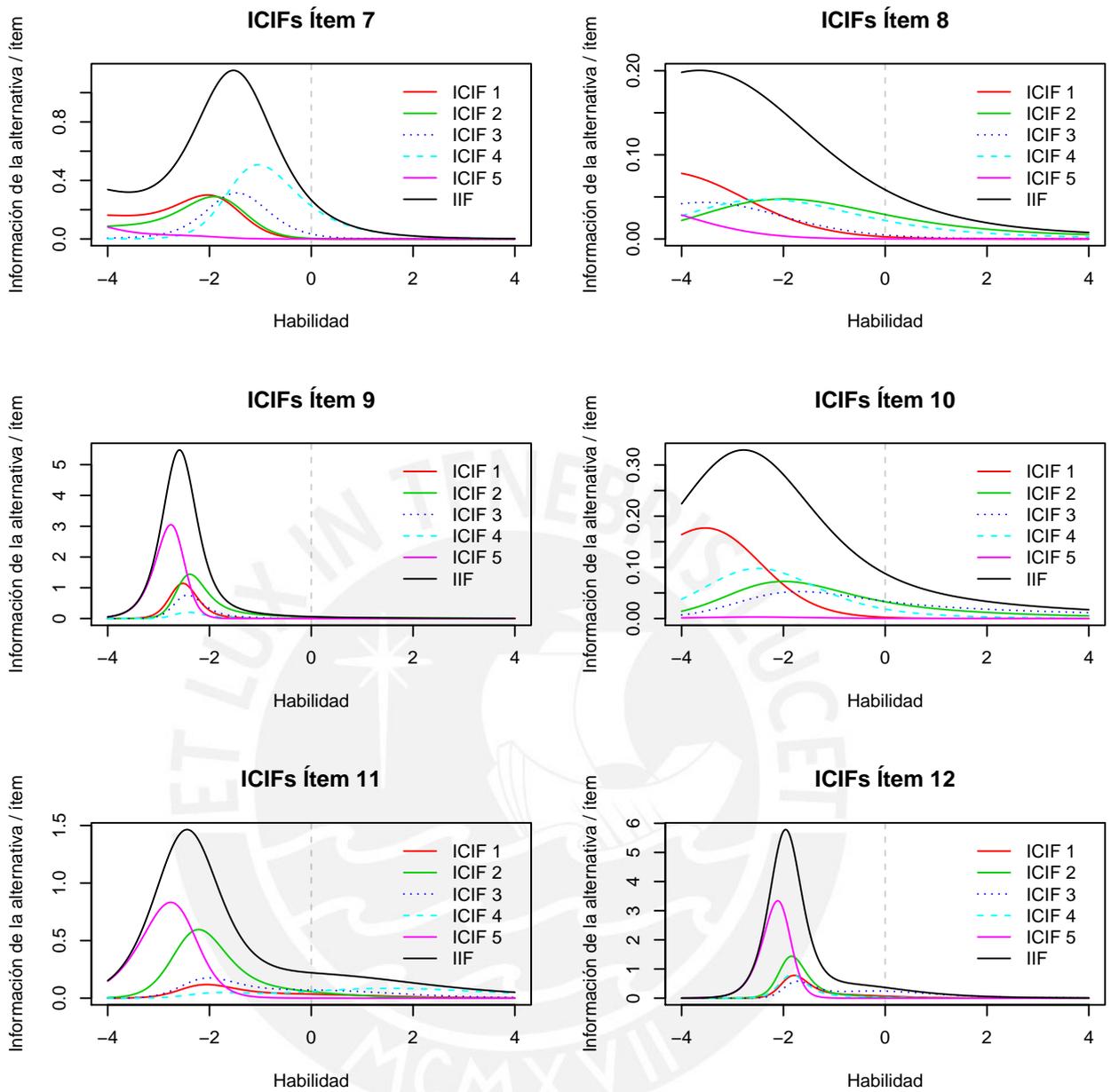


Figura 5.6: Funciones de información de las categorías de los ítems (ICIF) de los ítems 7-12 de la competencia de Comprensión Lectora de la evaluación de Nombramiento 2015.

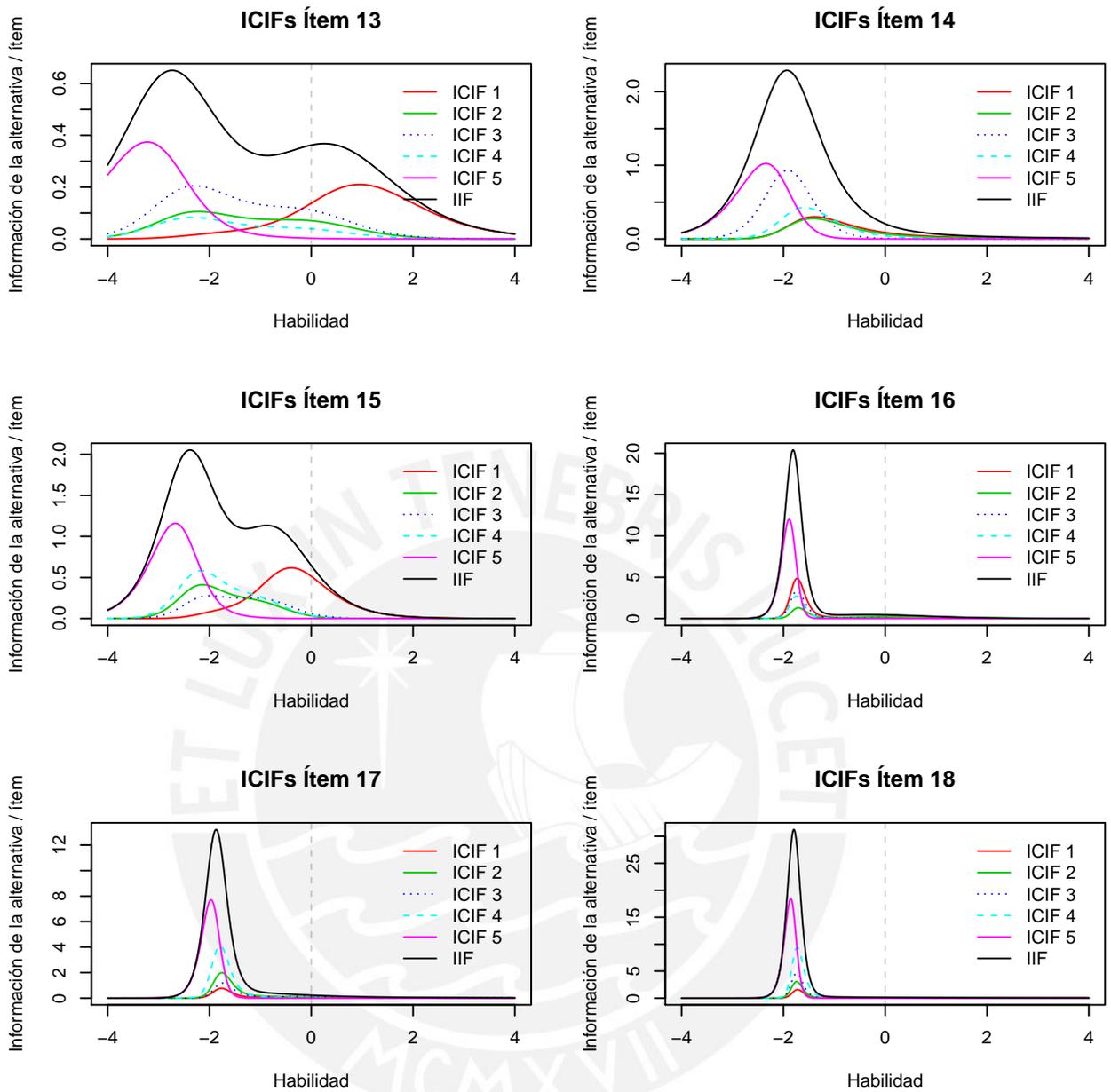


Figura 5.7: Funciones de información de las categorías de los ítems (ICIF) de los ítems 13-18 de la competencia de Comprensión Lectora de la evaluación de Nombramiento 2015.

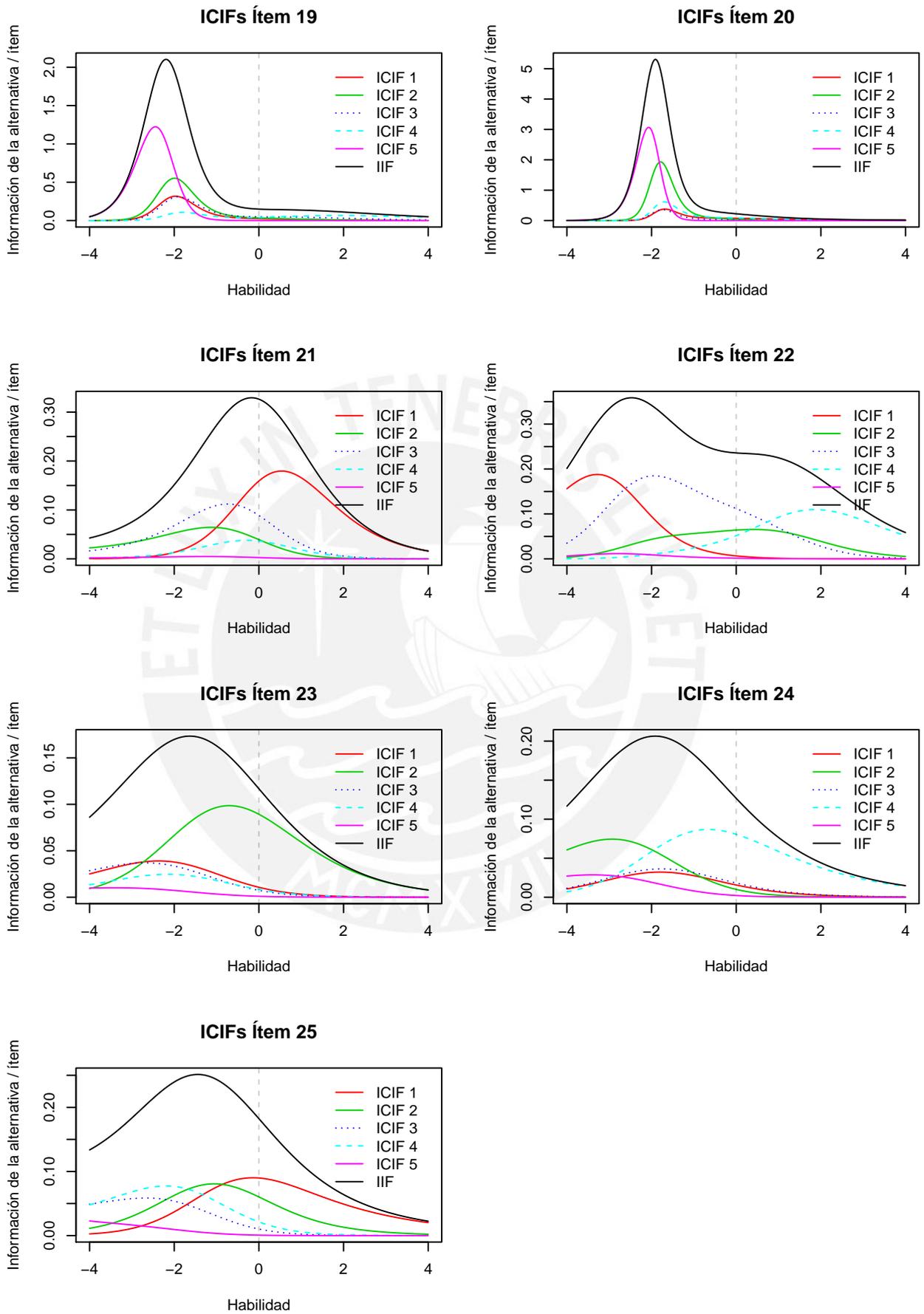


Figura 5.8: Funciones de información de las categorías de los ítems (ICIF) de los ítems 19-25 de la competencia de Comprensión Lectora de la evaluación de Nombramiento 2015.

Escala de Comprensión Lectora

Las características de la escala de *Comprensión lectora* se reportaron en la Tabla 5.7 y la Figura 5.9.

En relación a la escala se observa que: (a) la distribución que parece caracterizar a los individuos no es una distribución Normal, presenta su centro de gravedad alrededor de cero, pero tiene una asimetría positiva y curtosis negativa (distribución más apoyada en el lado izquierdo y más achatada); y (b) pese al elevado valor de la desviación de la media estimada (error estándar), no se observa la presencia de *outliers*.

Variable	media	desv.	med.	mad	min.	max.	rango	asim.	curt.	error
Media	0.005	0.886	-0.129	0.925	-2.227	2.494	4.721	0.447	-0.474	0.022
Desviación	0.382	0.092	0.367	0.085	0.137	0.665	0.529	0.434	-0.054	0.002
Error	0.009	0.002	0.008	0.002	0.003	0.015	0.012	0.434	-0.054	0.000

Cuadro 5.7: Características del estimador puntual de las competencias de los individuos en Comprensión Lectora

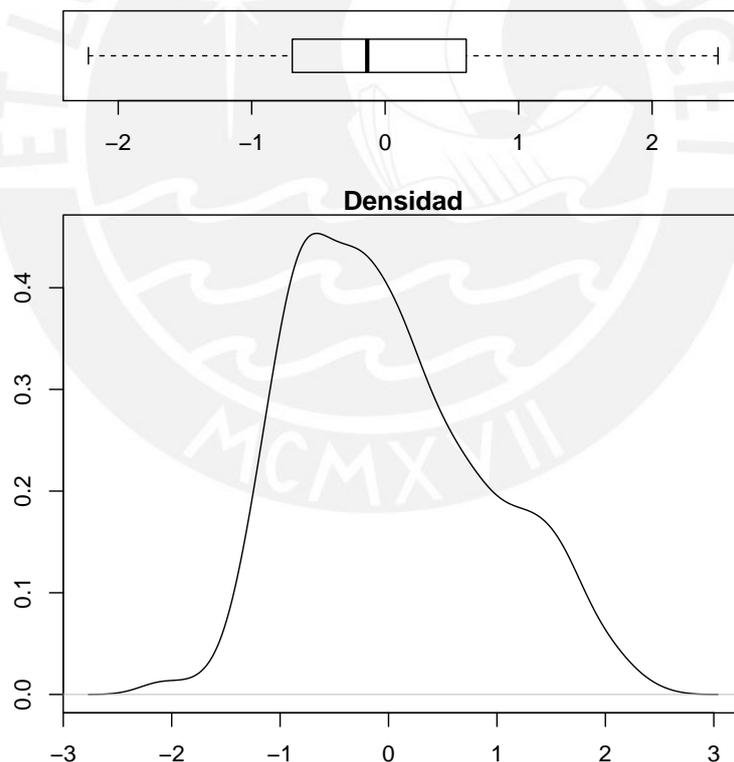


Figura 5.9: Distribución del estimado de las competencias de los individuos en Comprensión Lectora

Pese a que la distribución que parece caracterizar a los individuos no es una distribución Normal, se consideró que los intervalos de confianza para los estimados, aproximados mediante dicha distribución, no serían tan inapropiados. Así, los intervalos de confianza para la competencia de Comprensión Lectora de los individuos fueron definidos por $media \pm 1.96 \times desv..$

Tal como se observa en la Figura 5.10, podemos notar que, tal y como sucede en otros modelos IRT, las personas que muestran comportamientos “extremos” (scores nulos o completos) usualmente están asociados con un mayor intervalo de confianza (o menor precisión en la estimación), esto debido a que existe un menor número de individuos en esos rangos, lo que hace más difícil la estimación de sus competencias. Este resultado no presenta mayor soporte a la hipótesis de que los métodos bayesianos de estimación son mejores cuando los individuos tienen *puntajes* nulos o perfectos, o cuando presentan patrones de respuestas aberrantes (Azevedo, 2003); sin embargo, se considera que el presente trabajo no fue diseñado para realizar un análisis exhaustivo de la afirmación, por lo que quizás una investigación, más orientada a ese propósito, sea requerida.

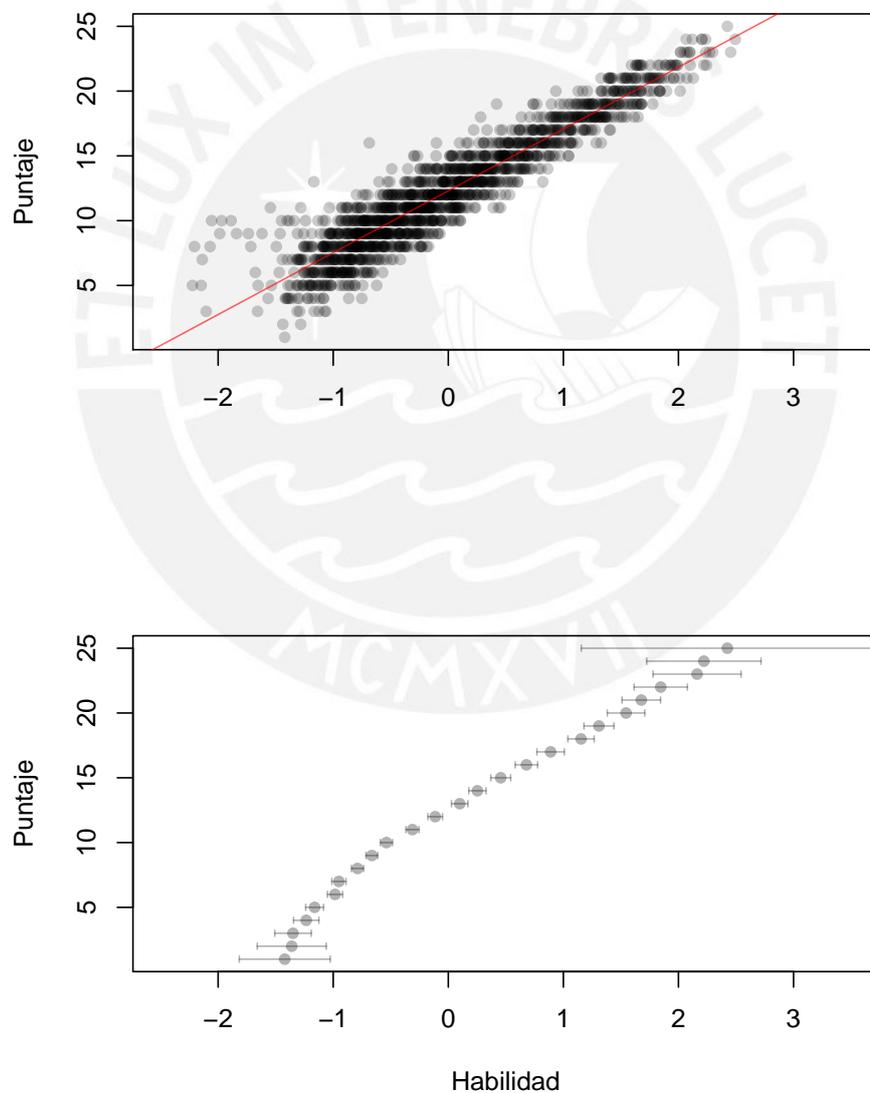


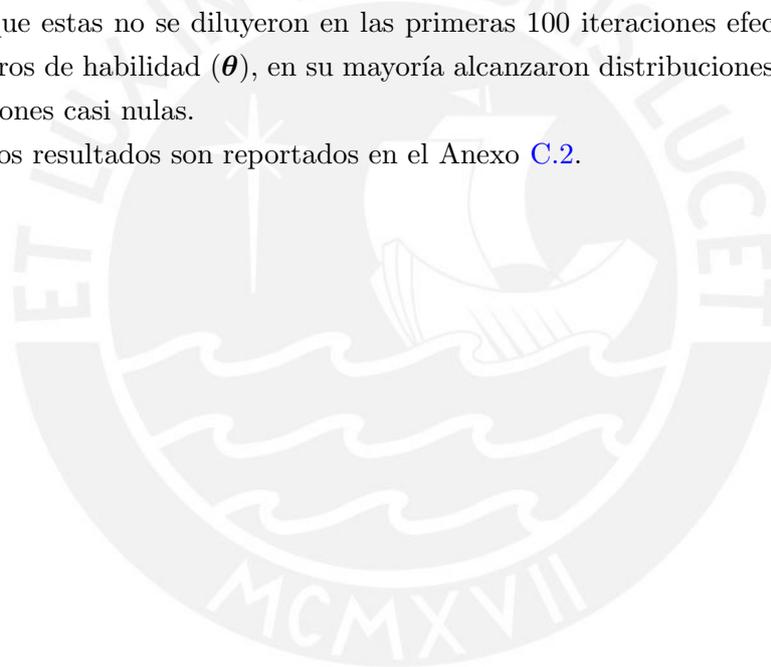
Figura 5.10: Estimador puntual e Intervalo de confianza al 95 % de las competencias de los individuos en Comprensión Lectora

5.2.2. Convergencia de la distribución a posteriori

Finalmente, al igual que en la simulación, se procedió a evaluar la convergencia de la distribución a posteriori de los parámetros hacia una distribución estacionaria. La convergencia se examinó de manera visual a través de diversos gráficos de diagnóstico obtenidos a través de la librería `coda` (Plummer et al., 2006).

Se observa que en dieciocho de los ítems (y sus respectivos parámetros) y las competencias, se observan cadenas que convergen a una distribución estacionaria, con densidades definidas y auto-correlaciones casi nulas. Por su parte, los siete ítems restantes (2, 5, 9, 12, 16, 18 y 20) no registraron similar comportamiento. Se observó que en los ítems 9 y 12 los parámetros alcanzaron estacionariedad, registrando cadenas con auto-correlaciones elevadas que rápidamente decrecen en las primeras 100 iteraciones efectivas resultantes. Similar resultado, pero con mayor magnitud, se observó en los ítems 2, 5, 16, 18 y 20. En estos, los niveles de auto-correlación de la cadena fue significativamente más alto, pero nuevamente, se observa que la cadena termina por alcanzar convergencia en las restantes iteraciones efectivas con algunos casos en los que estas no se diluyeron en las primeras 100 iteraciones efectivas. En relación a los parámetros de habilidad (θ), en su mayoría alcanzaron distribuciones estacionarias con auto-correlaciones casi nulas.

