

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ
ESCUELA DE POSGRADO



**MODELO DINA APLICADO A LA EVALUACIÓN DE
MATEMÁTICA EN ESTUDIANTES DE SEGUNDO GRADO
DE SECUNDARIA**

Tesis para optar el grado de Magíster en Estadística

AUTOR

Yuriko Kirilovna Sosa Paredes

ASESOR

Dr. Luis Hilmar Valdivieso Serrano

JURADO

Dr. Cristian Luis Bayes Rodriguez
Mg. Enver Gerald Tarazona Vargas

LIMA - PERÚ
2017



Dedicatoria

A Ena, mamá, que ahora es la luz que guía cada uno de mis pasos.

A Manuel, Benigno, Yitzhak, Fiorella y Antonino; cuya sola existencia me motiva a seguir cada día.



Agradecimientos

En primer lugar, toda mi admiración y agradecimiento a mi asesor el Dr. Luis Valdivieso por su paciencia, sus enseñanzas, su guía constante en el proceso de este trabajo; y por motivarme a continuar en el camino de la investigación.

A los docentes de la maestría en Estadística de la Pontificia Universidad Católica del Perú (PUCP), quienes me brindaron la mejor disposición para aprender de inicio a fin del programa, y por los comentarios que me ayudaron a seguir mejorando.

A Melissa Castillo, especialista en Matemática ex trabajadora de la Oficina de Medición de la Calidad de los Aprendizajes (UMC), por sus recomendaciones como experta y brindarme uno de los insumos principales para el desarrollo de este estudio.

A Jesús Gamboa, Mg. en Estadística, por su guía en el maravilloso mundo de la inferencia bayesiana y sobretodo por brindarme su amistad y apoyo en todo el proceso.

A todos los compañeros del trabajo, en especial al Equipo de Análisis de la UMC: Factores Asociados, Psicometría y Estadística; amigos de la vida que estuvieron siempre alentándome, cual hinchas de nuestra querida selección peruana, para no darme por vencida en este camino. Mención especial para Manuel Marcos, Juan Carlos Saravia, Carla Nizama e Ivonne Zurita.

¡Arriba Alianza, toda la vida!

Resumen

Los modelos de diagnóstico cognitivo (MDC) tienen como finalidad describir o diagnosticar el comportamiento de los evaluados por medio de clases o perfiles latentes, de tal manera que se obtenga información más específica acerca de las fortalezas y debilidades de ellos.

Uno de los modelos más populares de esta gran familia es el llamado modelo DINA, el cual tuvo su primera aparición en [Haertel \(1989\)](#) enfocado principalmente en el campo educacional. Este modelo considera solo respuestas observadas dicotómicas de parte de los individuos y tiene como restricción principal que ellos deben dominar necesariamente todas las habilidades requeridas por cada ítem; aquellas que se resumen en una matriz llamada **Q**. Asimismo, el modelo estima parámetros para los ítems, los cuales son denominados de “ruido”: Adivinación y Desliz.

En este trabajo desarrolla teóricamente el modelo expuesto; es decir, sus fundamentos y principales propiedades desde el enfoque bayesiano. Específicamente, las estimaciones se realizan mediante el Muestreador de Gibbs.

Se realizaron 8 estudios de simulación, cada uno de ellos con tres diferentes tamaños de población, donde se probaron combinaciones de los parámetros en estudio con el fin de comparar la recuperación de parámetros mediante el enfoque clásico y el bayesiano. El análisis de ambos enfoques se realizó con rutinas de código del software libre R, usando los paquetes *CDM* y *dina* para el enfoque clásico y el bayesiano, respectivamente.

En líneas generales, los resultados muestran estimaciones insesgadas y con valores pequeños de la raíz del error cuadrático medio (RMSE) para ambos enfoques. Incluso, conforme el tamaño de la población incrementa, las estimaciones no tienen mayores diferencias. Aunque en tamaños de población más pequeños el enfoque bayesiano obtiene ligeras ventajas con respecto al otro, especialmente en el parámetro de probabilidad de pertenencia a las clases (π). Además, es necesario mencionar que los parámetros de ruido de los ítems son estimados más precisamente con el enfoque clásico en varios de los estudios.

Finalmente, se presenta una aplicación enfocada en educación, donde se analiza una muestra de 3040 alumnos del 2do grado de secundaria, evaluados en una prueba de 48 ítems de la competencia matemática realizada por la Oficina de Medición de la Calidad de los Aprendizajes (UMC) en el 2015. A esta prueba se le aplica el modelo de Rasch y el modelo DINA bajo el enfoque bayesiano, con el fin de estudiar la correspondencia entre indicadores de ambos modelos, tanto para los parámetros de los alumnos (habilidad y perfiles latentes) como de los ítems (dificultad y parámetros de ruido).

Palabras clave: Modelo DINA, Modelos de Diagnóstico Cognitivo (MDC), Muestreador de Gibbs, modelo de Rasch, variables latentes, inferencia bayesiana, educación.

Índice general

Lista de Abreviaturas	VII
Índice de figuras	VIII
Índice de cuadros	IX
1. Introducción	1
1.1. Organización del Trabajo	2
2. El modelo DINA	3
2.1. Modelos de respuesta al ítem y modelos de diagnóstico cognitivo	3
2.2. Formulación del modelo DINA	6
3. Estimación del modelo	10
3.1. Consideraciones previas	10
3.2. Distribución de Dirichlet	11
3.2.1. Función de densidad de probabilidad	11
3.2.2. Propiedades	11
3.2.3. Distribución categórica	12
3.2.4. Conjugación a una distribución categórica y multinomial	12
3.3. Estimación clásica mediante el algoritmo EM	13
3.4. Estimación mediante el algoritmo de Metropolis-Hastings	15
3.5. Estimación mediante el muestreador de Gibbs	15
3.5.1. Distribución a priori	16
3.5.2. Distribución a posteriori	17
3.5.3. Distribuciones condicionales completas	17
4. Estudio de Simulación	20
4.1. Objetivos	21
4.2. Algoritmo para simular datos	22
4.3. Método para estimar los parámetros	22
4.4. Criterios para evaluar la simulación	22
4.5. Estudio 1	23
4.6. Estudio 2	28
4.7. Estudio 3	34
4.8. Estudio 4	40

4.9. Estudio 5	46
4.10. Estudio 6	52
4.11. Estudio 7	58
4.12. Estudio 8	64
5. Aplicación	71
5.1. Instrumento de medición	72
5.2. Muestra	73
5.3. Análisis de los datos mediante el modelo DINA	74
5.3.1. Estimación de los parámetros mediante el muestreador de Gibbs	75
5.4. Análisis de los datos mediante Rasch	77
5.4.1. Análisis de correspondencia entre Rasch y DINA	80
6. Conclusiones	89
6.1. Conclusiones	89
6.2. Sugerencias para futuras investigaciones	92
A. Código en R del estudio de simulación	94
B. Rutinas de código en R y Winbugs de la Aplicación	98
B.1. Modelo DINA	98
B.2. Modelo Rasch	100
B.2.1. Código en R	100
B.2.2. Código Winbugs	103
C. Ejemplos de ítems de la prueba de Matemática	104
C.1. Capacidad: matematiza situaciones	104
C.2. Capacidad: comunica y representa ideas matemáticas	105
C.3. Capacidad: elabora y usa estrategias	105
C.4. Capacidad: razona y argumenta generando ideas matemáticas	106
D. Análisis de convergencia	107
D.1. Convergencia del estudio de Simulación	107
D.1.1. Estudio de simulación 1	107
D.1.2. Estudio de simulación 2	109
D.1.3. Estudio de simulación 3	110
D.1.4. Estudio de simulación 4	111
D.1.5. Estudio de simulación 5	112
D.1.6. Estudio de simulación 6	113
D.1.7. Estudio de simulación 7	114
D.1.8. Estudio de simulación 8	115
D.2. Convergencia del modelo DINA	116
D.2.1. Comportamiento de la Cadena y densidad	116
D.2.2. Autocorrelaciones	118

D.3. Convergencia del modelo Rasch	120
Bibliografía	121



Lista de Abreviaturas

fdp	Función de densidad de probabilidad .
M-H	Metropolis-Hastings.
ECE	Evaluación Censal de Estudiantes.
TCT	Teoría Clásica del Test.
TRI	Teoría de Respuesta al Ítem.
MDC	Modelos de Diagnóstico Cognitivo.
CDM	<i>Cognitive Diagnosis Models.</i>
DINA	<i>Deterministic Input Noisy AND gate.</i>
Minedu	Ministerio de Educación.
EM	Esperanza - Maximización.
MCMC	<i>Markov Chain Monte Carlo.</i>
IE	Institución Educativa.
BC	Bloque común.
TIC	Tecnologías de la Información y Comunicación.
UMC	Oficina de Medición de la Calidad de los Aprendizajes.
RMSE	<i>Root Mean Squared Error.</i>
RSM	<i>Rule-space method.</i>
AHM	<i>Skill hierarchy method.</i>
BIN	<i>Bayesian inference network.</i>
HO-DINA	<i>Higher-order DINA.</i>
MS-DINA	<i>Multi-strategy DINA.</i>
LCDM	<i>Loglinear Cognitive Diagnosis Model.</i>
DINO	<i>Deterministic inputs, noisy 'or' gate.</i>
NIDA	<i>Noisy inputs, deterministic 'and' gate.</i>
NIDO	<i>Noisy inputs, deterministic 'or' gate .</i>
RUM	<i>Reparametrized unified model o Fusion model.</i>
C-RUM	<i>Compensatory RUM.</i>
NC-RUM	<i>Non-compensatory RUM.</i>
GDM	<i>General diagnostic model.</i>
MCLCM	<i>Multiple classification latent class model.</i>
Dir	Distribución Dirichlet.
Mul	Distribución Multinomial.

Índice de figuras

5.1. Histograma de las medidas de dificultad de los ítems de la prueba de matemática estimadas para la muestra de alumnos evaluados en 2do de secundaria	79
5.2. Histograma de las medidas de habilidad en matemática estimadas para la muestra de alumnos evaluados en 2do de secundaria	80
5.3. Biplot del análisis de correspondencias simple para los indicadores de los ítems	84
5.4. Gráfico de dispersión de la relación entre las estimaciones de dificultad y adivinación de los ítems	85
5.5. Gráfico de dispersión de la relación entre la correlación ítem-medida y las estimaciones del desliz de los ítems	85
5.6. Biplot del análisis de correspondencia simples para los indicadores de los alumnos	88
C.1. Ítem de la prueba modelo de matemática que requiere de la habilidad: Matematiza situaciones	104
C.2. Ítem de la prueba modelo de matemática que requiere de la habilidad: Comunica y representa ideas matemáticas	105
C.3. Ítem de la prueba modelo de matemática que requiere de la habilidad: Elabora y usa estrategias	105
C.4. Ítem de la prueba modelo de matemática que requiere de la habilidad: Razona y argumenta generando ideas matemáticas	106

Índice de cuadros

2.1. Taxonomía de los modelos de diagnóstico cognitivo	5
2.2. Resumen de matrices	8
4.1. Escenarios del Estudio 1 de simulación para una relación alta entre los parámetros de los ítems y las habilidades de los individuos, manteniendo el parámetro π constante	23
4.2. Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 1 con $N = 100$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	23
4.3. Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 1 con $N = 100$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	24
4.4. Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 1 con $N = 300$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	25
4.5. Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 1 con $N = 300$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	25
4.6. Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 1 con $N = 500$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	26
4.7. Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 1 con $N = 500$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	27
4.8. Resumen de medias de los valores RMSE para todos los parámetros estudiados .	27
4.9. Resumen de medias de los valores del sesgo para todos los parámetros estudiados	28
4.10. Escenarios del Estudio 2 de simulación para una relación alta entre los parámetros de los ítems y las habilidades de los individuos, manteniendo el parámetro π constante	28
4.11. Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 2 con $N = 100$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	29
4.12. Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 2 con $N = 100$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	30



4.13. Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 2 con $N = 300$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico 31

4.14. Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 2 con $N = 300$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico 31

4.15. Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 2 con $N = 500$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico 32

4.16. Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 2 con $N = 500$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico 33

4.17. Resumen de medias de los valores RMSE para todos los parámetros estudiados 33

4.18. Resumen de medias de los valores del sesgo para todos los parámetros estudiados 34

4.19. Escenarios del Estudio 3 de simulación para una relación fuerte para el parámetro de desliz (s); aunque una relación débil para el parámetro de adivinación (g), con las habilidades de los individuos, manteniendo el parámetro π constante 34

4.20. Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 3 con $N = 100$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico 35

4.21. Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 3 con $N = 100$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico 36

4.22. Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 3 con $N = 300$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico 37

4.23. Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 3 con $N = 300$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico 37

4.24. Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 3 con $N = 500$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico 38

4.25. Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 3 con $N = 500$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico 39

4.26. Resumen de medias de los valores RMSE para todos los parámetros estudiados 39

4.27. Resumen de medias de los valores del sesgo para todos los parámetros estudiados 40

4.28. Escenarios del Estudio 4 de simulación para una relación baja para el parámetro de desliz (s); aunque una relación fuerte para el parámetro de adivinación (g), con las habilidades de los individuos, manteniendo el parámetro π constante 40

4.29. Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 4 con $N = 100$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	41
4.30. Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 4 con $N = 100$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	42
4.31. Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 4 con $N = 300$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	43
4.32. Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 4 con $N = 300$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	43
4.33. Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 4 con $N = 500$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	44
4.34. Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 4 con $N = 500$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	45
4.35. Resumen de medias de los valores RMSE para todos los parámetros estudiados .	45
4.36. Resumen de medias de los valores del sesgo para todos los parámetros estudiados	46
4.37. Escenarios del Estudio 5 de simulación para una relación baja entre los parámetros de los ítems y las habilidades de los individuos, con el parámetro π manteniendo una relación de alto impacto	46
4.38. Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 5 con $N = 100$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	47
4.39. Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 5 con $N = 100$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	48
4.40. Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 5 con $N = 300$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	49
4.41. Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 5 con $N = 300$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	49
4.42. Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 5 con $N = 500$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	50
4.43. Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 5 con $N = 500$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	51
4.44. Resumen de medias de los valores RMSE para todos los parámetros estudiados .	51

4.45. Resumen de medias de los valores del sesgo para todos los parámetros estudiados	52
4.46. Escenarios del Estudio 6 de simulación para una relación alta entre los parámetros de los ítems y las habilidades de los individuos, con el parámetro π manteniendo una relación de alto impacto	52
4.47. Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 6 con $N = 100$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	53
4.48. Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 6 con $N = 100$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	54
4.49. Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 6 con $N = 300$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	55
4.50. Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 6 con $N = 300$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	55
4.51. Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 6 con $N = 500$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	56
4.52. Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 6 con $N = 500$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	57
4.53. Resumen de medias de los valores RMSE para todos los parámetros estudiados .	57
4.54. Resumen de medias de los valores del sesgo para todos los parámetros estudiados	58
4.55. Escenarios del Estudio 7 de simulación para una relación alta para el parámetro de desliz (s); aunque una relación débil para el parámetro de adivinación (g), con las habilidades de los individuos, con el parámetro π manteniendo una relación de alto impacto	58
4.56. Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 7 con $N = 100$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	59
4.57. Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 7 con $N = 100$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	60
4.58. Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 7 con $N = 300$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	61
4.59. Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 7 con $N = 300$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	61

4.60. Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 7 con $N = 500$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	62
4.61. Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 7 con $N = 500$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	63
4.62. Resumen de medias de los valores RMSE para todos los parámetros estudiados .	63
4.63. Resumen de medias de los valores del sesgo para todos los parámetros estudiados	64
4.64. Escenarios del Estudio 8 de simulación para una relación baja para el parámetro de desliz (s); aunque una relación fuerte para el parámetro de adivinación (g), con las habilidades de los individuos, con el parámetro π manteniendo una relación de alto impacto	64
4.65. Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 8 con $N = 100$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	65
4.66. Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 8 con $N = 100$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	66
4.67. Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 8 con $N = 300$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	67
4.68. Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 8 con $N = 300$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	67
4.69. Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 8 con $N = 500$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	68
4.70. Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 8 con $N = 500$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico	69
4.71. Resumen de medias de los valores RMSE para todos los parámetros estudiados .	69
4.72. Resumen de medias de los valores del sesgo para todos los parámetros estudiados	70
5.1. Matriz de diseño de bloques de Matemática para el día 1	72
5.2. Matriz de diseño de bloques de Matemática para el día 2	73
5.3. Descripción de las habilidades matemáticas que se evalúan en la prueba de matemática	74
5.4. Porción de la matriz Q para la prueba de matemática	75
5.5. Estimaciones de los parámetros g y s de los ítems de la prueba de matemática de 2do de secundaria	76
5.6. Descriptivos de la estimación de los parámetros de los ítems, del modelo DINA, en la muestra de alumnos evaluados en 2do de secundaria	77

5.7. Distribución de la probabilidad posterior de cada clase latente para la prueba de matemática de 2do de secundaria	77
5.8. Estimación de la dificultad de los ítems mediante el modelo de Rasch	78
5.9. Descriptivos de la medida de dificultad de la muestra de alumnos evaluados en 2do de secundaria	79
5.10. Descriptivos de la medida de habilidad de la muestra de alumnos evaluados en 2do de secundaria	80
5.11. Distribución de los ítems de acuerdo a los parámetros estimados por el modelo DINA y el modelo Rasch, junto a la matriz Q y el indicador de cuartiles	82
5.12. Tabla de contingencia entre las habilidades presentadas por la matriz Q y el indicador de cuartiles de la escala de dificultad estimada a partir del modelo Rasch	83
5.13. Tabla de contingencia entre los cuartiles de la escala de habilidad latente estimada a partir del modelo Rasch y el perfil latente estimado a partir del modelo DINA	86
5.14. Distribución de los 16 perfiles latentes de acuerdo al estrato al que pertenece la institución educativa del alumno	87
5.15. Distribución de los cuartiles de rendimiento de acuerdo al estrato al que pertenece la institución educativa del alumno	87
D.1. Gráficos de las cadenas y autocorrelaciones para los parámetros s , g y π de un ítem y una clase, respectivamente, elegido aleatoriamente para el estudio 1 de simulación	107
D.2. Gráficos de las cadenas y autocorrelaciones para los parámetros s , g y π de un ítem y una clase, respectivamente, elegido aleatoriamente para el estudio 2 de simulación	109
D.3. Gráficos de las cadenas y autocorrelaciones para los parámetros s , g y π de un ítem y una clase, respectivamente, elegido aleatoriamente para el estudio 3 de simulación	110
D.4. Gráficos de las cadenas y autocorrelaciones para los parámetros s , g y π de un ítem y una clase, respectivamente, elegido aleatoriamente para el estudio 4 de simulación	111
D.5. Gráficos de las cadenas y autocorrelaciones para los parámetros s , g y π de un ítem y una clase, respectivamente, elegido aleatoriamente para el estudio 5 de simulación	112
D.6. Gráficos de las cadenas y autocorrelaciones para los parámetros s , g y π de un ítem y una clase, respectivamente, elegido aleatoriamente para el estudio 6 de simulación	113
D.7. Gráficos de las cadenas y autocorrelaciones para los parámetros s , g y π de un ítem y una clase, respectivamente, elegido aleatoriamente para el estudio 7 de simulación	114
D.8. Gráficos de las cadenas y autocorrelaciones para los parámetros s , g y π de un ítem y una clase, respectivamente, elegido aleatoriamente para el estudio 8 de simulación	115

D.9. Gráficos de las cadenas para los parámetros s y g de una selección aleatoria de 5 ítems de los 48 usados para la aplicación 116

D.10. Gráficos de las cadenas para el parámetro π de una selección aleatoria de 5 clases de las 16 posibles en la aplicación 117

D.11. Gráficos de las autocorrelaciones para los parámetros s y g de una selección aleatoria de 5 ítems de los 48 usados para la aplicación 118

D.12. Gráficos de las autocorrelaciones para el parámetro π de una selección aleatoria de 5 clases de las 16 posibles en la aplicación 119

D.13. Gráficos de densidad, de las cadenas y de autocorrelación de una selección de 20 ítems aleatorios de los 48 usados para la aplicación 120



Capítulo 1

Introducción

En las últimas décadas, el campo de la medición ha experimentado un creciente auge con el estudio de los modelos de la Teoría de Respuesta al Ítem (TRI), los cuales nos brindan una medida continua en escala interválica de una variable latente que se pretende medir; pero que es no observable. Tal resultado es expresado como una cantidad numérica que podría carecer de valor conceptual si es que no se coloca en un contexto o bajo cierta descripción; por ejemplo, en el campo de la educación, el contexto sería el pedagógico.

En este trabajo consideraremos específicamente el campo de la educación enfocándonos en la competencia matemática; por lo que, más allá de asignar a los estudiantes un puntaje en una escala continua que represente la habilidad en la competencia matemática, como se trabaja en la TRI; se pondrá trabajar con los modelos de diagnóstico cognitivo (MDC).

Los modelos de diagnóstico cognitivo (MDC) son modelos psicométricos que pueden ser usados para evaluar las fortalezas y debilidades de los estudiantes ya que proveen información específica en forma de perfiles o clases que pueden permitir una efectiva medición del aprendizaje y progreso del estudiante, el diseño de una mejor instrucción y posiblemente una intervención para hacer frente a las necesidades individuales y grupales (Huebner, 2010). Los diferentes modelos de esta familia son ampliamente revisados en los trabajos de DiBello et al. (2007), Rupp y Templin (2008).

Uno de los modelos que pertenecen a la familia mencionada en el párrafo anterior es el llamado modelo DINA. El presente trabajo se enfocará en el desarrollo de este modelo desde el enfoque bayesiano orientándose principalmente en el trabajo de Culpepper (2015a).

El trabajo terminará con la aplicación del modelo a una muestra de 3040 alumnos de la Evaluación Censal de Estudiantes (ECE) 2015 del segundo grado de secundaria del Ministerio de Educación (Minedu) evaluados en la competencia matemática. Esta prueba tiene un diseño de bloques rotados, de la que se usará una combinación de dos cuadernillos, los cuales cuentan con 50 ítems y cuatro habilidades ($K = 4$) matemáticas necesarias para desarrollarlos: matematiza situaciones, comunica y representa, razona y argumenta, y elabora y usa estrategias (Minedu, 2016).

El objetivo general de la tesis es estudiar los fundamentos y propiedades del modelo DINA; así como estimar y aplicar este modelo a un conjunto de datos reales desde el punto de vista de la estadística bayesiana. De manera específica:

- Revisar la literatura acerca de las diferentes propuestas de modelos de diagnóstico cognitivo (MDC).

- Estudiar los fundamentos y las propiedades del modelo DINA.
- Estudiar el proceso de estimación de los parámetros en el modelo DINA con especial énfasis en el enfoque bayesiano.
- Estudiar las rutinas en el software libre R para la estimación del modelo DINA y el muestreador de Gibbs.
- Realizar un estudio de simulación a efectos de comparar la estimación bayesiana y clásica para el modelo DINA.
- Aplicar el modelo a un conjunto de datos educacionales del país que evalúan la competencia matemática.
- Analizar la correspondencia entre los perfiles encontrados por el modelo DINA y las agrupaciones por percentiles realizadas a partir de la estimación con un modelo de la TRI; así como analizar la relación entre las características de los ítems y los parámetros de estos en el modelo DINA.

1.1. Organización del Trabajo

En el capítulo 2 se presenta una breve introducción sobre los modelos de respuesta al ítem y los modelos de diagnóstico cognitivo; para luego desarrollar la formulación teórica del modelo DINA hasta llegar a su función de verosimilitud.

En el capítulo 3 se muestra inicialmente la estimación clásica del modelo para luego hacer énfasis en la estimación bajo el enfoque bayesiano propuesto por Culpepper (2015a) usando el muestreador de Gibbs. En el capítulo 4 se presentan 8 estudios de simulación donde se compara la recuperación de parámetros del modelo en estudio, desde el enfoque clásico y bayesiano. Luego, en el capítulo 5 se analizan los resultados de aplicar el modelo propuesto a una muestra de alumnos evaluados en la competencia matemática en la evaluación censal de estudiantes (ECE) en segundo de secundaria. Además, en este capítulo se presenta un análisis de correspondencia entre la posible clasificación por cuartiles mediante el modelo Rasch y la clasificación mediante probabilidades que brinda el modelo DINA. De la misma manera, se analiza la relación entre los cuartiles de la medida de dificultad de los ítems obtenida por el modelo Rasch y las cuatro habilidades requeridas para responderlos correctamente. Se estudiará también la relación entre la dificultad de cada ítem y la estimación del parámetro de adivinación; así como, la relación entre la correlación ítem-medida, que desde un punto de vista clásico podría acercarse al concepto de discriminación, y las estimaciones del parámetro de desliz.

Finalmente el capítulo 6 presenta algunas conclusiones y sugerencias para futuras investigaciones. Se incluyen cuatro anexos. En el primero (A) se presenta el código utilizado en R para la estimación bayesiana mediante el muestreador de Gibbs, en el segundo (B) se muestran las rutinas de código utilizadas para la aplicación en educación, y en el tercero (C) se presentan algunos ejemplos de ítems de las cuatro capacidades mencionadas que han sido evaluadas en la prueba de Matemática del Ministerio de Educación (Minedu). Para terminar, en el cuarto (D) se presentan los gráficos de cadenas, densidad y autocorrelaciones para verificar las convergencias en los estudios de simulación y aplicación desarrolladas.

Capítulo 2

El modelo DINA

2.1. Modelos de respuesta al ítem y modelos de diagnóstico cognitivo

En las ciencias sociales, especialmente en psicología y educación, uno de los objetivos fundamentales recae en la definición de constructos que ayuden a la toma de decisiones. Ello consiste en describir alguna característica de los sujetos mediante un puntaje numérico; es decir, en medir o cuantificar un atributo específico. Este trabajo es tarea fundamental de la psicometría.

Los orígenes de la psicometría, como lo explica [Jones y Thissen \(2007\)](#), vienen desde el siglo XIX con los primeros trabajos de Thurstone presentando a la psicología como una ciencia cuantitativa racional y fundando en 1935 “La Sociedad Psicométrica”, la cual patrocinó la revista “Psicométrica”. El desarrollo de la psicometría continuó con los trabajos de Brown y Thomson en 1921, quienes hicieron importantes contribuciones al desarrollo de la teoría del test y del análisis factorial. Además de ellos existieron otros grandes pioneros en los modelos psicométricos como Thorndike, Spearman, Binet, entre otros.

De igual forma, [Jones y Thissen \(2007\)](#) describe el estudio de la psicometría en dos partes: la primera, depende de la dimensionalidad de la escala, la cual puede estudiar la aproximación a la unidimensionalidad o asumir múltiples dimensiones; mientras que, la segunda parte está orientada al enfoque de medición, especialmente en la Teoría Clásica del Test (TCT) y la Teoría de Respuesta al Ítem (TRI).

Los últimos mencionados han sido ampliamente estudiados en los trabajos de [Van der Linden y Hambleton \(1997\)](#), [Reckase \(2009\)](#); donde exploran tanto los modelos unidimensionales como multidimensionales de la TRI y están enfocados principalmente a la obtención de una escala cuantitativa continua de la variable latente que se pretende medir, sea cual fuere el enfoque. Específicamente en el campo educativo esto quiere decir, conocer la habilidad de un alumno en una determinada competencia, por lo que, esta medición dependerá de los parámetros que se consideren necesarios en los modelos de la TRI.

A partir de esta escala continua, los tomadores de decisiones pueden dar mayores o menores consecuencias a estos resultados. Como discute [De la Torre y Minchen \(2014\)](#), a menudo los modelos de la TCT o la TRI se usan en las evaluaciones educativas tradicionales para una variedad de propósitos, como identificar el nivel de competencia de un estudiante, diferenciar a estudiantes que aprueban o no una competencia educativa, seleccionar candidatos para un programa de beneficios, admitir a los estudiantes a una universidad o determinar los beneficiarios de las becas.

Lo expuesto por [De la Torre y Minchen \(2014\)](#) indica que las evaluaciones educativas utili-

zadas para estos fines tienen como objetivo común determinar hasta qué punto los estudiantes poseen la competencia o rasgo de interés, lo cual funciona bien para evaluaciones a grandes escalas ya que pueden ser útiles para impulsar decisiones de política y programación de alto nivel; por ejemplo, la modificación del plan de estudios y la instrucción para los años subsiguientes. Sin embargo, el diseño de estas evaluaciones, no siempre resultan ser adecuadas para proporcionar información diagnóstica, debido a que ellas no brindan suficiente información que pueda ser utilizada para mejorar la enseñanza y el aprendizaje en el aula (De la Torre, 2009a).

Es por esto que, debido a la necesidad de ofrecer información de diagnóstico en forma oportuna, surgen los modelos de diagnóstico cognitivo (MDC) (Henson y Douglas, 2005), los cuales como describe Rupp y Templin (2008) son también conocidos, dependiendo del aspecto particular que se resalte del modelo, como modelos psicométricos cognitivos (Rupp, 2007); modelos de clasificación múltiple; modelos de respuesta latente; modelos de clases latentes restringidos (Haertel, 1989); entre otros. Es necesario enfatizar que estos modelos trabajan con variables cualitativas, tanto las dependientes como las independientes.

Precisamente, en el ámbito educativo estos modelos son mejor reconocidos como *modelos psicométricos de diagnóstico cognitivo*; donde la palabra “psicométrico” hace énfasis al estudio de las variables latentes; es decir, aquellas que no se pueden medir directamente; y a su vez, la palabra “diagnóstico” hace referencia a la aplicación de estos modelos a un problema en particular, el cual requiere brindar una valoración o juicio de un individuo. Por ejemplo, describir sus fortalezas y debilidades respecto a un área curricular.

Estos modelos han sido clasificados de distintas formas, Rupp y Templin (2008) plantea organizarlos de acuerdo a tres características:

1. Según la escala de medición de las variables de respuesta observadas: Dicotómicas (unos y ceros) o politómicas (más de dos categorías).
2. Según la escala de medición de las variables predictoras latentes: Dicotómicas o politómicas.
3. Según la combinación de las variables predictoras latentes: Esta clasificación hace referencia a la manera de cómo se combinan las variables predictoras latentes con las diferentes habilidades para obtener las respuestas observadas. De acuerdo a lo mencionado, se pueden dividir en: modelos compensatorios y no compensatorios. La principal diferencia entre ellos es que el primero permite que una habilidad requerida para responder un ítem puede ser compensada con otra adicional. Mientras, que los modelos no compensatorios requieren que cada una de las habilidades esté presente para producir una respuesta correcta al ítem (dicotómico) o un alto grado de respuesta (si fuera politómico).

Los modelos de diagnóstico cognitivo son presentados, bajo la clasificación mencionada, en el cuadro 2.1:

Cuadro 2.1: Taxonomía de los modelos de diagnóstico cognitivo

Variables de respuesta observadas	Variables predictoras latentes		Tipo de modelo
	Dicotómicas	Politómicas	
Dicotómicas	RSM		No compensatorio
	AHM		
	DINA		
	HO-DINA		
	MS-DINA		
	NIDA		
	BIN	BIN	
	MCLCM	MCLCM	
	NC-RUM	NC-RUM	
	RERUM		
DINO		Compensatorio	
NIDO			
BIN	BIN		
MCLCM	MCLCM		
C-RUM	C-RUM		
GDM	GDM		
LCDM	LCDM		
Politómico	RSM		No compensatorio
	AHM		
	BIN	BIN	
	MCLCM	MCLCM	
	NC-RUM	NC-RUM	Compensatorio
	BIN	BIN	
	MCLCM	MCLCM	
	C-RUM	C-RUM	
	GDM	GDM	
	LCDM	LCDM	

Fuente: Extraído de la sección 4, tabla 1: Una taxonomía de MDC (Rupp y Templin, 2008)

Las siglas mostradas en la clasificación de los modelos corresponden a los siguientes: RSM refiere a la regla del espacio; AHM se refiere al método de la jerarquía de habilidad; BIN es redes de inferencia bayesiana; DINA es el modelo en estudio, que puede ser traducido como el modelo de entrada determinística con ruido “AND” (Y) salida; HO-DINA es el modelo DINA pero con alto orden; MS-DINA es el DINA multi estratégico; LCDM es el modelo de diagnóstico cognitivo loglineal; DINO es el modelo de entrada determinística con ruido “OR” (O) salida; NIDA tiene los mismos atributos que el DINA pero con diferente orden de prioridad; NIDO lo mismo que el anterior pero con el DINO; RUM es el modelo unificado reparamétrizado o también llamado modelo de fusión. C-RUM es el modelo RUM pero de tipo compensatorio; NC-RUM es el modelo RUM pero de tipo no compensatorio; GDM se refiere al modelo general de diagnóstico; finalmente el MCLCM es el modelo de clasificación de clases latentes múltiple.

Cabe mencionar que los modelos no compensatorios pueden ser divididos en: conjuntivos y disjuntivos. Se habla de un modelo conjuntivo cuando todas las habilidades requeridas para responder un ítem, necesariamente tienen que ser dominadas por el individuo para dar con la respuesta correcta. Mientras que en los modelos disjuntivos, el evaluado solo necesitará dominar un subconjunto de habilidades para poder dar con la respuesta correcta al ítem (DiBello et al., 2007).

Es importante mencionar que los modelos disjuntivos no son lo mismo que los compensatorios, ya que para estos últimos, el dominio de una habilidad compensa a las otras que se necesitan; mientras que en los disjuntivos, el subconjunto de habilidades que pueda dominar el evaluado

no compensa a las otras (Lim, 2015).

Como se ha podido ver, existen múltiples clasificaciones y modelos estudiados para esta gran familia MDC. Es por esto, que este trabajo se enfocará solo en el desarrollo de las propiedades de uno de los primeros y más populares MDC como es el DINA, el cual es un modelo con variables observadas dicotómicas, variables predictoras latentes dicotómicas y además, se encuentra clasificado como no compensatorio conjuntivo.