Parte de los resultados son reportados en el Anexo [C.2](#).



Parámetro	Ítem	Alternativa	Media	EE	Mediana	2.5 %	97.5 %
Intercepto	1	1	-0.3685	0.0057	-0.3771	-0.3797	-0.3572
Intercepto	1	2	-0.7076	0.0060	-0.7157	-0.7194	-0.6958
Intercepto	1	3	2.8343	0.0045	2.8140	2.8255	2.8430
Intercepto	1	4	2.1096	0.0044	2.0935	2.1009	2.1183
Intercepto	1	5	-3.8678	0.0165	-3.7625	-3.9001	-3.8355
Intercepto	2	1	0.9689	0.0083	0.8973	0.9526	0.9852
Intercepto	2	2	0.3999	0.0083	0.3356	0.3836	0.4163
Intercepto	2	3	2.1469	0.0082	2.0810	2.1309	2.1630
Intercepto	2	4	2.1461	0.0083	2.0760	2.1299	2.1623
Intercepto	2	5	-5.6618	0.0325	-5.3890	-5.7254	-5.5981
Intercepto	3	1	0.4196	0.0021	0.4148	0.4155	0.4237
Intercepto	3	2	0.4557	0.0020	0.4527	0.4518	0.4597
Intercepto	3	3	1.1782	0.0019	1.1725	1.1745	1.1819
Intercepto	3	4	0.7364	0.0021	0.7305	0.7323	0.7404
Intercepto	3	5	-2.7899	0.0066	-2.7555	-2.8029	-2.7770
Intercepto	4	1	-0.0377	0.0058	-0.0481	-0.0491	-0.0263
Intercepto	4	2	3.4443	0.0049	3.4180	3.4347	3.4540
Intercepto	4	3	0.7244	0.0054	0.7127	0.7138	0.7349
Intercepto	4	4	-0.5840	0.0063	-0.5889	-0.5964	-0.5715
Intercepto	4	5	-3.5471	0.0179	-3.4490	-3.5822	-3.5120
Intercepto	5	1	2.9711	0.0097	2.9285	2.9520	2.9902
Intercepto	5	2	1.4626	0.0097	1.4270	1.4435	1.4816
Intercepto	5	3	2.8975	0.0097	2.8570	2.8785	2.9165
Intercepto	5	4	0.5248	0.0099	0.4937	0.5053	0.5443
Intercepto	5	5	-7.8560	0.0382	-7.7000	-7.9309	-7.7811
Intercepto	6	1	1.7104	0.0049	1.6980	1.7009	1.7200
Intercepto	6	2	0.3197	0.0052	0.3115	0.3095	0.3298
Intercepto	6	3	2.8178	0.0048	2.8015	2.8084	2.8272
Intercepto	6	4	-0.7985	0.0064	-0.8032	-0.8110	-0.7860
Intercepto	6	5	-4.0494	0.0178	-4.0090	-4.0843	-4.0145
Intercepto	7	1	-0.5262	0.0112	-0.5602	-0.5483	-0.5042
Intercepto	7	2	0.0096	0.0106	-0.0331	-0.0113	0.0304
Intercepto	7	3	2.1300	0.0099	2.0675	2.1106	2.1495
Intercepto	7	4	4.0218	0.0099	3.9645	4.0024	4.0412
Intercepto	7	5	-5.6352	0.0379	-5.4020	-5.7094	-5.5610
Intercepto	8	1	-0.4182	0.0038	-0.4254	-0.4256	-0.4108
Intercepto	8	2	1.9935	0.0030	1.9870	1.9877	1.9993
Intercepto	8	3	0.1910	0.0033	0.1835	0.1845	0.1976
Intercepto	8	4	1.7140	0.0030	1.7020	1.7082	1.7199
Intercepto	8	5	-3.4804	0.0112	-3.4390	-3.5023	-3.4584
Intercepto	9	1	-0.2012	0.0170	-0.2855	-0.2345	-0.1679
Intercepto	9	2	4.5918	0.0159	4.5105	4.5606	4.6230
Intercepto	9	3	2.5899	0.0159	2.5095	2.5588	2.6211
Intercepto	9	4	1.5174	0.0161	1.4385	1.4858	1.5490
Intercepto	9	5	-8.4979	0.0628	-8.1780	-8.6210	-8.3748
Intercepto	10	1	-1.1450	0.0038	-1.1410	-1.1525	-1.1376
Intercepto	10	2	1.4799	0.0020	1.4770	1.4760	1.4838
Intercepto	10	3	1.5119	0.0020	1.5090	1.5081	1.5158
Intercepto	10	4	0.9095	0.0020	0.9086	0.9055	0.9135
Intercepto	10	5	-2.7563	0.0063	-2.7345	-2.7687	-2.7440

Cuadro 5.1: Estadísticos de los estimados de los parámetros de Interceptos de los ítems 1 al 10 en el contexto del modelo NRM, bajo el método MCMC de inferencia bayesiana

Parámetro	Ítem	Alternativa	Media	EE	Mediana	2.5 %	97.5 %
Intercepto	11	1	0.5708	0.0033	0.5653	0.5643	0.5772
Intercepto	11	2	1.0504	0.0032	1.0410	1.0442	1.0566
Intercepto	11	3	1.2835	0.0032	1.2710	1.2773	1.2898
Intercepto	11	4	1.0807	0.0032	1.0670	1.0744	1.0870
Intercepto	11	5	-3.9854	0.0121	-3.9440	-4.0091	-3.9617
Intercepto	12	1	1.7671	0.0067	1.7475	1.7540	1.7801
Intercepto	12	2	1.1021	0.0067	1.0825	1.0890	1.1152
Intercepto	12	3	3.0562	0.0066	3.0310	3.0432	3.0692
Intercepto	12	4	0.7454	0.0067	0.7248	0.7322	0.7586
Intercepto	12	5	-6.6708	0.0257	-6.5795	-6.7211	-6.6204
Intercepto	13	1	1.2120	0.0021	1.2075	1.2079	1.2161
Intercepto	13	2	0.5405	0.0021	0.5382	0.5363	0.5447
Intercepto	13	3	0.9976	0.0020	0.9947	0.9936	1.0015
Intercepto	13	4	-0.0457	0.0024	-0.0491	-0.0503	-0.0410
Intercepto	13	5	-2.7044	0.0070	-2.6890	-2.7182	-2.6906
Intercepto	14	1	2.0779	0.0044	2.0610	2.0693	2.0864
Intercepto	14	2	1.8452	0.0044	1.8340	1.8366	1.8537
Intercepto	14	3	-0.5285	0.0050	-0.5346	-0.5382	-0.5187
Intercepto	14	4	1.3751	0.0043	1.3605	1.3667	1.3835
Intercepto	14	5	-4.7697	0.0158	-4.7055	-4.8006	-4.7387
Intercepto	15	1	3.1541	0.0050	3.1500	3.1443	3.1639
Intercepto	15	2	0.4070	0.0055	0.4081	0.3963	0.4178
Intercepto	15	3	1.0383	0.0052	1.0390	1.0281	1.0485
Intercepto	15	4	0.4977	0.0053	0.5017	0.4873	0.5082
Intercepto	15	5	-5.0972	0.0183	-5.1135	-5.1331	-5.0613
Intercepto	16	1	3.3967	0.0140	3.3140	3.3693	3.4242
Intercepto	16	2	4.2105	0.0140	4.1075	4.1829	4.2380
Intercepto	16	3	2.8991	0.0140	2.8070	2.8717	2.9265
Intercepto	16	4	1.6008	0.0143	1.5190	1.5728	1.6288
Intercepto	16	5	-12.1071	0.0557	-11.7100	-12.2163	-11.9978
Intercepto	17	1	1.6644	0.0126	1.6070	1.6396	1.6892
Intercepto	17	2	3.1793	0.0126	3.1285	3.1546	3.2040
Intercepto	17	3	3.4542	0.0126	3.4060	3.4295	3.4788
Intercepto	17	4	1.7643	0.0126	1.7225	1.7396	1.7890
Intercepto	17	5	-10.0622	0.0500	-9.8795	-10.1602	-9.9643
Intercepto	18	1	4.1984	0.0159	4.1365	4.1672	4.2297
Intercepto	18	2	2.4079	0.0159	2.3555	2.3768	2.4391
Intercepto	18	3	2.9576	0.0160	2.9005	2.9263	2.9889
Intercepto	18	4	5.2069	0.0159	5.1425	5.1757	5.2380
Intercepto	18	5	-14.7708	0.0632	-14.5500	-14.8946	-14.6469
Intercepto	19	1	0.6607	0.0032	0.6529	0.6544	0.6671
Intercepto	19	2	0.9857	0.0032	0.9779	0.9795	0.9919
Intercepto	19	3	1.3366	0.0032	1.3280	1.3303	1.3428
Intercepto	19	4	1.1617	0.0032	1.1580	1.1554	1.1680
Intercepto	19	5	-4.1447	0.0121	-4.1180	-4.1684	-4.1210
Intercepto	20	1	1.9646	0.0059	1.9485	1.9530	1.9761
Intercepto	20	2	1.3917	0.0059	1.3845	1.3802	1.4033
Intercepto	20	3	0.7012	0.0061	0.6912	0.6893	0.7131
Intercepto	20	4	2.1474	0.0059	2.1320	2.1359	2.1590
Intercepto	20	5	-6.2049	0.0232	-6.1540	-6.2503	-6.1596

Cuadro 5.2: Estadísticos de los estimados de parámetros de Interceptos de los ítems 11 al 20 en el contexto del modelo NRM, bajo el método MCMC de inferencia bayesiana

Parámetro	Ítem	Alternativa	Media	EE	Mediana	2.5 %	97.5 %
Intercepto	21	1	1.5099	0.0018	1.5075	1.5063	1.5134
Intercepto	21	2	0.1133	0.0022	0.1167	0.1090	0.1176
Intercepto	21	3	0.9271	0.0018	0.9249	0.9235	0.9307
Intercepto	21	4	0.0373	0.0022	0.0354	0.0330	0.0415
Intercepto	21	5	-2.5875	0.0060	-2.5665	-2.5993	-2.5758
Intercepto	22	1	-1.1328	0.0042	-1.1290	-1.1409	-1.1246
Intercepto	22	2	1.1440	0.0022	1.1430	1.1397	1.1483
Intercepto	22	3	1.7020	0.0022	1.6970	1.6977	1.7062
Intercepto	22	4	0.9294	0.0024	0.9274	0.9248	0.9340
Intercepto	22	5	-2.6426	0.0071	-2.6190	-2.6566	-2.6286
Intercepto	23	1	0.1679	0.0023	0.1656	0.1633	0.1725
Intercepto	23	2	2.2908	0.0017	2.2890	2.2874	2.2942
Intercepto	23	3	-0.2046	0.0026	-0.2021	-0.2096	-0.1995
Intercepto	23	4	-0.0713	0.0024	-0.0768	-0.0760	-0.0665
Intercepto	23	5	-2.1828	0.0058	-2.1675	-2.1942	-2.1715
Intercepto	24	1	0.1213	0.0018	0.1218	0.1178	0.1249
Intercepto	24	2	-0.3217	0.0022	-0.3246	-0.3260	-0.3174
Intercepto	24	3	0.2735	0.0017	0.2734	0.2702	0.2768
Intercepto	24	4	1.7610	0.0014	1.7590	1.7582	1.7637
Intercepto	24	5	-1.8340	0.0043	-1.8215	-1.8424	-1.8257
Intercepto	25	1	1.7734	0.0025	1.7670	1.7684	1.7784
Intercepto	25	2	1.3752	0.0025	1.3660	1.3703	1.3801
Intercepto	25	3	-0.3965	0.0032	-0.3991	-0.4028	-0.3902
Intercepto	25	4	0.3370	0.0028	0.3307	0.3315	0.3426
Intercepto	25	5	-3.0892	0.0091	-3.0435	-3.1070	-3.0713

Cuadro 5.3: Estadísticos de los estimados de parámetros de Interceptos de los ítems 21 al 25 en el contexto del modelo NRM, bajo el método MCMC de inferencia bayesiana

Parámetro	Ítem	Alternativa	Media	EE	Mediana	2.5 %	97.5 %
Pendientes	1	1	-0.6884	0.0059	-0.7001	-0.7000	-0.6768
Pendientes	1	2	-0.8789	0.0063	-0.8808	-0.8913	-0.8666
Pendientes	1	3	1.0241	0.0047	1.0190	1.0150	1.0333
Pendientes	1	4	-0.1213	0.0046	-0.1345	-0.1304	-0.1122
Pendientes	1	5	0.6644	0.0168	0.7286	0.6316	0.6973
Pendientes	2	1	0.7587	0.0061	0.7373	0.7467	0.7707
Pendientes	2	2	-0.4474	0.0062	-0.4544	-0.4596	-0.4353
Pendientes	2	3	0.4094	0.0059	0.3889	0.3977	0.4210
Pendientes	2	4	1.3776	0.0060	1.3510	1.3658	1.3894
Pendientes	2	5	-2.0982	0.0231	-1.9930	-2.1434	-2.0531
Pendientes	3	1	-0.4213	0.0024	-0.4216	-0.4260	-0.4166
Pendientes	3	2	-0.0709	0.0023	-0.0718	-0.0754	-0.0663
Pendientes	3	3	0.3042	0.0022	0.3007	0.2999	0.3084
Pendientes	3	4	0.8570	0.0023	0.8514	0.8525	0.8615
Pendientes	3	5	-0.6690	0.0074	-0.6468	-0.6835	-0.6545
Pendientes	4	1	-0.3422	0.0054	-0.3397	-0.3527	-0.3316
Pendientes	4	2	1.3914	0.0044	1.3880	1.3828	1.4001
Pendientes	4	3	-0.2055	0.0049	-0.2104	-0.2151	-0.1958
Pendientes	4	4	0.2056	0.0065	0.2097	0.1930	0.2183
Pendientes	4	5	-1.0494	0.0151	-1.0095	-1.0790	-1.0199
Pendientes	5	1	2.0671	0.0061	2.0545	2.0551	2.0790
Pendientes	5	2	0.5663	0.0061	0.5526	0.5543	0.5783
Pendientes	5	3	1.2571	0.0060	1.2400	1.2454	1.2688
Pendientes	5	4	0.0760	0.0065	0.0718	0.0632	0.0888
Pendientes	5	5	-3.9664	0.0227	-3.9000	-4.0110	-3.9218
Pendientes	6	1	0.7171	0.0043	0.7129	0.7087	0.7256
Pendientes	6	2	-0.1491	0.0046	-0.1428	-0.1581	-0.1401
Pendientes	6	3	1.3185	0.0042	1.3120	1.3102	1.3268
Pendientes	6	4	-0.5728	0.0058	-0.5748	-0.5842	-0.5613
Pendientes	6	5	-1.3138	0.0148	-1.3120	-1.3427	-1.2848
Pendientes	7	1	-0.6637	0.0086	-0.6591	-0.6805	-0.6468
Pendientes	7	2	-0.3730	0.0081	-0.3808	-0.3888	-0.3572
Pendientes	7	3	0.8414	0.0074	0.8365	0.8269	0.8558
Pendientes	7	4	1.9652	0.0073	1.9570	1.9508	1.9796
Pendientes	7	5	-1.7699	0.0267	-1.7650	-1.8222	-1.7176
Pendientes	8	1	-0.3010	0.0039	-0.3015	-0.3086	-0.2933
Pendientes	8	2	0.6181	0.0029	0.6149	0.6124	0.6239
Pendientes	8	3	0.0064	0.0034	0.0001	-0.0002	0.0130
Pendientes	8	4	0.4916	0.0029	0.4905	0.4859	0.4972
Pendientes	8	5	-0.8151	0.0107	-0.7961	-0.8360	-0.7942
Pendientes	9	1	-0.2048	0.0104	-0.2179	-0.2252	-0.1844
Pendientes	9	2	1.6533	0.0089	1.6280	1.6359	1.6706
Pendientes	9	3	1.0851	0.0090	1.0570	1.0675	1.1026
Pendientes	9	4	1.1886	0.0093	1.1570	1.1704	1.2067
Pendientes	9	5	-3.7221	0.0341	-3.6070	-3.7889	-3.6553
Pendientes	10	1	-0.8835	0.0041	-0.8804	-0.8915	-0.8756
Pendientes	10	2	0.3828	0.0023	0.3800	0.3782	0.3873
Pendientes	10	3	0.5735	0.0023	0.5678	0.5690	0.5780
Pendientes	10	4	-0.0038	0.0024	-0.0048	-0.0085	0.0010
Pendientes	10	5	-0.0689	0.0075	-0.0594	-0.0836	-0.0543

Cuadro 5.4: Estadísticos de los estimados de parámetros de Pendientes de los ítems 1 al 10 en el contexto del modelo NRM, bajo el método MCMC de inferencia bayesiana

Parámetro	Ítem	Alternativa	Media	EE	Mediana	2.5 %	97.5 %
Pendientes	11	1	0.4698	0.0029	0.4708	0.4642	0.4755
Pendientes	11	2	-0.0642	0.0027	-0.0698	-0.0695	-0.0590
Pendientes	11	3	0.6242	0.0026	0.6238	0.6190	0.6293
Pendientes	11	4	1.1936	0.0028	1.1870	1.1882	1.1990
Pendientes	11	5	-2.2234	0.0095	-2.2020	-2.2420	-2.2047
Pendientes	12	1	1.0894	0.0049	1.0775	1.0798	1.0990
Pendientes	12	2	0.3850	0.0048	0.3794	0.3756	0.3944
Pendientes	12	3	1.9923	0.0048	1.9830	1.9828	2.0017
Pendientes	12	4	0.5125	0.0050	0.5023	0.5028	0.5222
Pendientes	12	5	-3.9792	0.0172	-3.9420	-4.0130	-3.9454
Pendientes	13	1	1.2669	0.0022	1.2650	1.2626	1.2712
Pendientes	13	2	0.1548	0.0023	0.1534	0.1503	0.1592
Pendientes	13	3	0.0644	0.0021	0.0629	0.0603	0.0685
Pendientes	13	4	0.0037	0.0027	-0.0042	-0.0015	0.0090
Pendientes	13	5	-1.4898	0.0064	-1.4825	-1.5025	-1.4772
Pendientes	14	1	1.4648	0.0036	1.4560	1.4578	1.4717
Pendientes	14	2	1.3554	0.0036	1.3445	1.3484	1.3624
Pendientes	14	3	-0.7671	0.0043	-0.7638	-0.7756	-0.7587
Pendientes	14	4	0.7580	0.0035	0.7512	0.7512	0.7649
Pendientes	14	5	-2.8110	0.0115	-2.7665	-2.8336	-2.7884
Pendientes	15	1	2.2011	0.0042	2.2010	2.1928	2.2093
Pendientes	15	2	0.0336	0.0047	0.0345	0.0245	0.0428
Pendientes	15	3	0.5286	0.0044	0.5305	0.5201	0.5372
Pendientes	15	4	-0.0859	0.0045	-0.0827	-0.0947	-0.0771
Pendientes	15	5	-2.6773	0.0131	-2.6980	-2.7029	-2.6518
Pendientes	16	1	1.7201	0.0089	1.6700	1.7026	1.7376
Pendientes	16	2	2.9591	0.0089	2.9080	2.9417	2.9766
Pendientes	16	3	1.6460	0.0089	1.5890	1.6285	1.6635
Pendientes	16	4	1.0229	0.0093	0.9820	1.0045	1.0412
Pendientes	16	5	-7.3481	0.0348	-7.1250	-7.4162	-7.2800
Pendientes	17	1	1.4632	0.0082	1.4290	1.4471	1.4793
Pendientes	17	2	1.7808	0.0079	1.7510	1.7652	1.7963
Pendientes	17	3	2.2356	0.0080	2.2030	2.2199	2.2512
Pendientes	17	4	0.5744	0.0080	0.5446	0.5587	0.5901
Pendientes	17	5	-6.0539	0.0312	-5.9350	-6.1150	-5.9928
Pendientes	18	1	3.0375	0.0097	3.0050	3.0185	3.0564
Pendientes	18	2	1.6183	0.0098	1.5810	1.5991	1.6375
Pendientes	18	3	1.6955	0.0099	1.6765	1.6761	1.7149
Pendientes	18	4	2.5837	0.0096	2.5500	2.5649	2.6025
Pendientes	18	5	-8.9350	0.0377	-8.8215	-9.0089	-8.8611
Pendientes	19	1	0.3407	0.0028	0.3400	0.3353	0.3462
Pendientes	19	2	0.2251	0.0027	0.2229	0.2198	0.2303
Pendientes	19	3	0.7009	0.0026	0.6971	0.6957	0.7060
Pendientes	19	4	1.1817	0.0027	1.1780	1.1763	1.1871
Pendientes	19	5	-2.4484	0.0094	-2.4390	-2.4668	-2.4299
Pendientes	20	1	1.5227	0.0044	1.5170	1.5141	1.5312
Pendientes	20	2	0.2520	0.0043	0.2440	0.2436	0.2603
Pendientes	20	3	0.8170	0.0046	0.8132	0.8079	0.8261
Pendientes	20	4	1.3409	0.0043	1.3300	1.3325	1.3494
Pendientes	20	5	-3.9326	0.0163	-3.9140	-3.9646	-3.9005

Cuadro 5.5: Estadísticos de los estimados de parámetros de Pendientes de los ítems 11 al 20 en el contexto del modelo NRM, bajo el método MCMC de inferencia bayesiana