2.2. Formulación del modelo DINA

El modelo DINA plantea un contexto en el que N individuos, que pertenecen a C clases o perfiles latentes, son evaluados bajo J ítems los cuales requieren de K habilidades o atributos.

El modelo asume que existen K habilidades conocidas posibles que un individuo $i \in \{1, 2, \dots, N\}$ debería poseer para responder correctamente a un ítem $j \in \{1, 2, \dots, J\}$.

A partir de ello, el investigador debe construir inicialmente una matriz $J \times K$ llamada matriz \mathbf{Q} :

$$\mathbf{Q} = \begin{matrix} & \begin{matrix} 1 & 2 & \dots & K \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ \vdots \\ J \end{matrix} & \begin{pmatrix} q_{11} & q_{12} & \dots & q_{1k} \\ q_{21} & q_{22} & \dots & q_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ q_{J1} & q_{J2} & \dots & q_{JK} \end{pmatrix} \end{matrix}$$

la cual posee una estructura binaria; es decir, con elementos 0 y 1 en donde cada q_{jk} se define como:

$$q_{jk} = \begin{cases} 1 & , \text{ si la habilidad } k \in \{1, 2, \dots, K\} \text{ es necesaria para responder correctamente} \\ & \text{el ítem } j \\ 0 & , \text{ en caso contrario} \end{cases}$$

Para mejor ilustrar estas definiciones consideraremos el siguiente ejemplo hipotético. Supongamos que se tiene una prueba de matemática con 5 ítems que es aplicada a 10 individuos y además se establece que para poder desarrollar correctamente estos ítems se necesitarían dominar 3 habilidades. De acuerdo a los especialistas en el área, se sabe que para responder correctamente al ítem 1, el individuo debe dominar las tres habilidades; para responder correctamente al ítem 2 y el ítem 4 se requiere solo la segunda habilidad; asimismo se necesitará solo de la tercera habilidad para acertar el ítem 3; finalmente, el individuo debe poseer la primera habilidad para responder el último ítem planteado. Siguiendo este ejemplo, la matriz \mathbf{Q} quedará entonces expresada de la siguiente manera:

$$\mathbf{Q} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \end{matrix}$$

A partir de lo anterior, definamos $\mathbf{q}_j = (q_{j1}, \dots, q_{jK})$ como la j -ésima fila de $\mathbf{Q} = [\mathbf{q}_1, \dots, \mathbf{q}_J]^T$, el cual indica el vector de habilidades que son requisitos para responder el ítem j .

Hasta aquí tenemos el insumo principal, que es brindado por los especialistas en el área. Sin embargo, solo hemos creado la relación entre los J ítems y las K habilidades que debería poseer el individuo.

El modelo DINA estima la relación entre los N individuos y las K habilidades basándose en el patrón de respuestas de los individuos a los ítems, el cual es expresado como una matriz de unos y ceros, donde se le asigna 1 cuando responde correctamente a un ítem j y 0 en caso contrario. Tal relación se materializa en una matriz $N \times K$, \mathbf{A} , definida por:

$$\mathbf{A} = \begin{matrix} & \begin{matrix} 1 & 2 & \dots & K \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ \vdots \\ N \end{matrix} & \begin{pmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1k} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{N1} & \alpha_{N2} & \dots & \alpha_{NK} \end{pmatrix} \end{matrix}$$

la cual posee, al igual que \mathbf{Q} , una estructura binaria; es decir, con elementos 0 y 1, donde cada α_{ik} se define como:

$$\alpha_{ik} = \begin{cases} 1 & , \text{ si el individuo } i \text{ domina la habilidad } k \\ 0 & , \text{ en caso contrario} \end{cases}$$

Sin embargo, a diferencia de \mathbf{Q} , las entradas de esta matriz son latentes; es decir, no observables. Supongamos, retornando a nuestro ejemplo que hipotéticamente, los individuos 1, 5 y 7 solo dominan la primera habilidad; el individuo 4 y 10 solo dominan la tercera habilidad y el individuo 8 solo domina la segunda habilidad. Por otro lado, supongamos también de que existen individuos que dominan más de una habilidad, por citar que el sexto individuo domina la primera y segunda habilidad; el segundo individuo domina la segunda y tercera habilidad. Finalmente, asumamos que solo el tercer individuo posee las tres habilidades, mientras que el noveno no domina ninguna de ellas. Conforme a esta suposición, la matriz \mathbf{A} quedaría expresada por:

$$\mathbf{A} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \end{matrix} & \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{matrix}$$

En el mismo sentido como para \mathbf{Q} , se define aquí el vector latente $\alpha_i = (\alpha_{i1}, \dots, \alpha_{iK})$ que

indica todas las habilidades que domina o no el individuo i . Este vector es también llamado el “estado de conocimiento” del individuo i (Junker y Sijtsma, 2001).

Hasta este momento se tienen dos relaciones con las K habilidades, la primera con los ítems j representados con la matriz \mathbf{Q} y la segunda con los individuos i representado por la matriz \mathbf{A} donde importa principalmente el estado de conocimiento α_i .

A partir de lo anterior, podemos empezar a desarrollar el agrupamiento o conformación de las clases para los individuos. Si C es el número de clases latentes, el cual toma como valor máximo $C = 2^K$ cuando todo posible patrón de dominio es admisible, podemos definir una variable aleatoria Y con valores en $\{1, 2, \dots, C\}$ como aquella que nos indique a que clase pertenece un individuo. Denotaremos igualmente por $\alpha_c = (\alpha_{c1}, \alpha_{c2}, \dots, \alpha_{cK})$ al estado de conocimiento de cualquier individuo en la clase c , ya que por clase entenderemos a que todos los individuos en ella poseen idéntico estado de conocimiento.

Vale comentar que el modelo que estudiaremos se denomina conjuntivo pues tiene como asunción central que para que un individuo i responda correctamente al ítem j necesitará de todas las habilidades que conciernen a tal ítem. Ello hace que sea necesario el definir las siguientes variables dicotómicas:

$$\eta_{ij} = \begin{cases} 1 & , \text{ si el individuo } i \text{ domina todas las habilidades que se necesitan para responder} \\ & \text{correctamente el ítem } j \\ 0 & , \text{ en caso contrario} \end{cases}$$

Así, $\eta_{ij} = \prod_{k=1}^K \alpha_{ik}^{q_{jk}}$ o equivalentemente $\eta_{ij} = \prod_{k=1}^K \alpha_{ck}^{q_{jk}}$, si $Y_i = c$. Continuando con nuestro ejemplo, puede ser ilustrativo ahora resumir en el cuadro 2.2 los diferentes elementos hasta aquí considerados.

Cuadro 2.2: Resumen de matrices

\mathbf{Q}	α_c	$\mathbf{A} = [\alpha_{ik}]$	η_{ij}
$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$

Note que si bien el número de clases es de 7, pues no se presentó el patrón $(1, 0, 1)$, ello en la práctica no lo sabremos ya que tales clases son latentes; sin embargo, pudiera ocurrir que por razones teóricas o de jerarquías algunos patrones pudieran ser a priori inadmisibles, lo cual reduciría el número de clases C a considerar.

Al modelo DINA (del acrónimo en inglés *Deterministic Input Noisy AND gate*) se le atribuye este nombre por la naturaleza conjuntiva de su proceso; es decir, este tendrá como salidas 1 si el individuo posee todas las habilidades requeridas para responder correctamente al ítem o 0 en caso contrario. Además, la denominación de “ruido” es atribuido también al modelo ya que

para los ítems se estiman dos probabilidades de error (parámetros), los cuales de definirse la respuesta binaria del individuo i a un ítem j por X_{ij} se definen como:

- Adivinación (*guessing* g): Es la probabilidad de responder correctamente a un ítem j a pesar de no dominar las habilidades requeridas para hacerlo.

$$g_j = P(X_{ij} = 1 \mid \eta_{ij} = 0)$$

- Desliz (*slipping* s): Es la probabilidad de fallar a un ítem j a pesar de dominar las habilidades requeridas para hacerlo.

$$s_j = P(X_{ij} = 0 \mid \eta_{ij} = 1)$$

Además de la estimación de estos parámetros de ítems, el modelo DINA requiere la estimación de las probabilidades $\pi_c = P(Y_i = c) = P(\alpha_i = \alpha_c)$ de pertenencia de cada individuo i a la clase o perfil latente c .

La probabilidad de que un individuo i , que pertenece a la clase c , responda correctamente a un ítem j en el modelo está determinada por:

$$P_{jc} = P(X_{ij} = 1 \mid \alpha_i = \alpha_c, \theta_j) = (1 - s_j)^{\eta_{ij}} g_j^{1 - \eta_{ij}}, \quad (2.1)$$

donde $\theta_j = (g_j, s_j)$ es un vector fila de los probabilidades de error del ítem ya definidos. En general, asumiendo independencia condicional y usando (2.1), la probabilidad de observar la respuesta $x_{ij} \in \{0, 1\}$ en el individuo i , que pertenece a la clase c , viene dada por:

$$P(X_{ij} = x_{ij} \mid \alpha_i = \alpha_c, \theta_j) = P_{jc}^{x_{ij}} (1 - P_{jc})^{1 - x_{ij}}, \quad (2.2)$$

donde P_{jc} viene dada por la expresión (2.1).

Utilizando la ley de la probabilidad total y la ecuación (2.2), se deduce entonces que la probabilidad de obtener por parte del individuo i un patrón de respuestas $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{iJ}]$ dado los parámetros de todos los ítems, que los resumiremos en el vector θ , y las probabilidades de pertenencia a las clases latentes $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_C)$, estará dada por:

$$P(\mathbf{X}_i = \mathbf{x}_i \mid \alpha_i, \boldsymbol{\pi}, \theta) = \sum_{c=1}^C \pi_c \left[\prod_{j=1}^J P(X_{ij} = x_{ij} \mid \alpha_i = \alpha_c, \theta_j) \right] \quad (2.3)$$

Finalmente, la verosimilitud de observar una muestra de respuestas de N individuos a los J ítems estará dada por:

$$L(\boldsymbol{\alpha}, \boldsymbol{\pi}, \theta) = \prod_{i=1}^N P(\mathbf{X}_i = \mathbf{x}_i \mid \alpha_i, \boldsymbol{\pi}, \theta) = \prod_{i=1}^N \left\{ \sum_{c=1}^C \pi_c \left[\prod_{j=1}^J P_{jc}^{x_{ij}} (1 - P_{jc})^{1 - x_{ij}} \right] \right\}, \quad (2.4)$$

donde $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_C)$.

Capítulo 3

Estimación del modelo

3.1. Consideraciones previas

En la literatura se han propuesto diferentes métodos para estimar los parámetros del modelo DINA. En este capítulo daremos una breve revisión de estos centrandó nuestro interés en los métodos bayesianos y en particular en el uso del muestreador de Gibbs, para el cual [Culpepper \(2015a\)](#) ha desarrollado las distribuciones condicionales completas requeridas.

Como lo describe [De La Torre \(2009b\)](#), la estimación de los parámetros del modelo podría realizarse mediante el método de máxima verosimilitud en base a la ecuación (2.4). Sin embargo, al requerirse una estimación donde conjuntamente se estimen los estados de conocimiento α , los parámetros de los ítems θ y las pertenencias a las clases π , ello podría llevarnos a tener estimadores inconsistentes.

[De La Torre \(2009b\)](#) plantea, por tanto, estimar los parámetros mencionados por medio del estimador de máxima verosimilitud marginal en lugar de trabajar con la verosimilitud condicional de los datos. Para ello él propone discretizar la distribución conjunta del vector α . Esta manera de estimar los parámetros puede ser implementada mediante el algoritmo EM.

De otro lado desde el punto de vista bayesiano han habido, como lo discute [Culpepper \(2015a\)](#), la práctica de usar el algoritmo de Metropolis-Hastings. Sin embargo, la desventaja de este radica en la necesidad que tiene de ajustar ciertos parámetros, lo cual resulta bastante tedioso, lo hace ineficiente, y además altamente dependiente de los datos particulares en uso.

Otra opción, desde el enfoque bayesiano, es el muestreador de Gibbs que, con el hecho de que las distribuciones condicionales completas son explícitas, evita esta etapa de ajuste manual y, en general, tiende a mejorar la mezcla de la cadena.

Como se describe en [Hoff \(2009\)](#) los métodos bayesianos proveen en general estimadores adecuados que funcionan tanto para muestras pequeñas como grandes, a diferencia de los métodos clásicos, donde es importante tener un tamaño de muestra lo suficientemente grande. Más aún, el grado de incertidumbre se analiza en el enfoque clásico mediante intervalos de confianza mientras que en los métodos bayesianos se utilizan los llamados intervalos de credibilidad, los cuales son simples de obtener y proveen valores, en general, más precisos.

Finalmente, el enfoque bayesiano es flexible para generar nuevos procedimientos estadísticos en problemas o modelos más complejos ya que actualmente existen múltiples herramientas computacionales útiles para implementarlos.

Antes de entrar propiamente a detallar los métodos de estimación, daremos una breve revisión a las distribuciones categóricas y de Dirichlet, última que es conjugada de la primera y de la

distribución multinomial. Ellas son claves en el estudio del muestreador de Gibbs para nuestro modelo.

3.2. Distribución de Dirichlet

La distribución de Dirichlet, llamada así en honor a Peter Gustav Lejeune Dirichlet y denotada por $Dir(\boldsymbol{\alpha})$, es una familia de distribuciones de probabilidades continuas multivariadas, parametrizadas por un vector $\boldsymbol{\alpha}$ perteneciente al conjunto de los números reales positivos. Ella puede pensarse como una generalización multivariada de la distribución beta.

Esta distribución se utiliza frecuentemente como distribución a priori en estadística Bayesiana, como se hace en este trabajo, y tiene la propiedad de ser conjugada a la priori de una distribución categórica y de una distribución multinomial.

3.2.1. Función de densidad de probabilidad

La distribución Dirichlet de orden $K \geq 2$ con vector de parámetros no negativos $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_K)$ viene definida por la función de densidad conjunta:

$$f(x_1, \dots, x_K; \boldsymbol{\alpha}) = \frac{1}{B(\boldsymbol{\alpha})} \prod_{i=1}^K x_i^{\alpha_i - 1}, \quad (3.1)$$

donde el soporte de esta distribución recae sobre el simplejo de orden $K - 1$:

$$\Delta_{K-1} = \left\{ (x_1, x_2, \dots, x_K) \in [0, 1]^K \mid \sum_{i=1}^K x_i = 1 \right\}$$

La constante normalizadora en (3.1) conocida como la función multivariada Beta, está definida en términos de funciones Gamma como:

$$B(\boldsymbol{\alpha}) = \frac{\prod_{i=1}^K \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^K \alpha_i)}.$$

3.2.2. Propiedades

Si un vector aleatorio (X_1, X_2, \dots, X_K) tiene distribución Dirichlet, lo cual denotaremos por $\mathbf{X} = (X_1, \dots, X_K) \sim Dir(\boldsymbol{\alpha})$, definamos:

$$\alpha_0 = \sum_{i=1}^K \alpha_i.$$

Entonces el valor esperado de las distribuciones marginales de las primeras $K - 1$ componentes, vienen dadas por:

$$E(X_i) = \frac{\alpha_i}{\alpha_0}$$

mientras que sus varianzas por:

$$Var(X_i) = \frac{\alpha_i(\alpha_0 - \alpha_i)}{\alpha_0^2(\alpha_0 + 1)}.$$

Una propiedad más general de los momentos de estas variables aleatorias nos dice que:

$$E\left(\prod_{i=1}^K X_i^{\beta_i}\right) = \frac{B(\boldsymbol{\alpha} + \boldsymbol{\beta})}{B(\boldsymbol{\alpha})} = \frac{\Gamma(\sum_{i=1}^K \alpha_i)}{\Gamma(\sum_{i=1}^K (\alpha_i + \beta_i))} \times \prod_{i=1}^K \frac{\Gamma(\alpha_i + \beta_i)}{\Gamma(\alpha_i)},$$

donde $\boldsymbol{\beta}$ es un vector con entradas $\beta_i > 0$.

3.2.3. Distribución categórica

De la misma forma, es importante definir la distribución categórica, la cual es una distribución de probabilidad discreta cuyo rango es un conjunto de k items identificados individualmente. Esta distribución es también conocida como la generalización de la distribución Bernoulli a una variable aleatoria categórica.

En una de sus varias formulaciones, el rango de esta variable es tomado como una secuencia finita de números enteros. Estos pueden ser $\{1, 2, \dots, k\}$ o cualquier otro conjunto arbitrario de valores.

Se define entonces la función de probabilidad de la variable X con distribución categórica como:

$$f_X(i | \mathbf{p}) = p_i$$

donde $\mathbf{p} = (p_1, \dots, p_k)$, es una distribución de probabilidades; es decir que $\sum_{i=1}^k p_i = 1$.

Otra formulación que facilita generalmente la manipulación matemática y que utilizaremos en este trabajo, para esta misma función es:

$$f_X(x) = \prod_{i=1}^k p_i^{I_i(x)}, \quad (3.2)$$

donde $I_A(x)$ denotará en adelante a la función indicadora.

3.2.4. Conjugación a una distribución categórica y multinomial

La distribución Dirichlet es la distribución a priori conjugada de la distribución descrita en el acápite anterior y de la distribución multinomial.

Esto quiere decir que si una observación tiene una distribución categórica o multinomial, y la distribución a priori de sus parámetros es Dirichlet, entonces la distribución a posteriori del parámetro también será Dirichlet.

Formalmente para $\mathbf{x} = (x_1, \dots, x_k)$ y $\mathbf{p} = (p_1, \dots, p_k)$

$$\text{Si } \mathbf{X} = (X_1, X_2, \dots, X_k) | \mathbf{P} = \mathbf{p} \sim \text{Mul}(n; \mathbf{p}) \quad \text{y} \\ \mathbf{P} \sim \text{Dir}(\boldsymbol{\alpha})$$

entonces

$$\begin{aligned} P(\mathbf{P} = \mathbf{p} \mid \mathbf{X} = \mathbf{x}) &\propto P(\mathbf{X} = \mathbf{x} \mid \mathbf{P} = \mathbf{p}) \times P(\mathbf{P} = \mathbf{p}) \\ &\propto \frac{n!}{x_1! x_2! \dots x_k!} \times p_1^{x_1} \times \dots \times p_k^{x_k} \times \frac{1}{B(\boldsymbol{\alpha})} \times p_1^{\alpha_1-1} \times \dots \times p_k^{\alpha_k-1} \\ &\propto p_1^{x_1+\alpha_1-1} \times \dots \times p_k^{x_k+\alpha_k-1} \end{aligned}$$

Así,

$$\mathbf{P} \mid \mathbf{X} = \mathbf{x} \sim Dir(x_1 + \alpha_1, x_2 + \alpha_2, \dots, x_k + \alpha_k)$$

3.3. Estimación clásica mediante el algoritmo EM

Tal como se menciona en las consideraciones previas, [De La Torre \(2009b\)](#) sugiere que la estimación de los parámetros del modelo DINA, se haga mediante el estimador de máxima verosimilitud marginal a través del algoritmo EM.

En el trabajo de [George et al. \(2016\)](#), se describe el proceso iterativo del algoritmo EM para este modelo. Este consta de dos pasos: Esperanza y Maximización, donde los valores iniciales del vector de parámetros del ítem $\boldsymbol{\theta}$ y la distribución de las probabilidades de pertenencia de los individuos a las clases c tienen que ser escogidos. Dicho esto, se detalla el proceso en cada uno de los pasos de este algoritmo:

1. Paso E (esperanza)

- a) La distribución a posteriori individual puede deducirse por el Teorema de Bayes:

$$P(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_c \mid \mathbf{X}_i = \mathbf{x}_i, \boldsymbol{\theta}_j, \boldsymbol{\pi}) = \frac{P(\mathbf{X}_i = \mathbf{x}_i \mid \boldsymbol{\alpha}_i = \boldsymbol{\alpha}_c, \boldsymbol{\theta}_j), \boldsymbol{\pi}_c}{\sum_{m=1}^C P(\mathbf{X}_i = \mathbf{x}_i \mid \boldsymbol{\alpha}_i = \boldsymbol{\alpha}_m) \boldsymbol{\pi}_m}$$

donde $c \in \{1, \dots, C\}$ denota a la clase latente a la cual pertenece el individuo i .

- b) Dos tipos de valores esperados son derivados de la posteriori: El primero es el número esperado de estudiantes, los cuales son clasificados en la clase $\boldsymbol{\alpha}_c$ para los ítems $j = 1, \dots, J$.

$$T_{jc} = \sum_{i=1}^N P(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_c \mid \mathbf{X}_i = \mathbf{x}_i, \boldsymbol{\theta}_j, \boldsymbol{\pi})$$

El segundo es el número esperado de estudiantes que son clasificados en las clases $\boldsymbol{\alpha}_c$ si responden correctamente el ítem j .

$$R_{jc} = \sum_{i=1}^N x_{ij} \times P(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_c \mid \mathbf{X}_i = \mathbf{x}_i, \boldsymbol{\theta}_j, \boldsymbol{\pi})$$

2. Paso M (maximización).

- a) El vector de parámetros de los ítems $\boldsymbol{\theta}$ es actualizado. Las ecuaciones de estimación son obtenidas mediante el cálculo de la primera derivada de la log-verosimilitud esperada con respecto a los parámetros de los ítems cuando son iguales a cero. Esta derivada solo toma en cuenta los dos valores esperados obtenidos en el paso E.

Entonces, sea

$$T_j^{(0)} = \sum_{c=1}^C T_{jc} \times (1 - \eta_{cj})$$

el número esperado de estudiantes con la falta de al menos una de las habilidades requeridas para dominar el ítem j ($\eta_{cj} = 0$) y

$$R_j^{(0)} = \sum_{c=1}^C x_{ij} \times T_{jc} \times (1 - \eta_{cj})$$

es el número esperado de estudiantes entre $T_j^{(0)}$ quienes responden correctamente al ítem j . Además, si $T_j^{(1)}$ y $R_j^{(1)}$ tienen la misma interpretación excepto que ellos pertenecen a estudiantes quienes poseen todas las habilidades requeridas para un ítem j $\eta_{cj} = 1$. Basándonos en estas definiciones, los parámetros del ítem j son actualizados acorde a

$$\hat{g}_j = \frac{R_j^{(0)}}{T_j^{(0)}} \quad \text{y} \quad \hat{s}_j = 1 - \frac{R_j^{(1)}}{T_j^{(1)}}.$$

- b) La distribución de las clases $P(\alpha_i = \alpha_c \mid \mathbf{X}_i = \mathbf{x}_i, \theta_j, \pi)$ es actualizada. Por lo tanto, el número esperado η_c de los estudiantes en la clase latente α_c es calculado:

$$\eta_c = \sum_{i=1}^N P(\alpha_c \mid \mathbf{X} = \mathbf{x}_i; \theta)$$

Luego la distribución de las clases es actualizada por

$$P(\alpha_c \mid \pi) = \frac{\eta_c}{N}; \quad c = 1, \dots, C.$$

y las probabilidades de dominar las habilidades son definidas como

$$P(\alpha_k) = \sum_{c=1}^C \alpha_{ck} \times P(\alpha_c \mid \pi), \quad k = 1, \dots, K.$$

Ambos pasos funcionan alternadamente hasta lograr la convergencia. Esta convergencia es lograda si el máximo cambio entre el valor del parámetro estimado, o el cambio relativo en la devianza, está por debajo de un valor específico predefinido, o después de un número máximo de iteraciones.

Finalmente, la matriz de covarianza para los parámetros estimados $\hat{\theta}$ y $\hat{\pi}$; es decir, los errores estándar empíricos son calculados usando la estrategia *empírica de producto cruzado*; el cual como mencionan Paek y Cai (2014) funciona mejor si el tamaño de muestra es más grande que la cantidad de ítems en la prueba.

Todos estos métodos de estimación desde el enfoque clásico se encuentran desarrollados gracias a Robitzsch et al. (2016) en el paquete CDM del software libre R.

3.4. Estimación mediante el algoritmo de Metropolis-Hastings

El algoritmo de Metropolis-Hastings (M-H) se suele utilizar en situaciones donde la distribución conjugada a posteriori no está disponible, y en particular, las distribuciones condicionales completas de los parámetros no tienen una forma estándar definida, lo cual complicaría realizar la estimación mediante el muestreador de Gibbs que veremos en el siguiente capítulo.

Metropolis-Hastings es un método genérico de aproximación a la distribución posteriori mediante cualquier combinación de distribuciones a prioris propuestas y muestreo del modelo (Hoff, 2009).

El algoritmo se puede presentar de la siguiente manera (Flores, 2012):

Sea $f(x)$ una distribución objetivo para la que se desea generar una muestra de tamaño T . El algoritmo M-H genera iterativamente los valores $x^{(t)}$ de esta distribución como:

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

Independientemente de la distribución q elegida el algoritmo M-H debería de converger a su distribución en equilibrio. Sin embargo, en la práctica, la elección de q es importante dado que elecciones pobres demoran la convergencia a la distribución en equilibrio. El éxito de este algoritmo depende de no tener un ratio de aceptación para los valores propuestos demasiado bajos. Como lo menciona Gamerman y Lopes (2006), se han realizado diversos estudios de optimización para definir “ratios muy bajos”, entre los cuales Bennett, Racine-Poon y Wakefield en 1995; Besag et al. en 1995 y otros autores parecen indicar que el ratio de aceptación debería estar entre 20 % y 50 %. Culpepper (2015a) menciona que en estudios de simulación para este modelo se necesita de un considerable tiempo de ensayo y error para poder definir los valores de la distribución propuesta y alcanzar a lo más este rango del 20 % y el 50 % en ratio de aceptación. Esto se hace aún más crítico mientras el modelo tenga más dimensiones.

En ese sentido y tomando en cuenta que el modelo DINA presenta múltiples dimensiones (habilidades) y parámetros a ser estimados, el optar por este algoritmo hará que el proceso de búsqueda y calibración de candidatos para la estimación se vuelva tedioso y poco práctico para ser aplicado en diferentes bases de datos y estudios de simulación con diferentes características como se presentará más adelante en este trabajo.

3.5. Estimación mediante el muestreador de Gibbs

A lo largo de este subcapítulo se desarrollará secuencialmente el proceso de estimación bajo este método. Se empezará resumiendo los subíndices con los que hemos trabajado en el capítulo anterior, los vectores y la distribución muestral de los datos; así como especificar las distribuciones a priori propuestas por Culpepper (2015a), utilizar la función de verosimilitud (2.4) y mostrar las distribuciones condicionales completas a partir de la posteriori.

Este método es un caso especial del algoritmo M-H usando como función de densidad propuesta $q(\theta' | \theta^{(t)})$ a la distribución a posteriori condicional completa $f(\theta_j | \boldsymbol{\theta}_{-j})$, donde $\boldsymbol{\theta}_{-j} = (\theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_d)^T$. Estas distribuciones propuestas tienen una probabilidad de aceptación $\alpha = 1$, por lo tanto, serán aceptadas en todas las iteraciones.

Una de las ventajas del muestreador de Gibbs es que en cada paso se deben generar valores aleatorios de distribuciones unidimensionales, las cuales frecuentemente tienen una forma conocida, es así que se pueden generar fácilmente números aleatorios usando funciones estándar en softwares estadísticos y computacionales.

El algoritmo puede ser resumido en los siguientes pasos (Flores, 2012):

1. Dar valores iniciales a $x^{(0)}$.
2. Para $t = 1, \dots, T$ repetir los siguientes pasos:
 - Fijar $\theta = \theta^{(t-1)}$.
 - Para $j = 1, \dots, d$ actualizar la componente θ_j tomando una muestra de $f(\cdot | \boldsymbol{\theta}_{-j})$.
 - $\theta^{(t)} = \theta$ y guardarlos como un conjunto de valores generados en la iteración $t + 1$.

En función a lo descrito, trabajaremos con N individuos, donde cada individuo $i \in \{1, \dots, N\}$ responde a J ítems, de acuerdo a las K habilidades que pueda poseer, y que de acuerdo a ello estará clasificado bajo una probabilidad de pertenencia en cualquiera de las C clases.

Asimismo, si conociéramos los vectores, $\boldsymbol{\alpha}_i = (\alpha_1, \dots, \alpha_N)$ que describe los estados de conocimiento de los individuos; $\boldsymbol{\theta}_j = (s_j, g_j)$ que implica el conjunto de los dos parámetros a estimar para los ítems; y $\boldsymbol{\pi} = (\pi_1, \dots, \pi_C)$ que se refiere a la probabilidad de pertenencia de un individuo. Recordemos que por (2.4) la función de verosimilitud del modelo viene dada por:

$$\begin{aligned}
 L(\boldsymbol{\alpha}, \boldsymbol{\pi}, \boldsymbol{\theta}) &= \prod_{i=1}^N \left\{ \sum_{c=1}^C \pi_c \left[\prod_{j=1}^J P_{jc}^{x_{ij}} (1 - P_{jc})^{1-x_{ij}} \right] \right\} \\
 &= \prod_{i=1}^N \sum_{c=1}^C \pi_c \prod_{j=1}^J \left[(1 - s_j)^{\eta_{ij}} g_j^{1-\eta_{ij}} \right]^{x_{ij}} \left[s_j^{1-\eta_{ij}} (1 - g_j)^{\eta_{ij}} \right]^{1-x_{ij}}
 \end{aligned} \quad (3.3)$$

3.5.1. Distribución a priori

Se presentan las distribuciones priori, según las especificaciones del modelo. Estas han sido tomadas de Culpepper (2015a):

- Para los vectores o estados de conocimiento de cada individuo $\boldsymbol{\alpha}_i$, se presenta la siguiente distribución categórica, la cual fue definida en (3.2):

$$P(\boldsymbol{\alpha}_i | \boldsymbol{\pi}) = \prod_{c=1}^C \pi_c^{I(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_c)} \quad (3.4)$$

Entonces podemos decir que $\boldsymbol{\alpha}_i$ tiene una priori categórica con probabilidad de pertenencia a la clase c igual a $\pi_c = P(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_c)$.

- Para el vector de las probabilidades de pertenencia a las clases $\boldsymbol{\pi}$:

$$\boldsymbol{\pi} \propto \text{Dir}(\boldsymbol{\delta}_0) \quad (3.5)$$

donde $\boldsymbol{\delta}_0 = (\delta_{01}, \dots, \delta_{0C})$ es un vector de hiperparámetros.

- Para los parámetros de los ítems s_j y g_j expresado conjuntamente con $\boldsymbol{\theta}_j$:

$$P(\boldsymbol{\theta}_j | \boldsymbol{\alpha}, \boldsymbol{\pi}) = P(s_j, g_j) \propto s_j^{a_s-1} (1-s_j)^{b_s-1} g_j^{a_g-1} (1-g_j)^{b_g-1} I((s_j, g_j) \in \mathcal{P}) \quad (3.6)$$

Por lo que podemos decir que la distribución priori de $\boldsymbol{\theta}_j$ es un producto de dos densidades Beta, con hiperparámetros a_s y b_s para s_j ; y a_g y b_g para g_j . Esto se da puesto que Junker y Sijtsma (2001) presentaron la restricción de monotonocidad $0 \leq s_j + g_j < 1$, donde $\mathcal{P} = \{(s, g) : 0 \leq s + g < 1, 0 \leq s < 1, 0 \leq g < 1\}$ y I representa la función indicadora.