Parámetro	Ítem	Alternativa	Media	EE	Mediana	2.5 %	97.5 %
Pendientes	21	1	0.9371	0.0022	0.9321	0.9327	0.9415
Pendientes	21	2	-0.4512	0.0026	-0.4499	-0.4563	-0.4460
Pendientes	21	3	-0.1690	0.0022	-0.1734	-0.1734	-0.1646
Pendientes	21	4	0.2147	0.0027	0.2137	0.2094	0.2199
Pendientes	21	5	-0.5316	0.0070	-0.5140	-0.5454	-0.5178
Pendientes	22	1	-1.0357	0.0043	-1.0300	-1.0442	-1.0272
Pendientes	22	2	0.4903	0.0024	0.4876	0.4856	0.4951
Pendientes	22	3	0.0465	0.0023	0.0411	0.0420	0.0511
Pendientes	22	4	1.1132	0.0026	1.1070	1.1081	1.1182
Pendientes	22	5	-0.6143	0.0077	-0.5876	-0.6293	-0.5993
Pendientes	23	1	-0.0793	0.0027	-0.0832	-0.0846	-0.0741
Pendientes	23	2	0.6998	0.0020	0.6957	0.6958	0.7038
Pendientes	23	3	-0.2053	0.0030	-0.2025	-0.2112	-0.1995
Pendientes	23	4	0.0178	0.0028	0.0185	0.0123	0.0233
Pendientes	23	5	-0.4330	0.0065	-0.4196	-0.4457	-0.4203
Pendientes	24	1	0.1472	0.0021	0.1484	0.1431	0.1513
Pendientes	24	2	-0.3992	0.0025	-0.3963	-0.4042	-0.3943
Pendientes	24	3	0.1650	0.0020	0.1660	0.1610	0.1689
Pendientes	24	4	0.6633	0.0017	0.6631	0.6600	0.6666
Pendientes	24	5	-0.5763	0.0047	-0.5692	-0.5855	-0.5670
Pendientes	25	1	0.9139	0.0027	0.9096	0.9086	0.9191
Pendientes	25	2	0.4513	0.0026	0.4502	0.4461	0.4564
Pendientes	25	3	-0.3572	0.0035	-0.3590	-0.3640	-0.3504
Pendientes	25	4	-0.1687	0.0030	-0.1718	-0.1746	-0.1628
Pendientes	25	5	-0.8392	0.0091	-0.8232	-0.8570	-0.8214

Cuadro 5.6: Estadísticos de los estimados de parámetros de Pendientes de los ítems 21 al 25 en el contexto del modelo NRM, bajo el método MCMC de inferencia bayesiana

Capítulo 6

Conclusión

6.1. Conclusiones

1. Se presentó el modelo de Respuesta Nominal (NRM) propuesto por Bock (1972, 1997), se profundizó sus principales características y propiedades más importantes.
2. Se implementó la estimación de los parámetros del NRM a través de la inferencia bayesiana (MCMC) usando R y WinBUGS.
3. En el contexto del modelo, se realizó un estudio de simulación con la finalidad de comparar la capacidad de recuperación de los parámetros de los ítems y competencias de los individuos de los métodos clásicos (MML) frente a los bayesianos (MCMC). Para este propósito se consideraron dos escenarios casi idénticos diferenciados únicamente por el ratio *tamaño de muestra - número de ítems* (n/p). Los dos diseños quedaron definidos por tamaños de muestra de $n = 200$ y $n = 1600$.
4. Los parámetros de los ítems fueron recuperados de manera similarmente satisfactoria con los métodos bayesianos así como con su contraparte clásica. De esta forma, se podría afirmar que el NRM, bajo el método bayesiano, es un buen sustituto del método clásico.
5. Los parámetros de las pendientes fueron recuperados, consistentemente, de manera más precisa por el método bayesiano. Por su parte, la recuperación de los parámetros de los interceptos dependió del diseño implementado. Los resultados provenientes del método bayesiano fueron marginalmente inferiores cuando los tamaños de muestras fueron menores, mientras que fueron similares a los obtenidos por el método clásico cuando la muestra fue mayor; en particular, cuando tuvo el ratio *tamaño de muestra - número de ítems* fue de 20 a 1 ($n/p = 20 : 1$).
6. El proceso de estimación MCMC fue entre 620 a 14100 veces más lento que el MML, pese al énfasis en la paralelización de los procesos.
7. La aplicación del modelo se realizó en una muestra anónima, aleatoria y representativa de 1641 docentes del nivel secundario de la especialidad de Inglés, modalidad de Educación Básica Regular (E.B.R.), que fueron expuestos a la sub-prueba de Comprensión Lectora del Concurso de Nombramiento 2015.

8. Se logró aplicar satisfactoriamente el NRM, bajo el método bayesiano, a un conjunto de datos reales, que fue el principal propósito del presente estudio.
9. De los resultados se puede afirmar que (en) el modelo:
 - a) la ICRF asociada con la respuesta correcta siempre tiene una relación monotónicamente creciente con a la competencia requerida (Baker y Kim, 1992; Hambleton y Swaminathan, 1991);
 - b) Facilita la recuperación de una mayor proporción información disponible en los ítems, frente a los modelos de respuestas dicotómicas (Bock, 1972; Thissen, 1976; Levine y Drasgow, 1983; Thissen y Steinberg, 1984)
 - c) Permite hallar el ordenamiento implícito en datos categóricos inicialmente no ordenados (Samejima, 1988; Bock, 1997),
 - d) Brinda información relevante para la valoración de la calidad de un ítem (Thissen et al., 1989), que en este caso se tradujo en la labor de encontrar alternativas forzadas y la posibilidad de colapsar otras; y
 - e) Los resultados no ofrecieron medios para probar que el modelo devela perfiles de individuos asociados a las alternativas elegidas (De Ayala, 1993). Por el contrario, se encontró la capacidad del modelo, para develar perfiles de individuos asociados a las alternativas elegidas, depende ampliamente del diseño del instrumento.
10. Finalmente, en la mayoría de los ítems, así como en las competencias, se observa que los parámetros registran cadenas que convergen a una distribución estacionaria, con densidades definidas y auto-correlaciones casi nulas. Se señala que a algunos parámetros de los ítems les costó alcanzar dichas cadenas, pero finalmente las alcanzaron.

6.2. Sugerencias para Investigaciones Futuras

1. Pese a que se realizó una breve investigación empírica acerca del número de iteraciones que debía tomar la cadena para alcanzar parámetros precisos, se recomienda analizar la capacidad de recupero de los parámetros ante un menor número de iteraciones, intentando encontrar el óptimo. Incluso, se insta a realizar un investigación más minuciosa acerca de que factores harían que el modelo requiriera un mayor número de iteraciones para converger a distribuciones estacionarias.
2. Considerar la implementación de Funciones de Respuestas a las Categorías del Ítem (ICRFs) con enlaces asimétricos, que podrían ayudar en el recupero de parámetros de ítems donde la distribución de competencias no se encuentre pareada con la de los parámetros de los ítems. De este modo, el trabajo estaría en línea con corroborar la existencia de los problemas de sobre-parametrización y sobre-flexibilidad ya previstos por Thissen y Steinberg (1997).
3. Comparar los resultados obtenidos bajo el modelo clásico de Bock (1972, 1997) frente al modelo propuesto por Thissen et al. (2010) (sub-sección 2.5.2), e incluso, extender la implementación al caso Multidimensional, acorde con lo propuesto por el autor.

4. Intentar optimizar aún más el tiempo de estimación de los modelos bajo el método bayesiano. En el presente trabajo se consideraron métodos de computación paralela al exterior del proceso MCMC, pero en investigaciones futuras se podrían considerar similares métodos al interior del proceso, con paquetes como `bugsparell`.
5. Implementar el NRM en otras plataformas de simulación bayesiana, como `Stan` (Carpenter et al., 2016) e incluso `Python` (van Rossum, 1995), que según sus autores tienen implementaciones MCMC más robustas; y finalmente comparar resultados con lo obtenido en el presente trabajo, bajo `WinBUGS` (Lunn et al., 2000).



Apéndice A

Resultados teóricos

A.1. Desarrollo conceptual del NRM

Conceptualmente, la distribución de las respuestas de los examinados a través de alternativas mutuamente excluyentes y exhaustivas de ítems del tipo Nominal-Politémico (NP, acorde con Mellenberg, 1995) es categórico; de este modo, la suma de las probabilidades a través de todas las alternativas del ítem es uno.

Asumiendo que se poseen ítems con cuatro alternativas, donde P_1 , P_2 , P_3 y P_4 representan las probabilidades asociadas, uno puede determinar el ratio de Odds (también conocido como ratio de oportunidades) de una respuesta respecto a cualquier otra, por ejemplo, el ratio de oportunidad de la alternativa 2 respecto a la 1 es P_2/P_1 .

Debido a que las probabilidades siempre se encuentran en el rango $[0, 1]$, el ratio de Odds siempre se encontrará en el rango $[0, \infty]$, donde el valor de 1 refleja ninguna diferencia entre las probabilidades de ocurrencias de una alternativa y otra; en el ejemplo esto significaría que $P_1 = P_2$. Dada la asimetría de la escala de Odds, usualmente el ratio es transformado a través de una función monótonica, que preserva el orden original de los datos. La función usualmente usada es el logaritmo (natural), en donde la escala de Odds es transformada al rango $[-\infty, \infty]$.

En la escala logarítmica del ratio de Odds (también conocida por *log Odds* o *logit*) el valor de 0 refleja la igualdad en probabilidades de los eventos, valores positivos indican que la probabilidad de elegir una alternativa es mayor que otra y valores negativos representan lo opuesto. Mas importante aún, el logit permite que uno exprese los Odds en términos de la parametrización pendiente-intercepto considerando un predictor lineal X :

$$\log(P_2/P_1) = c_2 + a_2 \cdot X \quad (\text{A.1})$$

Similar ecuación se puede establecer para cada combinación de alternativas usando a la alternativa 1 como la categoría de respuesta *base*, alternativa que puede ser de particular interés o la que tenga la mayor frecuencia (Agresti, 2007), de la siguiente manera:

$$\log(P_3/P_1) = c_3 + a_3 \cdot X \quad (\text{A.2})$$

$$\log(P_4/P_1) = c_4 + a_4 \cdot X \quad (\text{A.3})$$

La representación gráfica de las ecuaciones [A.1](#), [A.2](#) y [A.3](#) consiste en $(4 - 1)$ líneas rectas

como en la Figura A.1, donde el parámetro c_k es conocido como el intercepto (valor del logit en $X = 0$) y refleja la propensión a elegir una alternativa frente a otra independiente de la influencia de X , mientras a_k conocida como la pendiente y refleja el cambio en el logit a medida que el predictor X cambia en una unidad.

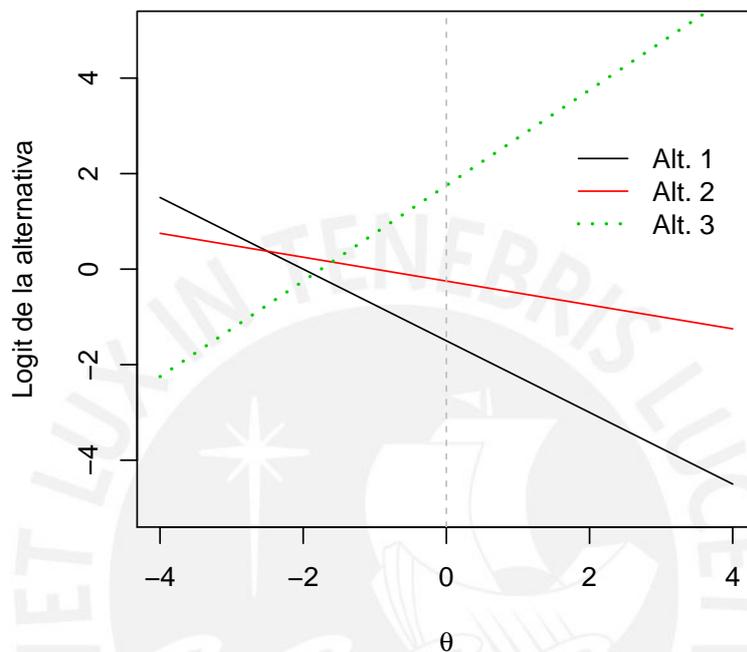


Figura A.1: Líneas logit de regresión para un ítem de tres categorías con $\mathbf{a}_j = (-0.75; -0.25; 1.0)$ y $\mathbf{c}_j = (-1.5; -0.25; 1.75)$. Tomado de De Ayala (1993).

Sin embargo, en vez de referirnos al log del ratio de posibilidad de una categoría respecto a otra, realizando algunos despejes algebraicos podemos expresar directamente la probabilidad de una categoría en particular,

$$\begin{aligned}
 P_2 &= P_1 e^{c_2 + a_2 \cdot X} \\
 P_3 &= P_1 e^{c_3 + a_3 \cdot X} \\
 P_4 &= P_1 e^{c_4 + a_4 \cdot X}
 \end{aligned}$$

Por su parte, dado que las alternativas del ítem son exhaustivas tenemos que $P_1 + P_2 + P_3 + P_4 = 1$. Por lo tanto, podemos expresar la probabilidad condicional de cada alternativa dado

X de la siguiente manera:

$$\begin{aligned} P(x = 1|X) &= \frac{1}{1 + \sum_{h=2}^4 e^{c_h + a_h \cdot X}} = \frac{e^{c_1 + a_1 \cdot X}}{\sum_{h=1}^4 e^{c_h + a_h \cdot X}} \\ P(x = 2|X) &= \frac{e^{c_2 + a_2 \cdot X}}{1 + \sum_{h=2}^4 e^{c_h + a_h \cdot X}} = \frac{e^{c_2 + a_2 \cdot X}}{\sum_{h=1}^4 e^{c_h + a_h \cdot X}} \\ P(x = 3|X) &= \frac{e^{c_3 + a_3 \cdot X}}{1 + \sum_{h=2}^4 e^{c_h + a_h \cdot X}} = \frac{e^{c_3 + a_3 \cdot X}}{\sum_{h=1}^4 e^{c_h + a_h \cdot X}} \\ P(x = 4|X) &= \frac{e^{c_4 + a_4 \cdot X}}{1 + \sum_{h=2}^4 e^{c_h + a_h \cdot X}} = \frac{e^{c_4 + a_4 \cdot X}}{\sum_{h=1}^4 e^{c_h + a_h \cdot X}} \end{aligned}$$

Si generalizamos dejando que m_j represente el número de categorías de respuesta del ítem, tenemos que la probabilidad condicional de elegir la categoría k dado X es

$$P(x = k|X) = \frac{e^{c_k + a_k \cdot X}}{\sum_{h=1}^{m_j} e^{c_h + a_h \cdot X}} \quad (\text{A.4})$$

La ecuación A.4 puede ser empelada en el contexto de IRT si se asume que en vez de una variable observada se posee una variable latente no observada y parámetros latentes del ítem. El resultado de tal re-parametrización nos llevaría al NRM propuesto por Bock (1972) acorde con la presentación en la sección 2.3.1

$$P_{jk}(\theta_i) = P_j(x = k|\theta_i, \mathbf{c}_j, \mathbf{a}_j) = \frac{e^{c_{jk} + a_{jk} \cdot \theta_i}}{\sum_{h=1}^{m_j} e^{c_{jh} + a_{jh} \cdot \theta_i}}$$

A.2. Función de Información del NRM

Las derivaciones para $P'_{jk}(\boldsymbol{\theta}) = \partial P_{jk}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$ y $P''_{jk}(\boldsymbol{\theta}) = \partial^2 P_{jk}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}^2$, necesarias para entender mejor las ICIFs y las IIFs fueron provistas por Vale y Weiss (1977).

De la ecuación (2.2) tenemos:

$$P_{jk}(\boldsymbol{\theta}) = \frac{e^{z_{jk}(\boldsymbol{\theta})}}{\sum_{h=1}^{m_j} e^{z_{jh}(\boldsymbol{\theta})}}$$

con el propósito de simplificar la ecuación tenemos que esta puede ser re-escrita de la siguiente manera:

$$P_{jk}(\boldsymbol{\theta}) = \frac{e^{z_{jk}(\boldsymbol{\theta})}}{D} \quad (\text{A.5})$$

donde,

$$D = \sum_{h=1}^{m_j} e^{z_{jh}(\boldsymbol{\theta})}$$

se tiene además, que en cualquier nivel de “habilidad”,

$$\sum_{k=1}^{m_j} P_{jk}(\boldsymbol{\theta}) = \sum_{k=1}^{m_j} \frac{e^{z_{jk}(\boldsymbol{\theta})}}{D} = \frac{1}{D} \sum_{k=1}^{m_j} e^{z_{jk}(\boldsymbol{\theta})} = \frac{D}{D} = 1$$

y su primera derivada esa definida por,

$$\frac{\partial D}{\partial \theta} = \sum_{h=1}^{m_j} a_{jh} \cdot e^{z_{jh}(\theta)}$$

Entonces, de la ecuación A.5, tenemos que la primera derivada de la probabilidad toma la siguiente forma,

$$\begin{aligned} \frac{\partial P_{jk}(\theta)}{\partial \theta} &= \frac{D \cdot (\partial e^{z_{jk}(\theta)} / \partial \theta) - e^{z_{jk}(\theta)} \cdot (\partial D / \partial \theta)}{D^2} \\ P'_{jk}(\theta) &= \frac{D \cdot a_{jk} \cdot e^{z_{jk}(\theta)} - e^{z_{jk}(\theta)} \cdot \sum_{h=1}^{m_j} a_{jh} \cdot e^{z_{jh}(\theta)}}{D^2} \\ &= \frac{a_{jk} \cdot e^{z_{jk}(\theta)} (e^{z_{j1}(\theta)} + \dots + e^{z_{j(m_j)}(\theta)}) - e^{z_{jk}(\theta)} \cdot (a_{j1} \cdot e^{z_{j1}(\theta)} + \dots + a_{j(m_j)} \cdot e^{z_{j(m_j)}(\theta)})}{D^2} \\ &= \frac{[a_{jk} \cdot e^{z_{j1}(\theta)} \cdot e^{z_{jk}(\theta)} - a_{j1} \cdot e^{z_{j1}(\theta)} \cdot e^{z_{jk}(\theta)}] + \dots + [a_{jk} \cdot e^{z_{j(m_j)}(\theta)} \cdot e^{z_{jk}(\theta)} - a_{j(m_j)} \cdot e^{z_{j(m_j)}(\theta)} \cdot e^{z_{jk}(\theta)}]}{D^2} \\ &= \frac{e^{z_{jk}(\theta)} \cdot [a_{jk} \cdot e^{z_{j1}(\theta)} - a_{j1} \cdot e^{z_{j1}(\theta)}] + \dots + e^{z_{jk}(\theta)} \cdot [a_{jk} \cdot e^{z_{j(m_j)}(\theta)} - a_{j(m_j)} \cdot e^{z_{j(m_j)}(\theta)}]}{D^2} \\ &= \frac{e^{z_{jk}(\theta)} \cdot [e^{z_{j1}(\theta)} \cdot (a_{jk} - a_{j1})] + \dots + e^{z_{jk}(\theta)} \cdot [e^{z_{j(m_j)}(\theta)} \cdot (a_{jk} - a_{j(m_j)})]}{D^2} \\ &= \frac{e^{z_{jk}(\theta)} \cdot \sum_{h=1}^{m_j} e^{z_{jh}(\theta)} \cdot (a_{jk} - a_{jh})}{D^2} \\ &= \frac{e^{z_{jk}(\theta)} \cdot \sum_{h=1}^{m_j} e^{z_{jh}(\theta)} \cdot (a_{jk} - a_{jh})}{D} \cdot \frac{1}{D} \\ P'_{jk}(\theta) &= P_{jk}(\theta) \cdot \frac{\sum_{h=1}^{m_j} e^{z_{jh}(\theta)} \cdot (a_{jk} - a_{jh})}{D} \end{aligned} \tag{A.6}$$

Dado este resultado, la segunda derivada estaría dada por la forma,

$$\begin{aligned} P''_{jk}(\theta) &= \frac{\partial P'_{jk}(\theta)}{\partial \theta} = \frac{\partial^2 P_{jk}(\theta)}{\partial \theta^2} \\ &= \frac{D^2 \cdot \frac{\partial}{\partial \theta} [e^{z_{jk}(\theta)} \cdot \sum_{h=1}^{m_j} e^{z_{jh}(\theta)} \cdot (a_{jk} - a_{jh})]}{D^4} - \frac{\frac{\partial D^2}{\partial \theta} \cdot [e^{z_{jk}(\theta)} \cdot \sum_{h=1}^{m_j} e^{z_{jh}(\theta)} \cdot (a_{jk} - a_{jh})]}{D^4} \end{aligned}$$

para la primera parte de la ecuación tenemos,

$$\begin{aligned}
 & \frac{\partial}{\partial \theta} \left[e^{z_{jk}(\theta)} \cdot \sum_{h=1}^{m_j} e^{z_{jh}(\theta)} \cdot (a_{jk} - a_{jh}) \right] \\
 &= e^{z_{jk}(\theta)} \cdot \frac{\partial}{\partial \theta} \left[\sum_{h=1}^{m_j} e^{z_{jk}(\theta)} \cdot (a_{jk} - a_{jh}) \right] + \sum_{h=1}^{m_j} e^{z_{jh}(\theta)} \cdot (a_{jk} - a_{jh}) \frac{\partial e^{z_{jk}(\theta)}}{\partial \theta} \\
 &= e^{z_{jk}(\theta)} \cdot \left[\sum_{h=1}^{m_j} a_{jk} \cdot a_{jh} \cdot e^{z_{jh}(\theta)} - \sum_{h=1}^{m_j} a_{jh}^2 \cdot e^{z_{jh}(\theta)} \right] + \left[\sum_{h=1}^{m_j} e^{z_{jh}(\theta)} \cdot (a_{jk} - a_{jh}) \right] a_{jk} \cdot e^{z_{jk}(\theta)} \\
 &= e^{z_{jk}(\theta)} \cdot \left[\sum_{h=1}^{m_j} a_{jk} \cdot a_{jh} \cdot e^{z_{jh}(\theta)} - \sum_{h=1}^{m_j} a_{jh}^2 \cdot e^{z_{jh}(\theta)} + \sum_{h=1}^{m_j} a_{jk}^2 \cdot e^{z_{jh}(\theta)} - \sum_{h=1}^{m_j} a_{jk} \cdot a_{jh} \cdot e^{z_{jh}(\theta)} \right] \\
 &= e^{z_{jk}(\theta)} \cdot \left[\sum_{h=1}^{m_j} e^{z_{jh}(\theta)} \cdot (a_{jk}^2 - a_{jh}^2) \right]
 \end{aligned}$$

para la segunda parte de la ecuación tenemos,

$$\begin{aligned}
 \frac{\partial D^2}{\partial \theta} &= 2D \cdot \frac{\partial D}{\partial \theta} \\
 &= 2D \cdot \sum_{h=1}^{m_j} a_{jh} \cdot e^{z_{jh}(\theta)}
 \end{aligned}$$

Combinando las partes obtenemos,

$$\begin{aligned}
 P''_{jk}(\theta) &= \frac{D^2 \cdot \frac{\partial}{\partial \theta} \left[e^{z_{jk}(\theta)} \cdot \sum_{h=1}^{m_j} e^{z_{jh}(\theta)} \cdot (a_{jk} - a_{jh}) \right]}{D^4} \\
 &\quad - \frac{\frac{\partial D^2}{\partial \theta} \cdot \left[e^{z_{jk}(\theta)} \cdot \sum_{h=1}^{m_j} e^{z_{jh}(\theta)} \cdot (a_{jk} - a_{jh}) \right]}{D^4} \\
 &= \frac{D^2 \cdot e^{z_{jk}(\theta)} \cdot \left[\sum_{h=1}^{m_j} e^{z_{jh}(\theta)} \cdot (a_{jk}^2 - a_{jh}^2) \right]}{D^4} \\
 &\quad - \frac{2D \cdot \sum_{h=1}^{m_j} a_{jh} \cdot e^{z_{jh}(\theta)} \cdot \left[e^{z_{jk}(\theta)} \cdot \sum_{h=1}^{m_j} e^{z_{jh}(\theta)} \cdot (a_{jk} - a_{jh}) \right]}{D^4} \\
 &= \frac{e^{z_{jk}(\theta)} \cdot \left[D \cdot \sum_{h=1}^{m_j} e^{z_{jh}(\theta)} \cdot (a_{jk}^2 - a_{jh}^2) - 2 \cdot \sum_{h=1}^{m_j} a_{jh} \cdot e^{z_{jh}(\theta)} \cdot \left[\sum_{h=1}^{m_j} e^{z_{jh}(\theta)} \cdot (a_{jk} - a_{jh}) \right] \right]}{D^3} \\
 &= \frac{e^{z_{jk}(\theta)}}{D} \cdot \frac{\left[D \cdot \sum_{h=1}^{m_j} e^{z_{jh}(\theta)} \cdot (a_{jk}^2 - a_{jh}^2) - 2 \cdot \sum_{h=1}^{m_j} a_{jh} \cdot e^{z_{jh}(\theta)} \cdot \left[\sum_{h=1}^{m_j} e^{z_{jh}(\theta)} \cdot (a_{jk} - a_{jh}) \right] \right]}{D^2} \\
 &= P_{jk}(\theta) \cdot \frac{\left[D \cdot \sum_{h=1}^{m_j} e^{z_{jh}(\theta)} \cdot (a_{jk}^2 - a_{jh}^2) - 2 \cdot \sum_{h=1}^{m_j} a_{jh} \cdot e^{z_{jh}(\theta)} \cdot \left[\sum_{h=1}^{m_j} e^{z_{jh}(\theta)} \cdot (a_{jk} - a_{jh}) \right] \right]}{D^2} \\
 P''_{jk}(\theta) &= P_{jk}(\theta) \cdot \left[\frac{\sum_{h=1}^{m_j} e^{z_{jh}(\theta)} \cdot (a_{jk}^2 - a_{jh}^2)}{D} - \frac{2 \cdot \sum_{h=1}^{m_j} a_{jh} \cdot e^{z_{jh}(\theta)} \cdot \left[\sum_{h=1}^{m_j} e^{z_{jh}(\theta)} \cdot (a_{jk} - a_{jh}) \right]}{D^2} \right]
 \end{aligned} \tag{A.7}$$

Reemplazando los resultados obtenidos en las ecuaciones (A.6) y (A.7) en la ecuación (2.13)

tenemos:

$$\begin{aligned}
 I_{jk}(\boldsymbol{\theta}) &= \left[\frac{P'_{jk}(\boldsymbol{\theta})}{P_{jk}(\boldsymbol{\theta})} \right]^2 - \left[\frac{P''_{jk}(\boldsymbol{\theta})}{P_{jk}(\boldsymbol{\theta})} \right] \\
 I_{jk}(\boldsymbol{\theta}) &= \left[\frac{\sum_{h=1}^{m_j} e^{z_{jh}(\boldsymbol{\theta})} \cdot (a_{jk} - a_{jh})}{D} \right]^2 \\
 &\quad - \frac{\sum_{h=1}^{m_j} e^{z_{jh}(\boldsymbol{\theta})} \cdot (a_{jk}^2 - a_{jh}^2)}{D} \\
 &\quad + \frac{2 \cdot \sum_{h=1}^{m_j} a_{jh} \cdot e^{z_{jh}(\boldsymbol{\theta})} \cdot \left[\sum_{h=1}^{m_j} e^{z_{jh}(\boldsymbol{\theta})} \cdot (a_{jk} - a_{jh}) \right]}{D^2}
 \end{aligned} \tag{A.8}$$

donde la ecuación (A.8) es igual a la (2.14) presentada en la sub-sección 2.3.2. Es importante señalar, que mientras estos términos son matemáticamente complejos, son fáciles de computar y optimizar.



Apéndice B

Códigos

B.1. Código en R para simular datos aleatorios del NRM

B.1.1. Función para la simulación de datos

Con el propósito de generar los datos politómicos, se utilizó la función `NRM.sim`, proveniente del paquete `mcIRT` (Reif, 2014) (versión beta), la cual fue alimentada con la información generada del código de la sección previa. La formulación del modelo, seguida por el paquete, es la siguiente:

$$P_{iqv} = \frac{\exp(\zeta_{ik} - \lambda_{ik}\theta_j)}{\sum_{v=1}^m \exp(\zeta_{iv} - \lambda_{iv}\theta_j)} \quad (\text{B.1})$$

Donde ζ_{ik} sería el intercepto (similar al c_{jk}), λ_{ik} representaría la pendiente del modelo (similar al a_{jk}) y el θ_j representa la habilidad del individuo j . Notar que el autor del paquete usa j como el índice que identifica al individuo, mientras que en el presente documento el índice i hace el mismo trabajo.

```
#####  
# Autor: Reif (2014)  
# Versión: 0.41  
# Fecha de inicio: 30-08-2014  
# Última Modificación: 20-02-2015  
# Función: NRM.sim()  
# Descripción: La función simula datos acorde al Modelo de Respuesta Nominal.  
#####
```

```
NRM.sim <- function (Parlist , pp){  
  
  if (!is.list(Parlist)) {  
    stop("Parlist has to be a list!\n")  
  }  
  
  lparl <- sapply(Parlist , function(count) length(count) %% 2)
```

```

if (any(lparl != 0)) {
stop(paste("Wrong number of categories for items:", which(lparl!=0)))
}

realis <- lapply(Parlist, function(ITEM) {
Pdist <- mapply(function(ze, la) exp(ze + la * pp),
ze = ITEM[1:(length(ITEM)/2)],
la = ITEM[(length(ITEM)/2 + 1):length(ITEM)])

Pd2 <- Pdist/rowSums(Pdist)

spalte <- sapply(1:length(pp), function(NEX) {
back <- sample(0:((length(ITEM)/2) - 1), 1, prob=Pd2[NEX,])
return(back)
}
}

dfsim <- data.frame(realis)
return(dfsim)

}

```

B.1.2. Estudio de simulación

```

#####
# Autor: José Manuel Rivera Espejo
# Versión: 1.1
# Fecha de inicio: 12-09-2016
# Última Modificación: 17-06-2017
# Asesor: Enver Gerald Tarazona Vargas
# Función: estudio_simulacion()
# Descripción: Establecimiento de parámetros preliminares y del modelo
# para realizar el estudio de simulación.
#####

# parámetros de pendientes
a_jk <- data.frame(a_01k = c(-1.14, -0.15, 0.15, 1.14),
a_02k = c(1.25, 0.07, 0.42, -1.74),
a_03k = c(1.09, 0.69, -0.2, -1.58), a_04k = c(0.3, -1.17, 1.01, -0.14),
a_05k = c(0.94, 0.42, -0.5, -0.86), a_06k = c(-0.21, -1.07, 0.39, 0.89),
a_07k = c(-0.89, -0.29, -0.51, 1.69), a_08k = c(-0.32, -0.8, 0.99, 0.13),

```

```
a_09k = c(-0.42, -0.62, 0.73, 0.31), a_10k = c(-0.53, 1.62, -0.55, -0.54))
```

```
# parámetros de interceptos
```

```
c_jk <- data.frame(c_01k = c(-1.65, 0.16, 0.92, 0.57),
```

```
c_02k = c(1.73, 0.38, 0.62, -2.73),
```

```
c_03k = c(1.2, -0.26, 0.34, -1.28), c_04k = c(0.20, -0.98, 0.74, 0.04),
```

```
c_05k = c(0.70, -0.73, 0.11, -0.08), c_06k = c(0.30, 0.10, -0.84, 0.44),
```

```
c_07k = c(-0.10, 0.34, 0.33, -0.57), c_08k = c(0.78, -0.87, -0.22, 0.31),
```

```
c_09k = c(0.50, 1.40, -0.97, -0.93), c_10k = c(0.02, -1.86, -0.67, 2.51))
```

```
# función de simulación #####
```

```
estudio_simulacion <- function(dist="Normal", p=10, ratio=c(2.5, 20),  
  m_j=4, R=20, a_jk, c_jk) {
```

```
# a. parámetros preliminares #####
```

```
# dist = distribuciones muestreadas
```

```
# p = una sola longitud de prueba
```

```
# ratio = dos ratios de muestras
```

```
# m_j = numero de alternativas para todos los ítems
```

```
# R = numero de replicas
```

```
# a_jk = data.frame con parámetros de pendientes, una columna por ítem
```

```
# c_jk = data.frame con parámetros de interceptos, una columna por ítem
```

```
n <- c(p*m_j*2*ratio[1], p*m_j*2*ratio[2]) # dos tamaños de muestras
```

```
# comprobamos si suman cero
```

```
# round(colSums(c_jk), 2)
```

```
# round(colSums(a_jk), 2)
```

```
# b. generación de lista de parametros #####
```

```

param <- list()

for(j in 1:p){
  param[[i]] <- c(c_jk[,j], a_jk[,j])
  names(param[[i]]) <- c(paste("c_",j,"_",1:m_j, sep=""),
  paste("a_",j,"_",1:m_j, sep=""))
}

names(param) <- paste("item", 1:p, sep="")

for(a in 1:length(n)){
  for(b in 1:R){

# generamos habilidades
hab <- rnorm(n[a], mean=0, sd=1)

# generamos respuestas
assign(paste0('sim_',n[a],'_',b), NRM.sim(param, hab))

# almacenamos habilidades
nam <- paste0('hab_d', dist, '_n', n[a], '_p', p, '_R', b, '.csv')
dir1 <- file.path(getwd(), '3.data/simulacion/ejercicio_3', nam)
write.table(hab, dir1, sep=',', dec='.', row.names=F, col.names=F)

# almacenamos respuestas
nam <- paste0('sim_d', dist, '_n', n[a], '_p', p, '_R', b, '.csv')
dir2 <- file.path(getwd(), '3.data/simulacion/ejercicio_3', nam)
write.table(get(paste0('sim_',n[a],'_',b)),
            dir2, sep=",", dec=".", row.names=F)
}
}
}

```

B.2. Código en R para realizar la estimación del modelo

Es importante mencionar, que tanto para el caso de la estimación clásica como bayesiana, la estrategia (4.2), se apalancó en la potencia de los paquetes `foreach` (Weston, 2015) y `doParallel` (Weston y Calaway, 2015) para paralelizar las estimaciones de los parámetros del modelo, reduciendo significativamente el tiempo de implementación del mismo.

Por su parte, para la aplicación se procedió a aplicar el modelo sin el uso de los mencionados paquetes, pues ya no se requería realizar estimaciones por réplicas, anulando la necesidad

del uso de mencionados paquetes.

A continuación se presentan las líneas de código utilizadas para realizar la implementación del modelo tanto en los datos simulados como en los datos de aplicación.

B.2.1. Función de las Categorías del Ítem

```
#####
# Autor: José Manuel Rivera Espejo
# Versión: 0.1
# Fecha de inicio: 06-04-2018
# Última Modificación: 06-04-2018
# Asesor: Enver Gerald Tarazona Vargas
# Función: probability_function()
# Descripción: La función crea las ICRFs para un continuo theta.
#####

probability_function <- function(a_j, c_j, theta=seq(-4,4, by=0.001)){

# librerías
library(tibble)

# a. vector logit multivariado #####
z_jk_mom <- outer(outer(theta, a_j, FUN='*'), c_j, FUN='+')

z_jk <- matrix(NA, ncol=length(a_j), nrow=length(theta))

for(i in 1:length(a_j)){
z_jk[,i] <- z_jk_mom[,i,i]
}
# head(z_jk)

# b. cálculos de probabilidades #####
P_jk <- as_tibble(exp(z_jk)/rowSums(exp(z_jk)))
```

```

# c. otras variables
P_jk$theta <- theta
P_jk <- P_jk[,c(ncol(P_jk), 1:length(a_j))]
names(P_jk) <- c('theta', paste0('Alt_', 1:length(a_jk)))
# head(P_jk)

return(P_jk)

}

```

B.2.2. Función de Información de las Alternativas e Ítem

```

#####
# Autor: José Manuel Rivera Espejo
# Versión: 0.1
# Fecha de inicio: 02-04-2018
# Última Modificación: 02-04-2018
# Asesor: Enver Gerald Tarazona Vargas
# Función: information_function()
# Descripción: La función crea las ICIFs y IIF para un continuo theta.
#####

information_function <- function(a_j, c_j, theta=seq(-4,4, by=0.001)){

# librerías
library(tibble)

# a. vector logit multivariado #####
z_jk_mom <- outer(outer(theta, a_j, FUN='*'), c_j, FUN='+')

z_jk <- matrix(NA, ncol=length(a_j), nrow=length(theta))

for(i in 1:length(a_j)){
z_jk[,i] <- z_jk_mom[,i,i]
}
# head(z_jk)

```

```

# b. cálculos de probabilidades #####
P_jk <- exp(z_jk)/rowSums(exp(z_jk))
# head(P_jk)

# c. matriz ajk_x-Wjk_x-ajk #####
aj_x-Wj_x-aj <- matrix(NA, ncol=length(a_j)^2, nrow=length(theta))
# str(aj_x-Wj_x-aj)

# vemos si aplicamos distintos cálculos
# i=3
# j=3
for(i in 1:length(a_j)){
  for(j in 1:length(a_j)){

    if(i==j){
      # calculamos valores momentáneos
      aj_x-Wj_x-aj_mom <- (a_j[i]*a_j[j])*P_jk[,i]*(1-P_jk[,j])

      # indice de ubicación en W_j
      index <- i + 3*(i-1)

    } else {
      # calculamos valores momentáneos
      aj_x-Wj_x-aj_mom <- (a_j[i]*a_j[j])*(-P_jk[,i]*P_jk[,j])

      # indice de ubicación en W_j
      index <- j + 3*(i-1)

    }

    # index

    # asignamos valores de la matriz W_j
    aj_x-Wj_x-aj[,index] <- aj_x-Wj_x-aj_mom

  }
}
# head(aj_x-Wj_x-aj)

```

```

# d. matriz W_j #####
rowsum_W_j <- rowSums(a_j_x-Wj_x-a_j)
# head(rowsum_W_j)

# e. ICIFs #####
ICIFs <- rowsum_W_j*P_jk
# head(ICIFs)

# f. IIF #####
IIF <- rowsum_W_j
# head(P_jk)
# head(IIF)

# g. unión de datos #####
Information <- as_tibble(cbind(theta, ICIFs, IIF))
names(Information)[1:(length(a_j)+1)] <- c('theta',
      paste0('ICIF', 1:length(a_j)))
# head(Information)

return(Information)

}

B.2.3. Estimación Clásica
#####
# Autor: José Manuel Rivera Espejo
# Versión: 1.3
# Fecha de inicio: 12-09-2016
# Última Modificación: 27-04-2017
# Asesor: Enver Gerald Tarazona Vargas
# Función: estimacion_clasica_sim()
# Descripción: Código en R para realizar la estimación clásica del modelo NRM,
# en los datos de simulación generados.
#####

# archivos a correr
in <- file.path(getwd(), "3.data/simulacion/ejercicio_3")
out <- file.path(getwd(), "7.resultados/Clasico/ejercicio_3")

```

```

# función estimación clásica #####
estimacion_clasica_sim <- function(R=20, in , out){

# a. librerías #####
library(stringr)
library(foreach)
library(doParallel)
library(mcIRT)

# b. Definición de parametros previos #####
# R = número de replicas a analizar
# in = dirección donde se encuentran los archivos a revisar para estimar
# out = dirección general donde se almacenarán los resultados

# c. Estimación Clasica #####
archivos <- list.files(in)
archivos <- archivos[str_detect(archivos , 'sim')]
total <- 1:length(archivos)

for(a in total){

# cargando datos
in2 <- file.path(in , archivos[a])
data_zip <- unzip(in2)

# realizamos la estimación paralelizada
cores <- detectCores() - 2      # nucleos disponibles menos 2
cl <- makeCluster(cores)      # activamos clusters
registerDoParallel(cl)        # registramos clusters
clusterSetRNGStream(cl)      # para hacer replicable el ejercicio

tiempo <- system.time(
res <- foreach(b=1:R, .packages="mcIRT") %dopar% {

```

```

# cargando replica
data <- read.csv(data_zip[b], sep="," , dec="." , stringsAsFactor=F)

# generando nombres para almacenamiento
nombres <- paste0(rep(names(data), each=8), "_",
rep(c("zet", "lam"), each=4), 1:4)

# Preparamos la data
data1 <- reshMG(data, items=1:ncol(data), correct=rep(0,ncol(data)))

# ajuste del modelo
res_mom <- nrm(data1)

# almacenamiento de parametros de items
parm <- unlist(res_mom$ZLpar$A)
SE_parm <- unlist(attr(res_mom$SE,"listform")[[1]])
params <- data.frame(nombres, parm, SE_parm)

# almacenamiento de parametros de personas
hab <- unlist(res_mom$EAPs)

# almacenamiento conjunto
res <- list(hab=hab, params)
res
}
)

stopCluster(cl)

# d. almacenando resultados ####
nam <- str_replace_all(archivos[a], '.zip', '')
nam <- str_replace_all(nam, 'sim_', '')

for(b in 1:R){
nam1 <- paste0('habilidad', b, "_", nam, '.csv')
out1 <- file.path(out, nam1)
write.table(res[[b]][1], out1, sep="," , dec="." , row.names=F)

nam2 <- paste0('parametros', b, "_", nam, '.csv')
out2 <- file.path(out, nam2)
write.table(res[[b]][2], out2, sep="," , dec="." , row.names=F)
}

```

```

}

# e. almacenamiento del tiempo de procesamiento #####
out3 <- file.path(out, paste0("tiempo_clasico", a, ".txt"))
write.table(unclass(tiempo[1:3]), out3, col.names=F)

}

}

```

B.2.4. Estimación Bayesiana

```

#####
# Autor: José Manuel Rivera Espejo
# Versión: 1.3
# Fecha de inicio: 12-09-2016
# Última Modificación: 09-05-2017
# Asesor: Enver Gerald Tarazona Vargas
# Función: estimacion_bayesiana.sim()
# Descripción: Código en R para realizar la estimación bayesiana del modelo NRM,
# usando MCMC en WINBUGS, en los datos de simulación generados.
#####

# archivos a correr
in <- file.path(getwd(), "3.data/simulacion/ejercicio_3")
out <- file.path(getwd(), "7.resultados/Bayesiano/ejercicio_3")

# modelo y directorio de Winbugs
model <- file.path(getwd(), "5.code/analisis_3", "20170309_NRM.bug")
winbug_dir <- "C:/Users/jrivera/Desktop/PortableApps/WinBUGS14/"

# función estimación bayesiana #####
estimacion_bayesiana_sim <- function(R=20, in, out, winbug_dir, model){

# a. librerías
library(stringr)
library(foreach)
library(doParallel)
library(mcIRT)

```

```

library(coda)
library(R2WinBUGS)

# b. Definición de parametros previos #####
# R = número de replicas a analizar
# in = dirección donde se encuentran los archivos a revisar para estimar
# out = dirección general donde se almacenarán los resultados
# model = dirección donde se encuentra el modelo a utilizar
# winbug_dir = dirección donde se encuentra el ejecutable de WINBUGS

# c. Estimacion Bayesiana #####
archivos <- list.files(in)
archivos <- archivos[str_detect(archivos, 'sim')]
total <- 1:length(archivos)

for(a in total){

# des-zipeamos datos
dir1 <- file.path(in, archivos[a])
data_zip <- unzip(dir1)

# realizamos la estimación paralelizada
cores <- detectCores() - 2      # nucleos disponibles menos 1
cl <- makeCluster(cores)      # activamos clusters
registerDoParallel(cl)      # registramos clusters
clusterSetRNGStream(cl)

tiempo <- system.time(
res <- foreach(b=1:R, .packages=c('R2WinBUGS', 'coda')) %dopar% {

# cargando datos
data <- read.csv(data_zip[b], sep="," , dec="." , stringsAsFactor=F)

# transformamos la data en matriz
X <- as.matrix(data + 1)