A partir de las postulaciones previas (3.4), (3.5) y (3.6); la distribución a priori conjunta se define por:

$$\begin{aligned} P(\boldsymbol{\alpha}, \boldsymbol{\pi}, \boldsymbol{\theta}) &= P(\boldsymbol{\theta} | \boldsymbol{\alpha}, \boldsymbol{\pi}) P(\boldsymbol{\alpha} | \boldsymbol{\pi}) P(\boldsymbol{\pi}) \\ &\propto \prod_{i=1}^N \left[\prod_{c=1}^C \pi_c^{I(\alpha_i = \alpha_c)} \right] \times \left[\pi_1^{\delta_{01}-1} \times \pi_2^{\delta_{02}-1} \times \dots \times \pi_C^{\delta_{0C}-1} \right] \times \\ &\quad \left[\prod_{j=1}^J s_j^{a_s-1} (1-s_j)^{b_s-1} g_j^{a_g-1} (1-g_j)^{b_g-1} I((s_j, g_j) \in \mathcal{P}) \right] \end{aligned} \quad (3.7)$$

3.5.2. Distribución a posteriori

De las ecuaciones (3.3) y (3.7) se plantea la distribución a posteriori. Según el teorema de Bayes:

$$\begin{aligned} P(\boldsymbol{\alpha}, \boldsymbol{\pi}, \boldsymbol{\theta} | \boldsymbol{x}) &= \frac{P(\boldsymbol{x} | \boldsymbol{\alpha}, \boldsymbol{\pi}, \boldsymbol{\theta}) P(\boldsymbol{\alpha}, \boldsymbol{\pi}, \boldsymbol{\theta})}{P(\boldsymbol{x})} \\ P(\boldsymbol{\alpha}, \boldsymbol{\pi}, \boldsymbol{\theta} | \boldsymbol{x}) &\propto L(\boldsymbol{\alpha}, \boldsymbol{\pi}, \boldsymbol{\theta}) P(\boldsymbol{\alpha}, \boldsymbol{\pi}, \boldsymbol{\theta}) \\ P(\boldsymbol{\alpha}, \boldsymbol{\pi}, \boldsymbol{\theta} | \boldsymbol{x}) &\propto \prod_{i=1}^N \sum_{c=1}^C \pi_c \prod_{j=1}^J \left[(1-s_j)^{\eta_{ij}} g_j^{1-\eta_{ij}} \right]^{x_{ij}} \left[s_j^{1-\eta_{ij}} (1-g_j)^{\eta_{ij}} \right]^{1-x_{ij}} \times \\ &\quad \prod_{i=1}^N \left[\prod_{c=1}^C \pi_c^{I(\alpha_i = \alpha_c)} \right] \times \left[\pi_1^{\delta_{01}-1} \times \pi_2^{\delta_{02}-1} \times \dots \times \pi_C^{\delta_{0C}-1} \right] \times \\ &\quad \left[\prod_{j=1}^J s_j^{a_s-1} (1-s_j)^{b_s-1} g_j^{a_g-1} (1-g_j)^{b_g-1} I((s_j, g_j) \in \mathcal{P}) \right] \end{aligned} \quad (3.8)$$

3.5.3. Distribuciones condicionales completas

De lo planteado en (3.8) se pueden calcular las distribuciones condicionales completas para cada uno de los parámetros, ignorando todos los demás parámetros que son constantes con respecto al parámetro elegido:

- Para α_i :

$$\begin{aligned}
 P(\alpha_i | \pi, \theta, \mathbf{x}) &= \frac{P(\alpha_i, \pi, \theta, \mathbf{x})}{P(\pi, \theta, \mathbf{x})} = \frac{P(\alpha_i, \pi, \theta | \mathbf{x})P(\mathbf{x})}{P(\pi, \theta, \mathbf{x})} \\
 P(\alpha_i | \pi, \theta, \mathbf{x}) &\propto P(\alpha_i, \pi, \theta | \mathbf{x}) \\
 P(\alpha_i | \pi, \theta, \mathbf{x}) &\propto \prod_{c=1}^C \pi_c^{I(\alpha_i = \alpha_c)} \tag{3.9}
 \end{aligned}$$

Más concretamente

$$P(\alpha_i = \alpha_c | \pi, \theta, \mathbf{x}) = \frac{\pi_c P(\mathbf{x} | \alpha_i = \alpha_c, \theta)}{\sum_{c=1}^C \pi_c P(\mathbf{x} | \alpha_i = \alpha_c, \theta)} \tag{3.10}$$

Entonces:

$$P(\alpha_i | \pi, \theta, \mathbf{x}) \propto \prod_{c=1}^C \left\{ \frac{\pi_c P(\mathbf{x} | \alpha_i = \alpha_c, \theta)}{\sum_{c=1}^C \pi_c P(\mathbf{x} | \alpha_i = \alpha_c, \theta)} \right\}^{I(\alpha_i = \alpha_c)} \tag{3.11}$$

para $i = \{1, \dots, N\}$. Por ejemplo, para el primer individuo:

$$P(\alpha_1 | \pi, \theta, \mathbf{x}) \propto \prod_{c=1}^C \left\{ \frac{\pi_c P(\mathbf{x} | \alpha_1 = \alpha_c, \theta)}{\sum_{c=1}^C \pi_c P(\mathbf{x} | \alpha_1 = \alpha_c, \theta)} \right\}^{I(\alpha_1 = \alpha_c)} \tag{3.12}$$

Asumiendo que $K = 1$ y $C = 2$:

$$\begin{aligned}
 P(\alpha_i | \pi, \theta, \mathbf{x}) &\propto \left[\frac{\pi_1 P(x_1 | \alpha_1 = \alpha_1, \theta)}{\pi_1 P(x_1 | \alpha_1 = \alpha_1, \theta) + \pi_2 P(x_1 | \alpha_1 = \alpha_2, \theta)} \right]^{I(\alpha_1 = \alpha_1)} \\
 &\quad \times \left[\frac{\pi_2 P(x_i | \alpha_1 = \alpha_2, \theta)}{\sum_{c=1}^2 \pi_c P(x_1 | \alpha_1 = \alpha_c, \theta)} \right]^{I(\alpha_1 = \alpha_2)}
 \end{aligned}$$

de la misma manera para cada individuo $i = 1, \dots, N$.

- Para π :

$$\begin{aligned}
 P(\pi | \alpha, \theta, \mathbf{x}) &= \frac{P(\alpha, \pi, \theta, \mathbf{x})}{P(\alpha, \theta, \mathbf{x})} = \frac{P(\alpha, \pi, \theta | \mathbf{x})P(\mathbf{x})}{P(\alpha, \theta, \mathbf{x})} \\
 P(\pi | \alpha, \theta, \mathbf{x}) &\propto P(\alpha, \pi, \theta | \mathbf{x}) \\
 P(\pi | \alpha, \theta, \mathbf{x}) &\propto \prod_{i=1}^N \left[\prod_{c=1}^C \pi_c^{I(\alpha_i = \alpha_c)} \right] \pi_1^{\delta_{01}-1} \times \pi_2^{\delta_{02}-1} \times \dots \times \pi_C^{\delta_{0C}-1} \\
 &= \left[\prod_{c=1}^C \pi_c^{\sum_{i=1}^N I(\alpha_i = \alpha_c)} \right] \pi_1^{\delta_{01}-1} \times \dots \times \pi_C^{\delta_{0C}-1} \\
 &= \pi_1^{\sum_{i=1}^N I(\alpha_i = \alpha_1)} \times \pi_2^{\sum_{i=1}^N I(\alpha_i = \alpha_2)} \times \dots \times \pi_C^{\sum_{i=1}^N I(\alpha_i = \alpha_C)} \times \pi_1^{\delta_{01}-1} \times \dots \times \pi_C^{\delta_{0C}-1} \\
 &= \pi_1^{\delta_{01} + \sum_{i=1}^N I(\alpha_i = \alpha_1) - 1} \times \pi_2^{\delta_{02} + \sum_{i=1}^N I(\alpha_i = \alpha_2) - 1} \times \dots \times \pi_C^{\delta_{0C} + \sum_{i=1}^N I(\alpha_i = \alpha_C) - 1} \\
 &= \pi_1^{\delta_{01} + \tilde{N}_1 - 1} \times \pi_2^{\delta_{02} + \tilde{N}_2 - 1} \times \dots \times \pi_C^{\delta_{0C} + \tilde{N}_C - 1}
 \end{aligned}$$

donde $\tilde{N}_c = \sum_{i=1}^N I(\alpha_i = \alpha_c)$. Entonces se define: $\tilde{\mathbf{N}} = \{\tilde{N}_1, \dots, \tilde{N}_C\}$ y $\delta_0 = \{\delta_{01}, \dots, \delta_{0C}\}$.

Finalmente, la distribución completa de $\boldsymbol{\pi}$ será también una Dirichlet con vector parámetro $\boldsymbol{\delta}_0 + \widetilde{\mathbf{N}}$.

$$\boldsymbol{\pi} \mid \boldsymbol{\alpha}, \boldsymbol{\theta}, \mathbf{x} \propto \text{Dir}(\boldsymbol{\delta}_0 + \widetilde{\mathbf{N}}) \quad (3.13)$$

■ Para $\boldsymbol{\theta}_j$:

$$\begin{aligned} P(\boldsymbol{\theta}_j \mid \boldsymbol{\alpha}, \boldsymbol{\pi}, \mathbf{x}) &= \frac{P(\boldsymbol{\alpha}, \boldsymbol{\pi}, \boldsymbol{\theta}_j, \mathbf{x})}{P(\boldsymbol{\alpha}, \boldsymbol{\pi}, \mathbf{x})} = \frac{P(\boldsymbol{\alpha}, \boldsymbol{\pi}, \boldsymbol{\theta}_j \mid \mathbf{x})P(\mathbf{x})}{P(\boldsymbol{\alpha}, \boldsymbol{\pi}, \mathbf{x})} \\ P(\boldsymbol{\theta}_j \mid \boldsymbol{\alpha}, \boldsymbol{\pi}, \mathbf{x}) &\propto P(\boldsymbol{\alpha}, \boldsymbol{\pi}, \boldsymbol{\theta}_j \mid \mathbf{x}) \\ P(\boldsymbol{\theta}_j \mid \boldsymbol{\alpha}, \boldsymbol{\pi}, \mathbf{x}) &\propto \prod_{i=1}^N \left[(1-s_j)^{\eta_{ij}} g_j^{1-\eta_{ij}} \right]^{x_{ij}} \left[s_j^{\eta_{ij}} (1-g_j)^{1-\eta_{ij}} \right]^{1-x_{ij}} \\ &\quad s_j^{a_s-1} (1-s_j)^{b_s-1} g_j^{a_g-1} (1-g_j)^{b_g-1} \times I((s_j, g_j) \in \mathcal{P}) \\ &\quad \propto (1-s_j)^{\sum_{i|x_{ij}=1} \eta_{ij}} g_j^{\sum_{i|x_{ij}=1} (1-\eta_{ij})} s_j^{\sum_{i|x_{ij}=0} \eta_{ij}} (1-g_j)^{\sum_{i|x_{ij}=0} (1-\eta_{ij})} \\ &\quad s_j^{a_s-1} (1-s_j)^{b_s-1} g_j^{a_g-1} (1-g_j)^{b_g-1} \times I((s_j, g_j) \in \mathcal{P}) \end{aligned} \quad (3.14)$$

donde $\sum_{i|x_{ij}=1} \eta_{ij}$ quiere decir que sumará todos los η_{ij} cuando $x_{ij} = 1$; de la misma forma para $\sum_{i|x_{ij}=0} \eta_{ij}$ que sumará los η_{ij} cuando $x_{ij} = 0$.

De acuerdo a ello, sea:

$$\begin{aligned} \sum_{i=1}^N \eta_{ij} &= \sum_{i|x_{ij}=1} \eta_{ij} + \sum_{i|x_{ij}=0} \eta_{ij} \\ \tau_j &= \sum_{i|x_{ij}=1} \eta_{ij} + S_j \end{aligned} \quad (3.15)$$

$$\begin{aligned} \sum_{i=1}^N (1-\eta_{ij}) &= \sum_{i|x_{ij}=1} (1-\eta_{ij}) + \sum_{i|x_{ij}=0} (1-\eta_{ij}) \\ N - \tau_j &= \sum_{i|x_{ij}=0} (1-\eta_{ij}) + G_j \end{aligned} \quad (3.16)$$

Finalmente, reemplazando (3.15) y (3.16) en (3.14), la distribución condicional de $\boldsymbol{\theta}_j$:

$$\begin{aligned} P(\boldsymbol{\theta}_j \mid \boldsymbol{\alpha}, \boldsymbol{\pi}, \mathbf{x}) &= P(s_j, g_j \mid \boldsymbol{\alpha}, \boldsymbol{\pi}, \mathbf{x}) \propto (1-s_j)^{b_s+\tau_j-S_j-1} s_j^{a_s+S_j-1} (1-g_j)^{b_g+N-\tau_j-G_j-1} g_j^{a_g+G_j-1} \\ &\quad \propto (1-s_j)^{\widetilde{b}_s-1} g_j^{\widetilde{a}_g-1} s_j^{\widetilde{a}_s-1} (1-g_j)^{\widetilde{b}_g-1} \times I((s_j, g_j) \in \mathcal{P}) \end{aligned} \quad (3.17)$$

donde:

$$\begin{aligned} \widetilde{a}_s &= a_s + S_j \\ \widetilde{b}_s &= b_s + \tau_j - S_j \\ \widetilde{a}_g &= a_g + G_j \\ \widetilde{b}_g &= b_g + N - \tau_j - G_j \end{aligned}$$

Esta distribución sigue siendo el producto de dos densidades Beta pero restringidas.

Capítulo 4

Estudio de Simulación

En el presente capítulo se desarrollan ocho estudios de simulación donde se presentan combinaciones de valores para los parámetros: s , g y π ; en tres diferentes tamaños de población.

Se mencionarán dos tipos de relación entre los parámetros del ítem (s y g) y las habilidades de los individuos; la cual está referida al diagnóstico. Una alta relación o un fuerte impacto entre las habilidades y los parámetros de los ítems se dará cuando los valores de estos sean pequeños, ya que al reducir este ruido se generará una mejor evaluación del perfil latente que pueda obtener el evaluado.

Del mismo modo, si los valores de los parámetros del ítem son mayores, la relación se denomina débil, ya que el diagnóstico del evaluado estará influenciado por un mayor ruido en los ítems, haciendo que se genere un bajo impacto en la descripción del perfil latente de habilidades.

Asimismo, los valores del parámetro π tendrán una relación uniforme con los perfiles latentes de habilidades, cuando se coloque un valor como constante. Mientras que, se considerará una relación fuerte cuando se tenga una mayor probabilidad de que los individuos pertenezcan a una clase de mayor dominio de habilidades.

Todos los estudios bajo el muestreador de Gibbs se desarrollarán con $m = 50$ cadenas de 50000 iteraciones quemando las primeras 1000 en cada una y con saltos de 5.

A continuación se describen brevemente los estudios que se presentarán:

- El primero se enfoca en estudiar la precisión en la recuperación de los parámetros de los ítems (s y g) cuando estos tienen un impacto fuerte con las habilidades latentes; manteniendo una relación uniforme del parámetro π .
- El segundo se enfoca en estudiar la precisión en la recuperación de los parámetros de los ítems (s y g) cuando estos tienen un impacto débil con las habilidades latentes; manteniendo una relación uniforme del parámetro π .
- El tercero se enfoca en estudiar la precisión en la recuperación de los parámetros de los ítems cuando el parámetro de desliz (s) tiene un alto impacto, mientras que el de adivinación (g) una relación débil o baja con las habilidades latentes; manteniendo una relación uniforme del parámetro π .
- El cuarto se enfoca en estudiar la precisión en la recuperación de los parámetros de los ítems cuando el parámetro de desliz (s) tiene un bajo impacto, mientras que el de adivinación (g) una relación fuerte o de alto diagnóstico, con las habilidades latentes; manteniendo una relación uniforme del parámetro π .

- El quinto se enfoca en estudiar la precisión en la recuperación de los parámetros de los ítems (s y g) cuando estos tienen un impacto fuerte con las habilidades latentes; mientras que el parámetro π deja de ser constante, para mantener una relación fuerte. Se fija una probabilidad para la pertenencia a las clases conformadas por 2 o menos habilidades, y el doble de probabilidad para clases con 3 a más habilidades.
- El sexto se enfoca en estudiar la precisión en la recuperación de los parámetros de los ítems (s y g) cuando estos tienen un impacto débil con las habilidades latentes; mientras que el parámetro π deja de ser constante, para mantener una relación fuerte. Se fija una probabilidad para la pertenencia a las clases conformadas por 2 o menos habilidades, y el doble de probabilidad para clases con 3 a más habilidades.
- El séptimo se enfoca en estudiar la precisión en la recuperación de los parámetros de los ítems cuando el parámetro de desliz (s) tiene un alto impacto, mientras que el de adivinación (g) una relación débil o baja con las habilidades latentes; mientras que el parámetro π deja de ser constante, para mantener una relación fuerte. Se fija una probabilidad para la pertenencia a las clases conformadas por 2 o menos habilidades, y el doble de probabilidad para clases con 3 a más habilidades.
- El octavo se enfoca en estudiar la precisión en la recuperación de los parámetros de los ítems cuando el parámetro de desliz (s) tiene un bajo impacto, mientras que el de adivinación (g) una relación fuerte o de alto diagnóstico, con las habilidades latentes; mientras que el parámetro π deja de ser constante, para mantener una relación fuerte. Se fija una probabilidad para la pertenencia a las clases conformadas por 2 o menos habilidades, y el doble de probabilidad para clases con 3 a más habilidades.

Todos los estudios descritos son probados con tres tamaños de población ($N = 100$, $N = 300$ y $N = 500$).

Los escenarios que se plantearán para este estudio son guiados por los trabajos de [De la Torre y Douglas \(2004\)](#) y las adaptaciones de [Culpepper \(2015a\)](#), adecuándolos a nuestro contexto. Es por esto que se utilizarán como valores fijos $K = 4$ habilidades, $J = 48$ ítems y la matriz \mathbf{Q} de la evaluación de matemática para alumnos del segundo grado de secundaria del país ([Minedu, 2016](#)).

En resumen, los objetivos de este capítulo son los siguientes:

4.1. Objetivos

- Simular una matriz de respuestas dicotómicas, por cada repetición, generada a partir de las características de los parámetros que se presenten en cada escenario estudiado mediante el modelo DINA.
- Evaluar la precisión en la recuperación de los parámetros de los ítems (s y g) y las probabilidades de las clases latentes (π); mediante el método de estimación del muestreador de Gibbs.
- Evaluar la precisión en la recuperación de los parámetros de los ítems (s y g) y las probabilidades de las clases latentes (π); mediante el método de estimación del algoritmo

EM.

4.2. Algoritmo para simular datos

Para simular una matriz de respuestas dicotómicas a partir del modelo DINA se siguió el siguiente procedimiento:

- Definir la cantidad de ítems J , el número de habilidades K y la matriz \mathbf{Q} .
- Establecer los valores del tamaño de población, los parámetros de los ítems (s y g), y las probabilidades de pertenencia a cada clase latente ($\boldsymbol{\pi}$).
- Generar la matriz $\mathbf{A}_{N \times K}$ que representa los estados o perfiles de conocimiento de cada individuo i .
- Simular para cada uno de los N individuos la respuestas en cada uno de los J ítemes. Considerando los parámetros calculados en los puntos anteriores.

El código para simular la matriz de respuestas dicotómicas para el modelo DINA se muestra en el Apéndice A.

4.3. Método para estimar los parámetros

La estimación de los parámetros se realiza mediante la estimación MCMC usando dos enfoques. El primero a través de la inferencia bayesiana implementada en el paquete *dina* (Culpepper, 2015b) con el muestreador de Gibbs; mientras que el segundo a través del enfoque clásico usando el paquete *CDM* (Robitzsch et al., 2016) con el algoritmo EM.

Con estos dos enfoques se calcularán las medias estimadas de las distribuciones a posteriori de los parámetros a estimar. Se estiman al menos $m(2J + 2^K)$ parámetros en cada replicación de la simulación.

La implementación computacional para la inferencia bayesiana y clásica se encuentra también en el Apéndice A.

4.4. Criterios para evaluar la simulación

La precisión en la recuperación de los parámetros es evaluada usando la raíz del error cuadrático medio (RMSE) y el Sesgo. El RMSE es definido como la raíz cuadrada del promedio de las diferencias al cuadrado entre los valores reales y los estimados a lo largo de todas las m réplicas:

$$RMSE = \sqrt{\sum_{s=1}^m \frac{(\hat{\Theta}_s - \Theta)^2}{m}}, \quad (4.1)$$

donde Θ representa a los parámetros de los ítems (adivinación y desliz) como a las probabilidades de cada clase latente ($\boldsymbol{\pi}$). Mientras que el $\hat{\Theta}$ es el estimado del parámetro anterior y m representa el número de réplicas en la simulación.

Del mismo modo, el Sesgo es definido como:

$$Sesgo = \sum_{s=1}^m \frac{\hat{\Theta}_s - \Theta}{m}, \quad (4.2)$$

En base a lo descrito, se presentan los ocho estudios de simulación que se desarrollaron.

4.5. Estudio 1

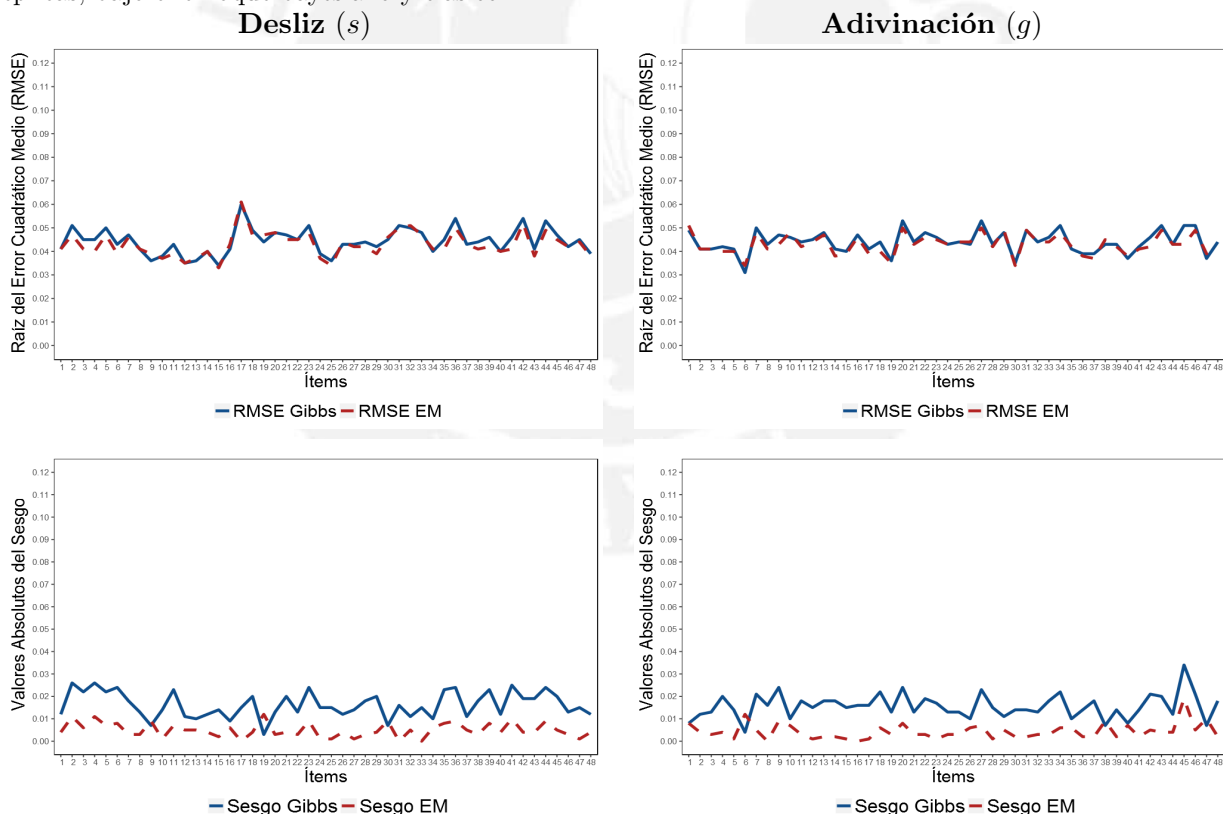
Para este estudio se presentan tres escenarios que se detallan en el cuadro 4.1, los cuales tienen como principal objetivo estudiar, bajo tres tamaños de población, la recuperación de los parámetros de los ítems cuando estos presentan una alta o fuerte relación con las habilidades de los individuos:

Cuadro 4.1: Escenarios del Estudio 1 de simulación para una relación alta entre los parámetros de los ítems y las habilidades de los individuos, manteniendo el parámetro π constante

Escenarios	Tamaño de población (N)	Adivinación (s)	Desliz (g)	Probabilidad de pertenencia (π)
1	100	0.1	0.1	0.0625
2	300	0.1	0.1	0.0625
3	500	0.1	0.1	0.0625

Se presenta en el cuadro 4.2 con gráficos de líneas, la lista de los 48 ítems con los valores de RMSE y el sesgo para cada parámetro de acuerdo a los dos enfoques estudiados, y cuando el tamaño de la población es $N = 100$:

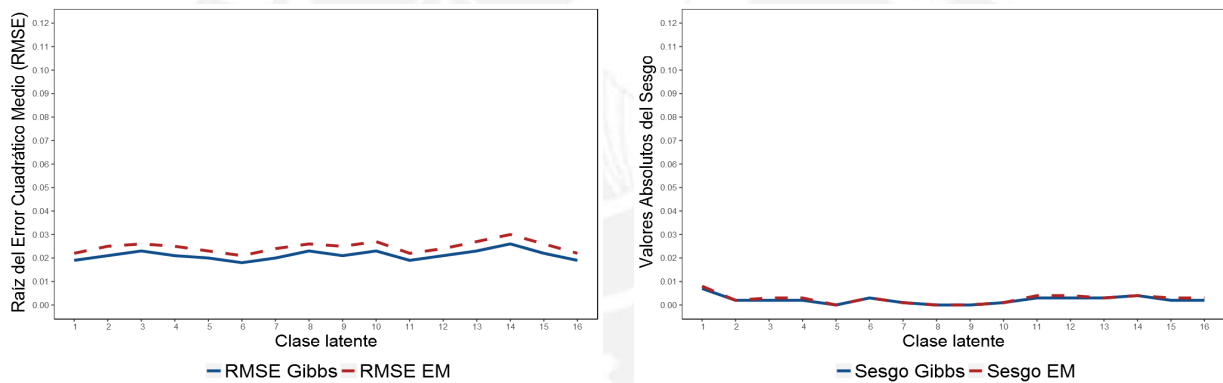
Cuadro 4.2: Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 1 con $N = 100$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico



En el siguiente cuadro 4.3 se presenta también la precisión de recuperación de parámetros para las probabilidades de pertenencia a las 2^K clases latentes, cuando se fija como $\pi = 2^{-K} = 0.0625$ y estimado a partir de los parámetros de los ítems y tamaño de población definidos anteriormente:

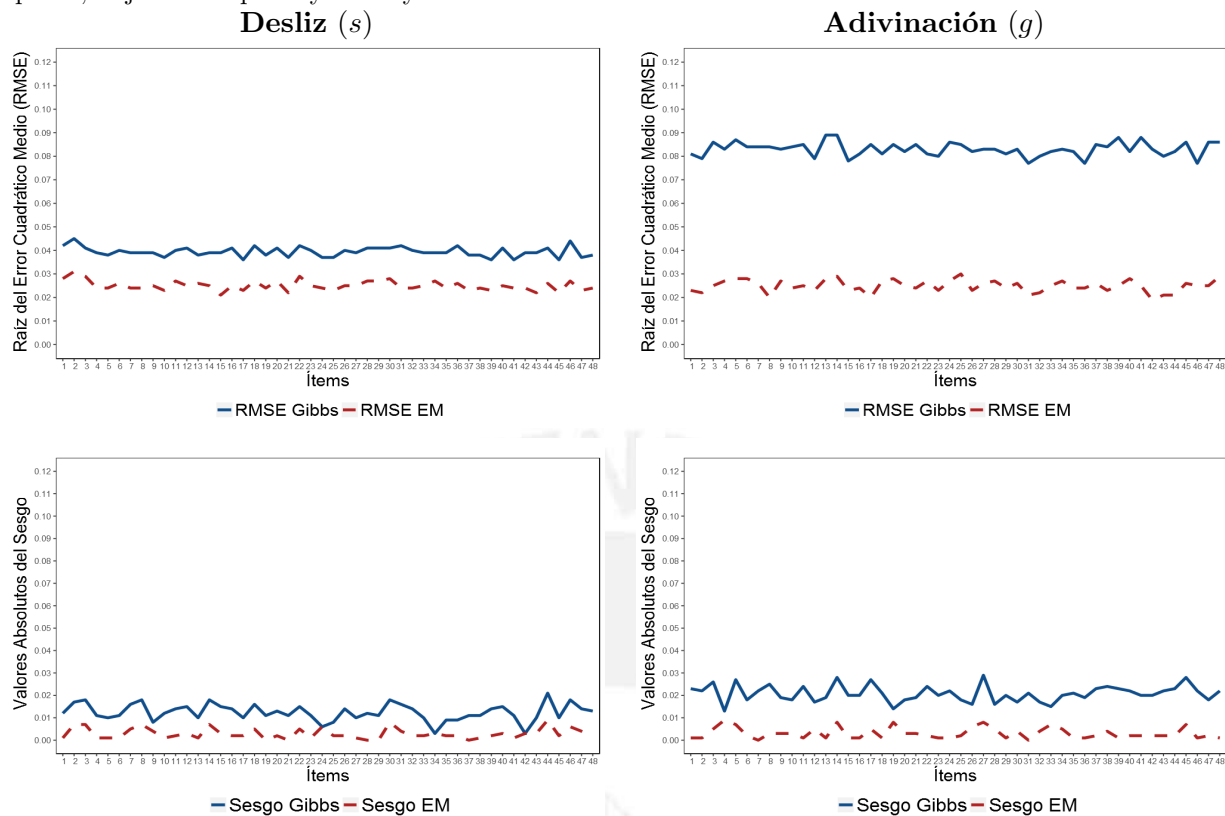
Cuadro 4.3: Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 1 con $N = 100$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico

Clases latentes	Perfil de habilidades				Muestreador de Gibbs	Algoritmo EM		
	α_1	α_2	α_3	α_4	RMSE	Sesgo	RMSE	Sesgo
1	0	0	0	0	0.019	-0.007	0.022	-0.008
2	1	0	0	0	0.021	0.002	0.025	0.002
3	0	1	0	0	0.023	0.002	0.026	0.003
4	0	0	1	0	0.021	-0.002	0.025	-0.003
5	0	0	0	1	0.02	0	0.023	0
6	1	1	0	0	0.018	-0.003	0.021	-0.003
7	1	0	1	0	0.02	-0.001	0.024	-0.001
8	1	0	0	1	0.023	0	0.026	0
9	0	1	1	0	0.021	0	0.025	0
10	0	1	0	1	0.023	0.001	0.027	0.001
11	0	0	1	1	0.019	0.003	0.022	0.004
12	1	1	1	0	0.021	0.003	0.024	0.004
13	1	1	0	1	0.023	0.003	0.027	0.003
14	1	0	1	1	0.026	0.004	0.03	0.004
15	0	1	1	1	0.022	-0.002	0.026	-0.003
16	1	1	1	1	0.019	-0.002	0.022	-0.003



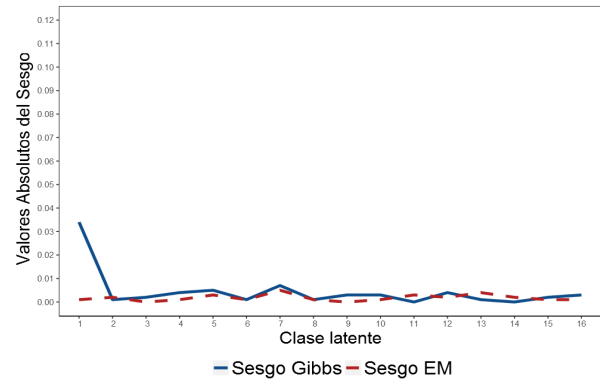
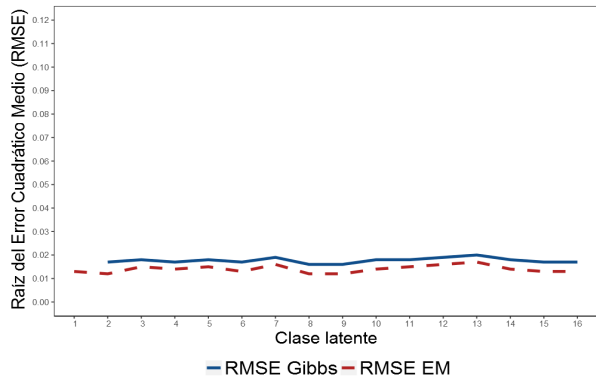
De la misma forma se presenta el resumen de resultados en los cuadros 4.4 y 4.5, para los parámetros de los ítems y las probabilidades de pertenencia con un tamaño de población $N = 300$.

Cuadro 4.4: Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 1 con $N = 300$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico



Cuadro 4.5: Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 1 con $N = 300$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico

Clases latentes	Perfil de habilidades				Muestreador de Gibbs		Algoritmo EM	
	α_1	α_2	α_3	α_4	RMSE	Sesgo	RMSE	Sesgo
1	0	0	0	0	0.178	0.034	0.013	-0.001
2	1	0	0	0	0.017	-0.001	0.012	0.002
3	0	1	0	0	0.018	-0.002	0.015	0
4	0	0	1	0	0.017	-0.004	0.014	-0.001
5	0	0	0	1	0.018	-0.005	0.015	-0.003
6	1	1	0	0	0.017	-0.001	0.013	0.001
7	1	0	1	0	0.019	-0.007	0.016	-0.005
8	1	0	0	1	0.016	-0.001	0.012	0.001
9	0	1	1	0	0.016	-0.003	0.012	0
10	0	1	0	1	0.018	-0.003	0.014	-0.001
11	0	0	1	1	0.018	0	0.015	0.003
12	1	1	1	0	0.019	-0.004	0.016	-0.002
13	1	1	0	1	0.02	0.001	0.017	0.004
14	1	0	1	1	0.018	0	0.014	0.002
15	0	1	1	1	0.017	-0.002	0.013	0.001
16	1	1	1	1	0.017	-0.003	0.013	-0.001

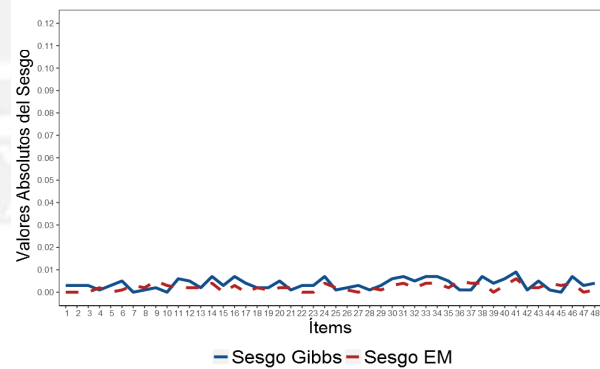
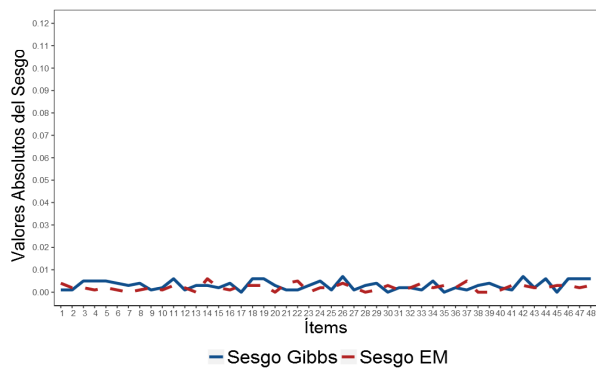
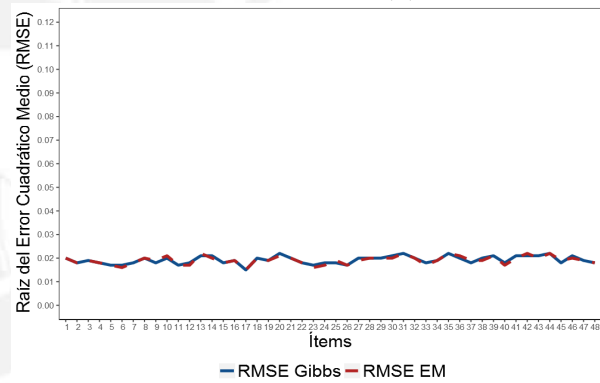
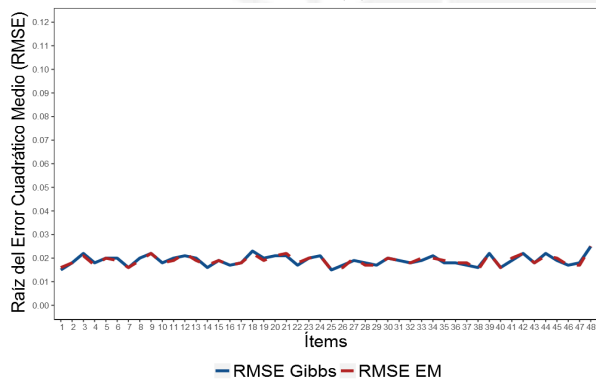


Finalmente, se presenta el resumen de resultados para el último escenario, en los cuadros 4.6 y 4.7, para los parámetros de los ítems y las probabilidades de pertenencia con un tamaño de población $N = 500$.