```

```

# definimos parametros previos
n <- nrow(X)
p <- ncol(X)
m <- apply(X, 2, max)

# Definiendo parametros para prioris
# habilidades de personas - N(0,1)
mu_theta <- 0.0
tau_theta <- 1.0

# interceptos - N(0,100), en bugs N(0,1/100)
mu_int <- 0.0
tau_int <- 1/100

# pendientes - N(0,100), en bugs N(0,1/100)
mu_pen <- 0.0
tau_pen <- 1/100

# data necesaria para el modelo
datos <- list(X=X, n=n, p=p, m=m,
mu_theta=mu_theta, tau_theta=tau_theta,
mu_int=mu_int, tau_int=tau_int,
mu_pen=mu_pen, tau_pen=tau_pen)

monitor <- c("theta", "int", "pen")

# parametros de la corrida
n.burn <- 3000
n.thin <- 17
n.sim <- 2000*n.thin + n.burn

# Estimacion
nrm.out <- bugs(data=datos, inits=NULL, model.file=model,
parameters.to.save=monitor, bugs.directory=winbug_dir,
debug=F, DIC=T, n.chains=1,
n.iter=n.sim, n.thin=n.thin, n.burnin=n.burn)

nrm.out <- list(sim=data.frame(nrm.out$sims.array[,1,]),

```

```

data.frame(pD=nrm.out$pD, DIC=nrm.out$DIC))

nrm.out

}
)

stopCluster(cl)

# almacenando resultados
for(b in 1:R){

nam <- str_replace_all(archivos[a], '.zip', '')
nam <- str_replace_all(nam, 'sim_', '')
nam <- paste0("estimados", b, "_", nam, '.csv')
out1 <- file.path(out, nam)
write.table(unclass(res[[b]][1]), out1, sep=",", dec=".", row.names=F)

nam <- str_replace_all(nam, 'estimados', 'indic')
out2 <- file.path(out, nam)
write.table(unclass(res[[b]][2]), out2, sep=",", dec=".", row.names=F)

}

# almacenamiento del tiempo de procesamiento
out3 <- file.path(out, paste0("tiempo_bayesiano", a, ".txt"))
write.table(unclass(tiempo[1:3]), out3, col.names=F)

}

}

```

B.2.5. Código en R para la aplicación en el análisis de datos educacionales

Estimación Bayesiana

```

#####
# Autor: José Manuel Rivera Espejo
# Versión: 1.3
# Fecha de inicio: 21-03-2017
# Última Modificación: 22-05-2017
# Asesor: Enver Gerald Tarazona Vargas
# Función: NRM_aplicacion_bayesiana()
# Descripción : Código en R para realizar la estimación bayesiana del modelo NRM,

```

```

# de los datos educacionales usando MCMC en WINBUGS.
#####

# archivos a correr
out <- file.path(getwd(), "7.resultados/Bayesiano/ejercicio_3")

# modelo y directorio de Winbugs
model <- file.path(getwd(), '5.code/analisis_4', '20170309v1752_NRM.bug')
dir <- "C:/Users/jrivera/Desktop/PortableApps/WinBUGS14/"

# transformamos la data en matriz
X <- as.matrix(data)

# definimos parametros previos
n <- nrow(X)
p <- ncol(X)
m <- apply(X, 2, max)

# Definiendo parametros para prioris
# habilidades de personas - N(0,1)
mu_theta <- 0.0
tau_theta <- 1.0

# interceptos - N(0,100), en bugs N(0,1/100)
mu_int <- 0.0
tau_int <- 1/100

# pendientes - N(0,100), en bugs N(0,1/100)
mu_pen <- 0.0
tau_pen <- 1/100

# data necesaria para el modelo
datos <- list(X=X,
n=n,
p=p,
m=m,

```

```

mu_theta=mu_theta ,
tau_theta=tau_theta ,
mu_int=mu_int ,
tau_int=tau_int ,
mu_pen=mu_pen ,
tau_pen=tau_pen)

monitor <- c("theta", "int", "pen")

# parametros de la corrida
n.burn <- 2000*7*2
n.thin <- 28
n.sim <- 2000*n.thin + n.burn

# modelo
time <- system.time(
nrm.out <- bugs(data=datos ,
inits=NULL,
parameters.to.save=monitor ,
model.file=model,
bugs.directory=dir ,
DIC=T,
debug=T,
n.iter=n.sim ,
n.thin=n.thin ,
n.burnin=n.burn ,
n.chains=1))

# almacenamiento
dir <- file.path(out, 'parametros_bayesianos1.csv')
write.table(data.frame(nrm.out$sims.array[,1,]),
dir, sep="," , dec="." , row.names=F)

dir <- file.path(out, 'indic1.csv')
write.table(data.frame(pD=nrm.out$pD, DIC=nrm.out$DIC),
out, sep="," , dec="." , row.names=F)

dir1 <- file.path(out, 'tiempo_bayesiano1.txt')

```

```
write.table(unclass(time[1:3]), dir1, col.names=F)

dir <- file.path(out, 'aplicacion_20190604v1630.RData')
save.image(file=dir)

}
```

B.3. Código BUGS para la estimación del Modelo

```
#####
# Autor: José Manuel Rivera Espejo
# Versión: 1.3
# Fecha de inicio: 12-09-2016
# Última Modificación: 09-03-2017
# Asesor: Enver Gerald Tarazona Vargas
# Función: 20170309_NRM.bug
# Descripción : Función en BUGS para realizar la inferencia del Modelo NRM.
#####
```

```
model{
#-----
# Prioris
#-----
#-----
# Habilidades
#-----
for (i in 1:n){
  theta[i] ~ dnorm(mu_theta, tau_theta)
}

#-----
# Parametros de los ítems
#-----
for (j in 1:p){
  for (k in 2:m[j]){
    int[j,k] ~ dnorm(mu_int, tau_int)
    pen[j,k] ~ dnorm(mu_pen, tau_pen)
  }

  int[j,1] <- (-sum(int[j,2:m[j]]))
  pen[j,1] <- (-sum(pen[j,2:m[j]]))
}
```

```
#####  
# Modelo  
#####  
for (i in 1:n){  
  for (j in 1:p){  
    for (k in 1:m[j]){  
      # vector logit multivariado  
      z[i,j,k] <- int[j,k] + pen[j,k]*theta[i]  
    }  
  
    for (k in 1:m[j]){  
      # numerador  
      tt[i,j,k] <- exp(z[i,j,k])  
    }  
  
    for (k in 1:m[j]){  
      # calculo de probabilidades  
      P[i,j,k] <- tt[i,j,k] / sum(tt[i,j,])  
    }  
  }  
  
  for (j in 1:p){  
    # Respuestas  
    X[i,j] ~ dcat(P[i,j,])  
  }  
}  
}
```

Apéndice C

Diagnósticos de convergencia

C.1. Simulación: Diagnósticos de Convergencia

C.1.1. Diseño 1:

Distribución Normal Estándar, $p = 10$, $m_j = 4$, $n/p = 2.5 : 1$ y $n = 200$

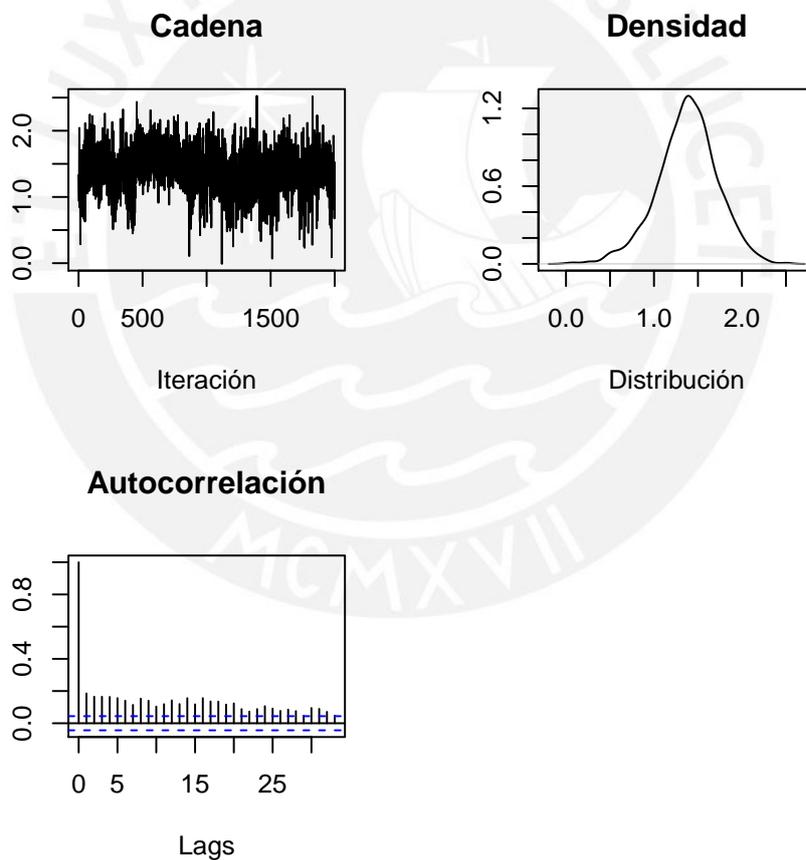


Figura C.1: Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la habilidad del individuo 48, réplica 7, estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0, 1)$, $n = 200$ y $p=10$

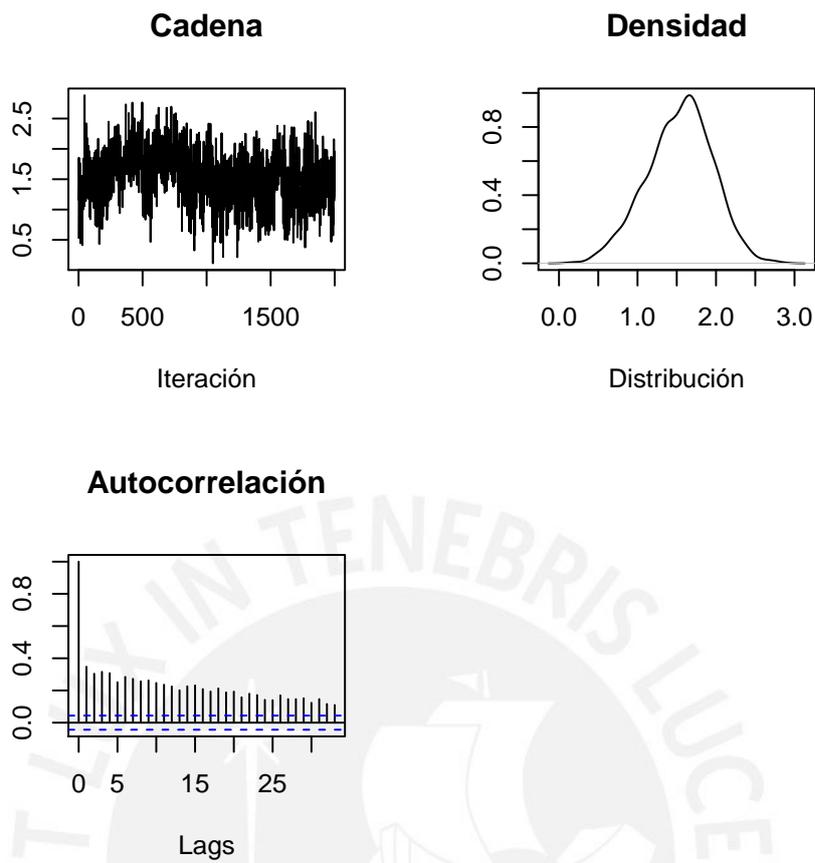


Figura C.2: Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la habilidad del individuo 194, réplica 8, estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0,1)$, $n = 200$ y $p=10$

Ejemplos de densidades estacionarias de parámetros de ítems, con auto-correlaciones bajas:

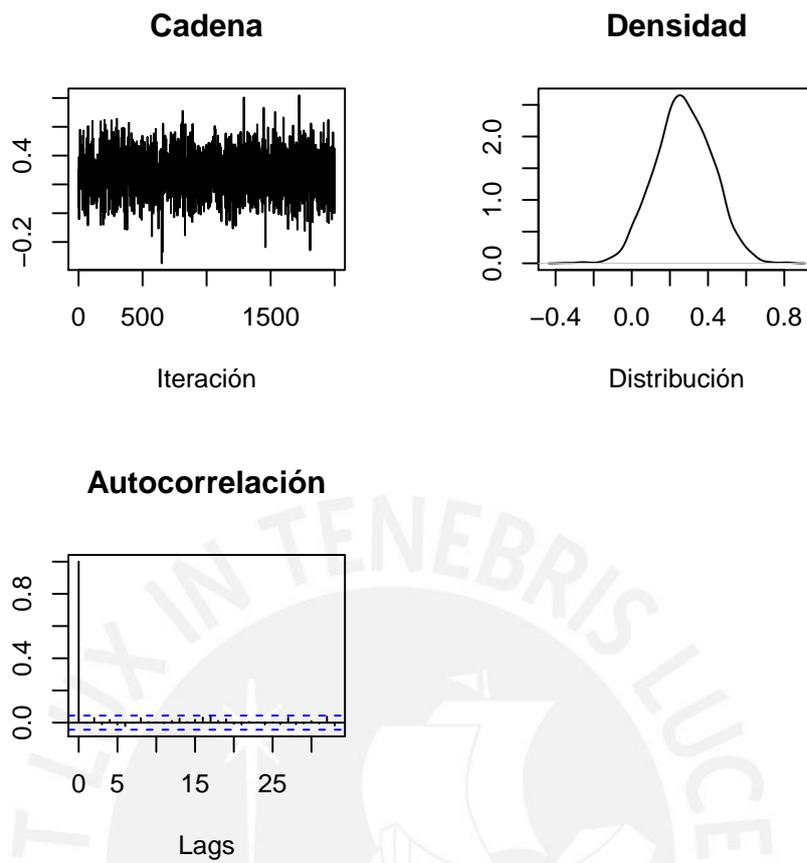


Figura C.3: Iteraciones efectivas, gráfica de densidad y auto-correlaciones para el intercepto del ítem 6, alternativa 1, réplica 1, estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0,1)$, $n = 200$ y $p=10$

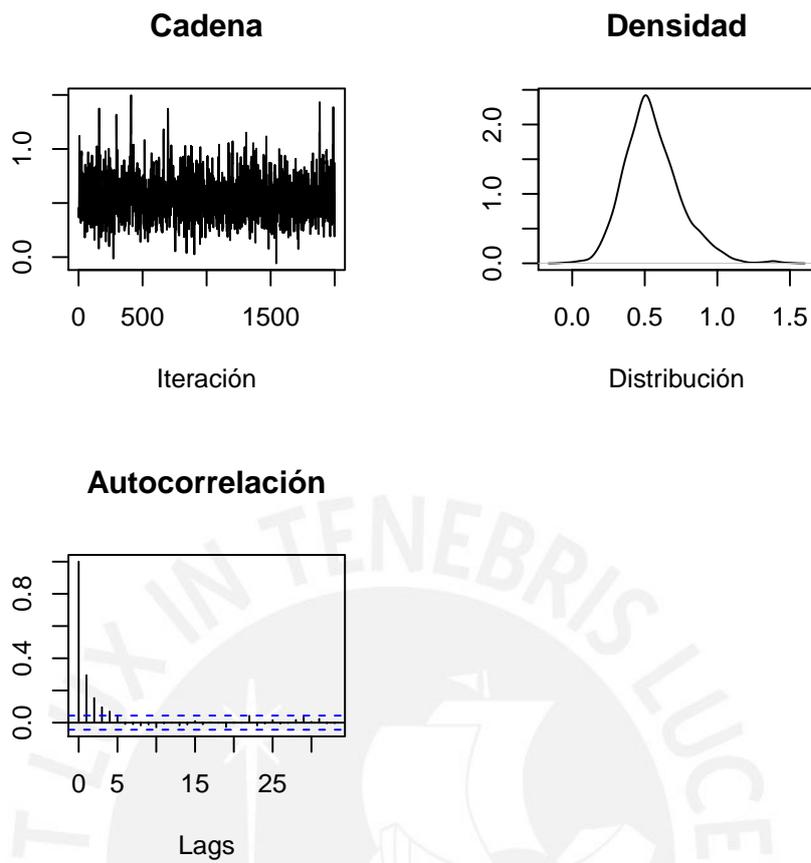


Figura C.4: Iteraciones efectivas, gráfica de densidad y auto-correlaciones para el intercepto del ítem 7, alternativa 2, réplica 4, estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0,1)$, $n = 200$ y $p=10$

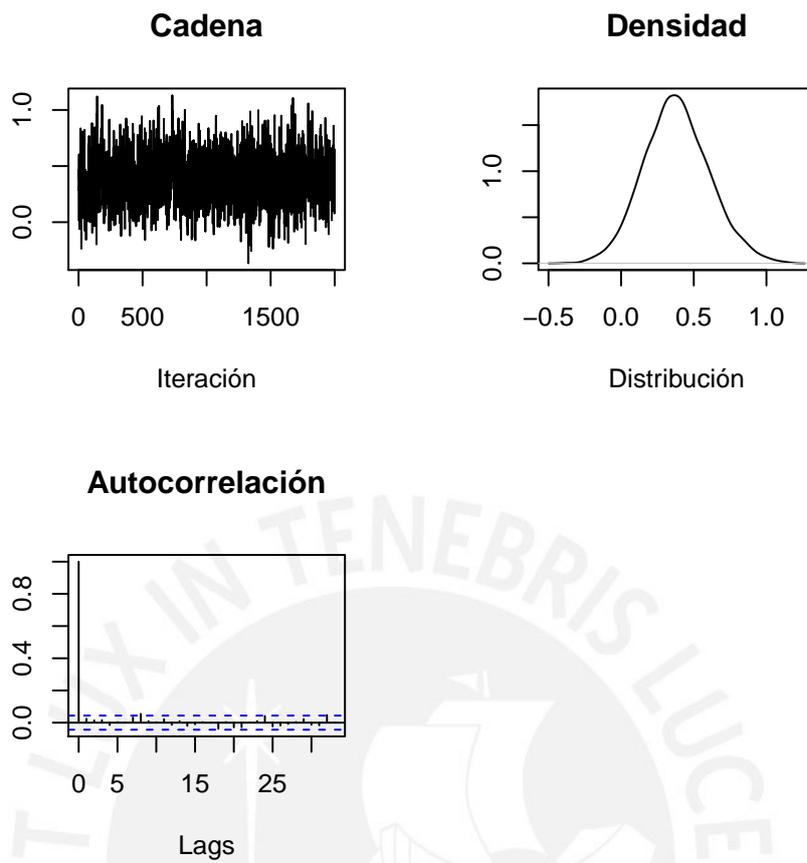


Figura C.5: Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la pendiente del ítem 4, alternativa 1, réplica 2, estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0,1)$, $n = 200$ y $p=10$

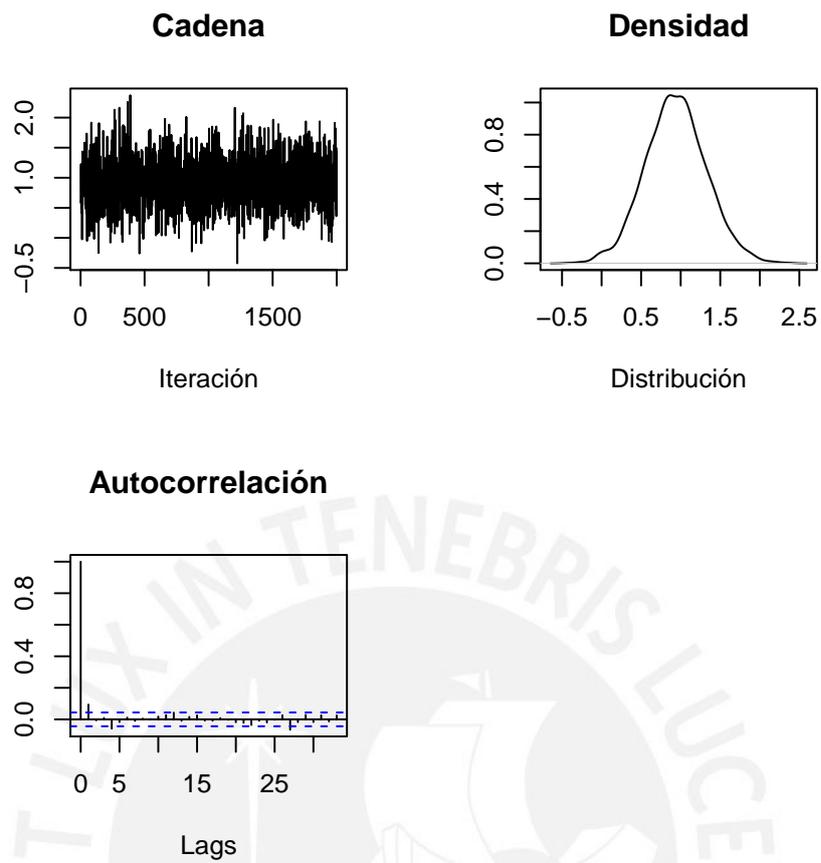


Figura C.6: Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la pendiente del ítem 5, alternativa 2, réplica 3 estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0, 1)$, $n = 200$ y $p=10$

Ejemplos de densidades estacionarias de parámetros, con auto-correlaciones medias y altas:

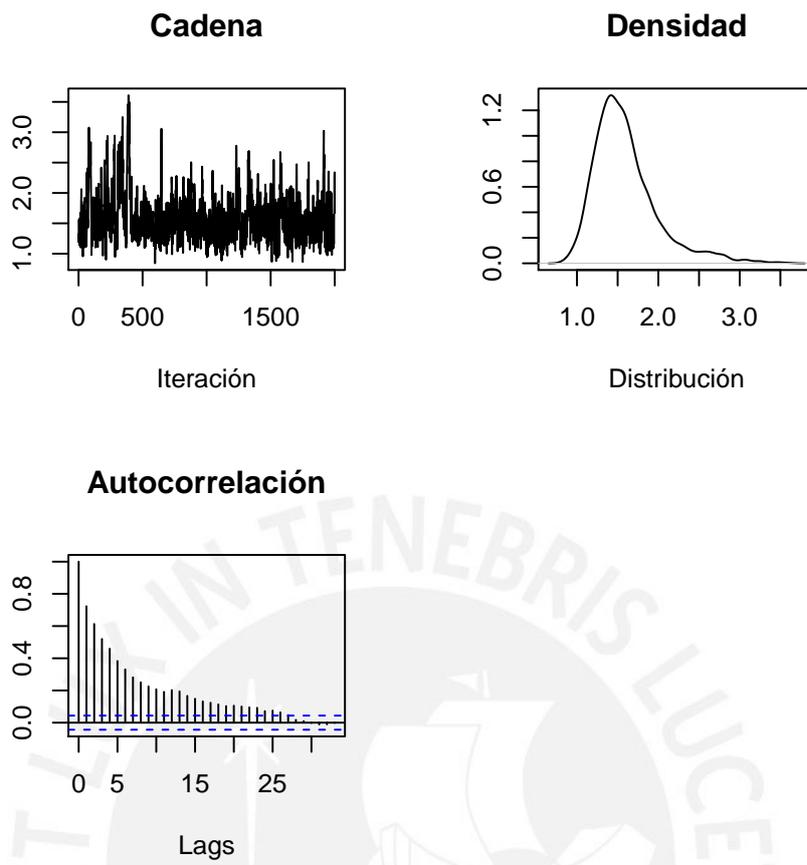


Figura C.7: Iteraciones efectivas, gráfica de densidad y auto-correlaciones para el intercepto del ítem 3, alternativa 1, réplica 9, estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0, 1)$, $n = 200$ y $p=10$

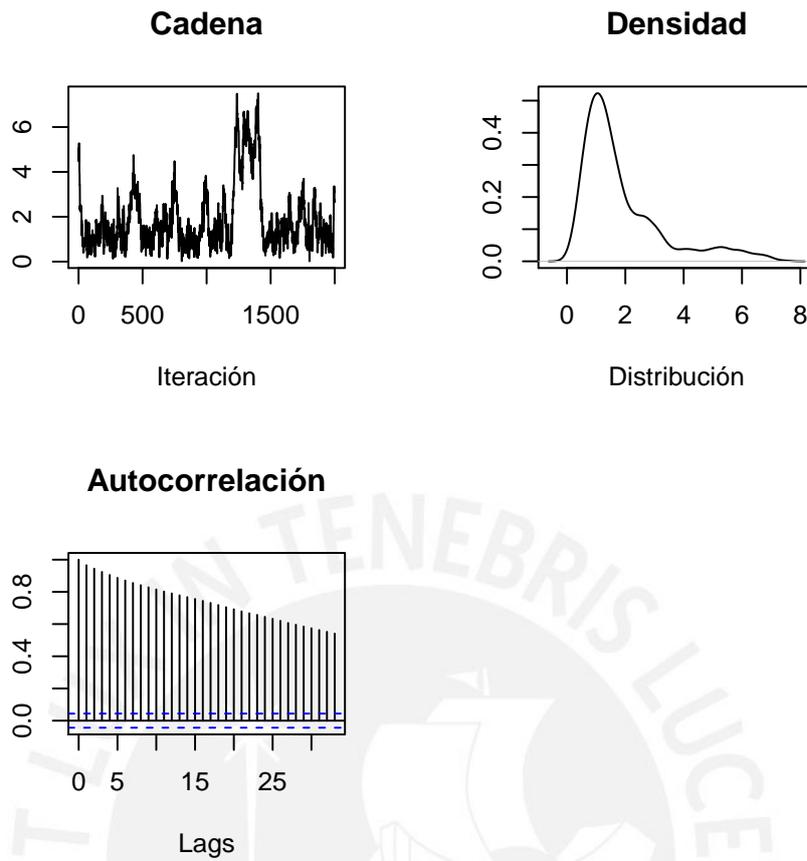


Figura C.8: Iteraciones efectivas, gráfica de densidad y auto correlaciones para el intercepto del ítem 2, alternativa 3, réplica 10, estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0,1)$, $n = 200$ y $p=10$

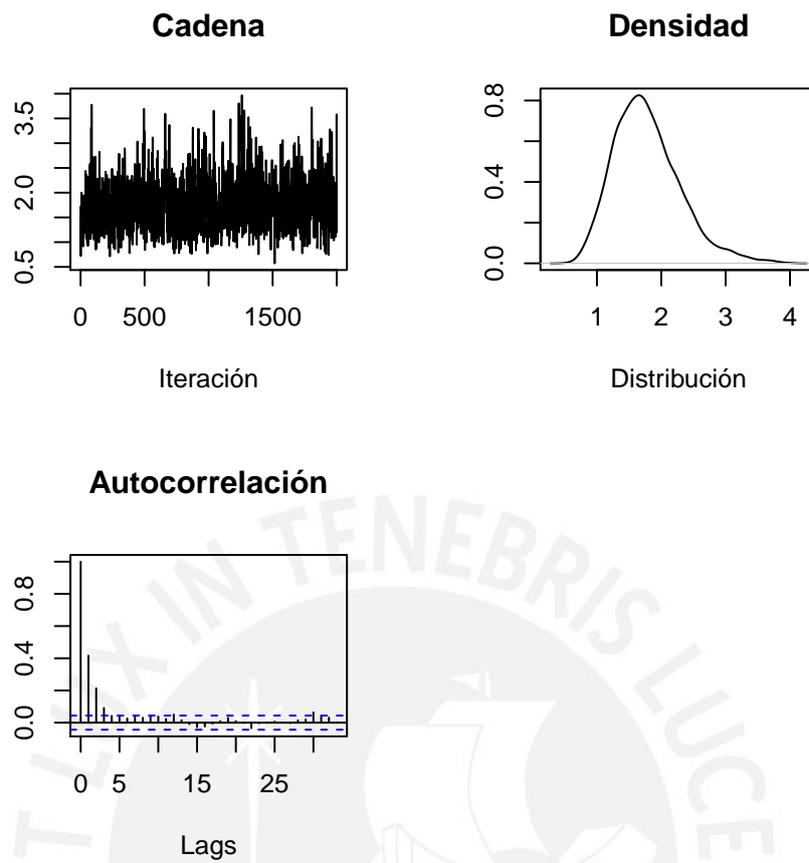


Figura C.9: Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la pendiente del ítem 3, alternativa 4, réplica 5, estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0, 1)$, $n = 200$ y $p=10$

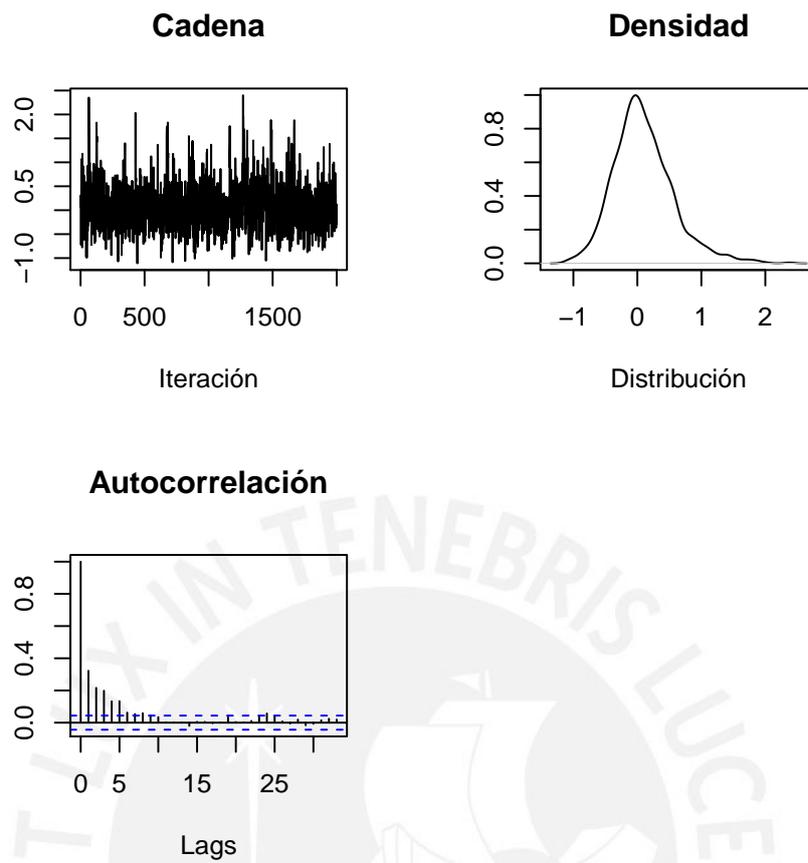


Figura C.10: Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la pendiente del ítem 10, alternativa 1, réplica 6 estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0,1)$, $n = 200$ y $p=10$

C.1.2. Diseño 2:

Distribución Normal Estándar, $p = 10$, $m_j = 4$, $n/p = 20 : 1$ y $n = 1600$

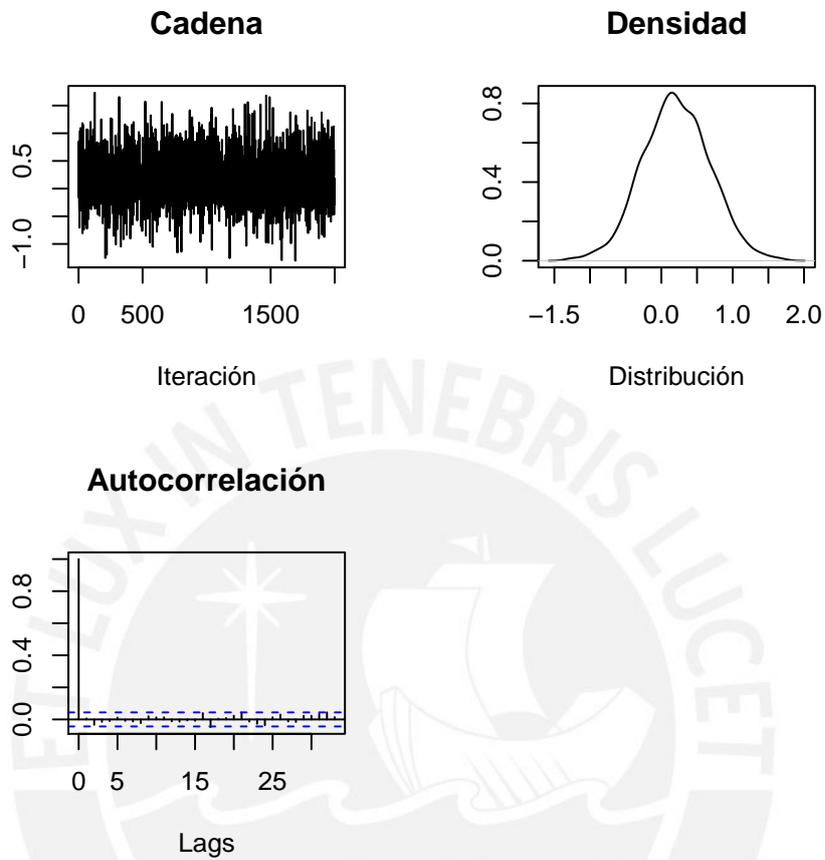


Figura C.11: Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la habilidad del individuo 83, réplica 13, estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0, 1)$, $n = 1600$ y $p=10$

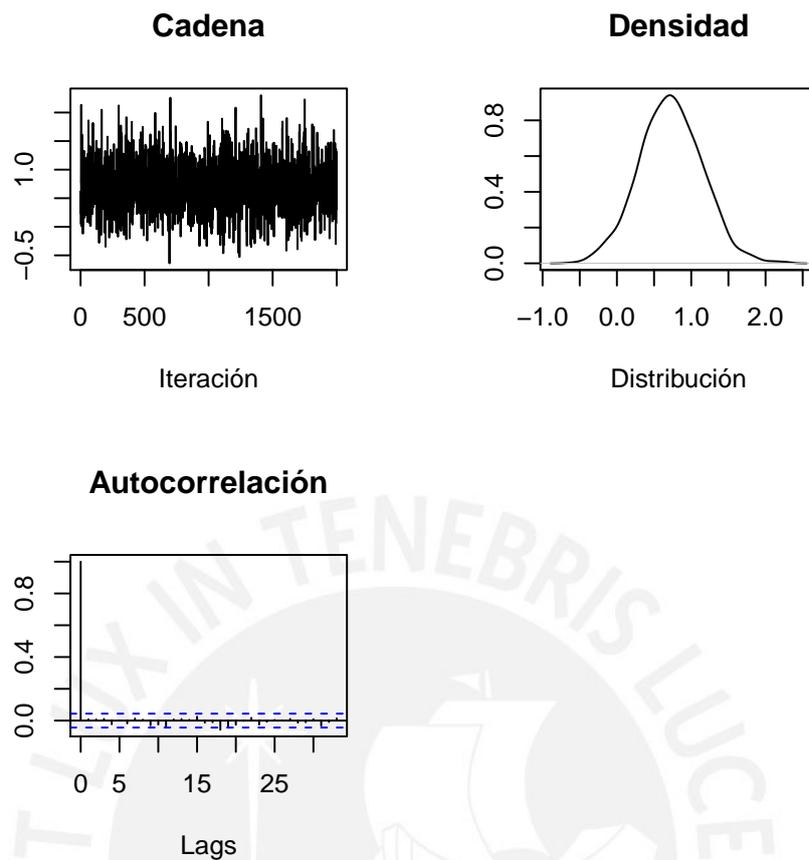


Figura C.12: Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la habilidad del individuo 342, réplica 2, estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0,1)$, $n = 1600$ y $p=10$

Ejemplos de densidades estacionarias de parámetros de ítems, con auto-correlaciones bajas:

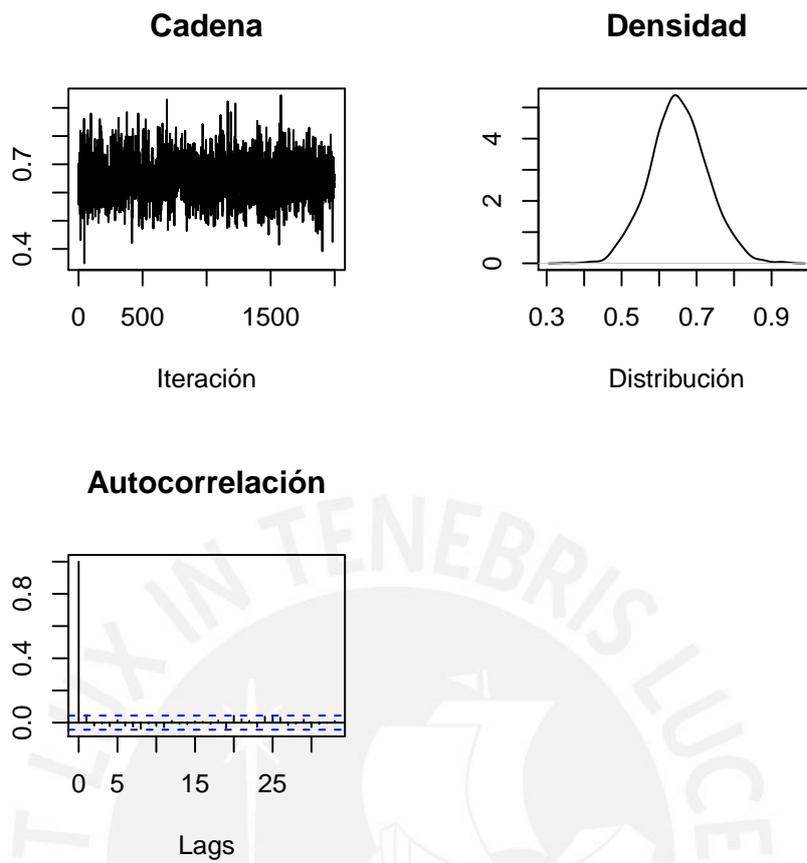


Figura C.13: Iteraciones efectivas, gráfica de densidad y auto-correlaciones para el intercepto del ítem 1, alternativa 4, réplica 9, estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0,1)$, $n = 1600$ y $p=10$

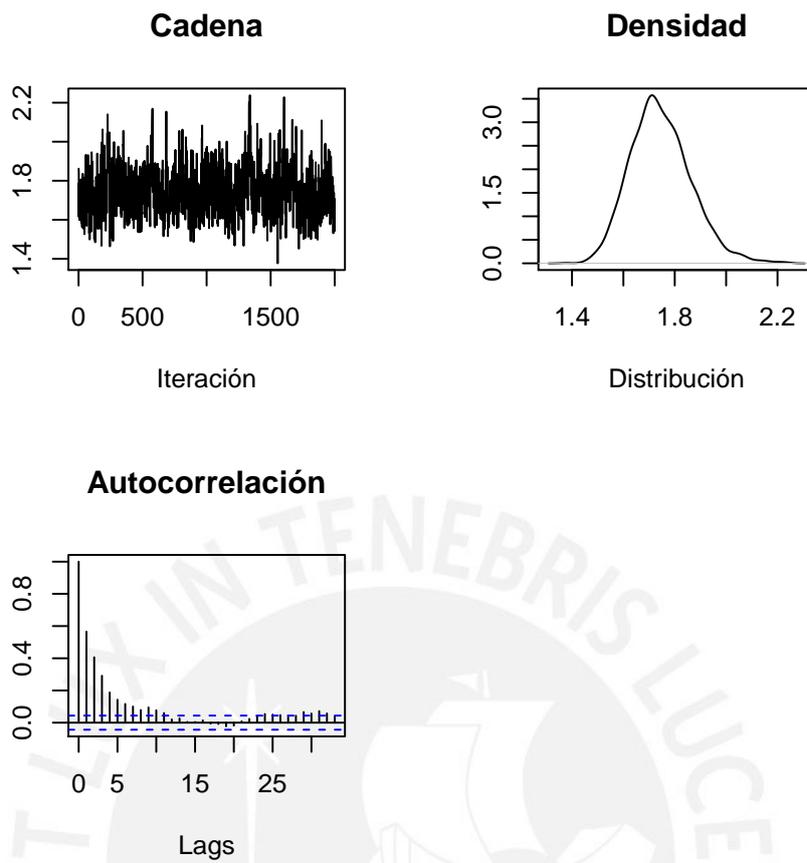


Figura C.14: Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la pendiente del ítem 2, alternativa 1, réplica 19, estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0,1)$, $n = 1600$ y $p=10$

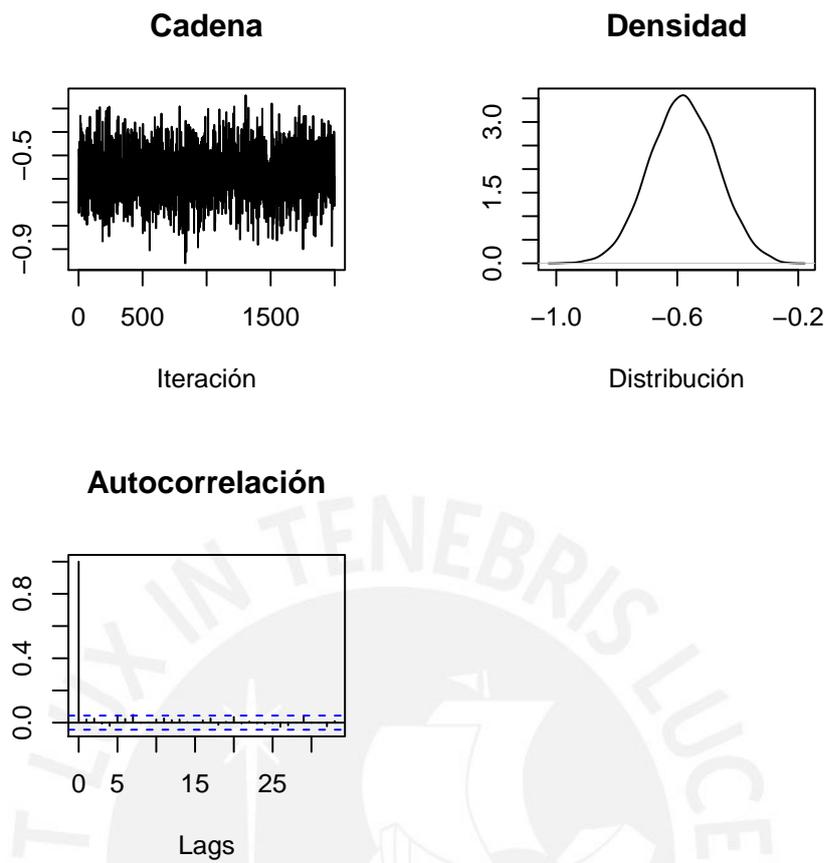


Figura C.15: Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la pendiente del ítem 5, alternativa 2, réplica 12, estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0,1)$, $n = 1600$ y $p=10$

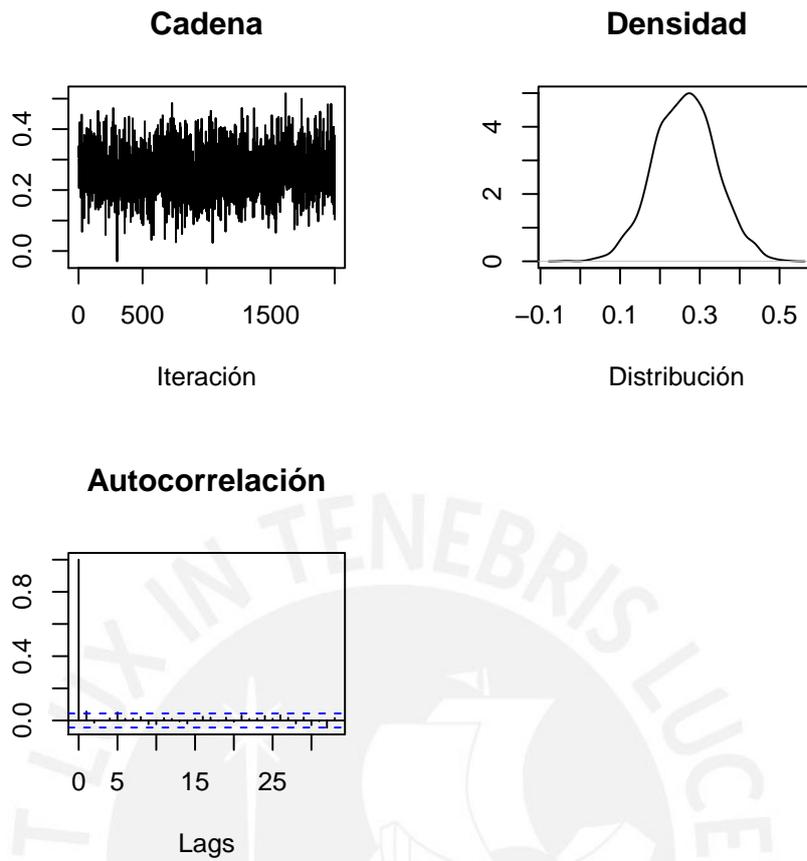


Figura C.16: Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la pendiente del ítem 3, alternativa 3, réplica 14, estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0,1)$, $n = 1600$ y $p=10$

Ejemplos de densidades estacionarias de parámetros, con auto-correlaciones medias y altas:

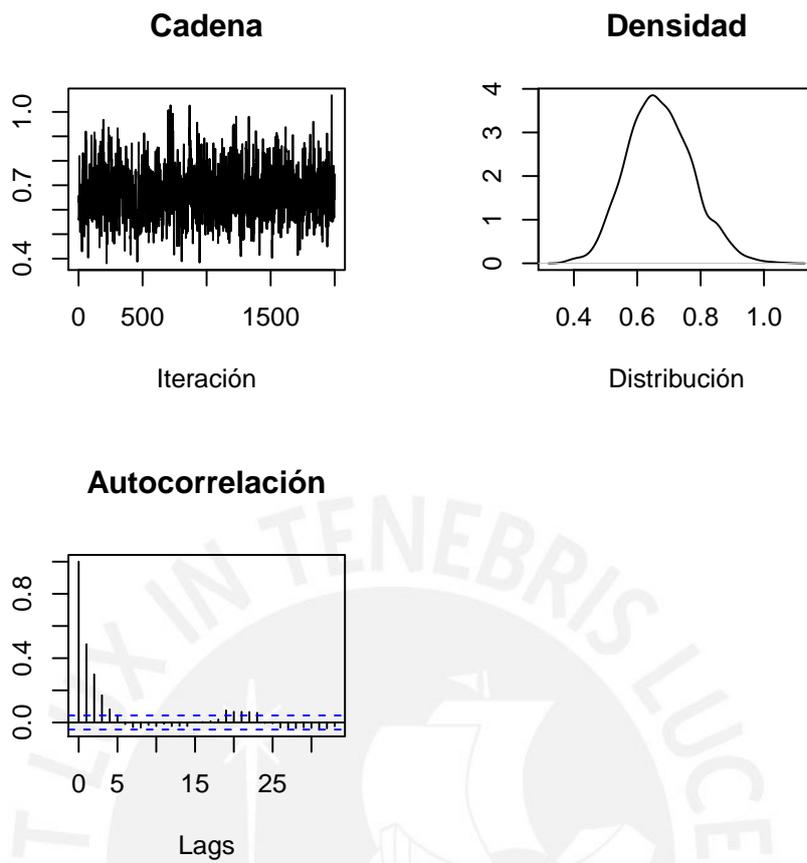


Figura C.17: Iteraciones efectivas, gráfica de densidad y auto-correlaciones para el intercepto del ítem 2, alternativa 3, réplica 10, estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0,1)$, $n = 1600$ y $p=10$

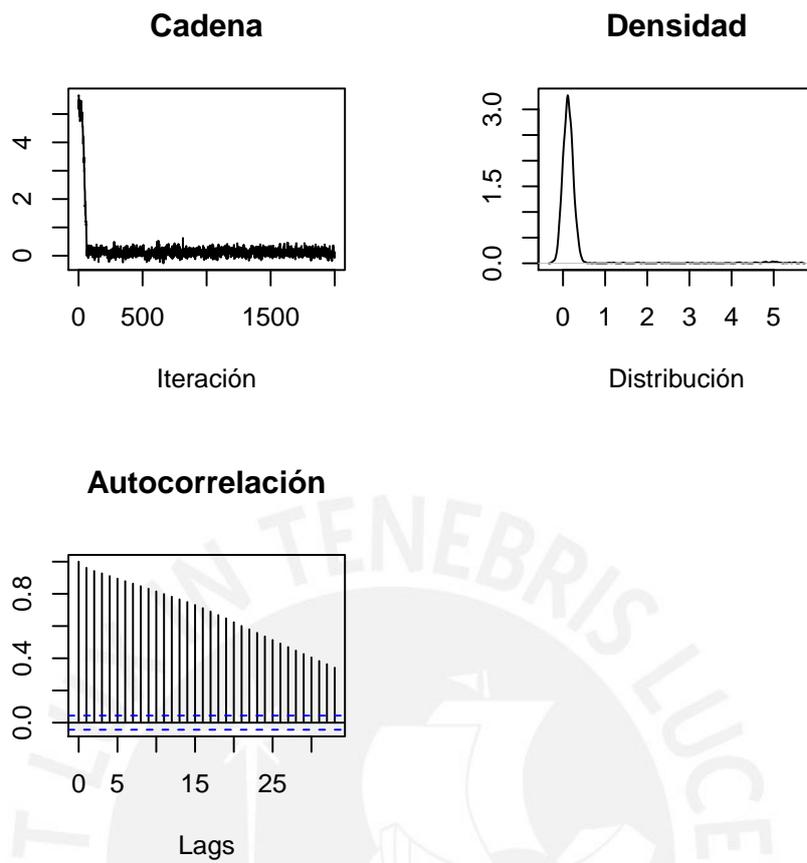


Figura C.18: Iteraciones efectivas, gráfica de densidad y auto-correlaciones para el intercepto del ítem 10, alternativa 1, réplica 15, estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0,1)$, $n = 1600$ y $p=10$

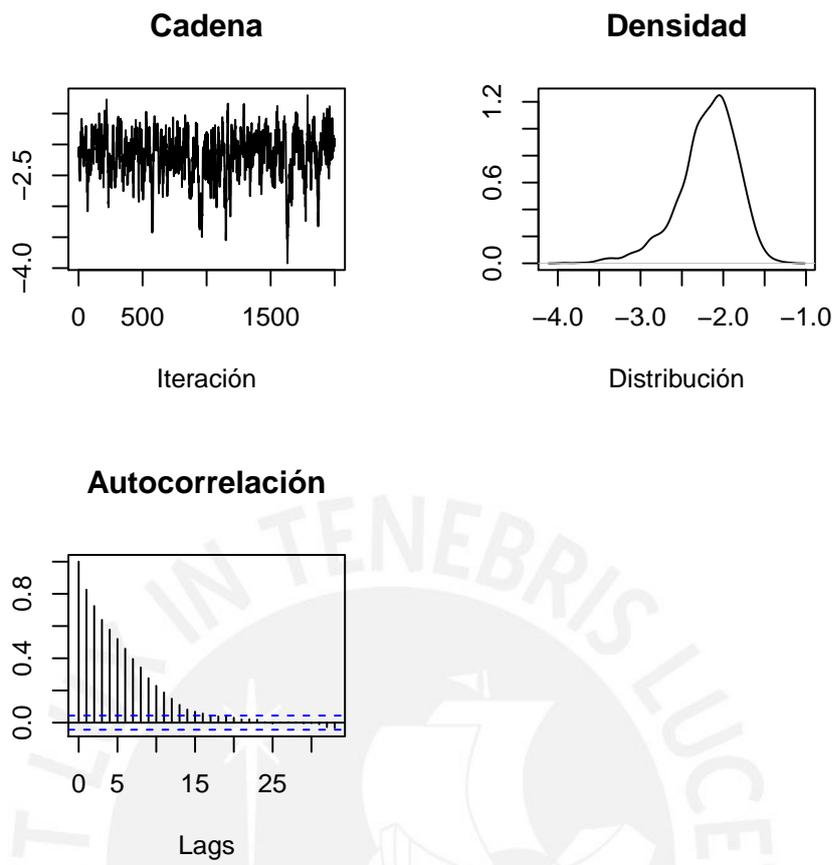


Figura C.19: Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la pendiente del ítem 2, alternativa 4, réplica 16, estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0,1)$, $n = 1600$ y $p=10$

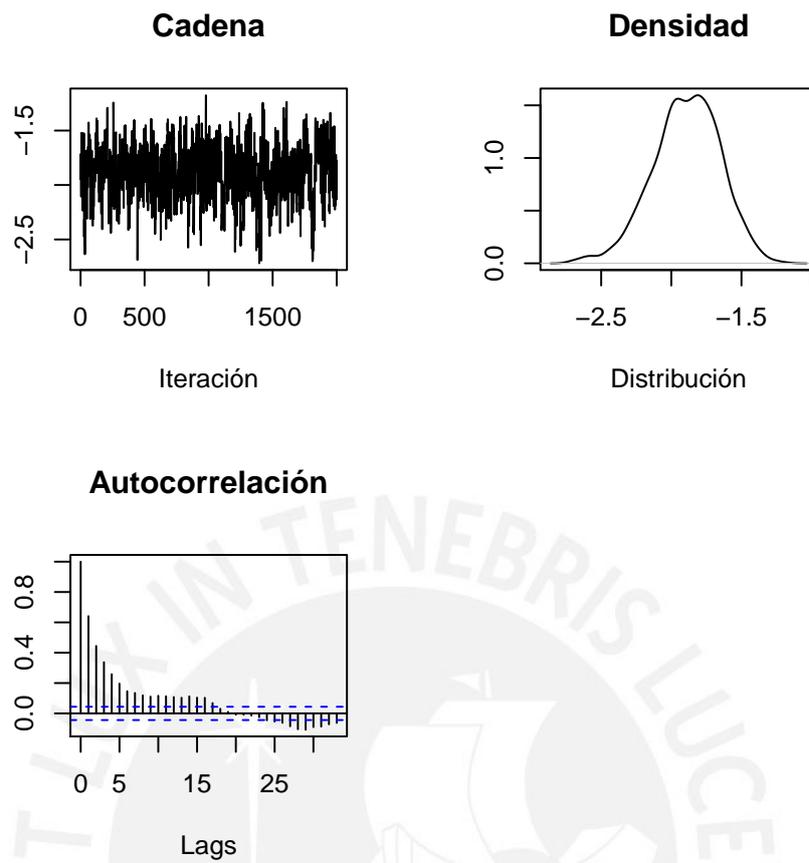


Figura C.20: Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la pendiente del ítem 2, alternativa 4, réplica 18, estimado con el Modelo NRM, para un conjunto de datos caracterizado por habilidades con distribución $N(0,1)$, $n = 1600$ y $p=10$

C.2. Aplicación: Diagnósticos de Convergencia

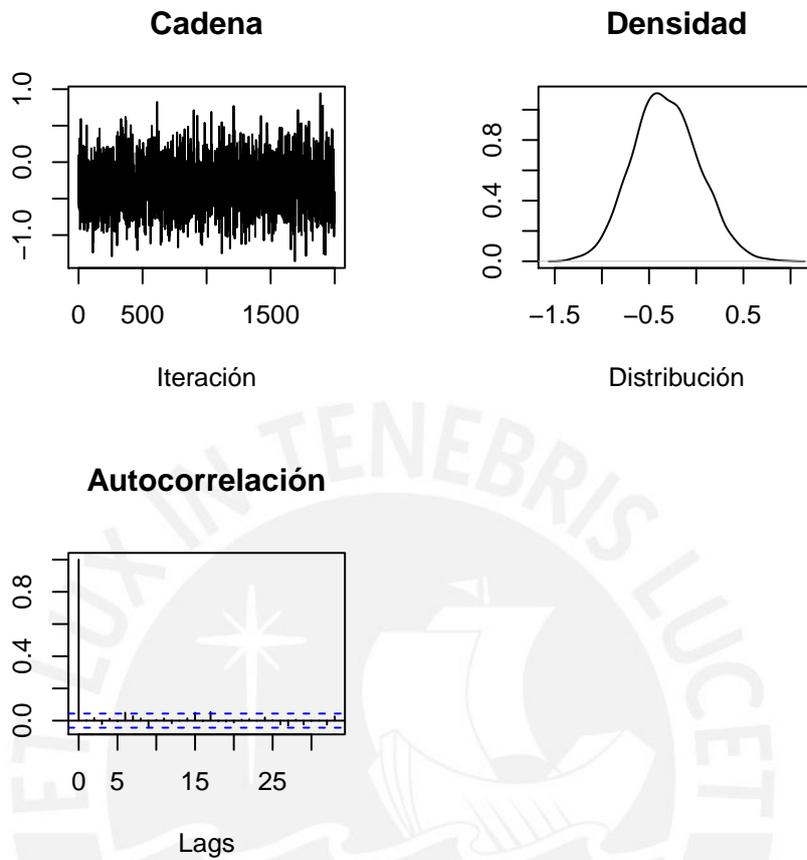


Figura C.21: Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la habilidad del individuo 406, estimado con el Modelo NRM, para los datos de aplicación

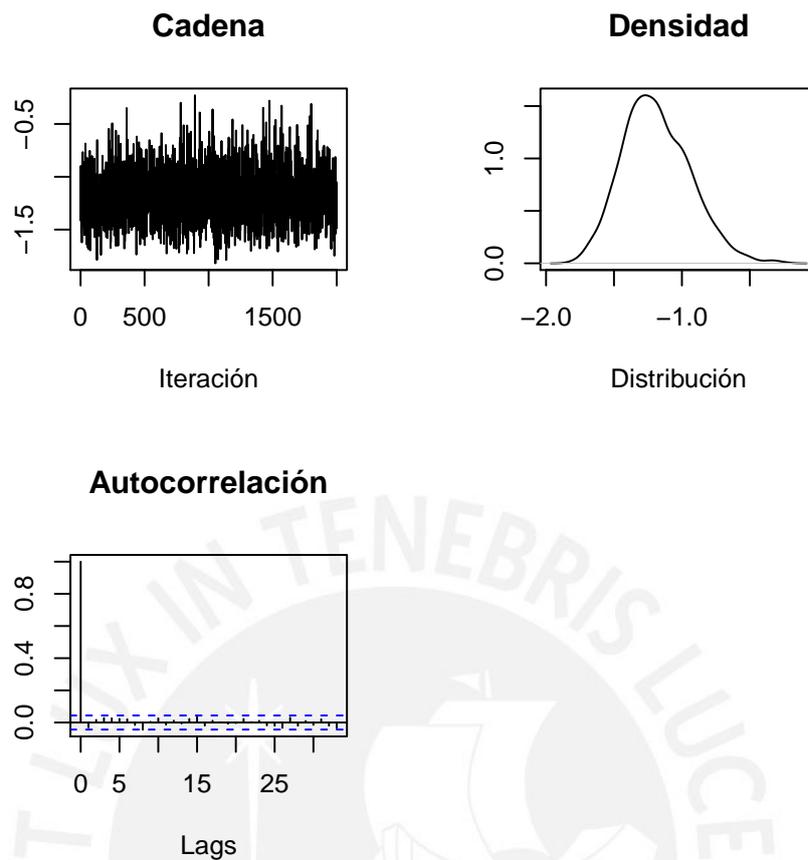


Figura C.22: Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la habilidad del individuo 1255, estimado con el Modelo NRM, para los datos de aplicación

Ejemplos de densidades estacionarias de parámetros de ítems, con auto-correlaciones bajas:

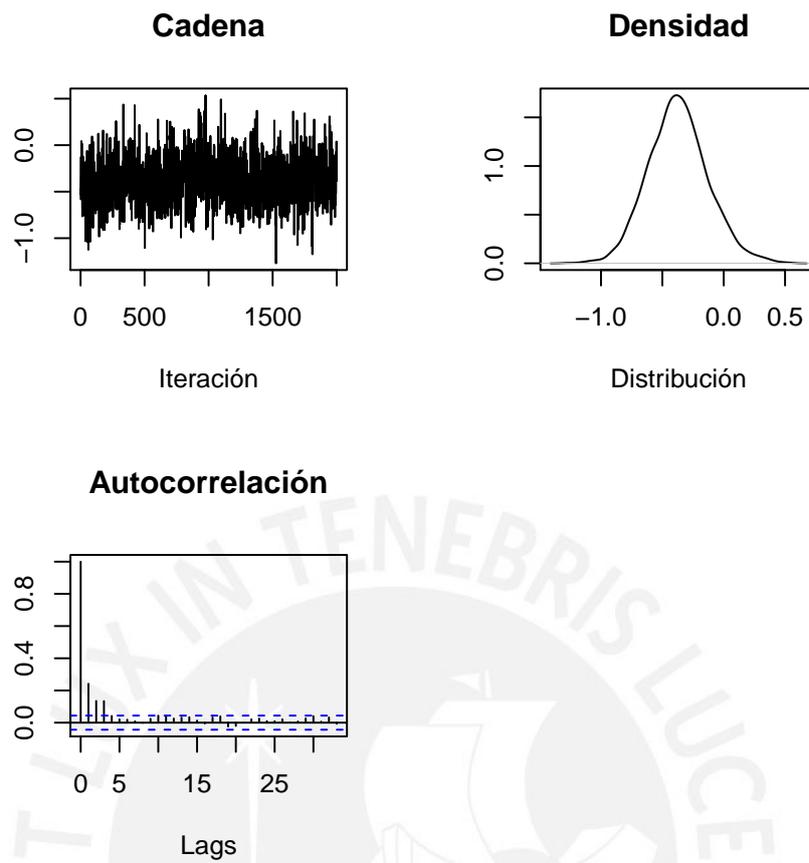


Figura C.23: Iteraciones efectivas, gráfica de densidad y auto-correlaciones para el intercepto del ítem 1, alternativa 1, estimado con el Modelo NRM, para los datos de aplicación

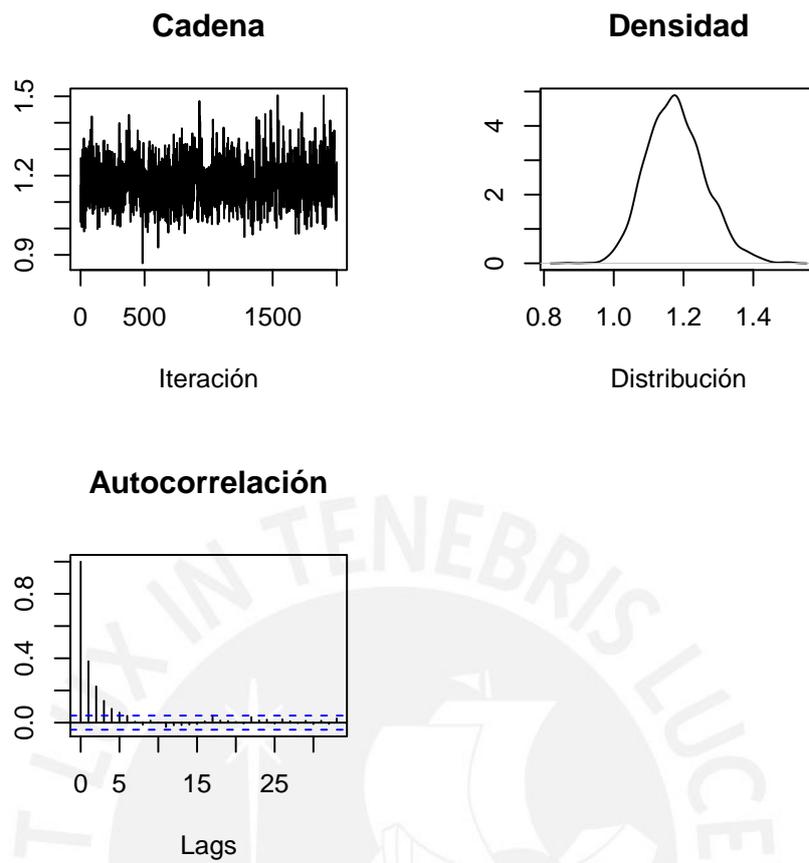


Figura C.24: Iteraciones efectivas, gráfica de densidad y auto-correlaciones para el intercepto del ítem 3, alternativa 3, estimado con el Modelo NRM, para los datos de aplicación

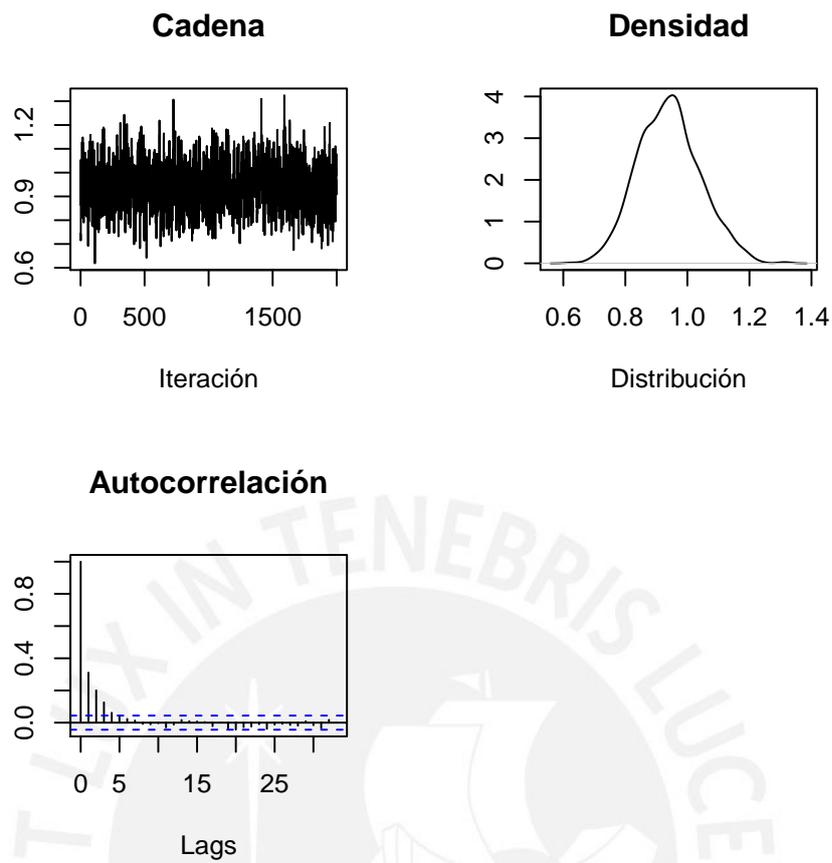


Figura C.25: Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la pendiente del ítem 21, alternativa 1, estimado con el Modelo NRM, para los datos de aplicación