Cuadro 4.6: Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 1 con $N = 500$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico

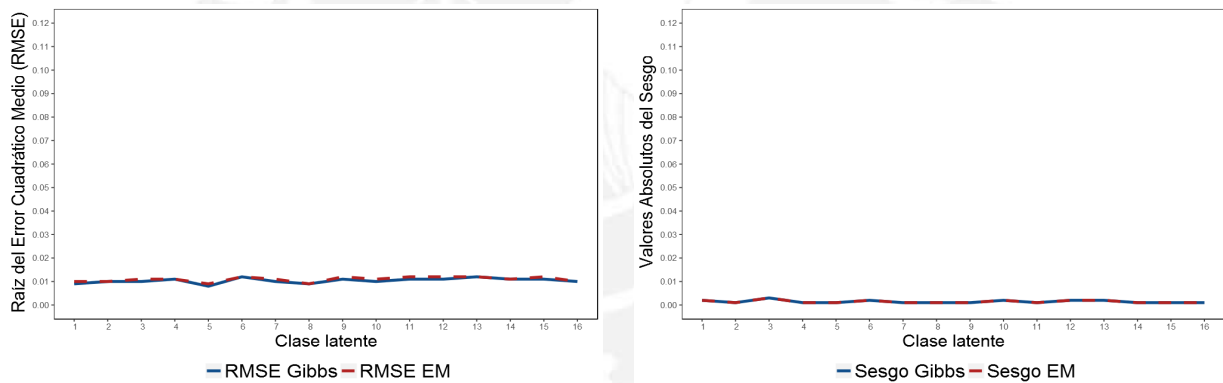
Desliz (s)

Adivinación (g)



Cuadro 4.7: Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 1 con $N = 500$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico

Clases latentes	Perfil de habilidades				Muestreador de Gibbs		Algoritmo EM	
	α_1	α_2	α_3	α_4	RMSE	Sesgo	RMSE	Sesgo
1	0	0	0	0	0.009	-0.002	0.01	-0.002
2	1	0	0	0	0.01	0.001	0.01	0.001
3	0	1	0	0	0.01	-0.003	0.011	-0.003
4	0	0	1	0	0.011	0.001	0.011	0.001
5	0	0	0	1	0.008	0.001	0.009	0.001
6	1	1	0	0	0.012	0.002	0.012	0.002
7	1	0	1	0	0.01	0.001	0.011	0.001
8	1	0	0	1	0.009	0.001	0.009	0.001
9	0	1	1	0	0.011	0.001	0.012	0.001
10	0	1	0	1	0.01	0.002	0.011	0.002
11	0	0	1	1	0.011	-0.001	0.012	-0.001
12	1	1	1	0	0.011	0.002	0.012	0.002
13	1	1	0	1	0.012	-0.002	0.012	-0.002
14	1	0	1	1	0.011	-0.001	0.011	-0.001
15	0	1	1	1	0.011	-0.001	0.012	-0.001
16	1	1	1	1	0.01	-0.001	0.01	-0.001



De acuerdo a lo observado en los tres escenarios, se presentan los cuadros 4.8 y 4.9 que resume lo hallado por medio de comparación de medias de los valores RMSE y el Sesgo (en valores absolutos), respectivamente:

Cuadro 4.8: Resumen de medias de los valores RMSE para todos los parámetros estudiados

Escenarios	Tamaño de la población (N)	Desliz (s)		Adivinación (g)		Prob de cada clase latente (π)	
		Gibbs	EM	Gibbs	EM	Gibbs	EM
1	100	0.044	0.043	0.044	0.043	0.021	0.025
2	300	0.039	0.025	0.081	0.025	0.028	0.014
3	500	0.019	0.019	0.019	0.019	0.010	0.011

Cuadro 4.9: Resumen de medias de los valores del sesgo para todos los parámetros estudiados

Escenarios	Tamaño de la población (N)	Desliz (s)		Adivinación (g)		Prob de cada clase latente (π)	
		Gibbs	EM	Gibbs	EM	Gibbs	EM
1	100	0.016	0.005	0.016	0.004	0.002	0.003
2	300	0.013	0.003	0.020	0.003	0.004	0.002
3	500	0.003	0.002	0.004	0.002	0.001	0.001

Cuando los parámetros de los ítems, desliz (s) y adivinación (g), tienen una alta o fuerte influencia en las habilidades de los alumnos; manteniendo constante la distribución de los individuos en todas las clases latentes; es decir, manteniendo uniforme el parámetro π . Se puede decir que ambos métodos producen estimaciones insesgadas para los tres parámetros en estudio puesto que el sesgo en promedio es cero; y además, los valores del RMSE son muy bajos.

Sin embargo, se revisa a detalle cada uno de los escenarios, cuando la población es pequeña ($N = 100$), el algoritmo EM muestra estimaciones más cercanas a los parámetros de los ítems. Mientras que, cuando analizamos los valores del RMSE casi no se encuentran diferencias entre ambos métodos para ellos; sin embargo, hay una ventaja a favor del muestreador de Gibbs en la estimación de π .

Por otro lado, cuando el tamaño de la población es mediano ($N = 300$), el algoritmo EM es claramente mejor con respecto al muestreador de Gibbs, con ambos indicadores evaluados. Aunque para el parámetro π las diferencias entre ambos enfoques son muy pequeñas.

Finalmente, cuando la población es grande ($N = 500$), no existen diferencias, en promedio, entre un enfoque respecto al otro.

4.6. Estudio 2

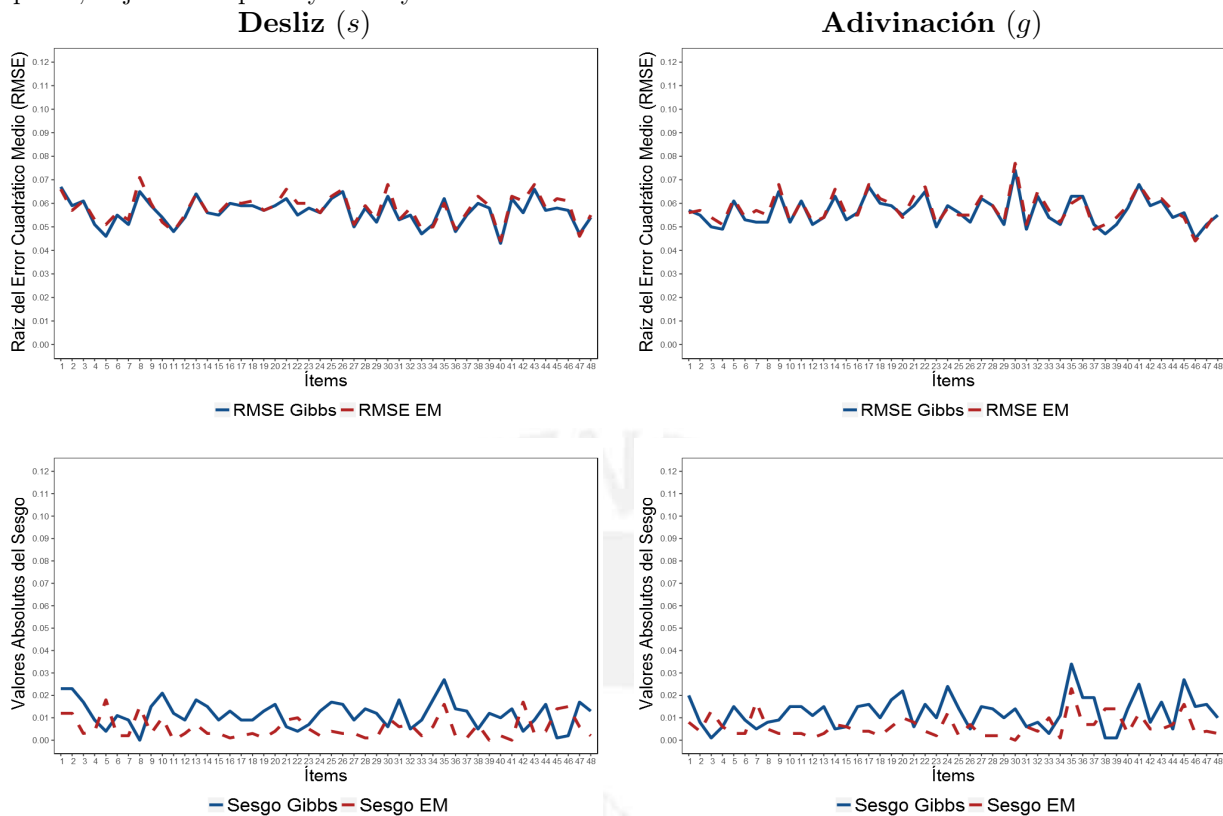
Para este estudio se presentan tres escenarios que se detallan en el cuadro 4.10, los cuales tienen como principal objetivo estudiar, bajo tres tamaños de población, la recuperación de los parámetros de los ítems cuando estos presentan una baja relación con las habilidades de los individuos:

Cuadro 4.10: Escenarios del Estudio 2 de simulación para una relación alta entre los parámetros de los ítems y las habilidades de los individuos, manteniendo el parámetro π constante

Escenarios	Tamaño de población (N)	Adivinación (s)	Desliz (g)	Probabilidad de pertenencia (π)
1	100	0.2	0.2	0.0625
2	300	0.2	0.2	0.0625
3	500	0.2	0.2	0.0625

Se presenta en el cuadro 4.11 con gráficos de líneas, la lista de los 48 ítems con los valores de RMSE y sesgo para cada parámetro y con tamaño de población $N = 100$, de acuerdo a los dos enfoques estudiados:

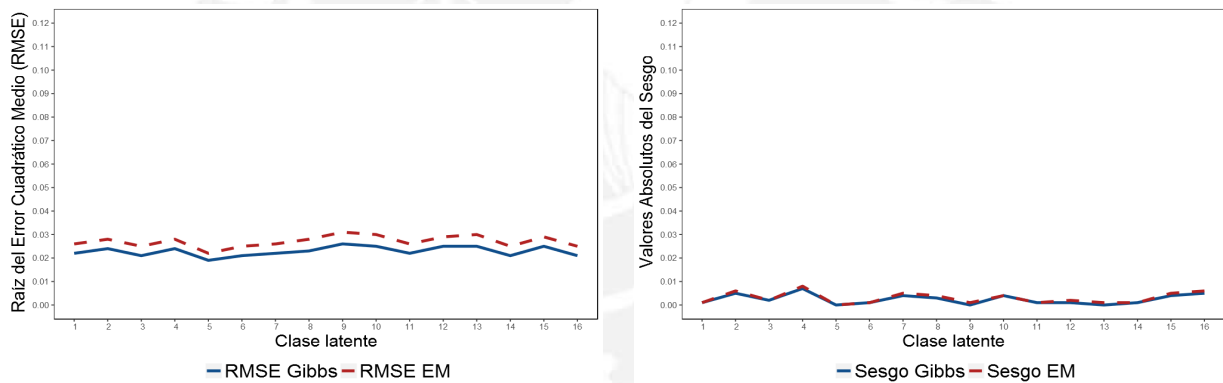
Cuadro 4.11: Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 2 con $N = 100$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico



En el siguiente cuadro 4.12 se presenta también la precisión de recuperación de parámetros para las probabilidades de pertenencia a las C clases latentes, cuando se fija como $\pi = 2^{-K}$ y estimado a partir de los parámetros de los ítems definidos en el cuadro 4.10, con un impacto débil en las habilidades y con tamaño de población $N = 100$:

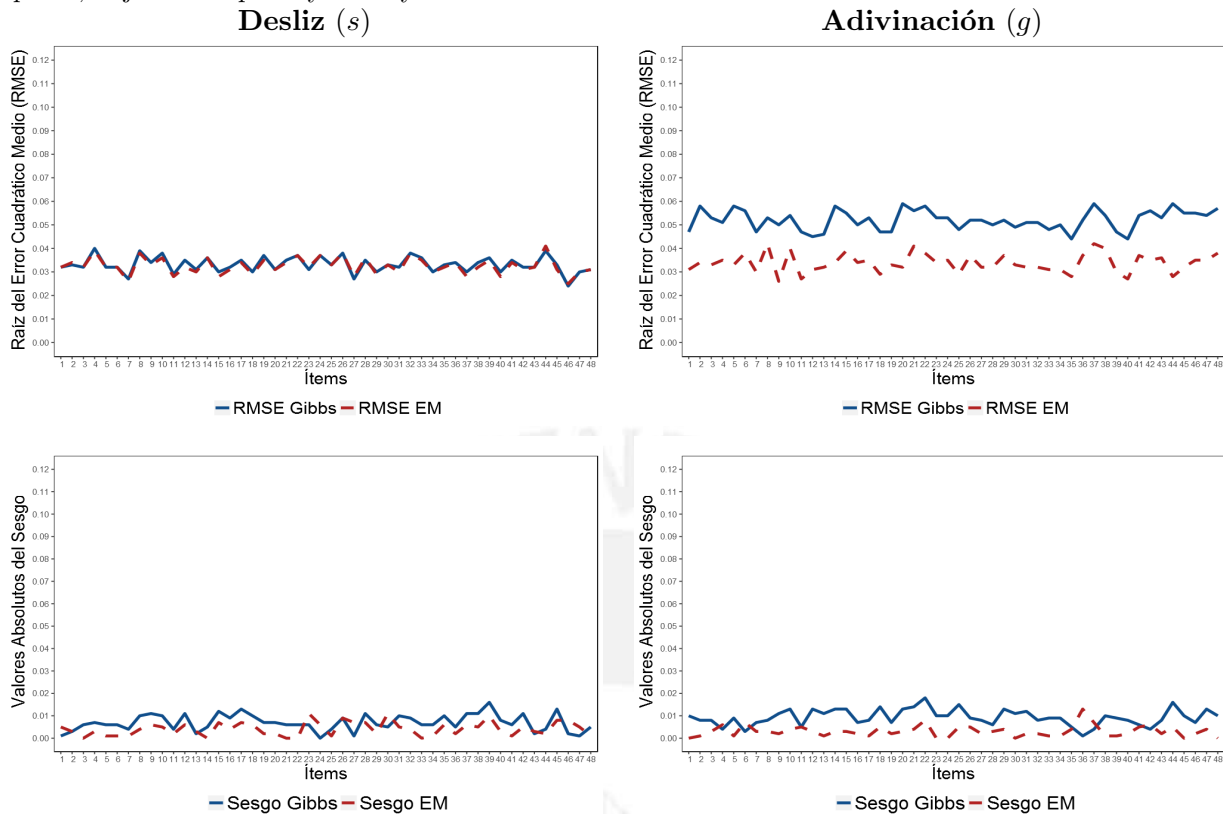
Cuadro 4.12: Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 2 con $N = 100$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico

Clases latentes	Perfil de habilidades				Muestreador de Gibbs	Algoritmo EM		
	α_1	α_2	α_3	α_4	RMSE	Sesgo	RMSE	Sesgo
1	0	0	0	0	0.022	-0.001	0.026	-0.001
2	1	0	0	0	0.024	0.005	0.028	0.006
3	0	1	0	0	0.021	-0.002	0.025	-0.002
4	0	0	1	0	0.024	-0.007	0.028	-0.008
5	0	0	0	1	0.019	0	0.022	0
6	1	1	0	0	0.021	0.001	0.025	0.001
7	1	0	1	0	0.022	0.004	0.026	0.005
8	1	0	0	1	0.023	0.003	0.028	0.004
9	0	1	1	0	0.026	0	0.031	0.001
10	0	1	0	1	0.025	-0.004	0.03	-0.004
11	0	0	1	1	0.022	0.001	0.026	0.001
12	1	1	1	0	0.025	-0.001	0.029	-0.002
13	1	1	0	1	0.025	0	0.03	-0.001
14	1	0	1	1	0.021	0.001	0.025	0.001
15	0	1	1	1	0.025	-0.004	0.029	-0.005
16	1	1	1	1	0.021	0.005	0.025	0.006



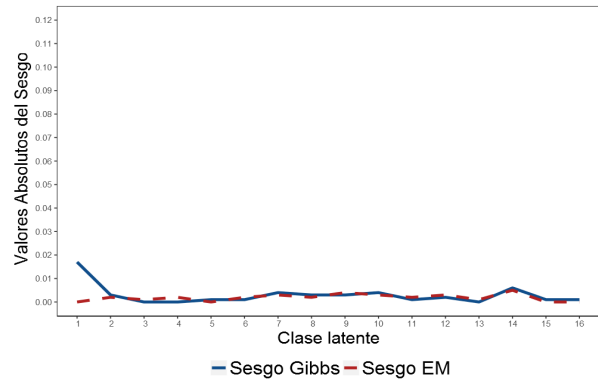
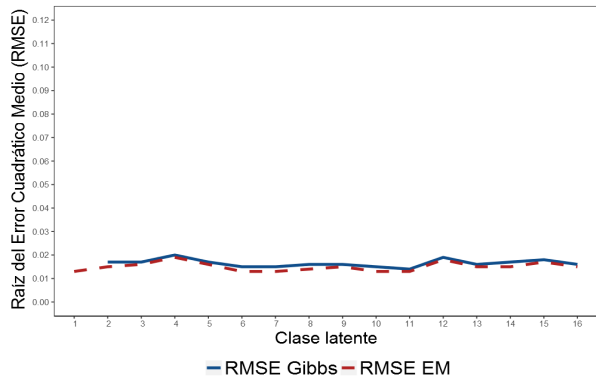
De la misma forma se presenta el resumen de resultados en los cuadros 4.13 y 4.14, para los parámetros de los ítems y las probabilidades de pertenencia con un tamaño de población $N = 300$.

Cuadro 4.13: Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 2 con $N = 300$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico



Cuadro 4.14: Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 2 con $N = 300$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico

Clases latentes	Perfil de habilidades				Muestreador de Gibbs		Algoritmo EM	
	α_1	α_2	α_3	α_4	RMSE	Sesgo	RMSE	Sesgo
1	0	0	0	0	0.127	0.017	0.013	0
2	1	0	0	0	0.017	-0.003	0.015	-0.002
3	0	1	0	0	0.017	0	0.016	0.001
4	0	0	1	0	0.02	0	0.019	0.002
5	0	0	0	1	0.017	-0.001	0.016	0
6	1	1	0	0	0.015	0.001	0.013	0.002
7	1	0	1	0	0.015	-0.004	0.013	-0.003
8	1	0	0	1	0.016	-0.003	0.014	-0.002
9	0	1	1	0	0.016	0.003	0.015	0.004
10	0	1	0	1	0.015	-0.004	0.013	-0.003
11	0	0	1	1	0.014	0.001	0.013	0.002
12	1	1	1	0	0.019	0.002	0.018	0.003
13	1	1	0	1	0.016	0	0.015	0.001
14	1	0	1	1	0.017	-0.006	0.015	-0.005
15	0	1	1	1	0.018	-0.001	0.017	0
16	1	1	1	1	0.016	-0.001	0.015	0

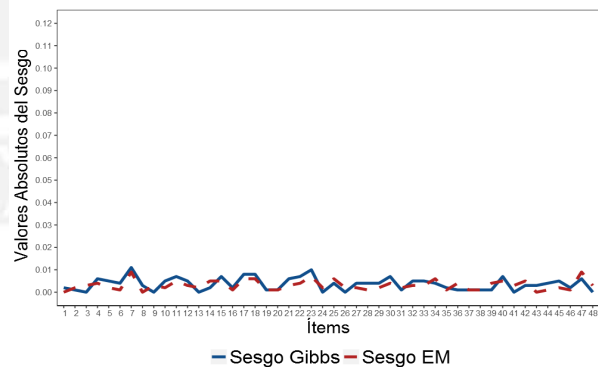
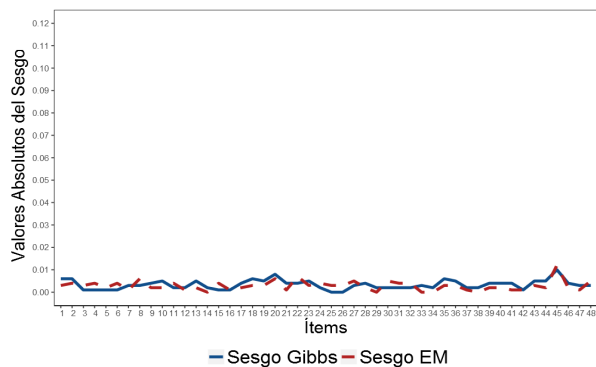
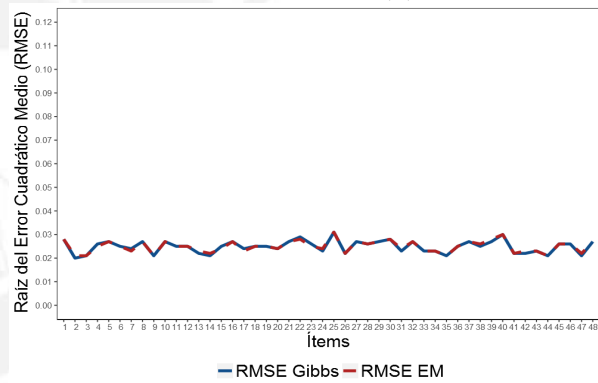
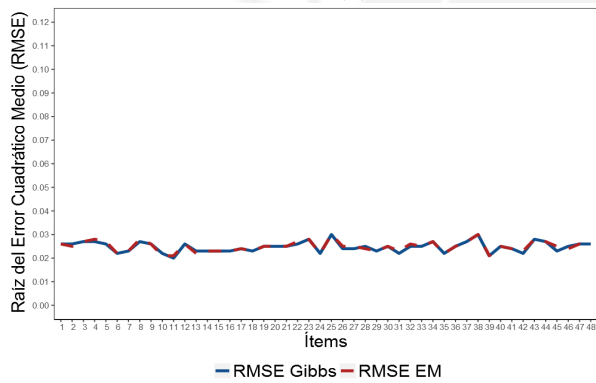


Finalmente, se presenta el resumen de resultados para el último escenario, en los cuadros 4.15 y 4.16, para los parámetros de los ítems y las probabilidades de pertenencia con un tamaño de población $N = 500$.

Cuadro 4.15: Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 2 con $N = 500$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico

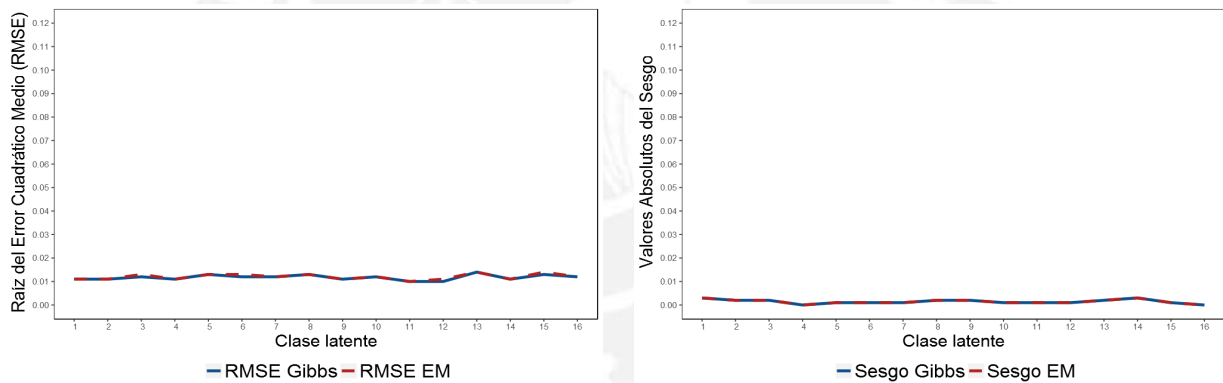
Desliz (s)

Adivinación (g)



Cuadro 4.16: Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 2 con $N = 500$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico

Clases latentes	Perfil de habilidades				Muestreador de Gibbs		Algoritmo EM	
	α_1	α_2	α_3	α_4	RMSE	Sesgo	RMSE	Sesgo
1	0	0	0	0	0.011	-0.003	0.011	-0.003
2	1	0	0	0	0.011	-0.002	0.011	-0.002
3	0	1	0	0	0.012	0.002	0.013	0.002
4	0	0	1	0	0.011	0	0.011	0
5	0	0	0	1	0.013	-0.001	0.013	-0.001
6	1	1	0	0	0.012	0.001	0.013	0.001
7	1	0	1	0	0.012	0.001	0.012	0.001
8	1	0	0	1	0.013	0.002	0.013	0.002
9	0	1	1	0	0.011	-0.002	0.011	-0.002
10	0	1	0	1	0.012	0.001	0.012	0.001
11	0	0	1	1	0.01	-0.001	0.01	-0.001
12	1	1	1	0	0.01	0.001	0.011	0.001
13	1	1	0	1	0.014	0.002	0.014	0.002
14	1	0	1	1	0.011	0.003	0.011	0.003
15	0	1	1	1	0.013	-0.001	0.014	-0.001
16	1	1	1	1	0.012	0.000	0.012	0.000



De acuerdo a lo observado en los tres escenarios, se presentan los cuadros 4.17 y 4.18 que resume lo hallado por medio de comparación de medias de los valores RMSE y del sesgo (en valores absolutos), respectivamente:

Cuadro 4.17: Resumen de medias de los valores RMSE para todos los parámetros estudiados

Escenarios	Tamaño de la población (N)	Desliz (s)		Adivinación (g)		Prob de cada clase latente (π)	
		Gibbs	EM	Gibbs	EM	Gibbs	EM
2	300	0.033	0.033	0.052	0.034	0.023	0.015
3	500	0.025	0.025	0.025	0.025	0.012	0.012

Cuadro 4.18: Resumen de medias de los valores del sesgo para todos los parámetros estudiados

Escenarios	Tamaño de la población (N)	Desliz (s)		Adivinación (g)		Prob de cada clase latente (π)	
		Gibbs	EM	Gibbs	EM	Gibbs	EM
1	100	0.012	0.006	0.012	0.006	0.002	0.003
2	300	0.007	0.004	0.009	0.003	0.003	0.002
3	500	0.003	0.003	0.004	0.003	0.001	0.001

Cuando la influencia de los parámetros de los ítems es débil en relación a las habilidades de los alumnos; y al igual que en el primer estudio, se mantienen constantes los valores para el parámetro π ; los resultados muestran que las estimaciones son insesgadas y con valores pequeños para el RMSE en ambos enfoques.

Sin embargo, en el escenario con un tamaño de población pequeño, con respecto a los parámetros de los ítems, existe una ligera ventaja a favor del algoritmo EM, cuando se toma en cuenta los valores del sesgo. Mientras que, para el π el muestreador de Gibbs presenta mejores estimaciones porque cuenta con valores más bajos de RMSE y sesgo.

En el segundo escenario, con un tamaño de población mediano en todos los parámetros estimados, ambos enfoques producen estimaciones muy similares cuando se comparan por los valores de RMSE, a excepción del parámetro de adivinación donde el algoritmo EM es claramente superior. Por otro lado, si solo tomáramos en cuenta el sesgo se observa un comportamiento irregular en ambos enfoques.

Finalmente, en el escenario con 500 individuos, prácticamente ambos enfoques tienen las mismas estimaciones para los parámetros en estudio.

4.7. Estudio 3

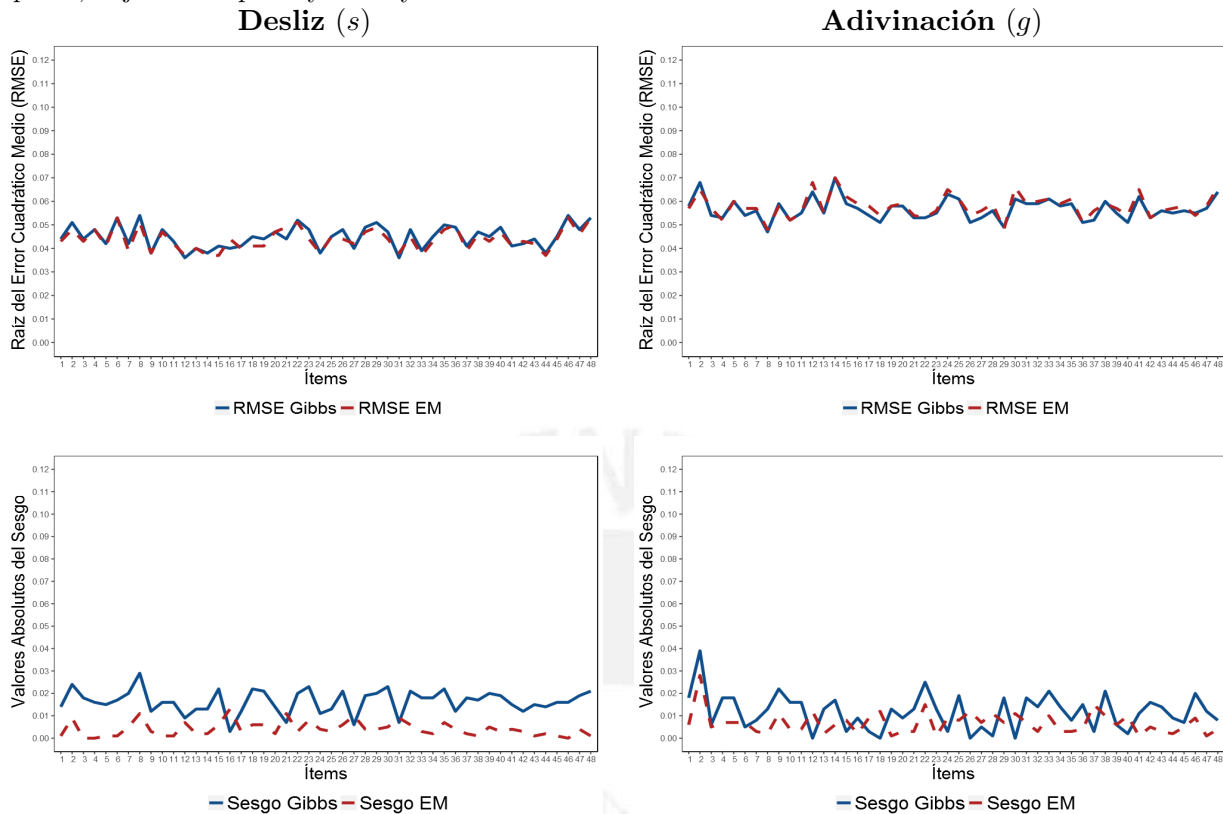
Para este estudio se presentan tres escenarios que se detallan en el cuadro 4.19, los cuales tienen como principal objetivo estudiar, bajo tres tamaños de población, la recuperación de los parámetros de los ítems. En este caso, se considerará el parámetro desliz ($s = 0.1$) con alto impacto y el parámetro adivinación ($g = 0.2$) de bajo impacto, con las habilidades de los individuos:

Cuadro 4.19: Escenarios del Estudio 3 de simulación para una relación fuerte para el parámetro de desliz (s); aunque una relación débil para el parámetro de adivinación (g), con las habilidades de los individuos, manteniendo el parámetro π constante

Escenarios	Tamaño de población (N)	Adivinación (s)	Desliz (g)	Probabilidad de pertenencia π
1	100	0.1	0.2	0.0625
2	300	0.1	0.2	0.0625
3	500	0.1	0.2	0.0625

Se presenta en el cuadro 4.20 con gráficos de líneas la lista de los 48 ítems con los valores de RMSE y sesgo para cada parámetro y con tamaño de población $N = 100$, de acuerdo a los dos enfoques estudiados:

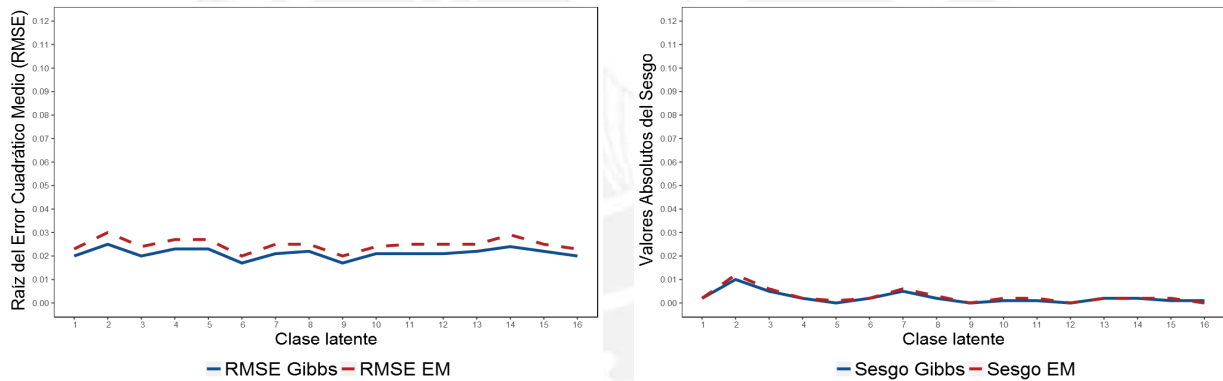
Cuadro 4.20: Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 3 con $N = 100$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico



En el siguiente cuadro 4.21 se presenta también la precisión de recuperación de parámetros para las probabilidades de pertenencia a las 2^K clases latentes, cuando se fija como $\pi = 2^{-K}$, manteniendo las características de los parámetros de los ítems y con $N = 100$.