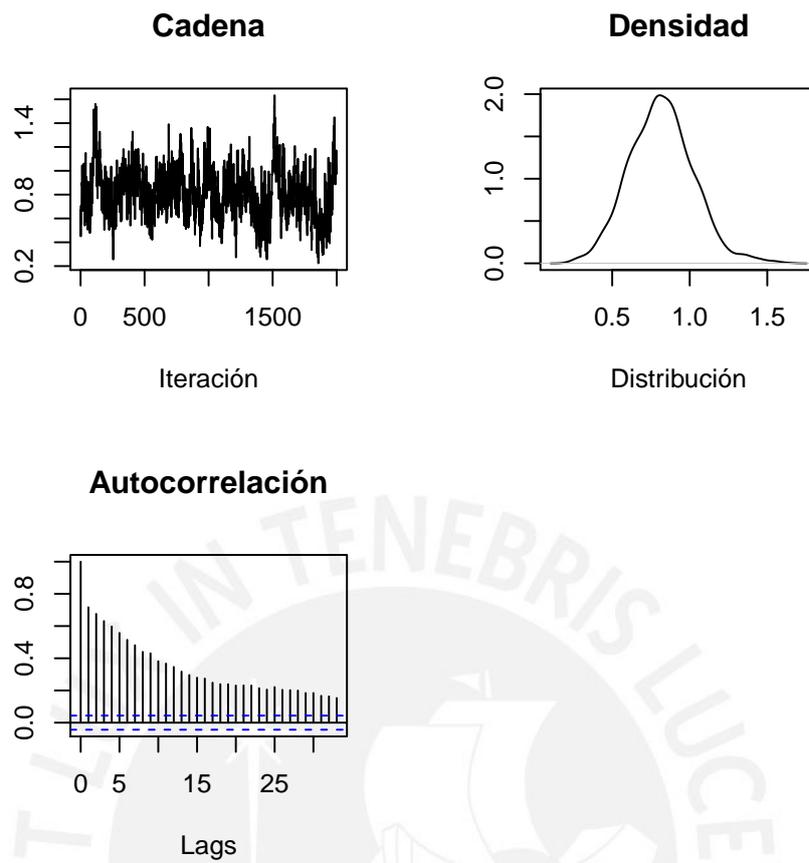


Figura C.26: Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la pendiente del ítem 20, alternativa 3, estimado con el Modelo NRM, para los datos de aplicación

Ejemplos de densidades estacionarias de parámetros de ítems, con auto-correlaciones altas:

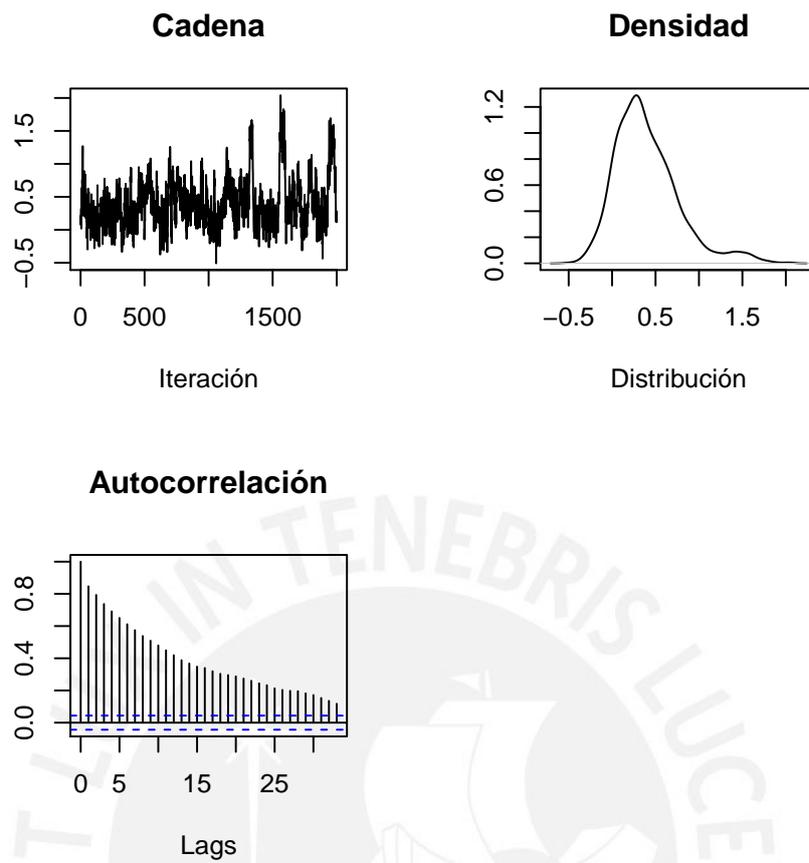


Figura C.27: Iteraciones efectivas, gráfica de densidad y auto-correlaciones para el intercepto del ítem 2, alternativa 2, estimado con el Modelo NRM, para los datos de aplicación

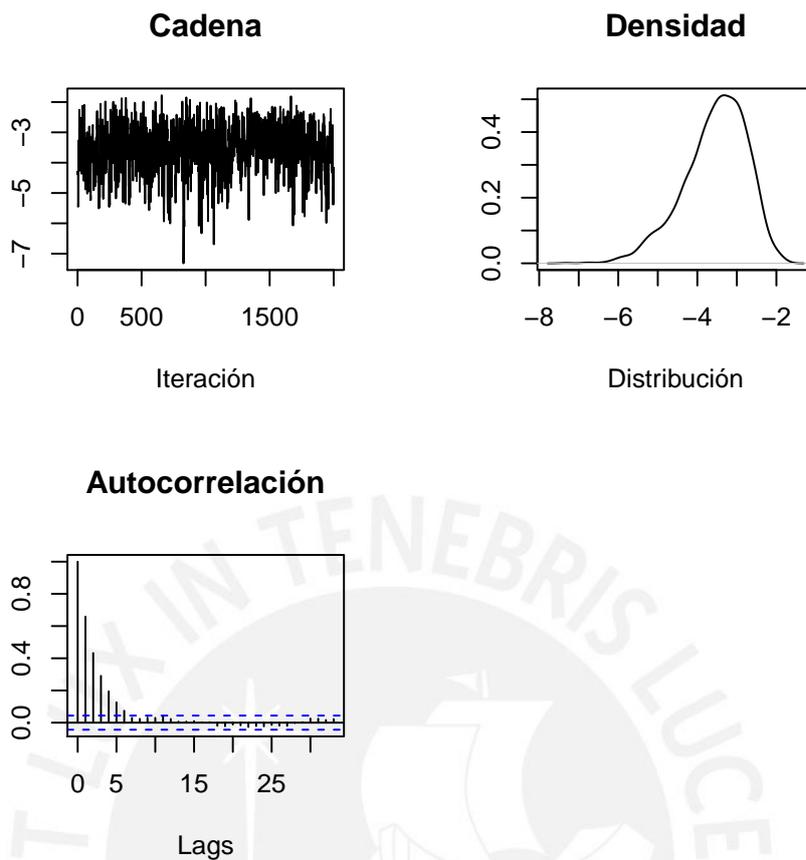


Figura C.28: Iteraciones efectivas, gráfica de densidad y auto-correlaciones para el intercepto del ítem 4, alternativa 5, estimado con el Modelo NRM, para los datos de aplicación

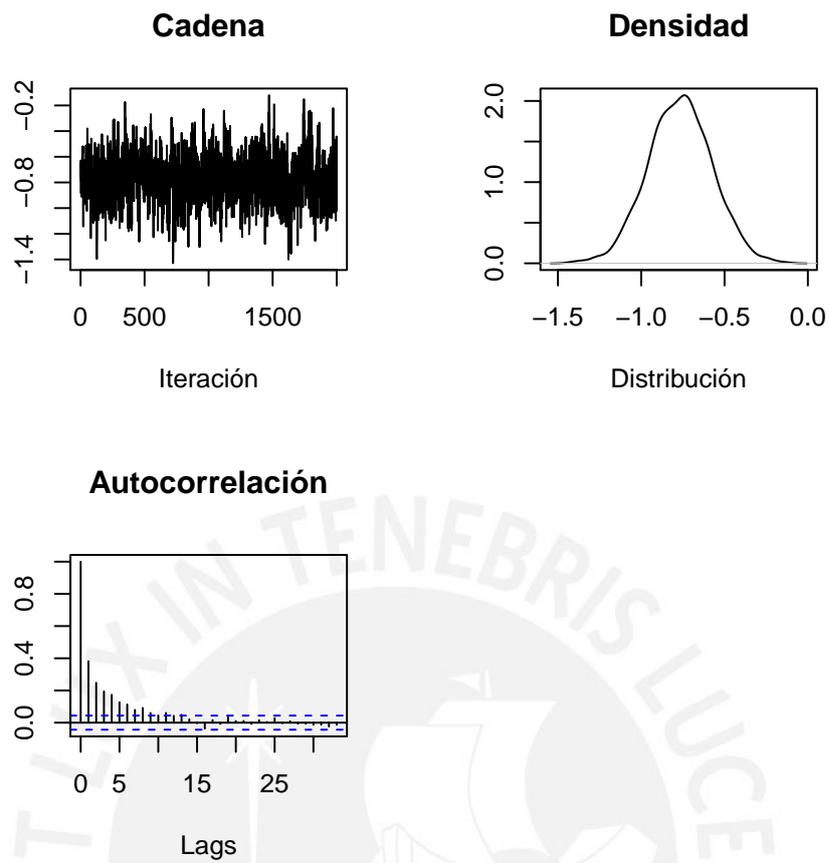


Figura C.29: Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la pendiente del ítem 14, alternativa 3, estimado con el Modelo NRM, para los datos de aplicación

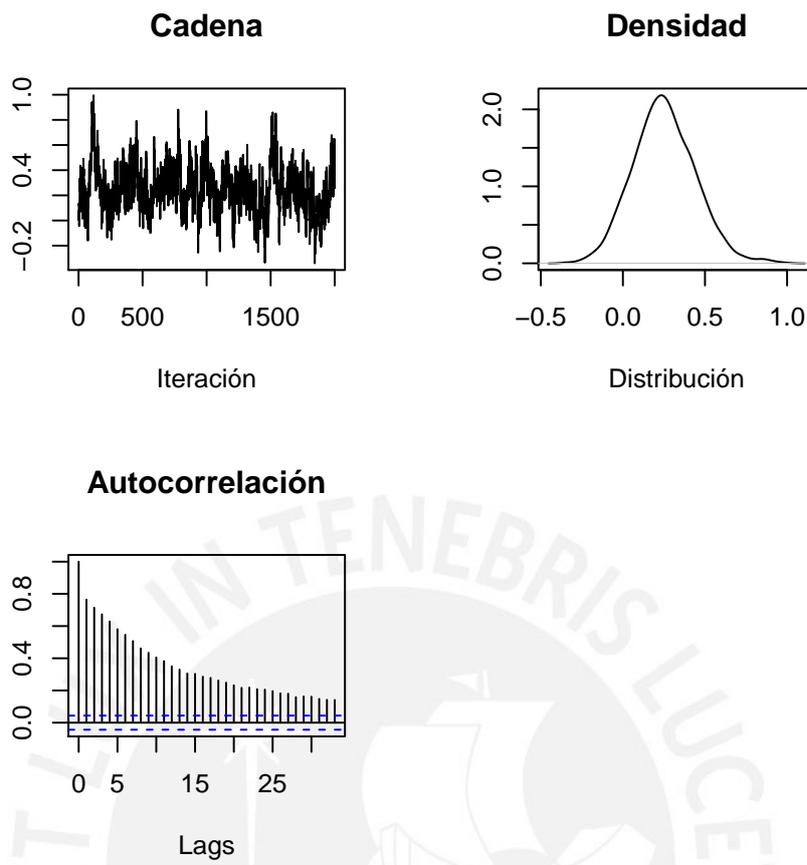


Figura C.30: Iteraciones efectivas, gráfica de densidad y auto-correlaciones para la pendiente del ítem 20, alternativa 2, estimado con el Modelo NRM, para los datos de aplicación

Bibliografía

- Agresti, A. (2007). *An introduction to categorical data analysis*, John Wiley and Sons, Inc.
- Ayala, R. J. D. y Sava-Bolesta, M. (1999). Item parameter recovery for the nominal response model, *Applied Psychological Measurement* **23**(1): 3–19.
- Azevedo, C. (2003). *Métodos de estimação na teoria de resposta ao item*, Master's thesis, Universidade de São Paulo (USP).
- Baker, F. (2001). The basic of item response theory, *Technical report*, ERIC Clearinghouse on Assessment and Evaluation.
- Baker, F. B. (1998). An investigation of the item parameter recovery characteristics of a gibbs sampling procedure, *Applied Psychological Measurement* **22**(22): 153–169.
- Baker, F. y Kim, S. (1992). *Item Response Theory: Parameter Estimation Techniques*, Statistics for the Social and Behavioral Sciences, CRC Press, Taylor and Francis Group.
- Bazán, J. (2005). *Uma Família de Modelos de Resposta ao Ítem Normal Assimétrica*, PhD thesis, Univversidade de São Paulo (USP).
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability, in F. Lord, M. Novick y A. Birnbaum (eds), *Statistical theories of mental test scores*, Oxford, England: Addison-Wesley, pp. 397–472.
- Bock, R. (1997). The nominal categories model, in R. Hambleton y W. Linden (eds), *Handbook of Modern Item Response Theory*, Springer Science+Business Media, LLC, chapter 2, pp. 33–49.
- Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories, *Psychometrika* **37**(1): 29–51.
- Carpenter, B., Gelman, A., Hoffman, H., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P. y Riddell, A. (2016). Stan: A probabilistic programming language, *Journal of Statistical Software* **76**.
- Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment, *Journal of Statistical Software* **48**(6): 1–29.
URL: <http://www.jstatsoft.org/v48/i06/>
- De Ayala, R. (1993). An introduction to polytomous item response theory models, *Measurement and Evaluation in Counseling and Development* **25**(4): 172–189.
- Fox, J. (2010). *Bayesian Item Response Modeling, Theory and Applications*, Statistics for Social and Behavioral Sciences, fienberg, s. and van der linden, w. edn, Springer Science+Business Media, LLC.
- Hambleton, R. y Swaminathan, H. (1991). *Item Response Theory*, Springer Science+Business Media, LLC.

- Hambleton, R., Swaminathan, H. y Rogers, H. (1991). *Fundamentals of Item Response Theory*, SAGE Publications Inc.
- Harwell, M., Stone, C. A., Hsu, T.-C. y Kirisci, L. (1996). Monte carlo studies in item response theory, *Applied Psychological Measurement* **20**(2): 101–125.
- Hsieh, M., Proctor, T., Hou, J. y Teo, K. (2010). A comparison of bayesian mcmc and marginal maximum likelihood methods in estimating the item parameters for the 2pl irt model, *International Journal of Innovative Management, Information and Production* **1**(1): 81–89.
- Kieftenbeld, V. y Natesan, P. (2012). Recovery of graded response model parameters: A comparison of marginal maximum likelihood and markov chain monte carlo estimation, *Applied Psychological Measurement* **36**(5): 399–419.
- Kim, S. y Cohen, A. (1999). Accuracy of parameter estimation in gibbs sampling under the two-parameter logistic model, *Annual Meeting of the American Educational Research Association*.
- Levine, M. V. y Drasgow, F. (1983). The relation between incorrect option choice and estimated ability, *Educational and Psychological Measurement* **43**: 675–685.
- Lord, F. y Novik, M. (2008). *Statistical Theories of Mental Test Scores*, Information Age Publishing.
- Lunn, D., Spiegelhalter, D., Thomas, A. y Best, N. (2009). The bugs project: Evolution, critique, and future directions, *Statistics in Medicine* **28**(28): 3049–3067.
- Lunn, D., Thomas, A., Best, N. y Spiegelhalter, D. (2000). Winbugs: A bayesian modelling framework, *Statistics and Computing* (10): 325–337.
- Masters, G. y Wright, B. (1984). The essential process in a family of measurement models, *Psychometrika* (49): 529–244.
- McDonald, R. (1967). Non-linear factor analysis, *Psychometric Monographs* (15).
- Mellenberg, G. (1995). Conceptual notes on models for discrete polytomous item responses, *Applied Psychological Measurement* (19): 91–100.
- Ostini, R. y Nering, M. (2006). *Polytomous item response theory models*, Vol. 07 of *Quantitative applications in the social sciences*, Sage Publications Inc.
- Ostini, R. y Nering, M. (2010). *Handbook of Polytomous Item Response Theory Models*, Routledge, Taylor and Francis Group, LLC.
- Patz, R. J. y Junker, B. W. (1999). A straightforward approach to markov chain monte carlo methods for item response models, *Journal of Educational and Behavioral Statistics* **24**(2): 146–178.
- Plummer, M., Best, N., Cowles, K. y Vines, K. (2006). Coda: Convergence diagnosis and output analysis for mcmc, *R News* **6**(1): 7–11.
URL: <https://journal.r-project.org/archive/>
- R Core Team (2015). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: <http://www.R-project.org/>
- Rasch, G. (1980). *Probabilistic Models for Some Intelligence and Attainment Tests*, University of Chicago Press.

- Reif, M. (2014). *mcIRT: IRT models for multiple choice items*. R package version 0.41.
URL: <https://github.com/manuelreif/mcIRT>
- Samejima, F. (1970). Erratum estimation of latent ability using a response pattern of graded scores, *Psychometrika* **35**(17): 139–139.
- Samejima, F. (1972). A general model for free-response data, *Psychometric Society* (18).
- Samejima, F. (1979). A new family of models for the multiple choice item, *Technical Report 79-4*, University of Tennessee, Department of Psychology, Knoxville, Tennessee, 37916.
- Samejima, F. (1988). Comprehensive latent trait theory, *Behaviourmetrika* **15**(24): 1–24.
- Samejima, F. (1991). Graded response model, in R. Hambleton y W. Linden (eds), *Handbook of Modern Item Response Theory*, Springer Science+Business Media, LLC, chapter 5, pp. 85–100.
- Stone, C. y Zhu, X. (2006). Bayesian analysis of item response theory models using sas, *Technical report*, SAS Institute Inc.
- Sturtz, S., Ligges, U. y Gelman, A. (2005). R2winbugs: A package for running winbugs from r, *Journal of Statistical Software* **12**(3): 1–16.
URL: <http://www.jstatsoft.org>
- Tarazona, E. (2013). *Modelos alternativos de respuesta graduada con aplicaciones en la calidad de servicios*, Master's thesis, Pontificia Universidad Católica del Perú (PUCP).
- Thissen, D. (1976). Information in wrong responses to the raven progressive matrices, *Journal of Educational Measurement* **13**(3): 201–214.
- Thissen, D., Cai, L. y Bock, R. (2010). The nominal categories item response model, in R. Ostini y M. Nering (eds), *Handbook of Polytomous Item Response Theory Models*, Routledge, Taylor and Francis Group, LLC, chapter 3.
- Thissen, D. y Steinberg, L. (1984). A response model for multiple choice items, *Psychometrika* **49**(4): 501–519.
- Thissen, D. y Steinberg, L. (1986). A taxonomy of item response models, *Psychometrika* **51**(4): 567–577.
- Thissen, D. y Steinberg, L. (1997). A response model for multiple-choice items, in R. Hambleton y W. Linden (eds), *Handbook of Modern Item Response Theory*, Springer Science+Business Media, LLC, chapter 3, pp. 51–65.
- Thissen, D., Steinberg, L. y Fitzpatrick, A. R. (1989). Multiple-choice models: The distractors are also part of the item, *Journal of Educational Measurement* **26**(2): 161–176.
- Thissen, D. (1991). *Multilog 6.0*, Scientific Software Incorporated.
- Vale, C. y Weiss, D. (1977). A comparison of information functions of multiple-choice and free-response vocabulary items, *Research Report 77-2*, Minnesota University, Minneapolis. Department of Psychology.
- van Rossum, G. (1995). *Python tutorial: Technical Report CS-R9526*. Centrum voor Wiskunde en Informatica (CWI), Amsterdam.
URL: <https://www.python.org/>

Weston, S. (2015). *Using The foreach Package*.

URL: <https://cran.r-project.org/web/packages/foreach/vignettes/foreach.pdf>

Weston, S. y Calaway, R. (2015). *Getting Started with doParallel and foreach*.

URL: <https://cran.r-project.org/web/packages/doParallel/vignettes/gettingstartedParallel.pdf>

Wollack, J. A., Bolt, D. M., Cohen, A. S. y Lee, Y.-S. (2002). Recovery of item parameters in the nominal response model: A comparison of marginal maximum likelihood estimation and markov chain monte carlo estimation, *Applied Psychological Measurement* **26**(3): 339–352.

Yen, W. M. (1987). A comparison of the efficiency and accuracy of bilog and logist, *Psychometrika* **52**(2): 275–291.