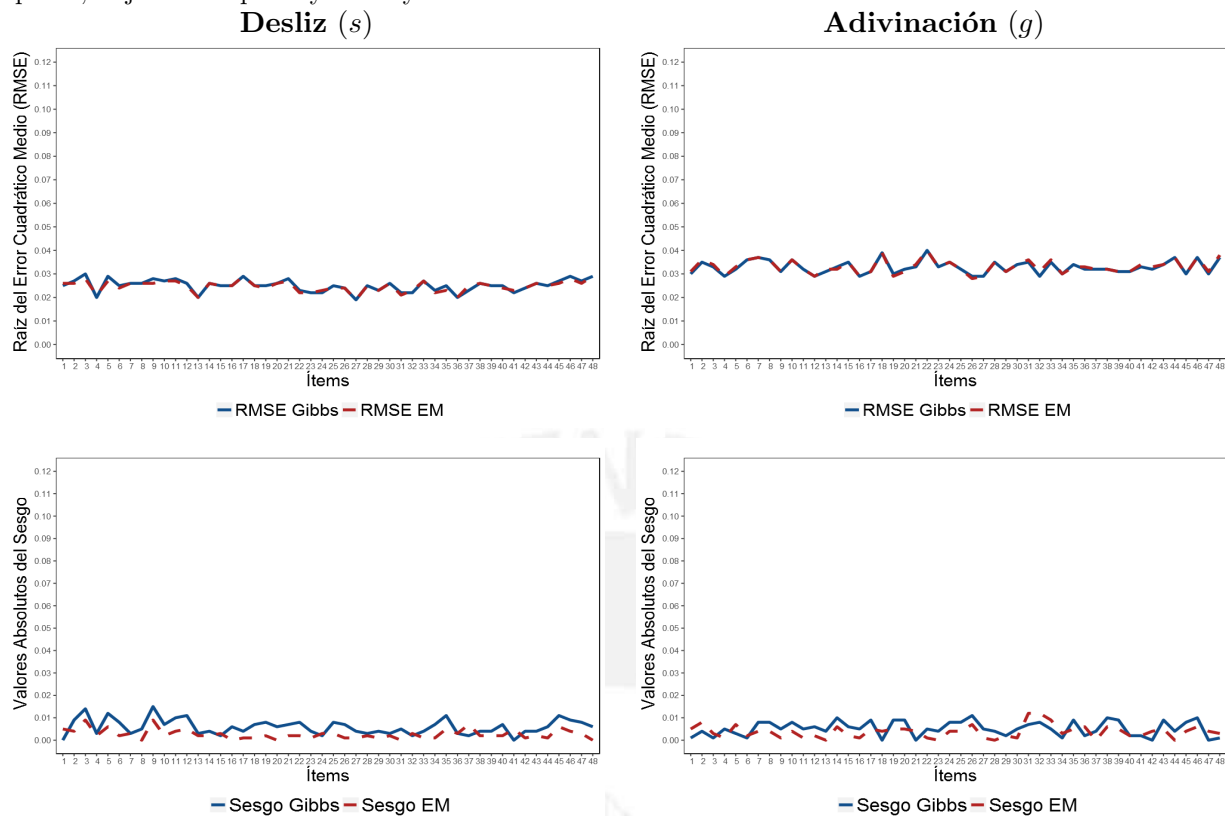
Cuadro 4.21: Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 3 con $N = 100$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico

Clases latentes	Perfil de habilidades				Muestreador de Gibbs	Algoritmo EM		
	α_1	α_2	α_3	α_4	RMSE	Sesgo	RMSE	Sesgo
1	0	0	0	0	0.02	-0.002	0.023	-0.002
2	1	0	0	0	0.025	0.01	0.03	0.012
3	0	1	0	0	0.02	-0.005	0.024	-0.006
4	0	0	1	0	0.023	0.002	0.027	0.002
5	0	0	0	1	0.023	0	0.027	0.001
6	1	1	0	0	0.017	-0.002	0.02	-0.002
7	1	0	1	0	0.021	-0.005	0.025	-0.006
8	1	0	0	1	0.022	0.002	0.025	0.003
9	0	1	1	0	0.017	0	0.02	0
10	0	1	0	1	0.021	0.001	0.024	0.002
11	0	0	1	1	0.021	-0.001	0.025	-0.002
12	1	1	1	0	0.021	0	0.025	0
13	1	1	0	1	0.022	0.002	0.025	0.002
14	1	0	1	1	0.024	-0.002	0.029	-0.002
15	0	1	1	1	0.022	-0.001	0.025	-0.002
16	1	1	1	1	0.02	0.001	0.023	0



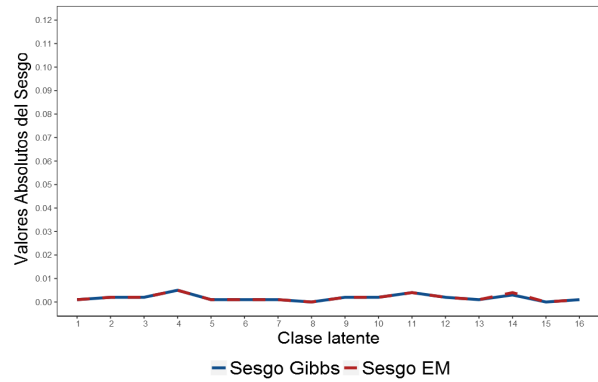
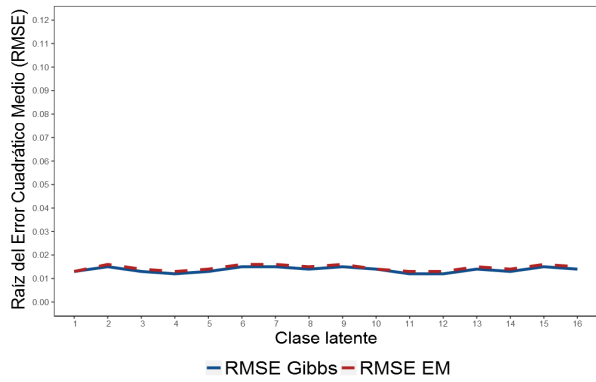
De la misma forma se presenta el resumen de resultados en los cuadros 4.22 y 4.23, para los parámetros de los ítems y las probabilidades de pertenencia, con un tamaño de población $N = 300$.

Cuadro 4.22: Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 3 con $N = 300$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico



Cuadro 4.23: Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 3 con $N = 300$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico

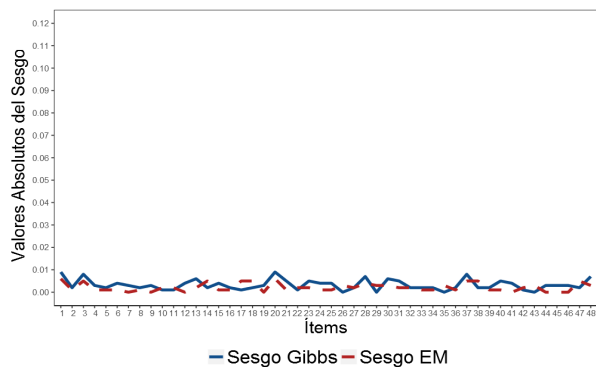
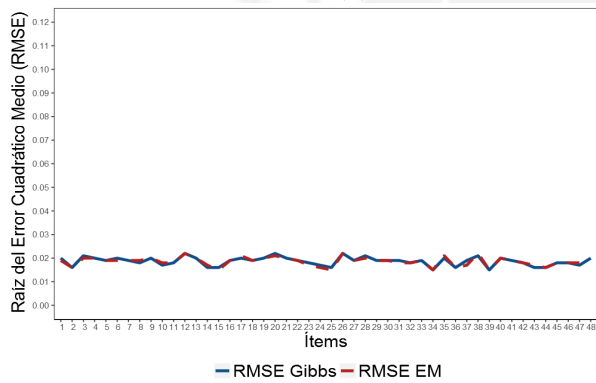
Clases latentes	Perfil de habilidades				Muestreador de Gibbs		Algoritmo EM	
	α_1	α_2	α_3	α_4	RMSE	Sesgo	RMSE	Sesgo
1	0	0	0	0	0.013	0.001	0.013	0.001
2	1	0	0	0	0.015	0.002	0.016	0.002
3	0	1	0	0	0.013	-0.002	0.014	-0.002
4	0	0	1	0	0.012	-0.005	0.013	-0.005
5	0	0	0	1	0.013	0.001	0.014	0.001
6	1	1	0	0	0.015	-0.001	0.016	-0.001
7	1	0	1	0	0.015	-0.001	0.016	-0.001
8	1	0	0	1	0.014	0	0.015	0
9	0	1	1	0	0.015	0.002	0.016	0.002
10	0	1	0	1	0.014	0.002	0.014	0.002
11	0	0	1	1	0.012	0.004	0.013	0.004
12	1	1	1	0	0.012	-0.002	0.013	-0.002
13	1	1	0	1	0.014	-0.001	0.015	-0.001
14	1	0	1	1	0.013	0.003	0.014	0.004
15	0	1	1	1	0.015	0	0.016	0
16	1	1	1	1	0.014	-0.001	0.015	-0.001



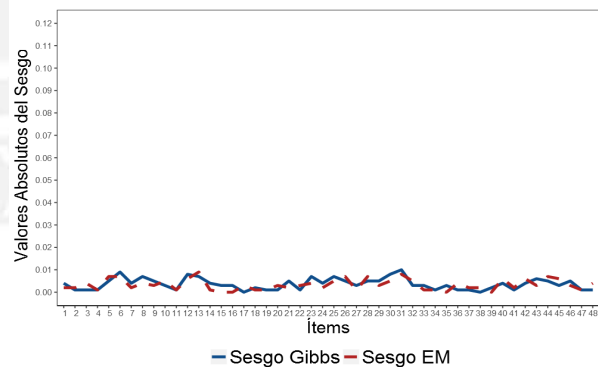
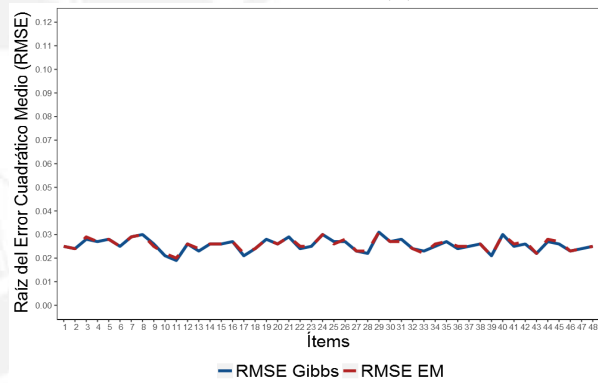
Finalmente, se presenta el resumen de resultados para el último escenario, en los cuadros 4.24 y 4.25, para los parámetros de los ítems y las probabilidades de pertenencia con un tamaño de población $N = 500$.

Cuadro 4.24: Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 3 con $N = 500$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico

Desliz (s)

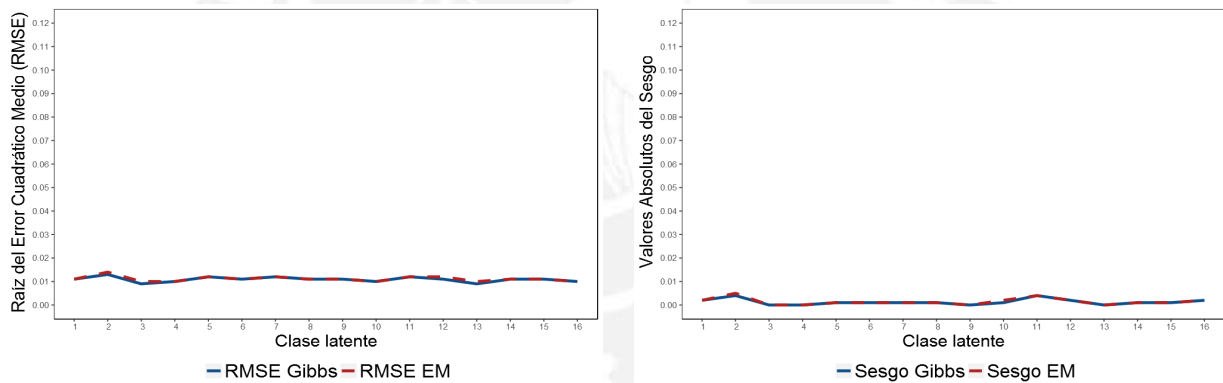


Adivinación (g)



Cuadro 4.25: Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 3 con $N = 500$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico

Clases latentes	Perfil de habilidades				Muestreador de Gibbs		Algoritmo EM	
	α_1	α_2	α_3	α_4	RMSE	Sesgo	RMSE	Sesgo
1	0	0	0	0	0.011	0.002	0.011	0.002
2	1	0	0	0	0.013	-0.004	0.014	-0.005
3	0	1	0	0	0.009	0	0.01	0
4	0	0	1	0	0.01	0	0.01	0
5	0	0	0	1	0.012	0.001	0.012	0.001
6	1	1	0	0	0.011	0.001	0.011	0.001
7	1	0	1	0	0.012	0.001	0.012	0.001
8	1	0	0	1	0.011	-0.001	0.011	-0.001
9	0	1	1	0	0.011	0	0.011	0
10	0	1	0	1	0.01	-0.001	0.01	-0.002
11	0	0	1	1	0.012	0.004	0.012	0.004
12	1	1	1	0	0.011	0.002	0.012	0.002
13	1	1	0	1	0.009	0	0.01	0
14	1	0	1	1	0.011	-0.001	0.011	-0.001
15	0	1	1	1	0.011	0.001	0.011	0.001
16	1	1	1	1	0.01	-0.002	0.01	-0.002



Se presenta el resumen de los tres escenarios por medio de los cuadros 4.26 y 4.27 que resume lo hallado por medio de comparación de medias de los valores RMSE y del sesgo (en valores absolutos), respectivamente:

Cuadro 4.26: Resumen de medias de los valores RMSE para todos los parámetros estudiados

Escenarios	Tamaño de la población (N)	Desliz (s)		Adivinación (g)		Prob de cada clase latente (π)	
		Gibbs	EM	Gibbs	EM	Gibbs	EM
2	300	0.025	0.025	0.033	0.033	0.014	0.015
3	500	0.019	0.019	0.026	0.026	0.011	0.011

Cuadro 4.27: Resumen de medias de los valores del sesgo para todos los parámetros estudiados

Escenarios	Tamaño de la población (N)	Desliz (s)		Adivinación (g)		Prob de cada clase latente (π)	
		Gibbs	EM	Gibbs	EM	Gibbs	EM
1	100	0.016	0.004	0.012	0.007	0.002	0.003
2	300	0.006	0.003	0.005	0.004	0.002	0.002
3	500	0.003	0.002	0.004	0.003	0.001	0.001

Cuando la influencia del parámetro de adivinación es baja y la del desliz es alta, con respecto a las habilidades de los alumnos; y manteniendo constantes los valores para el parámetro π . Las estimaciones en los tres escenarios son insesgadas y con valores pequeños para el RMSE en ambos enfoques.

Cabe mencionar que para el parámetro de desliz y adivinación, tomando en cuenta un tamaño de población pequeño, el algoritmo EM genera estimaciones con un menor sesgo que los que se consiguen con el muestreador de Gibbs. Por otro lado, este último obtiene mejores estimaciones, considerando los valores del RMSE, para el parámetro π en el escenario 1 ($N = 100$) y 2 ($N = 300$).

En las otras combinaciones de características, no hay mayores diferencias entre ambos enfoques.

4.8. Estudio 4

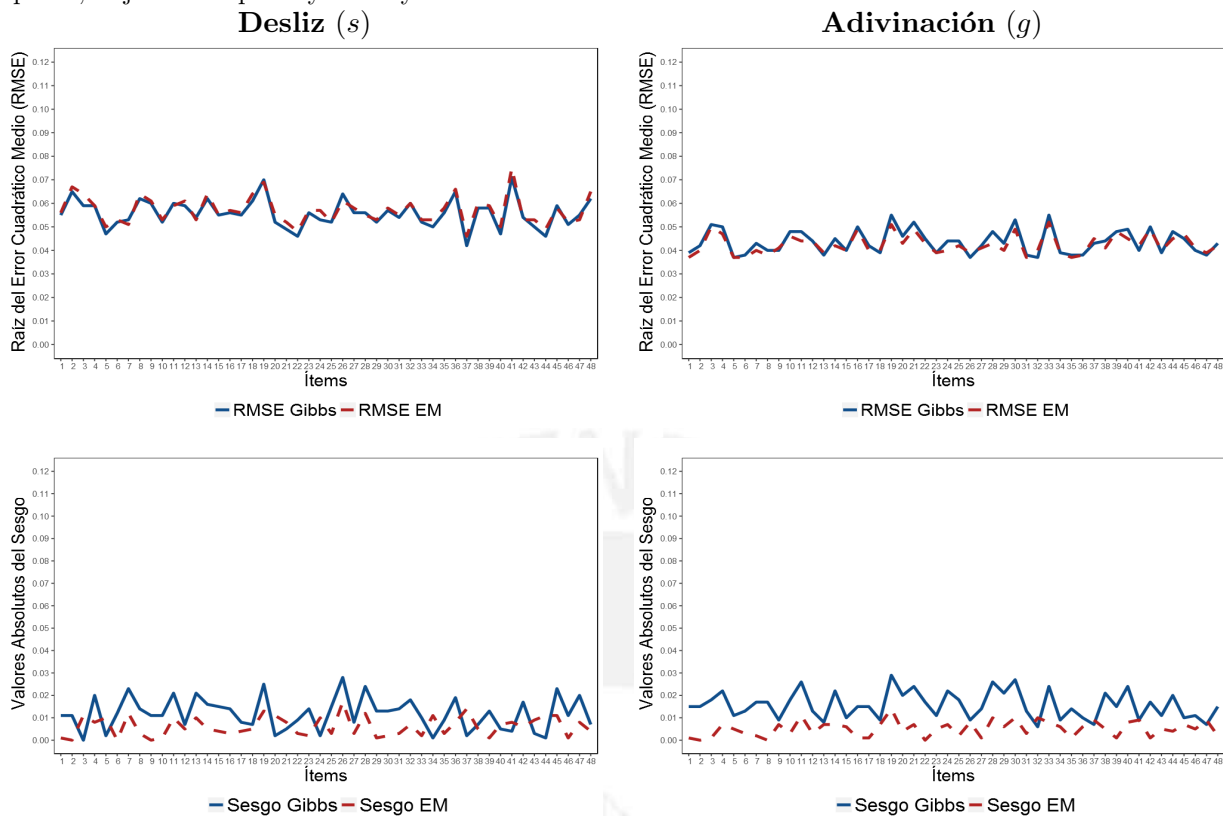
Para este estudio se presentan tres escenarios que se detallan en el cuadro 4.28, los cuales tienen como principal objetivo estudiar, bajo tres tamaños de población, la recuperación de los parámetros de los ítems. En este caso, se considerará la relación para el parámetro desliz ($s = 0.2$) con bajo impacto y el parámetro adivinación ($g = 0.1$) de alto impacto, con las habilidades de los individuos. Además, manteniendo constante el parámetro $\pi = 2^{-K} = 0.0625$, lo que denota una relación de poco impacto:

Cuadro 4.28: Escenarios del Estudio 4 de simulación para una relación baja para el parámetro de desliz (s); aunque una relación fuerte para el parámetro de adivinación (g), con las habilidades de los individuos, manteniendo el parámetro π constante

Escenarios	Tamaño de población (N)	Adivinación (s)	Desliz (g)	Probabilidad de pertenencia (π)
1	100	0.2	0.1	0.0625
2	300	0.2	0.1	0.0625
3	500	0.2	0.1	0.0625

Se presenta en el cuadro 4.29 con gráficos de líneas la lista de los 48 ítems con los valores de RMSE y sesgo para cada parámetro y con tamaño de población $N = 100$, de acuerdo a los dos enfoques estudiados:

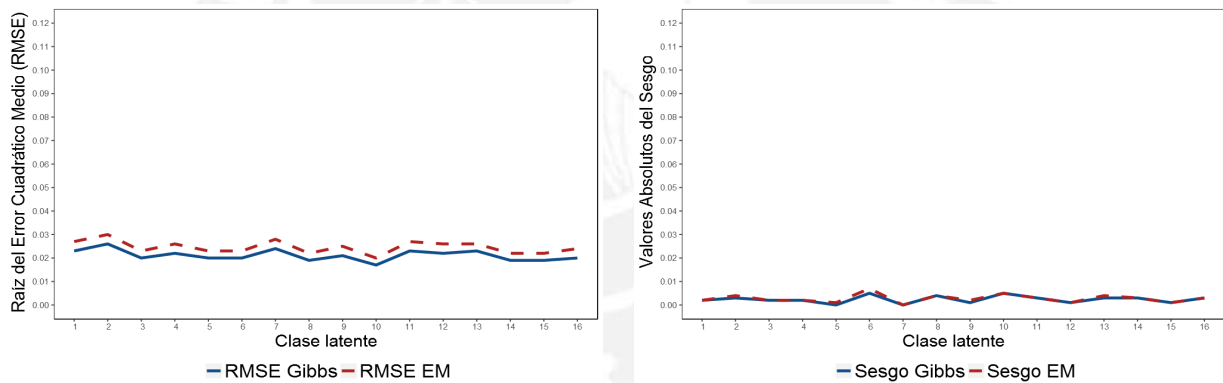
Cuadro 4.29: Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 4 con $N = 100$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico



En el siguiente cuadro 4.30 se presenta también la precisión de recuperación de parámetros para las probabilidades de pertenencia a las 2^K clases latentes, cuando se fija como $\pi = 2^{-K}$, manteniendo las características de los parámetros de los ítems mencionadas, y con $N = 100$.

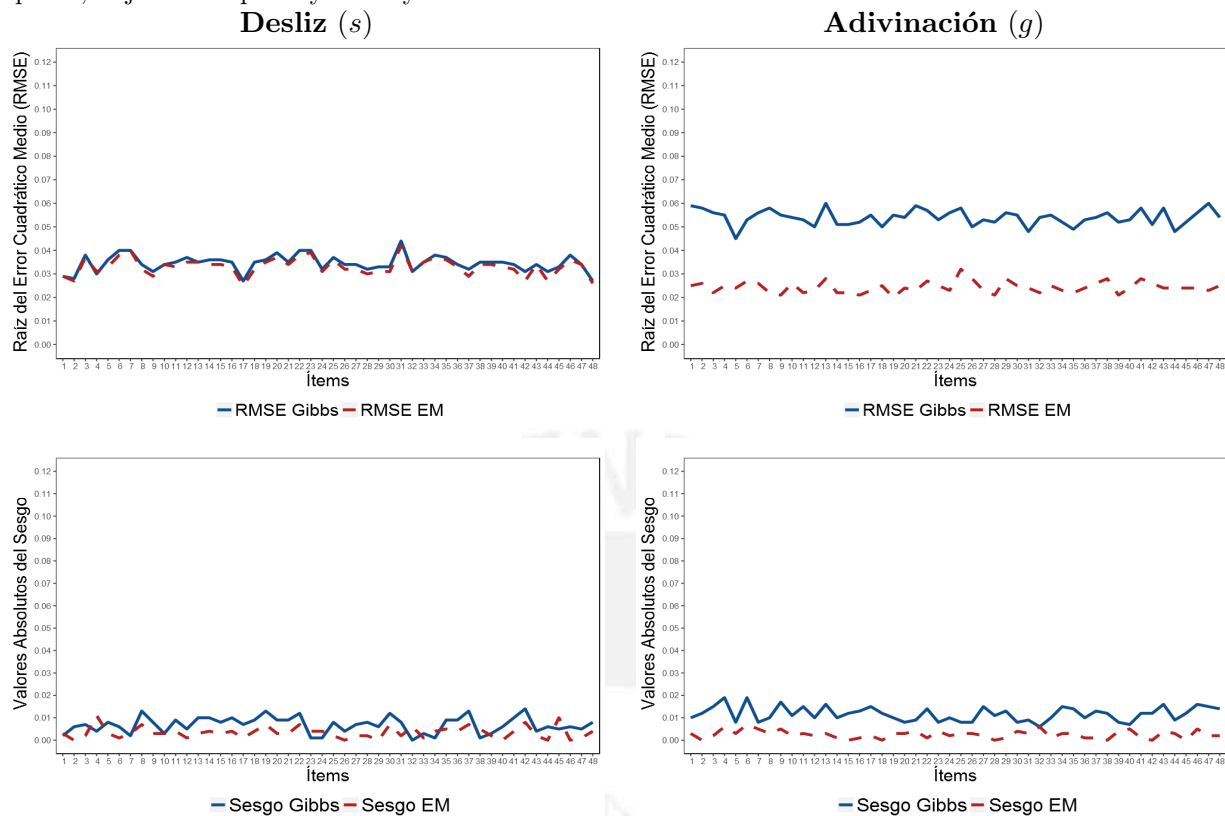
Cuadro 4.30: Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 4 con $N = 100$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico

Clases latentes	Perfil de habilidades				Muestreador de Gibbs	Algoritmo EM		
	α_1	α_2	α_3	α_4	RMSE	Sesgo	RMSE	Sesgo
1	0	0	0	0	0.023	0.002	0.027	0.002
2	1	0	0	0	0.026	0.003	0.03	0.004
3	0	1	0	0	0.02	0.002	0.023	0.002
4	0	0	1	0	0.022	-0.002	0.026	-0.002
5	0	0	0	1	0.02	0	0.023	-0.001
6	1	1	0	0	0.02	-0.005	0.023	-0.007
7	1	0	1	0	0.024	0	0.028	0
8	1	0	0	1	0.019	-0.004	0.022	-0.004
9	0	1	1	0	0.021	0.001	0.025	0.002
10	0	1	0	1	0.017	0.005	0.02	0.005
11	0	0	1	1	0.023	0.003	0.027	0.003
12	1	1	1	0	0.022	-0.001	0.026	-0.001
13	1	1	0	1	0.023	0.003	0.026	0.004
14	1	0	1	1	0.019	-0.003	0.022	-0.003
15	0	1	1	1	0.019	-0.001	0.022	-0.001
16	1	1	1	1	0.02	-0.003	0.024	-0.003



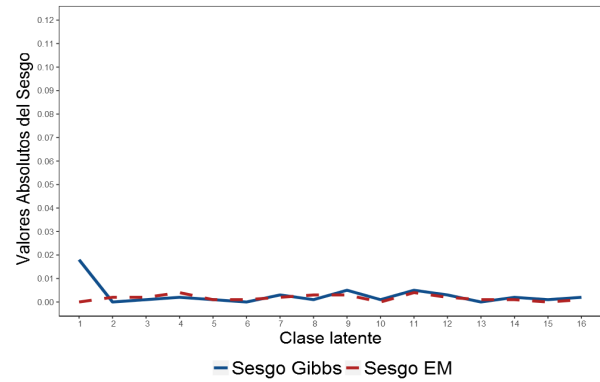
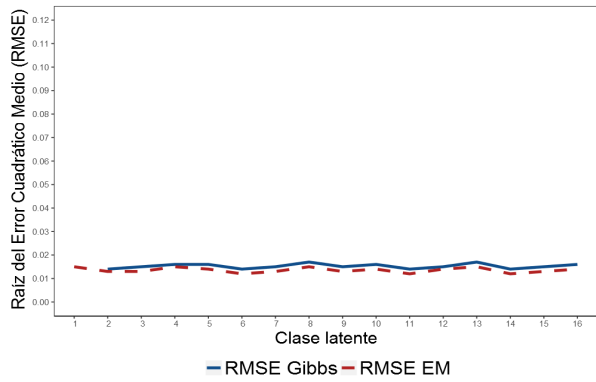
De la misma forma se presenta el resumen de resultados en los cuadros 4.31 y 4.32, para los parámetros de los ítems y las probabilidades de pertenencia, con un tamaño de población $N = 300$.

Cuadro 4.31: Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 4 con $N = 300$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico



Cuadro 4.32: Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 4 con $N = 300$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico

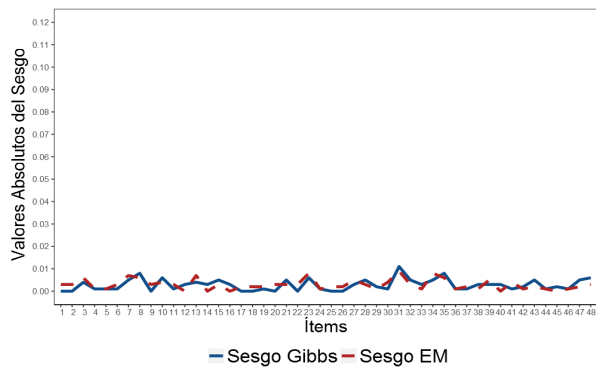
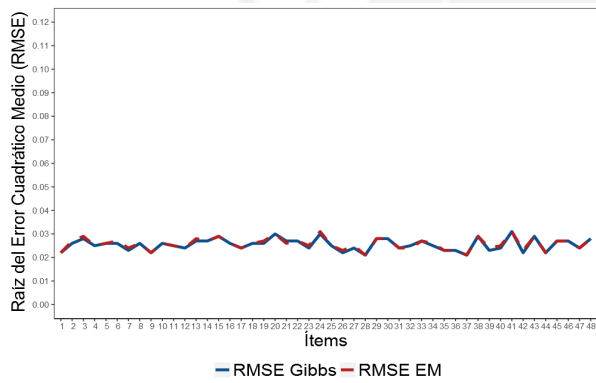
Clases latentes	Perfil de habilidades				Muestreador de Gibbs		Algoritmo EM	
	α_1	α_2	α_3	α_4	RMSE	Sesgo	RMSE	Sesgo
1	0	0	0	0	0.127	0.018	0.015	0
2	1	0	0	0	0.014	0	0.013	0.002
3	0	1	0	0	0.015	0.001	0.013	0.002
4	0	0	1	0	0.016	0.002	0.015	0.004
5	0	0	0	1	0.016	-0.001	0.014	0.001
6	1	1	0	0	0.014	0	0.012	0.001
7	1	0	1	0	0.015	-0.003	0.013	-0.002
8	1	0	0	1	0.017	0.001	0.015	0.003
9	0	1	1	0	0.015	-0.005	0.013	-0.003
10	0	1	0	1	0.016	-0.001	0.014	0
11	0	0	1	1	0.014	-0.005	0.012	-0.004
12	1	1	1	0	0.015	-0.003	0.014	-0.002
13	1	1	0	1	0.017	0	0.015	0.001
14	1	0	1	1	0.014	-0.002	0.012	-0.001
15	0	1	1	1	0.015	-0.001	0.013	0
16	1	1	1	1	0.016	-0.002	0.014	-0.001



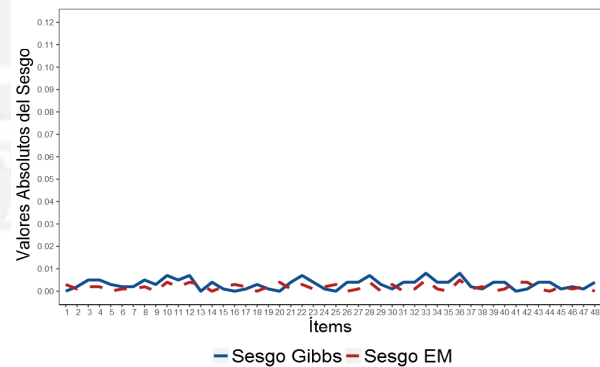
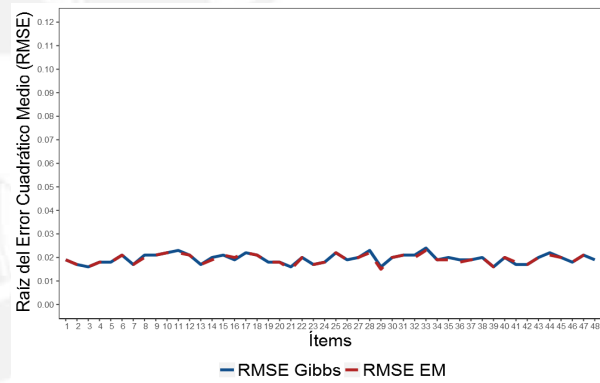
Finalmente, se presenta el resumen de resultados para el último escenario, en los cuadros 4.33 y 4.34, para los parámetros de los ítems y las probabilidades de pertenencia con un tamaño $N = 500$.

Cuadro 4.33: Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Deslíz (s) y Adivinación (g) en el Estudio 4 con $N = 500$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico

Deslíz (s)

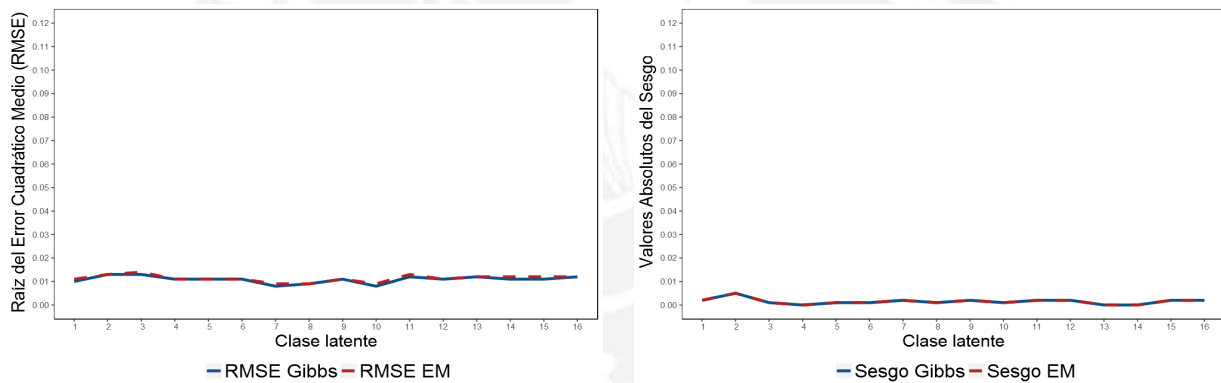


Adivinación (g)



Cuadro 4.34: Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 4 con $N = 500$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico

Clases latentes	Perfil de habilidades				Muestreador de Gibbs		Algoritmo EM	
	α_1	α_2	α_3	α_4	RMSE	Sesgo	RMSE	Sesgo
1	0	0	0	0	0.01	-0.002	0.011	-0.002
2	1	0	0	0	0.013	-0.005	0.013	-0.005
3	0	1	0	0	0.013	-0.001	0.014	-0.001
4	0	0	1	0	0.011	0	0.011	0
5	0	0	0	1	0.011	0.001	0.011	0.001
6	1	1	0	0	0.011	0.001	0.011	0.001
7	1	0	1	0	0.008	0.002	0.009	0.002
8	1	0	0	1	0.009	-0.001	0.009	-0.001
9	0	1	1	0	0.011	-0.002	0.011	-0.002
10	0	1	0	1	0.008	-0.001	0.009	-0.001
11	0	0	1	1	0.012	0.002	0.013	0.002
12	1	1	1	0	0.011	0.002	0.011	0.002
13	1	1	0	1	0.012	0	0.012	0
14	1	0	1	1	0.011	0	0.012	0
15	0	1	1	1	0.011	0.002	0.012	0.002
16	1	1	1	1	0.012	0.002	0.012	0.002



Se presenta el resumen de los tres escenarios por medio de los cuadros 4.35 y 4.36 que resume lo hallado por medio de comparación de medias de los valores RMSE y del sesgo (en valores absolutos), respectivamente:

Cuadro 4.35: Resumen de medias de los valores RMSE para todos los parámetros estudiados

Escenarios	Tamaño de la población (N)	Desliz (s)		Adivinación (g)		Prob de cada clase latente (π)	
		Gibbs	EM	Gibbs	EM	Gibbs	EM
1	100	0.056	0.057	0.044	0.042	0.021	0.025
2	300	0.035	0.033	0.053	0.024	0.022	0.014
3	500	0.026	0.026	0.020	0.019	0.011	0.011

Cuadro 4.36: Resumen de medias de los valores del sesgo para todos los parámetros estudiados

Escenarios	Tamaño de la población (N)	Desliz (s)		Adivinación (g)		Prob de cada clase latente (π)	
		Gibbs	EM	Gibbs	EM	Gibbs	EM
1	100	0.012	0.006	0.016	0.005	0.002	0.003
2	300	0.007	0.003	0.012	0.003	0.003	0.002
3	500	0.003	0.003	0.003	0.002	0.002	0.002

A diferencia del estudio anterior, en este se considera una alta influencia del parámetro de adivinación mientras que una débil relación para el parámetro desliz, con respecto a las habilidades de los alumnos. Manteniendo constantes los valores para el parámetro π . Los resultados son insesgados y con pequeños valores de RMSE en todos los escenarios.

Las mayores diferencias se encuentran, considerando los valores del sesgo, en la adivinación para los escenarios 1 y 2 a favor del algoritmo EM; mientras que, considerando los valores del RMSE, el muestreador de Gibbs obtiene mejores estimaciones para π en el escenario 1 y ligeramente en el escenario 3.

4.9. Estudio 5

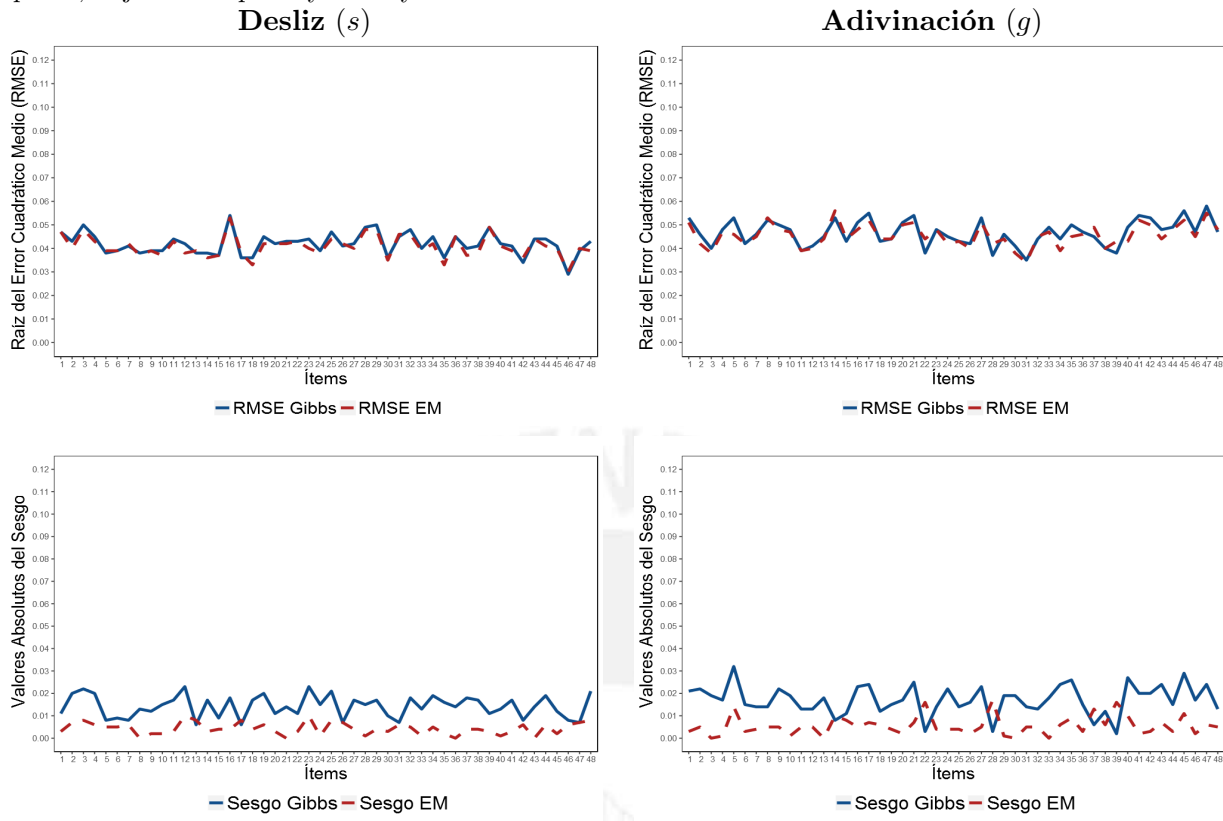
Para este estudio se presentan tres escenarios que se detallan en el cuadro 4.37, los cuales tienen como principal objetivo estudiar, bajo tres tamaños de población, la recuperación de los parámetros de los ítems cuando estos presentan una alta o fuerte relación con las habilidades de los individuos; pero con el parámetro π manteniendo una relación de alto impacto. Este parámetro ahora contará con dos valores: 0.021 y 0.042, cuando la clase o perfil latente está conformada por el dominio de dos o menos habilidades, y cuando lo está por tres a más habilidades; respectivamente:

Cuadro 4.37: Escenarios del Estudio 5 de simulación para una relación baja entre los parámetros de los ítems y las habilidades de los individuos, con el parámetro π manteniendo una relación de alto impacto

Escenarios	Tamaño de población (N)	Adivinación (s)	Desliz (g)	Probabilidad de pertenencia π
1	100	0.1	0.1	0.021 o 0.042
2	300	0.1	0.1	0.021 o 0.042
3	500	0.1	0.1	0.021 o 0.042

Se presenta en el cuadro 4.38 con gráficos de líneas, la lista de los 48 ítems con los valores de RMSE y sesgo para cada parámetro y con tamaño de población $N = 100$, de acuerdo a los dos enfoques estudiados:

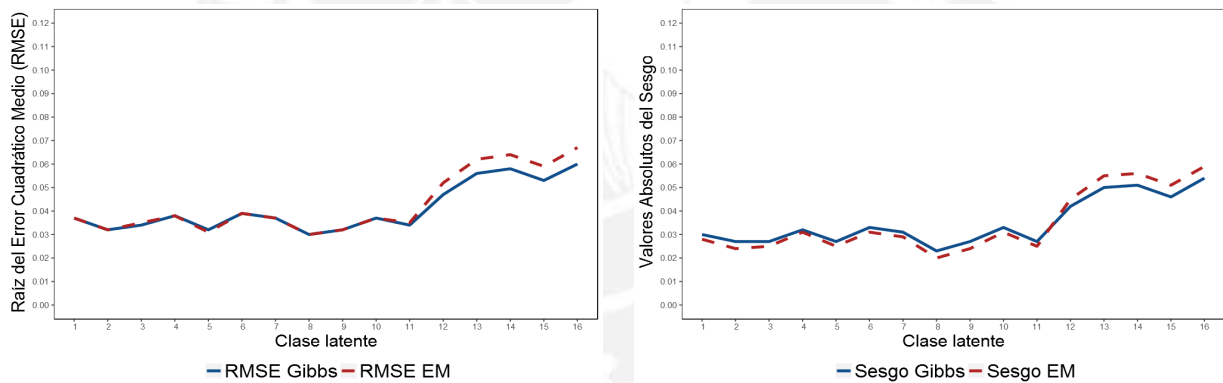
Cuadro 4.38: Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 5 con $N = 100$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico



En el siguiente cuadro 4.39 se presenta también la precisión de recuperación de parámetros para las probabilidades de pertenencia a las 2^K clases latentes, cuando se fija como $\pi = 2^{-K}$, manteniendo las características de los parámetros de los ítems mencionadas, y con $N = 100$.

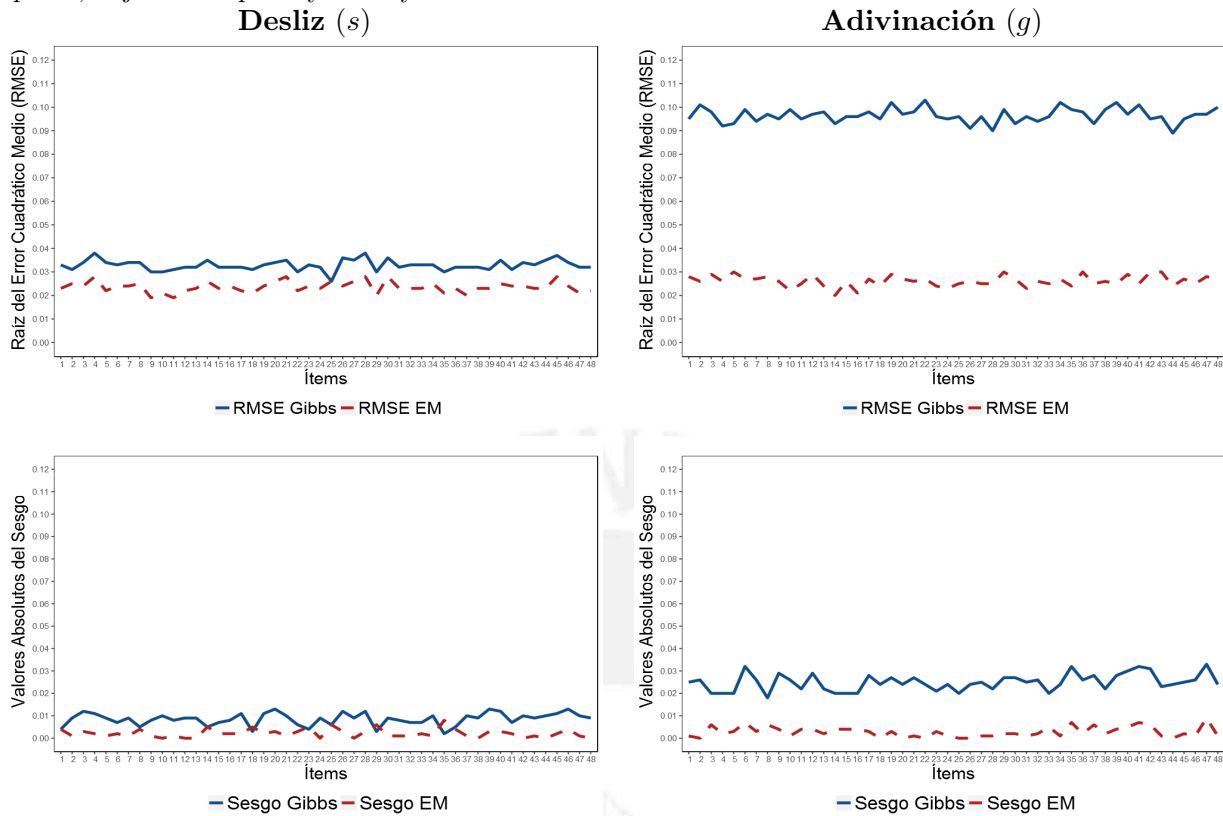
Cuadro 4.39: Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 5 con $N = 100$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico

Clases latentes	Perfil de habilidades				Muestreador de Gibbs	Algoritmo EM		
	α_1	α_2	α_3	α_4	RMSE	Sesgo	RMSE	Sesgo
1	0	0	0	0	0.037	0.03	0.037	0.028
2	1	0	0	0	0.032	0.027	0.032	0.024
3	0	1	0	0	0.034	0.027	0.035	0.025
4	0	0	1	0	0.038	0.032	0.038	0.031
5	0	0	0	1	0.032	0.027	0.031	0.025
6	1	1	0	0	0.039	0.033	0.039	0.031
7	1	0	1	0	0.037	0.031	0.037	0.029
8	1	0	0	1	0.03	0.023	0.03	0.02
9	0	1	1	0	0.032	0.027	0.032	0.024
10	0	1	0	1	0.037	0.033	0.037	0.031
11	0	0	1	1	0.034	0.027	0.035	0.025
12	1	1	1	0	0.047	0.042	0.052	0.045
13	1	1	0	1	0.056	0.05	0.062	0.055
14	1	0	1	1	0.058	0.051	0.064	0.056
15	0	1	1	1	0.053	0.046	0.059	0.051
16	1	1	1	1	0.06	0.054	0.067	0.059



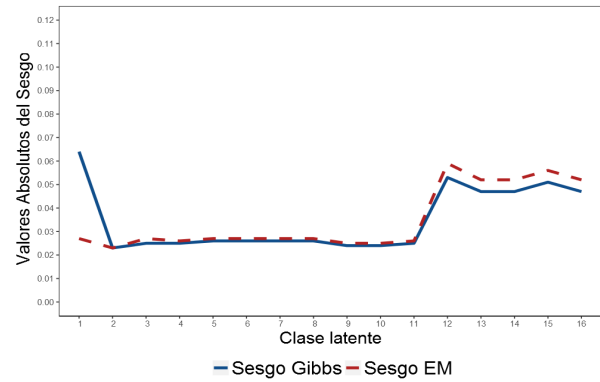
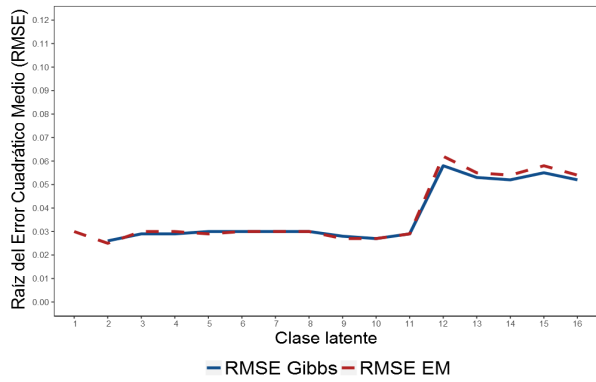
De la misma forma se presenta el resumen de resultados en los cuadros 4.40 y 4.41, para los parámetros de los ítems y las probabilidades de pertenencia, con un tamaño $N = 300$.

Cuadro 4.40: Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 5 con $N = 300$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico



Cuadro 4.41: Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 5 con $N = 300$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico

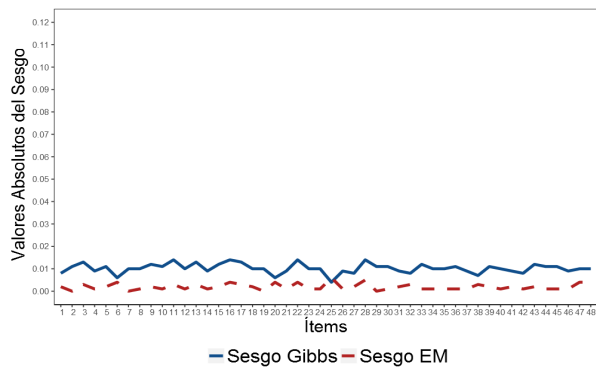
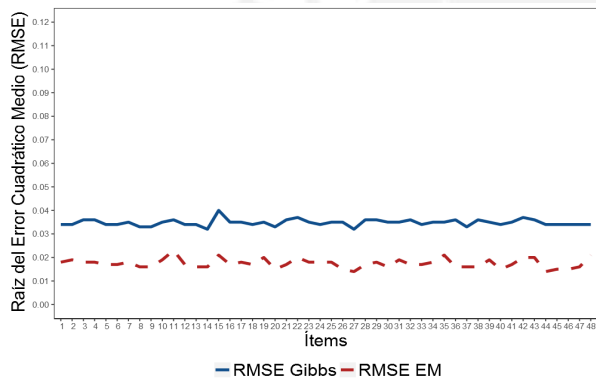
Clases latentes	Perfil de habilidades				Muestreador de Gibbs		Algoritmo EM	
	α_1	α_2	α_3	α_4	RMSE	Sesgo	RMSE	Sesgo
1	0	0	0	0	0.189	0.064	0.03	0.027
2	1	0	0	0	0.026	0.023	0.025	0.023
3	0	1	0	0	0.029	0.025	0.03	0.027
4	0	0	1	0	0.029	0.025	0.03	0.026
5	0	0	0	1	0.03	0.026	0.029	0.027
6	1	1	0	0	0.03	0.026	0.03	0.027
7	1	0	1	0	0.03	0.026	0.03	0.027
8	1	0	0	1	0.03	0.026	0.03	0.027
9	0	1	1	0	0.028	0.024	0.027	0.025
10	0	1	0	1	0.027	0.024	0.027	0.025
11	0	0	1	1	0.029	0.025	0.029	0.026
12	1	1	1	0	0.058	0.053	0.062	0.059
13	1	1	0	1	0.053	0.047	0.055	0.052
14	1	0	1	1	0.052	0.047	0.054	0.052
15	0	1	1	1	0.055	0.051	0.058	0.056
16	1	1	1	1	0.052	0.047	0.054	0.052



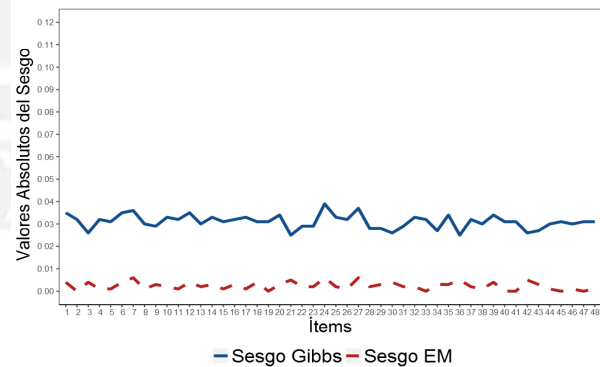
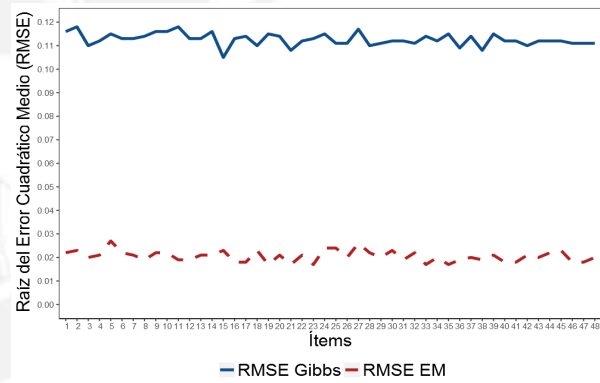
Finalmente, se presenta el resumen de resultados para el último escenario, en los cuadros 4.42 y 4.43, para los parámetros de los ítems y las probabilidades de pertenencia con un tamaño $N = 500$.

Cuadro 4.42: Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Deslíz (s) y Adivinación (g) en el Estudio 5 con $N = 500$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico

Deslíz (s)

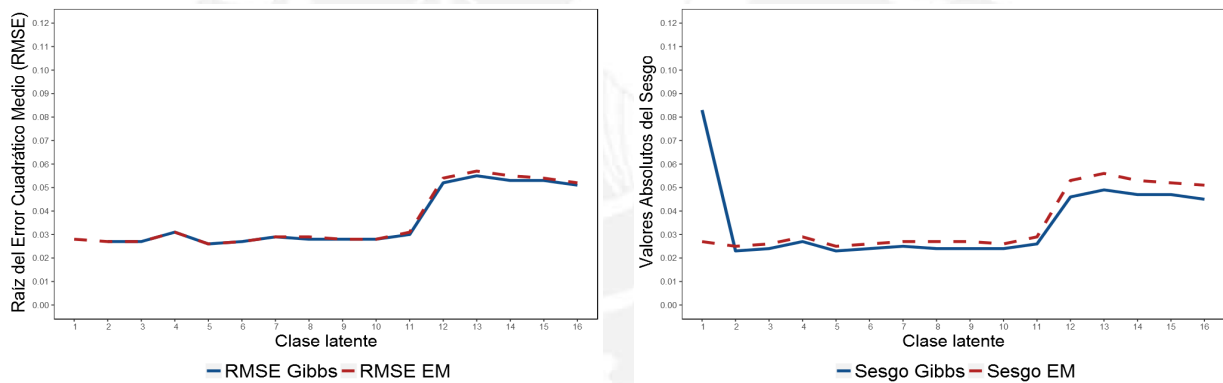


Adivinación (g)



Cuadro 4.43: Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 5 con $N = 500$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico

Clases latentes	Perfil de habilidades				Muestreador de Gibbs		Algoritmo EM	
	α_1	α_2	α_3	α_4	RMSE	Sesgo	RMSE	Sesgo
1	0	0	0	0	0.234	0.083	0.028	0.027
2	1	0	0	0	0.027	0.023	0.027	0.025
3	0	1	0	0	0.027	0.024	0.027	0.026
4	0	0	1	0	0.031	0.027	0.031	0.029
5	0	0	0	1	0.026	0.023	0.026	0.025
6	1	1	0	0	0.027	0.024	0.027	0.026
7	1	0	1	0	0.029	0.025	0.029	0.027
8	1	0	0	1	0.028	0.024	0.029	0.027
9	0	1	1	0	0.028	0.024	0.028	0.027
10	0	1	0	1	0.028	0.024	0.028	0.026
11	0	0	1	1	0.03	0.026	0.031	0.029
12	1	1	1	0	0.052	0.046	0.054	0.053
13	1	1	0	1	0.055	0.049	0.057	0.056
14	1	0	1	1	0.053	0.047	0.055	0.053
15	0	1	1	1	0.053	0.047	0.054	0.052
16	1	1	1	1	0.051	0.045	0.052	0.051



De acuerdo a lo observado en los tres escenarios, se presentan los cuadros 4.44 y 4.45 que resume lo hallado por medio de comparación de medias de los valores RMSE y del sesgo (en valores absolutos), respectivamente:

Cuadro 4.44: Resumen de medias de los valores RMSE para todos los parámetros estudiados

Escenarios	Tamaño de la población (N)	Desliz (s)		Adivinación (g)		Prob de cada clase latente (π)	
		Gibbs	EM	Gibbs	EM	Gibbs	EM
1	100	0.042	0.041	0.047	0.046	0.041	0.043
2	300	0.033	0.024	0.093	0.026	0.047	0.038
3	500	0.035	0.017	0.113	0.021	0.049	0.036

Cuadro 4.45: Resumen de medias de los valores del sesgo para todos los parámetros estudiados

Escenarios	Tamaño de la población (N)	Desliz (s)		Adivinación (g)		Prob de cada clase latente (π)	
		Gibbs	EM	Gibbs	EM	Gibbs	EM
1	100	0.014	0.004	0.017	0.006	0.035	0.035
2	300	0.008	0.002	0.024	0.003	0.035	0.035
3	500	0.01	0.002	0.031	0.002	0.035	0.035

Cuando los parámetros de los ítems, desliz (s) y adivinación (g), tienen una alta o fuerte influencia en las habilidades de los alumnos; y además, considerando que el parámetro π tiene dos valores de acuerdo a la cantidad de habilidades que conforman la clase latente; donde esta diferenciación la definiremos como un impacto alto del parámetro respecto al diagnóstico de las habilidades de los alumnos.

A partir de este estudio, notamos que las estimaciones obtenidas por el algoritmo EM, considerando ambos indicadores, son claramente mejores para los parámetros de los ítems, en los tres escenarios.

Sin embargo, esta superioridad no se evidencia en el parámetro π donde el muestreador de Gibbs viene mostrando ligeras ventajas cuando la población es pequeña; y en los demás escenarios los resultados son similares con ambos enfoques.

4.10. Estudio 6

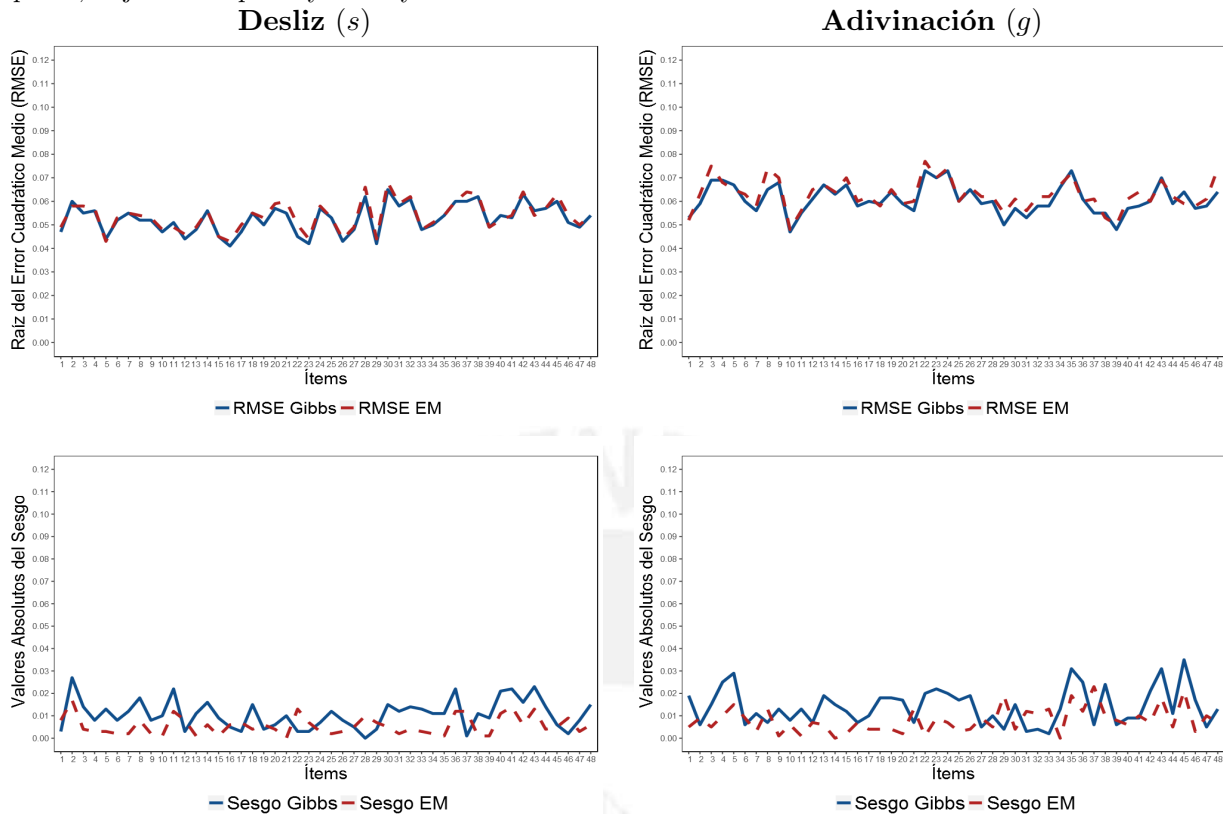
Para este estudio se presentan tres escenarios que se detallan en el cuadro 4.46, los cuales tienen como principal objetivo estudiar, bajo tres tamaños de población, la recuperación de los parámetros de los ítems cuando estos presentan una baja relación con las habilidades de los individuos; pero con el parámetro π manteniendo una relación de alto impacto. Este parámetro ahora contará con dos valores: 0.021 y 0.042, cuando la clase o perfil latente está conformada por el dominio de dos o menos habilidades, y cuando lo está por tres a más habilidades; respectivamente:

Cuadro 4.46: Escenarios del Estudio 6 de simulación para una relación alta entre los parámetros de los ítems y las habilidades de los individuos, con el parámetro π manteniendo una relación de alto impacto

Escenarios	Tamaño de población (N)	Adivinación (s)	Desliz (g)	Probabilidad de pertenencia π
1	100	0.2	0.2	0.021 o 0.042
2	300	0.2	0.2	0.021 o 0.042
3	500	0.2	0.2	0.021 o 0.042

Se presenta en el cuadro 4.47 con gráficos de línea, la lista de los 48 ítems con los valores de RMSE y sesgo para cada parámetro y con tamaño de población $N = 100$, de acuerdo a los dos enfoques estudiados:

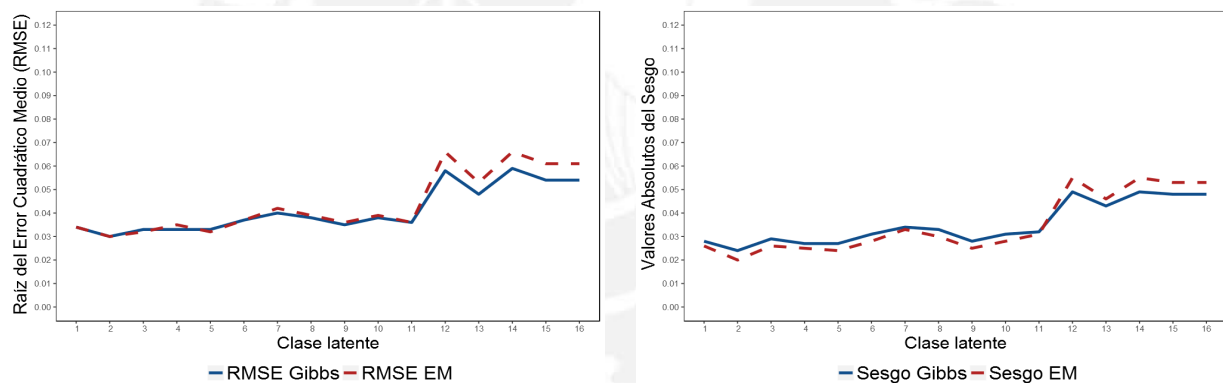
Cuadro 4.47: Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 6 con $N = 100$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico



En el siguiente cuadro 4.48 se presenta también la precisión de recuperación de parámetros para las probabilidades de pertenencia a las 2^K clases latentes, cuando se fija como $\pi = 2^{-K}$, manteniendo las características de los parámetros de los ítems mencionadas, y con $N = 100$.

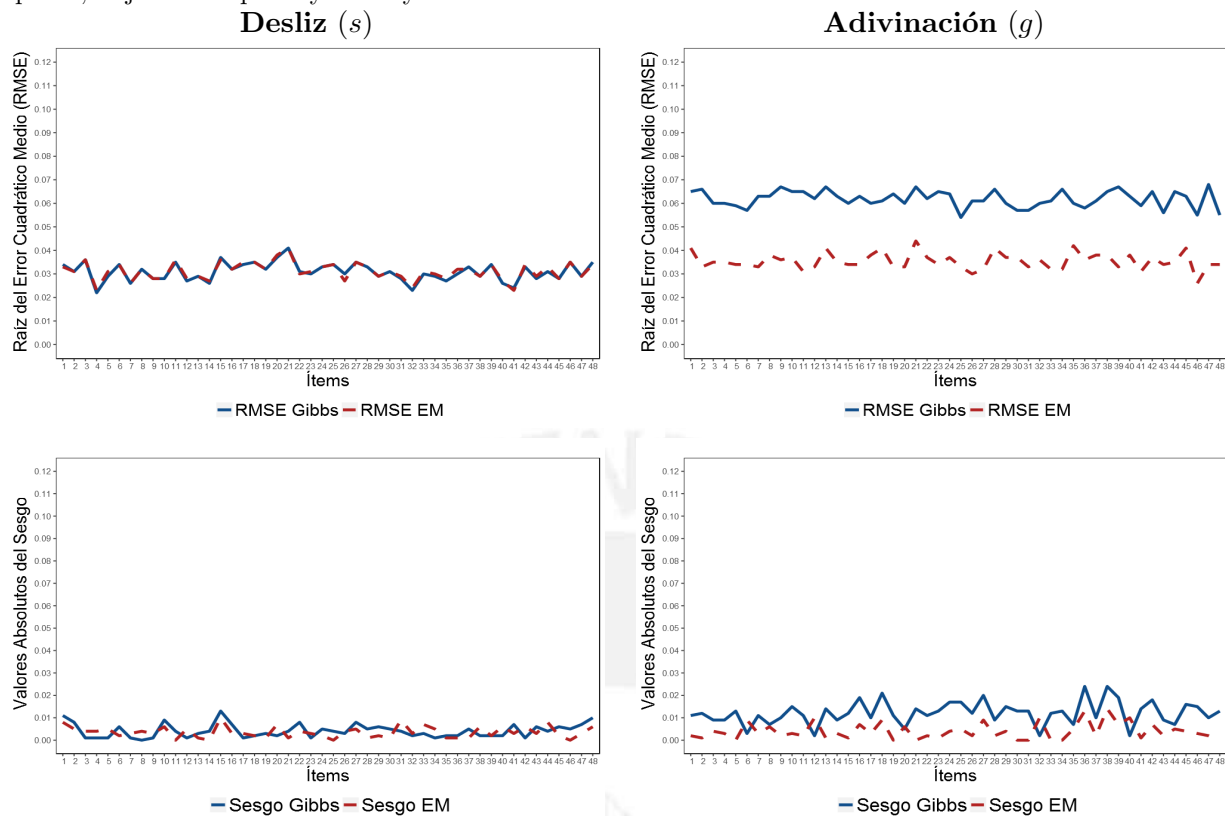
Cuadro 4.48: Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 6 con $N = 100$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico

Clases latentes	Perfil de habilidades				Muestreador de Gibbs	Algoritmo EM		
	α_1	α_2	α_3	α_4	RMSE	Sesgo	RMSE	Sesgo
1	0	0	0	0	0.034	0.028	0.034	0.026
2	1	0	0	0	0.03	0.024	0.03	0.02
3	0	1	0	0	0.033	0.029	0.032	0.026
4	0	0	1	0	0.033	0.027	0.035	0.025
5	0	0	0	1	0.033	0.027	0.032	0.024
6	1	1	0	0	0.037	0.031	0.037	0.028
7	1	0	1	0	0.04	0.034	0.042	0.033
8	1	0	0	1	0.038	0.033	0.039	0.03
9	0	1	1	0	0.035	0.028	0.036	0.025
10	0	1	0	1	0.038	0.031	0.039	0.028
11	0	0	1	1	0.036	0.032	0.036	0.031
12	1	1	1	0	0.058	0.049	0.066	0.055
13	1	1	0	1	0.048	0.043	0.053	0.046
14	1	0	1	1	0.059	0.049	0.066	0.055
15	0	1	1	1	0.054	0.048	0.061	0.053
16	1	1	1	1	0.054	0.048	0.061	0.053



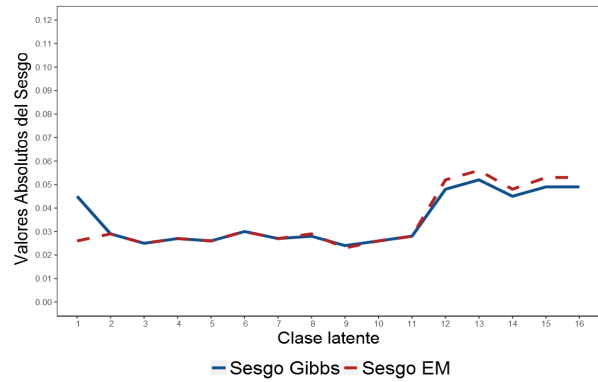
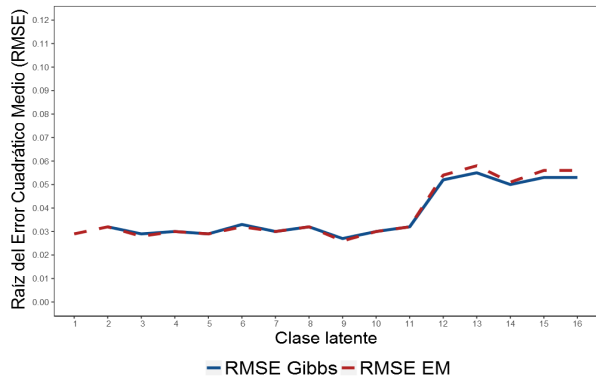
De la misma forma se presenta el resumen de resultados en los cuadros 4.49 y 4.50, para los parámetros de los ítems y las probabilidades de pertenencia, con un tamaño $N = 300$.

Cuadro 4.49: Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 6 con $N = 300$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico



Cuadro 4.50: Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 6 con $N = 300$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico

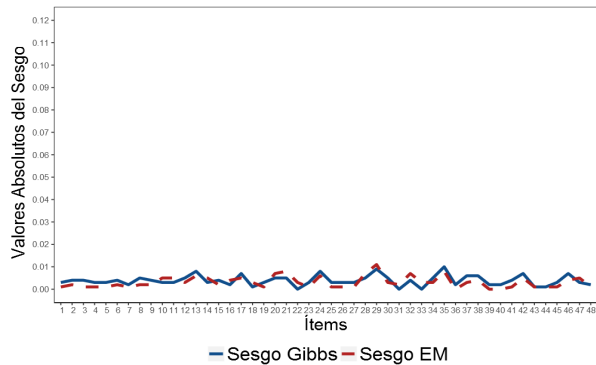
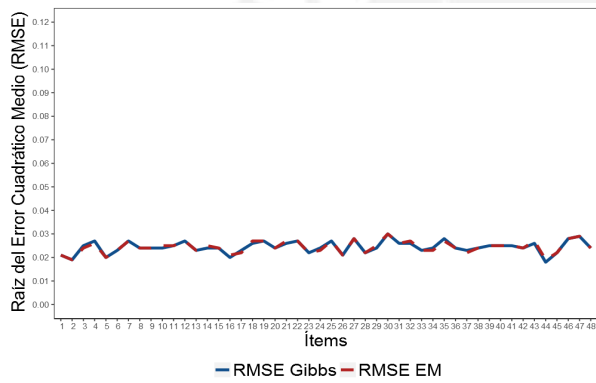
Clases latentes	Perfil de habilidades				Muestreador de Gibbs		Algoritmo EM	
	α_1	α_2	α_3	α_4	RMSE	Sesgo	RMSE	Sesgo
1	0	0	0	0	0.135	0.045	0.029	0.026
2	1	0	0	0	0.032	0.029	0.032	0.029
3	0	1	0	0	0.029	0.025	0.028	0.025
4	0	0	1	0	0.03	0.027	0.03	0.027
5	0	0	0	1	0.029	0.026	0.029	0.026
6	1	1	0	0	0.033	0.03	0.032	0.03
7	1	0	1	0	0.03	0.027	0.03	0.027
8	1	0	0	1	0.032	0.028	0.032	0.029
9	0	1	1	0	0.027	0.024	0.026	0.023
10	0	1	0	1	0.03	0.026	0.03	0.026
11	0	0	1	1	0.032	0.028	0.032	0.028
12	1	1	1	0	0.052	0.048	0.054	0.052
13	1	1	0	1	0.055	0.052	0.058	0.056
14	1	0	1	1	0.05	0.045	0.051	0.048
15	0	1	1	1	0.053	0.049	0.056	0.053
16	1	1	1	1	0.053	0.049	0.056	0.053



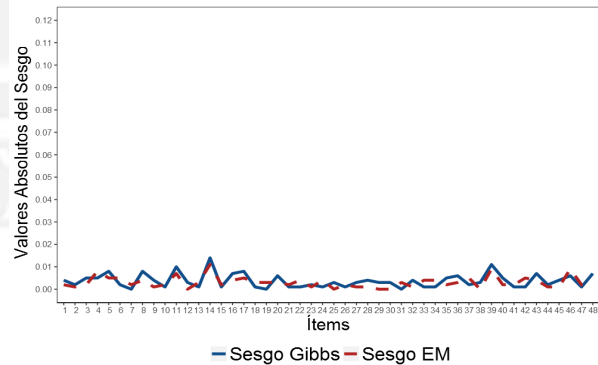
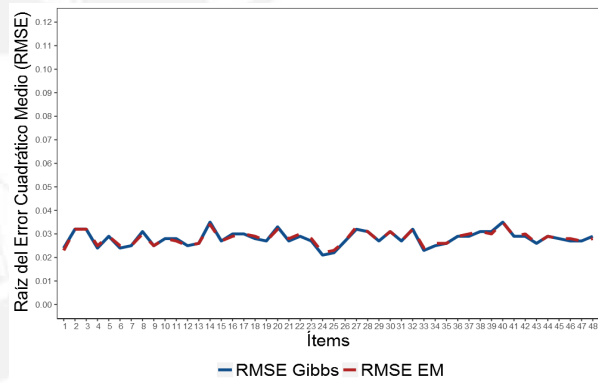
Finalmente, se presenta el resumen de resultados para el último escenario, en los cuadros 4.51 y 4.52, para los parámetros de los ítems y las probabilidades de pertenencia con un tamaño $N = 500$.

Cuadro 4.51: Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Deslíz (s) y Adivinación (g) en el Estudio 6 con $N = 500$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico

Deslíz (s)

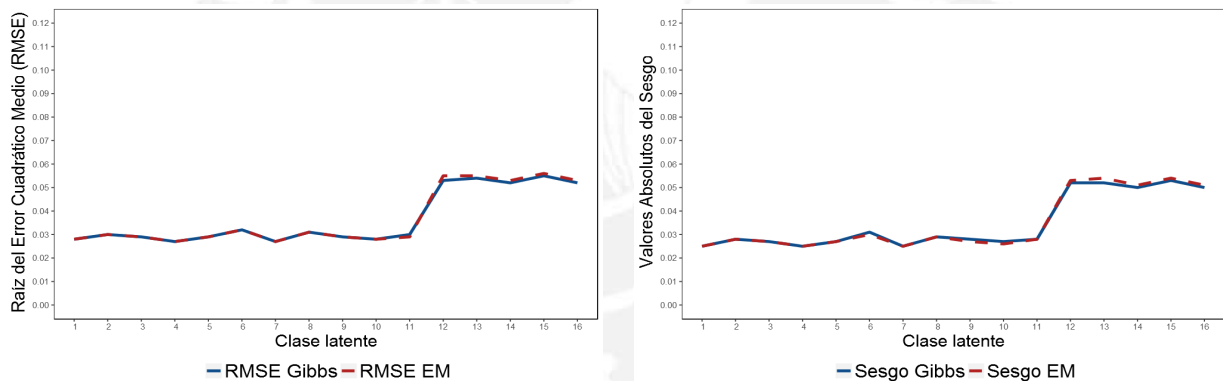


Adivinación (g)



Cuadro 4.52: Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 6 con $N = 500$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico

Clases latentes	Perfil de habilidades				Muestreador de Gibbs		Algoritmo EM	
	α_1	α_2	α_3	α_4	RMSE	Sesgo	RMSE	Sesgo
1	0	0	0	0	0.028	0.025	0.028	0.025
2	1	0	0	0	0.03	0.028	0.03	0.028
3	0	1	0	0	0.029	0.027	0.029	0.027
4	0	0	1	0	0.027	0.025	0.027	0.025
5	0	0	0	1	0.029	0.027	0.029	0.027
6	1	1	0	0	0.032	0.031	0.032	0.03
7	1	0	1	0	0.027	0.025	0.027	0.025
8	1	0	0	1	0.031	0.029	0.031	0.029
9	0	1	1	0	0.029	0.028	0.029	0.027
10	0	1	0	1	0.028	0.027	0.028	0.026
11	0	0	1	1	0.03	0.028	0.029	0.028
12	1	1	1	0	0.053	0.052	0.055	0.053
13	1	1	0	1	0.054	0.052	0.055	0.054
14	1	0	1	1	0.052	0.05	0.053	0.051
15	0	1	1	1	0.055	0.053	0.056	0.054
16	1	1	1	1	0.052	0.05	0.053	0.051



De acuerdo a lo observado en los tres escenarios, se presentan los cuadros 4.53 y 4.54 que resume lo hallado por medio de comparación de medias de los valores RMSE y del sesgo (en valores absolutos), respectivamente:

Cuadro 4.53: Resumen de medias de los valores RMSE para todos los parámetros estudiados

Escenarios	Tamaño de la población (N)	Desliz (s)		Adivinación (g)		Prob de cada clase latente (π)	
		Gibbs	EM	Gibbs	EM	Gibbs	EM
2	300	0.031	0.031	0.062	0.035	0.044	0.038
3	500	0.024	0.024	0.028	0.028	0.037	0.037

Cuadro 4.54: Resumen de medias de los valores del sesgo para todos los parámetros estudiados

Escenarios	Tamaño de la población (N)	Desliz (s)		Adivinación (g)		Prob de cada clase latente (π)	
		Gibbs	EM	Gibbs	EM	Gibbs	EM
1	100	0.011	0.006	0.014	0.008	0.035	0.035
2	300	0.004	0.004	0.012	0.004	0.035	0.035
3	500	0.004	0.003	0.004	0.003	0.035	0.035

Cuando la influencia de los parámetros de los ítems es débil en relación a las habilidades de los alumnos; y al igual que en el estudio 5, considerando que al parámetro π se le asignan dos valores de acuerdo a la cantidad de habilidades que conforman la clase latente; esta diferenciación la definiremos como un efecto alto del parámetro respecto a las habilidades de los alumnos.

Los resultados muestran que las estimaciones son insesgadas y con valores pequeños para el RMSE en ambos enfoques.

Como en la mayoría de los estudios previos, las diferencias suelen ser en los parámetros de los ítems, donde las diferencias más resaltantes para los valores del sesgo, se da para el parámetro de adivinación en el escenario 2, a favor del algoritmo EM. De la misma forma, las estimaciones para π son similares para todos los escenarios y considerando ambos indicadores. No hay claras ventajas de un enfoque respecto al otro.

4.11. Estudio 7

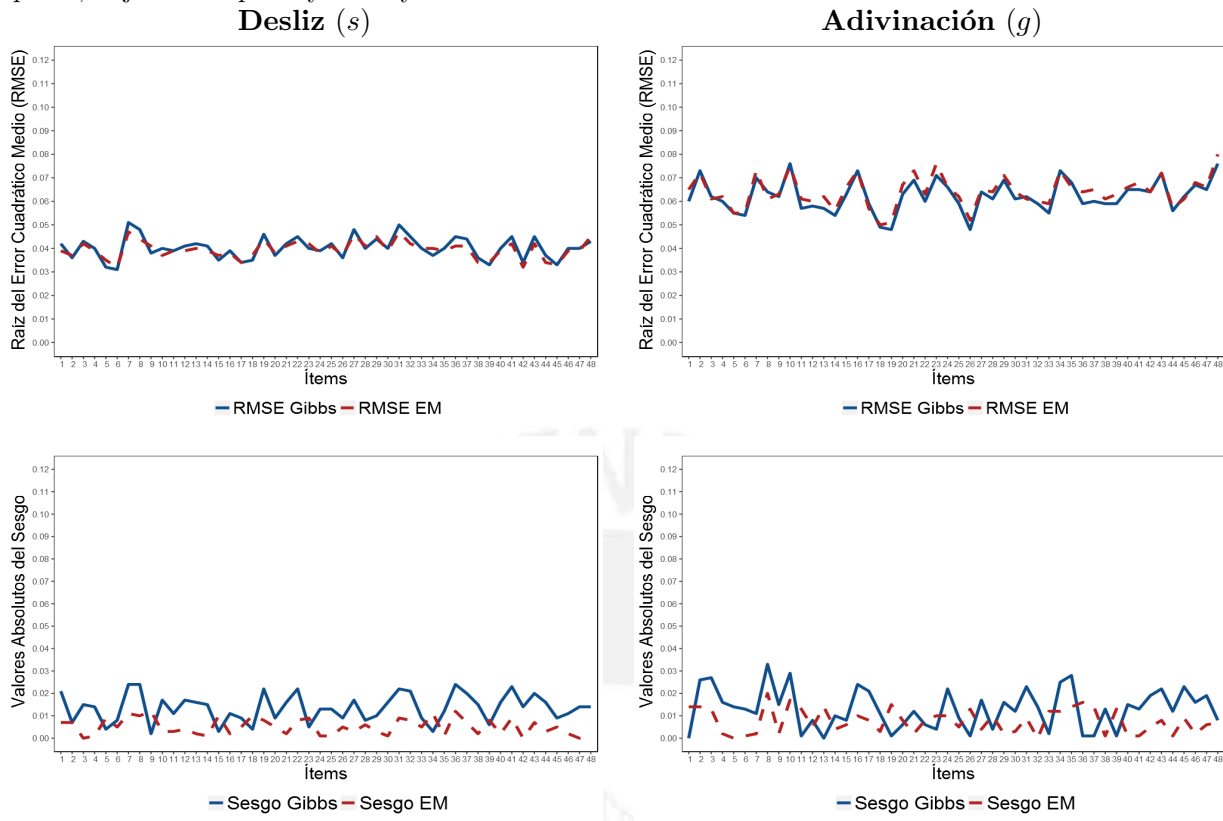
Para este estudio se presentan tres escenarios que se detallan en el cuadro 4.55, los cuales tienen como principal objetivo estudiar, bajo tres tamaños de población, la recuperación de los parámetros de los ítems. En este caso, se considerará el parámetro desliz ($s = 0.1$) con alto impacto y el parámetro adivinación ($g = 0.2$) de bajo impacto, con las habilidades de los individuos. Además, el parámetro π manteniendo una relación de alto impacto. Este parámetro ahora contará con dos valores: 0.021 y 0.042, cuando la clase o perfil latente está conformada por el dominio de dos o menos habilidades, y cuando lo está por tres a más habilidades; respectivamente:

Cuadro 4.55: Escenarios del Estudio 7 de simulación para una relación alta para el parámetro de desliz (s); aunque una relación débil para el parámetro de adivinación (g), con las habilidades de los individuos, con el parámetro π manteniendo una relación de alto impacto

Escenarios	Tamaño de población (N)	Adivinación (s)	Desliz (g)	Probabilidad de pertenencia π
1	100	0.1	0.2	0.021 o 0.042
2	300	0.1	0.2	0.021 o 0.042
3	500	0.1	0.2	0.021 o 0.042

Se presenta en el cuadro 4.56 con gráficos de línea, la lista de los 48 ítems con los valores de RMSE y sesgo para cada parámetro y con tamaño de población $N = 100$, de acuerdo a los dos enfoques estudiados:

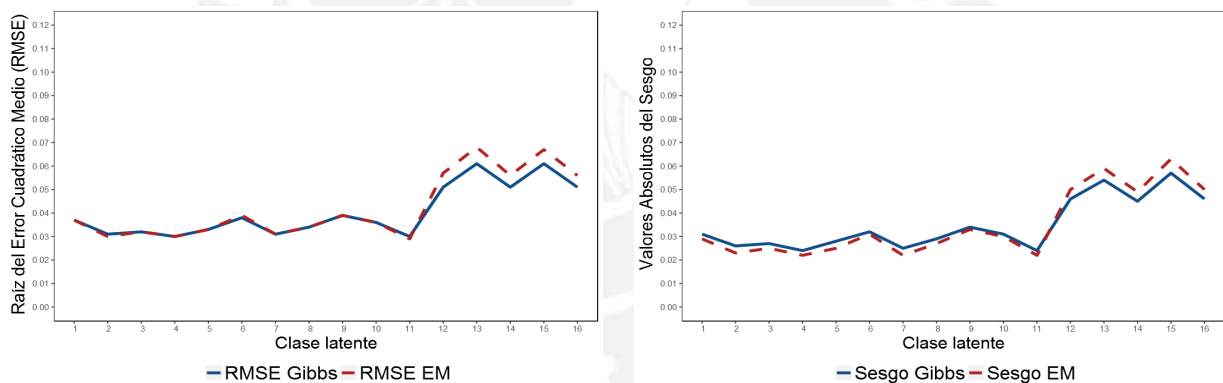
Cuadro 4.56: Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 7 con $N = 100$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico



En el siguiente cuadro 4.57 se presenta también la precisión de recuperación de parámetros para las probabilidades de pertenencia a las 2^K clases latentes, cuando se fija como $\pi = 2^{-K}$, manteniendo las características de los parámetros de los ítems mencionadas, y con $N = 100$.

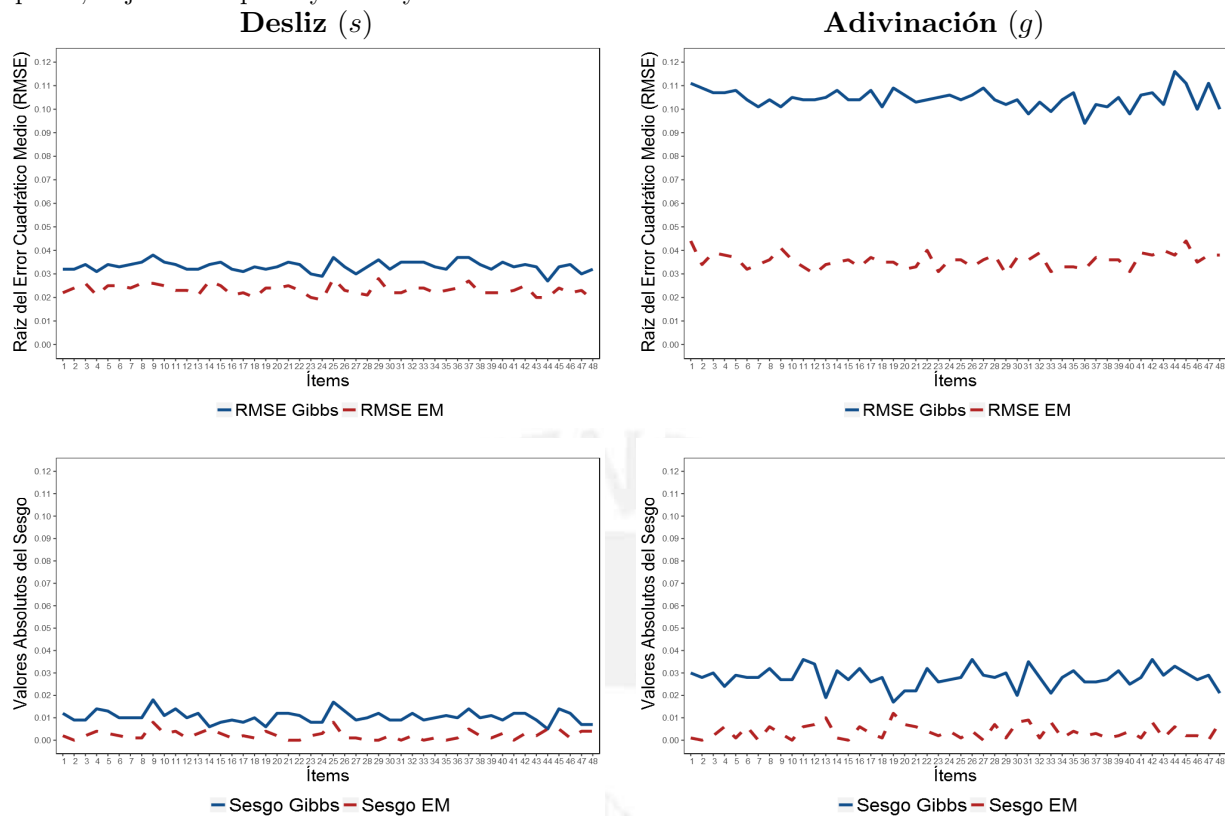
Cuadro 4.57: Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 7 con $N = 100$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico

Clases latentes	Perfil de habilidades				Muestreador de Gibbs	Algoritmo EM		
	α_1	α_2	α_3	α_4	RMSE	Sesgo	RMSE	Sesgo
1	0	0	0	0	0.037	0.031	0.037	0.029
2	1	0	0	0	0.031	0.026	0.03	0.023
3	0	1	0	0	0.032	0.027	0.032	0.025
4	0	0	1	0	0.03	0.024	0.03	0.022
5	0	0	0	1	0.033	0.028	0.033	0.025
6	1	1	0	0	0.038	0.032	0.039	0.031
7	1	0	1	0	0.031	0.025	0.031	0.022
8	1	0	0	1	0.034	0.029	0.034	0.027
9	0	1	1	0	0.039	0.034	0.039	0.033
10	0	1	0	1	0.036	0.031	0.036	0.03
11	0	0	1	1	0.03	0.024	0.029	0.022
12	1	1	1	0	0.051	0.046	0.057	0.05
13	1	1	0	1	0.061	0.054	0.068	0.059
14	1	0	1	1	0.051	0.045	0.056	0.049
15	0	1	1	1	0.061	0.057	0.067	0.063
16	1	1	1	1	0.051	0.046	0.056	0.05



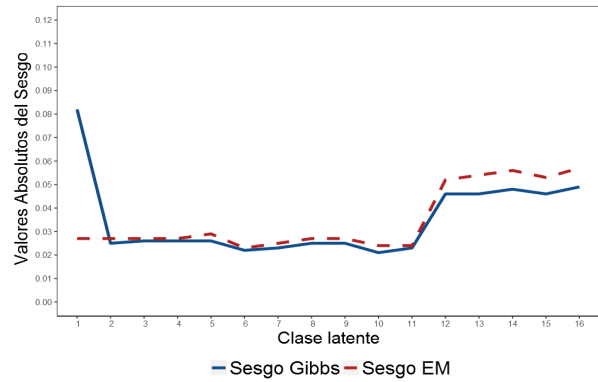
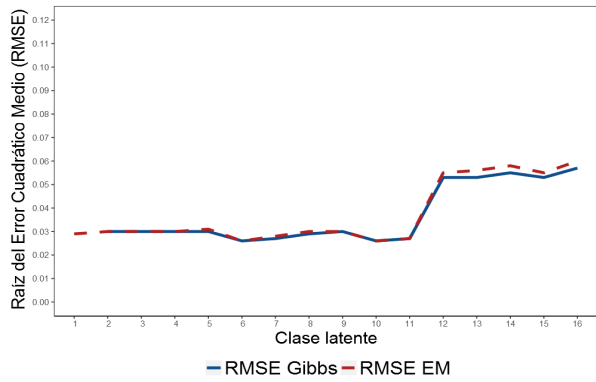
De la misma forma se presenta el resumen de resultados en los cuadros 4.58 y 4.59, para los parámetros de los ítems y las probabilidades de pertenencia, con un tamaño $N = 300$.

Cuadro 4.58: Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 7 con $N = 300$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico



Cuadro 4.59: Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 7 con $N = 300$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico

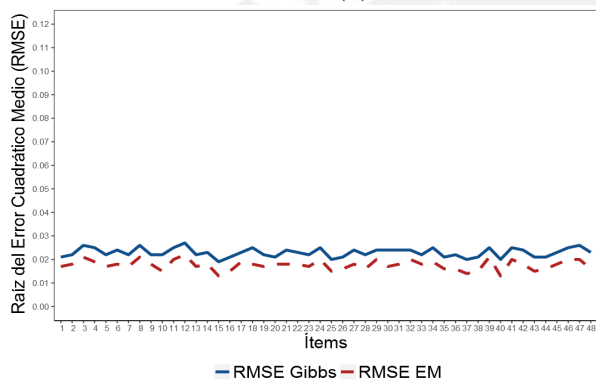
Clases latentes	Perfil de habilidades				Muestreador de Gibbs		Algoritmo EM	
	α_1	α_2	α_3	α_4	RMSE	Sesgo	RMSE	Sesgo
1	0	0	0	0	0.23	0.082	0.029	0.027
2	1	0	0	0	0.03	0.025	0.03	0.027
3	0	1	0	0	0.03	0.026	0.03	0.027
4	0	0	1	0	0.03	0.026	0.03	0.027
5	0	0	0	1	0.03	0.026	0.031	0.029
6	1	1	0	0	0.026	0.022	0.026	0.023
7	1	0	1	0	0.027	0.023	0.028	0.025
8	1	0	0	1	0.029	0.025	0.03	0.027
9	0	1	1	0	0.03	0.025	0.03	0.027
10	0	1	0	1	0.026	0.021	0.026	0.024
11	0	0	1	1	0.027	0.023	0.027	0.024
12	1	1	1	0	0.053	0.046	0.055	0.052
13	1	1	0	1	0.053	0.046	0.056	0.054
14	1	0	1	1	0.055	0.048	0.058	0.056
15	0	1	1	1	0.053	0.046	0.055	0.053
16	1	1	1	1	0.057	0.049	0.06	0.057



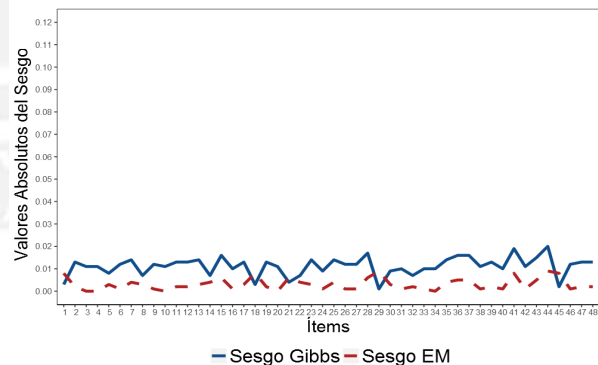
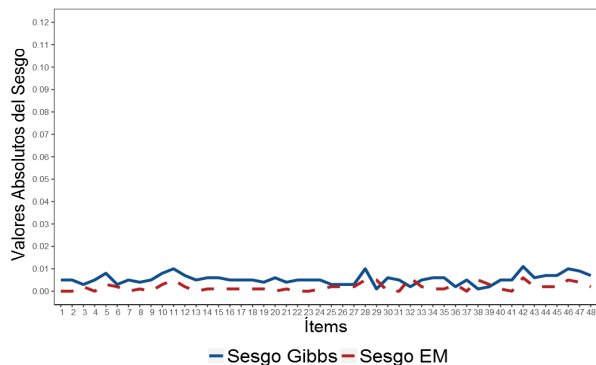
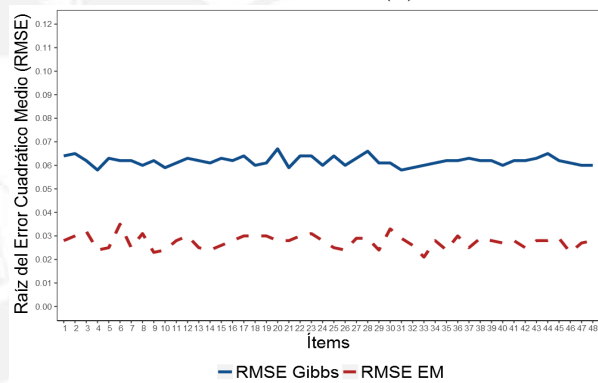
Finalmente, se presenta el resumen de resultados para el último escenario, en los cuadros 4.60 y 4.61, para los parámetros de los ítems y las probabilidades de pertenencia con un tamaño $N = 500$.

Cuadro 4.60: Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Deslíz (s) y Adivinación (g) en el Estudio 7 con $N = 500$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico

Deslíz (s)

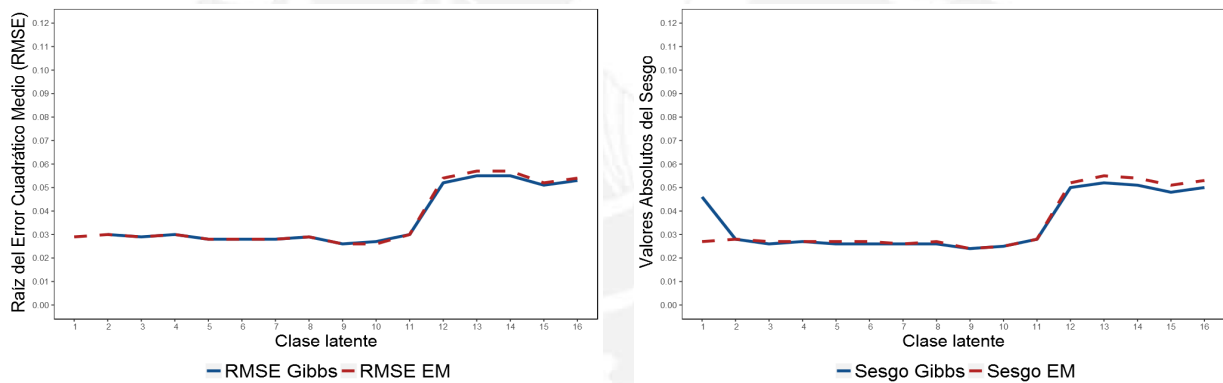


Adivinación (g)



Cuadro 4.61: Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 7 con $N = 500$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico

Clases latentes	Perfil de habilidades				Muestreador de Gibbs		Algoritmo EM	
	α_1	α_2	α_3	α_4	RMSE	Sesgo	RMSE	Sesgo
1	0	0	0	0	0.137	0.046	0.029	0.027
2	1	0	0	0	0.03	0.028	0.03	0.028
3	0	1	0	0	0.029	0.026	0.029	0.027
4	0	0	1	0	0.03	0.027	0.03	0.027
5	0	0	0	1	0.028	0.026	0.028	0.027
6	1	1	0	0	0.028	0.026	0.028	0.027
7	1	0	1	0	0.028	0.026	0.028	0.026
8	1	0	0	1	0.029	0.026	0.029	0.027
9	0	1	1	0	0.026	0.024	0.026	0.024
10	0	1	0	1	0.027	0.025	0.026	0.025
11	0	0	1	1	0.03	0.028	0.03	0.028
12	1	1	1	0	0.052	0.05	0.054	0.052
13	1	1	0	1	0.055	0.052	0.057	0.055
14	1	0	1	1	0.055	0.051	0.057	0.054
15	0	1	1	1	0.051	0.048	0.052	0.051
16	1	1	1	1	0.053	0.05	0.054	0.053



De acuerdo a lo observado en los tres escenarios, se presentan los cuadros 4.62 y 4.63 que resume lo hallado por medio de comparación de medias de los valores RMSE y del sesgo (en valores absolutos), respectivamente:

Cuadro 4.62: Resumen de medias de los valores RMSE para todos los parámetros estudiados

Escenarios	Tamaño de la población (N)	Desliz (s)		Adivinación (g)		Prob de cada clase latente (π)	
		Gibbs	EM	Gibbs	EM	Gibbs	EM
1	100	0.040	0.039	0.062	0.064	0.040	0.042
2	300	0.033	0.023	0.105	0.036	0.049	0.038
3	500	0.023	0.018	0.062	0.028	0.043	0.037

Cuadro 4.63: Resumen de medias de los valores del sesgo para todos los parámetros estudiados

Escenarios	Tamaño de la población (N)	Desliz (s)		Adivinación (g)		Prob de cada clase latente (π)	
		Gibbs	EM	Gibbs	EM	Gibbs	EM
1	100	0.014	0.005	0.013	0.008	0.035	0.035
2	300	0.010	0.002	0.028	0.004	0.035	0.035
3	500	0.005	0.002	0.011	0.003	0.035	0.035

Cuando la influencia del parámetro de adivinación es baja y la del desliz es alta, con respecto a las habilidades de los alumnos; y considerando también una alta influencia del parámetro π con las habilidades.

Notamos que en este estudio se manejan valores diferentes para cada uno de los parámetros mencionados, lo cual puede acercarse más a lo que podría ocurrir en una prueba.

De acuerdo a ello, los resultados se mantienen con buenos indicadores; es decir, son insesgados y con valores bajos del RMSE. Aunque las diferencias entre ambos enfoques son más notorias en los escenarios 2 y 3 cuando analizamos los parámetros de los ítems.

Finalmente, ha sido casi una constante que las estimaciones para el π son similares utilizando cualquiera de los dos enfoques. No obstante, este a veces muestra ligeras ventajas en tamaños de población pequeños a favor del muestreador de Gibbs.

4.12. Estudio 8

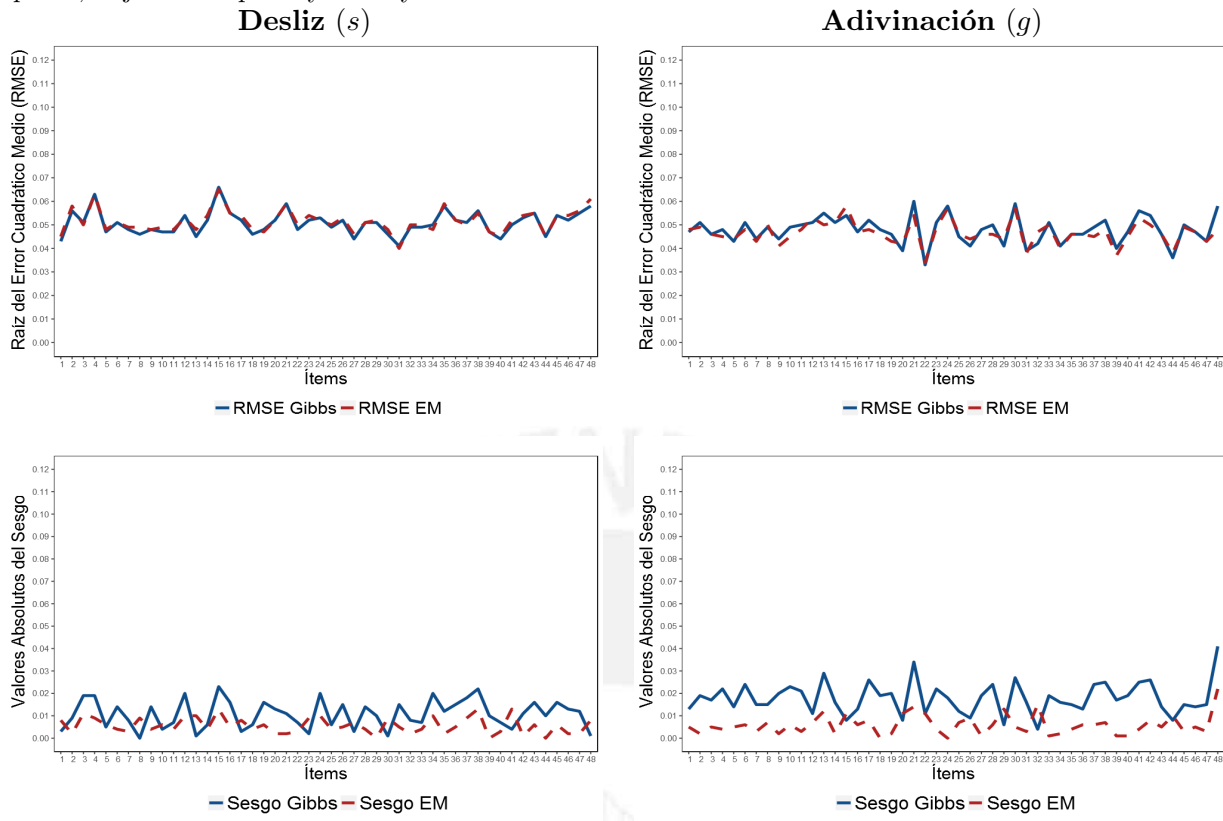
Para este estudio se presentan tres escenarios que se detallan en el cuadro 4.64, los cuales tienen como principal objetivo estudiar, bajo tres tamaños de población, la recuperación de los parámetros de los ítems. En este caso, se considerará la relación para el parámetro desliz ($s = 0.2$) con bajo impacto y el parámetro adivinación ($g = 0.1$) de alto impacto, con las habilidades de los individuos. Además, el parámetro π manteniendo una relación de alto impacto. Este parámetro ahora contará con dos valores: 0.021 y 0.042, cuando la clase o perfil latente está conformada por el dominio de dos o menos habilidades, y cuando lo está por tres a más habilidades; respectivamente:

Cuadro 4.64: Escenarios del Estudio 8 de simulación para una relación baja para el parámetro de desliz (s); aunque una relación fuerte para el parámetro de adivinación (g), con las habilidades de los individuos, con el parámetro π manteniendo una relación de alto impacto

Escenarios	Tamaño de población (N)	Adivinación (s)	Desliz (g)	Probabilidad de pertenencia π
1	100	0.2	0.1	0.021 o 0.042
2	300	0.2	0.1	0.021 o 0.042
3	500	0.2	0.1	0.021 o 0.042

Se presenta en el cuadro 4.65 con gráficos de líneas, la lista de los 48 ítems con los valores de RMSE y sesgo para cada parámetro y con tamaño de población $N = 100$, de acuerdo a los dos enfoques estudiados:

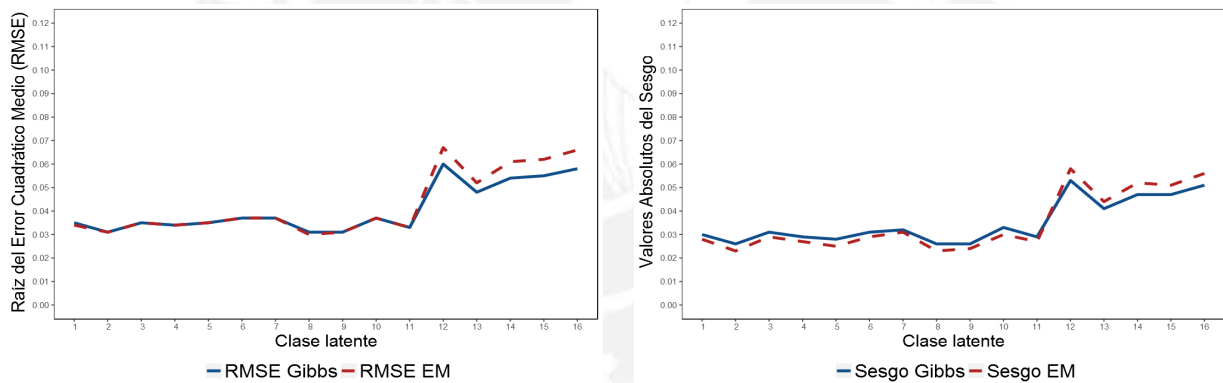
Cuadro 4.65: Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 8 con $N = 100$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico



En el siguiente cuadro 4.66 se presenta también la precisión de recuperación de parámetros para las probabilidades de pertenencia a las 2^K clases latentes, cuando se fija como $\pi = 2^{-K}$, manteniendo las características de los parámetros de los ítems mencionadas, y con $N = 100$.

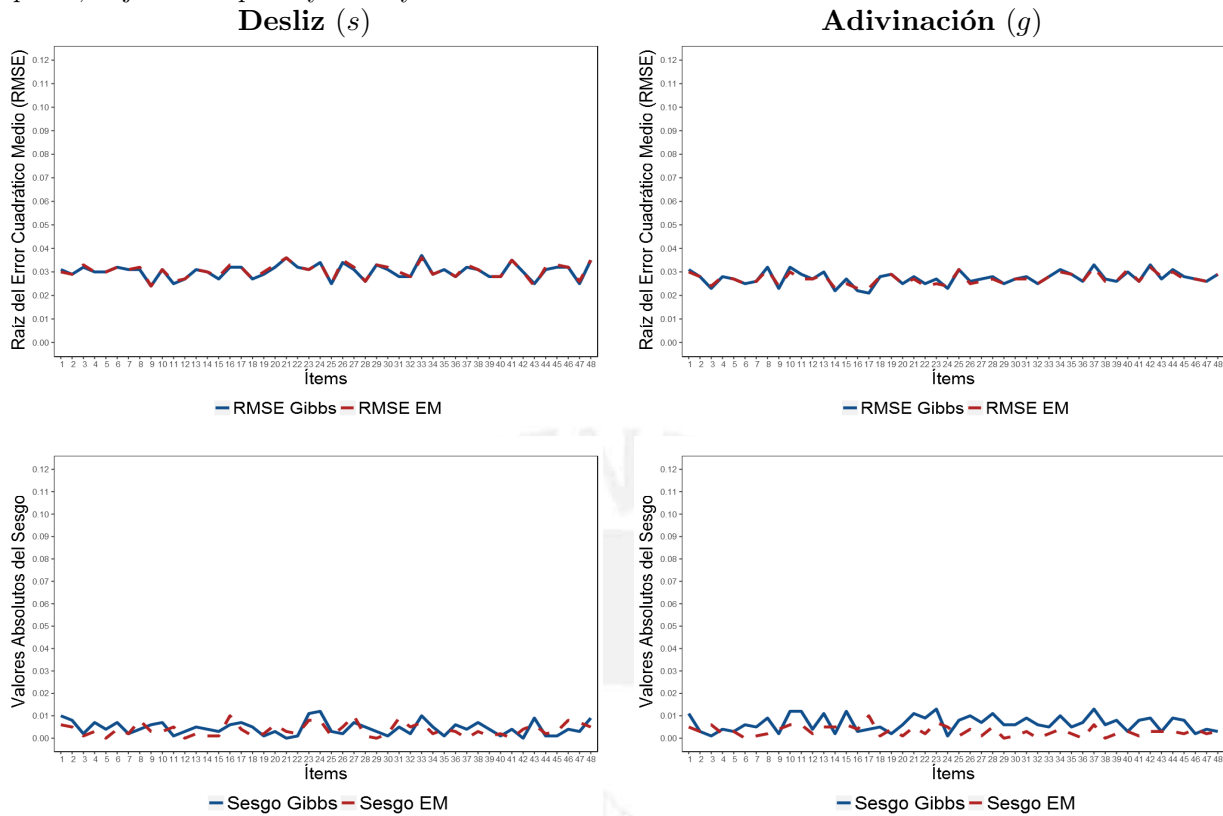
Cuadro 4.66: Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 8 con $N = 100$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico

Clases latentes	Perfil de habilidades				Muestreador de Gibbs		Algoritmo EM	
	α_1	α_2	α_3	α_4	RMSE	Sesgo	RMSE	Sesgo
1	0	0	0	0	0.035	0.03	0.034	0.028
2	1	0	0	0	0.031	0.026	0.031	0.023
3	0	1	0	0	0.035	0.031	0.035	0.029
4	0	0	1	0	0.034	0.029	0.034	0.027
5	0	0	0	1	0.035	0.028	0.035	0.025
6	1	1	0	0	0.037	0.031	0.037	0.029
7	1	0	1	0	0.037	0.032	0.037	0.031
8	1	0	0	1	0.031	0.026	0.03	0.023
9	0	1	1	0	0.031	0.026	0.031	0.024
10	0	1	0	1	0.037	0.033	0.037	0.03
11	0	0	1	1	0.033	0.029	0.033	0.027
12	1	1	1	0	0.06	0.053	0.067	0.058
13	1	1	0	1	0.048	0.041	0.052	0.044
14	1	0	1	1	0.054	0.047	0.061	0.052
15	0	1	1	1	0.055	0.047	0.062	0.051
16	1	1	1	1	0.058	0.051	0.066	0.056



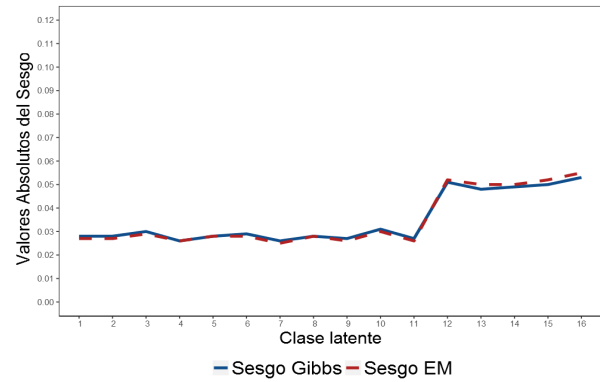
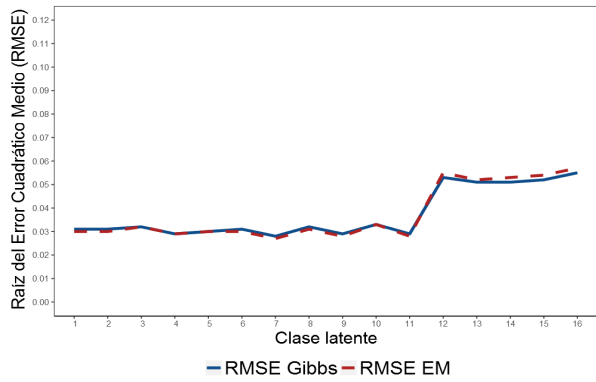
De la misma forma se presenta el resumen de resultados en los cuadros 4.67 y 4.68, para los parámetros de los ítems y las probabilidades de pertenencia, con un tamaño $N = 300$.

Cuadro 4.67: Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Desliz (s) y Adivinación (g) en el Estudio 8 con $N = 300$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico



Cuadro 4.68: Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 8 con $N = 300$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico

Clases latentes	Perfil de habilidades				Muestreador de Gibbs		Algoritmo EM	
	α_1	α_2	α_3	α_4	RMSE	Sesgo	RMSE	Sesgo
1	0	0	0	0	0.031	0.028	0.03	0.027
2	1	0	0	0	0.031	0.028	0.03	0.027
3	0	1	0	0	0.032	0.03	0.032	0.029
4	0	0	1	0	0.029	0.026	0.029	0.026
5	0	0	0	1	0.03	0.028	0.03	0.028
6	1	1	0	0	0.031	0.029	0.03	0.028
7	1	0	1	0	0.028	0.026	0.027	0.025
8	1	0	0	1	0.032	0.028	0.031	0.028
9	0	1	1	0	0.029	0.027	0.028	0.026
10	0	1	0	1	0.033	0.031	0.033	0.03
11	0	0	1	1	0.029	0.027	0.028	0.026
12	1	1	1	0	0.053	0.051	0.055	0.052
13	1	1	0	1	0.051	0.048	0.052	0.05
14	1	0	1	1	0.051	0.049	0.053	0.05
15	0	1	1	1	0.052	0.05	0.054	0.052
16	1	1	1	1	0.055	0.053	0.057	0.055

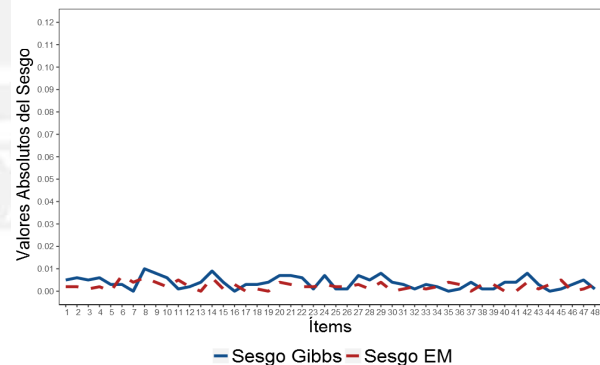
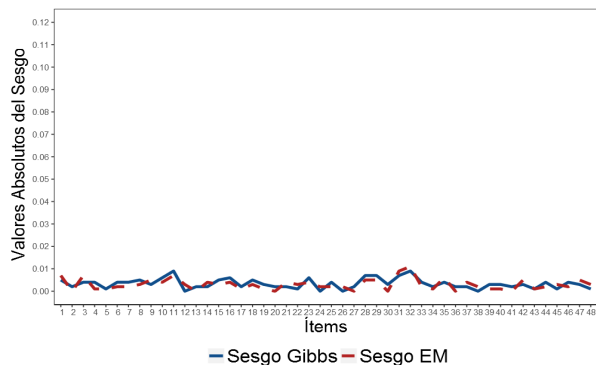
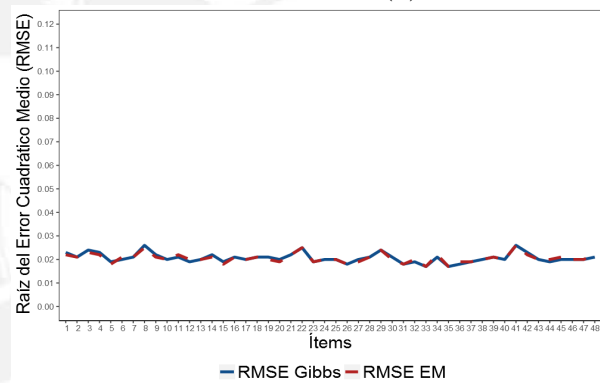
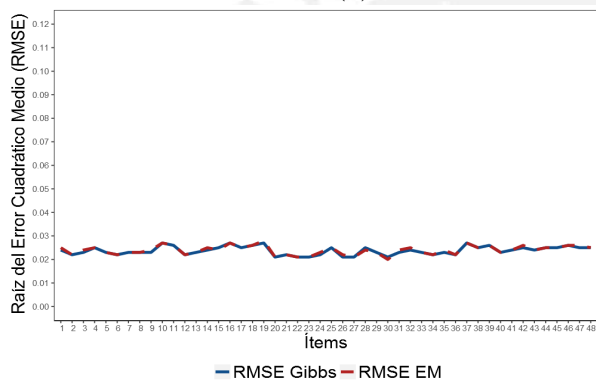


Finalmente, se presenta el resumen de resultados para el último escenario, en los cuadros 4.69 y 4.70, para los parámetros de los ítems y las probabilidades de pertenencia con un tamaño $N = 500$.

Cuadro 4.69: Raíz del error cuadrático (RMSE) y el Sesgo de las estimaciones de los parámetros de los ítems: Deslíz (s) y Adivinación (g) en el Estudio 8 con $N = 500$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico

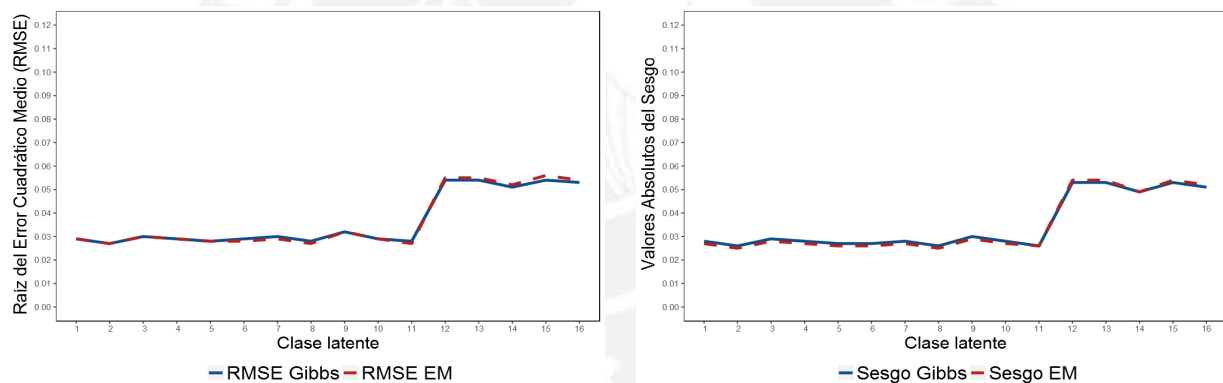
Deslíz (s)

Adivinación (g)



Cuadro 4.70: Raíz del error cuadrático (RMSE) y el sesgo de las estimaciones de las probabilidades de pertenencia de cada clase (π) en el Estudio 8 con $N = 500$ del modelo DINA considerando $m = 50$ réplicas, bajo el enfoque bayesiano y clásico

Clases latentes	Perfil de habilidades				Muestreador de Gibbs		Algoritmo EM	
	α_1	α_2	α_3	α_4	RMSE	Sesgo	RMSE	Sesgo
1	0	0	0	0	0.029	0.028	0.029	0.027
2	1	0	0	0	0.027	0.026	0.027	0.025
3	0	1	0	0	0.03	0.029	0.03	0.028
4	0	0	1	0	0.029	0.028	0.029	0.027
5	0	0	0	1	0.028	0.027	0.028	0.026
6	1	1	0	0	0.029	0.027	0.028	0.026
7	1	0	1	0	0.03	0.028	0.029	0.027
8	1	0	0	1	0.028	0.026	0.027	0.025
9	0	1	1	0	0.032	0.03	0.032	0.029
10	0	1	0	1	0.029	0.028	0.029	0.027
11	0	0	1	1	0.028	0.026	0.027	0.026
12	1	1	1	0	0.054	0.053	0.055	0.054
13	1	1	0	1	0.054	0.053	0.055	0.054
14	1	0	1	1	0.051	0.049	0.052	0.049
15	0	1	1	1	0.054	0.053	0.056	0.054
16	1	1	1	1	0.053	0.051	0.054	0.052



De acuerdo a lo observado en los tres escenarios, se presentan los cuadros 4.71 y 4.72 que resume lo hallado por medio de comparación de medias de los valores RMSE y del sesgo (en valores absolutos), respectivamente:

Cuadro 4.71: Resumen de medias de los valores RMSE para todos los parámetros estudiados

Escenarios	Tamaño de la población (N)	Desliz (s)		Adivinación (g)		Prob de cada clase latente (π)	
		Gibbs	EM	Gibbs	EM	Gibbs	EM
1	100	0.051	0.052	0.048	0.047	0.041	0.043
2	300	0.024	0.024	0.021	0.021	0.037	0.037
3	500	0.030	0.031	0.027	0.027	0.037	0.037

Cuadro 4.72: Resumen de medias de los valores del sesgo para todos los parámetros estudiados

Escenarios	Tamaño de la población (N)	Desliz (s)		Adivinación (g)		Prob de cada clase latente (π)	
		Gibbs	EM	Gibbs	EM	Gibbs	EM
1	100	0.011	0.006	0.018	0.006	0.035	0.035
2	300	0.005	0.004	0.007	0.003	0.035	0.035
3	500	0.003	0.003	0.004	0.002	0.035	0.035

A diferencia del estudio anterior, en este se tiene en cuenta una alta influencia del parámetro de adivinación, y por el contrario, una débil entre el parámetro desliz, y las habilidades de los alumnos. Asimismo, se considera al parámetro π con una alta relación, tal como se describió en el estudio previo.

Este estudio también es más cercano a lo que podría ocurrir en una prueba regular, ya que se están considerando diferentes valores para los parámetros mencionados.

En líneas generales, los resultados obtenidos muestran que no existen mayores diferencias entre un enfoque y otro, tomando en cuenta los tres tamaños de población.

Finalmente, como una recomendación general, es importante no solo analizar los cuadros 4.71 y 4.72 donde se muestran las medias, sino también los cuadros 4.65, 4.66, 4.67, 4.68, 4.69, 4.70, donde se detallan las estimaciones por cada uno de los 48 ítems expuestos y la distribución de las 16 clases, en los tres escenarios.



Capítulo 5

Aplicación

Como se ha mencionado anteriormente, la aplicación de este trabajo se orienta al ámbito educativo; principalmente a la evaluación de la competencia matemática, en una muestra aleatoria de alumnos del 2do grado de secundaria.

Esta muestra ha sido extraída de la Evaluación Censal de Estudiantes (ECE) del 2do grado de secundaria, la cual fue ejecutada por el Ministerio de Educación (Minedu) a través de la Oficina de Medición de la Calidad de los Aprendizajes (UMC) en el año 2015. Cabe precisar que solo son incluidas en el marco muestral las instituciones educativas que tienen 5 o más estudiantes matriculados.

Vale comentar que esta evaluación ha sido aplicada por primera vez a alumnos de este grado en nuestro país, priorizándose dos competencias: lectura y matemática. Ella es considerada un indicador importante para la toma de decisiones de políticas educativas a nivel nacional y brinda información a nivel regional, área geográfica de la institución educativa (IE), gestión con que se maneja la escuela sea estatal (pública) o no estatal (privada), y sexo del estudiante.

Es importante resaltar que el indicador mencionado es la medida de habilidad latente del estudiante en el área de interés; es decir, rendimiento en lectura y en matemática. Esta medida es obtenida en el análisis psicométrico de las pruebas aplicadas en la ECE, el cual se basó en el modelo Rasch para ítems dicotómicos. Este relaciona la probabilidad de respuesta de una persona a un ítem en términos de la diferencia entre la medida de rasgo o habilidad latente que la persona posee (θ) y la medida de dificultad (β) del ítem utilizado.

En relación con lo anterior, esta aplicación se dividirá principalmente en dos partes. La primera consistirá en la estimación de los parámetros del modelo propuesto utilizando el paquete *dina* (Culpepper, 2015b) del software libre R (R Core Team, 2016). Mientras que la segunda será la aplicación del modelo Rasch que, siguiendo las recomendaciones de Chincaro (2010), es más preciso realizar la estimación de los parámetros bajo el enfoque bayesiano. Para el análisis se utilizará también el software libre R conjuntamente con el Winbugs 1.4.3 (Lunn et al., 2007).

Las rutinas de código de la aplicación de ambos modelos se encuentran en el apéndice B.

Los insumos obtenidos a partir de lo anterior serán necesarios para realizar un análisis de correspondencias entre los perfiles latentes de los alumnos que genera el DINA y las medidas de habilidad latente de los mismos que da como resultado el modelo de Rasch. Así como estudiar la relación de los parámetros del ítem: Adivinación y Desliz, con la dificultad y un indicador cercano al concepto de discriminación del ítem; respectivamente.

En resumen, este trabajo busca complementar, dar otra visión o brindar mayor información acerca de la descripción de las habilidades matemáticas de los estudiantes evaluados, a través de

los grupos latentes conformados no solamente por una sino por varias dimensiones de habilidades en la competencia matemática.

5.1. Instrumento de medición

En líneas generales, el instrumento de medición es una prueba integrada de lectura y matemática que se aplicó en dos días. Esta prueba seguía un diseño de bloques rotados; es decir, los alumnos no rindieron el mismo conjunto de ítems sino que cada uno desarrollaba un subconjunto del total propuesto para cada competencia (el detalle del diseño usado para matemática se puede observar en los cuadros 5.1 y 5.2).

Se planteó este diseño debido a que en secundaria los aprendizajes que alcanzan los estudiantes amplían y profundizan lo logrado en primaria; es necesario cubrir un amplio espectro de capacidades y conocimientos, de manera que los resultados sean más precisos a nivel conceptual de lo que realmente pueden hacer los estudiantes. Sin embargo, las evaluaciones de carácter censal suelen enfrentar la restricción de medir con una prueba única (y las consiguientes limitaciones del tiempo de evaluación y el número de ítems) las habilidades de los estudiantes. Para lidiar con esta restricción, se estudió la posibilidad de utilizar diseños que contengan bloques de ítems comunes a todos los estudiantes y bloques no comunes que se distribuirían aleatoriamente entre la población (Minedu, 2016).

Como se mencionó en el primer párrafo, este tipo de diseño implica que el estudiante resuelva un subconjunto de ítems de todos los posibles, a los cuales llamaremos bloques, y estos a su vez se agrupan en cuadernillos. Finalmente, un tipo de prueba está conformado por dos cuadernillos.

Específicamente para matemática, se elaboraron 10 cuadernillos; lo que genera, 5 tipos de prueba; cada una de ellas conformada por 50 ítems repartidos en dos días. Es así que un alumno se enfrenta a 25 preguntas de matemática tanto el primer como el segundo día de aplicación; ya que los otros 25 son de lectura.

Gráficamente, la matriz 5.1 presenta el diseño de bloques de matemática para el primer día de aplicación y la matriz 5.2 para el segundo día de aplicación.

Cuadro 5.1: Matriz de diseño de bloques de Matemática para el día 1

Cuadernillo	Bloque	Ítems	
		Por bloque	Por cuadernillo
1	BC1	20	25
	B1	5	
2	BC1	20	25
	B2	5	
3	BC1	20	25
	B3	5	
4	BC1	20	25
	B4	5	
5	BC1	20	25
	B5	5	

Cuadro 5.2: Matriz de diseño de bloques de Matemática para el día 2

Cuadernillo	Bloque	Ítems	
		Por bloque	Por cuadernillo
6	BC1	20	25
	B6	5	
7	BC1	20	25
	B7	5	
8	BC1	20	25
	B8	5	
9	BC1	20	25
	B9	5	
10	BC1	20	25
	B10	5	

Todos los cuadernillos han sido construidos con ítems ya probados en evaluaciones piloto anteriores; por lo que, se ha tenido mucha consideración de que, en promedio, todos los tipos de prueba tengan la misma dificultad. Es por ello, que por razones prácticas solo se eligió uno de los cinco tipos de prueba como marco poblacional, la combinación elegida fue: cuadernillo 1 y cuadernillo 6.

5.2. Muestra

La combinación de cuadernillos elegida fue aplicada a 97168 estudiantes, de los cuales 93773 asistieron a ambos días de aplicación. En total, estos alumnos rindieron 50 preguntas, de las cuales 48 fueron cerradas (dicotómicas) y 2 de respuesta abierta (con crédito parcial). Los ítems cerrados constaban de cuatro alternativas (A, B, C y D) donde solo una de ellas era la respuesta correcta y equivalía a un puntaje de 1, por lo que el resto tenían puntaje 0. Por otro lado, las preguntas abiertas eran aquellas donde el estudiante tenía que llegar a la respuesta mediante un procedimiento sustentado. Por la naturaleza del modelo, solo trabajaremos con los ítems dicotómicos.

Además de la variable sexo, se incluyó una variable denominada “estrato” que comprendía tres categorías:

- Urbano Público: asignado a todos los alumnos evaluados que estudian en una escuela de gestión estatal en un área geográfica urbana.
- Urbano Privada: asignado a todos los alumnos evaluados que estudian en una escuela de gestión no estatal en un área geográfica urbana.
- Rural: asignado a todos los alumnos evaluados que estudian en una escuela en un área geográfica rural sin importar que sea de gestión estatal o no estatal.

Se realizó un muestreo aleatorio simple para cada estrato. La fórmula utilizada para hallar el tamaño de muestra n_i en cada estrato para un error de 0.03, y una confianza del 95 % en las estimación de cualquier proporción (Lohr, 1999) fue la siguiente:

$$n_i = \frac{(Z_{1-\alpha/2}^2 p(1-p) + e^2) N_i}{Z_{1-\alpha/2}^2 p(1-p) + e^2 N_i}, \quad (5.1)$$

donde N_i es el tamaño del estrato i .

En suma, la muestra total $n = n_1 + n_2 + n_3$ resultó en 3040 casos.

5.3. Análisis de los datos mediante el modelo DINA

Es fundamental que antes de realizar las estimaciones de los parámetros que propone el modelo, ya sea por la vía clásica o bajo la perspectiva bayesiana, se definan pedagógicamente las capacidades y/o habilidades matemáticas que alumno podría dominar para poder resolver los ítems planteados en la evaluación. En otras palabras, definir uno de los dos insumos principales que se utilizan como input en este análisis. Para esto es necesario caracterizar las 4 habilidades que se plantean para esta evaluación. Según el documento disponible Rutas del Aprendizaje de Matemática del [Minedu \(2015\)](#), estas se presentan en la siguiente tabla 5.3:

Cuadro 5.3: Descripción de las habilidades matemáticas que se evalúan en la prueba de matemática

Parámetro	Habilidad	Descripción
α_1	Matematiza situaciones (MA)	Es la capacidad de expresar un problema, reconocido en una situación, en un modelo matemático. En su desarrollo se usa, interpreta y evalúa el modelo matemático, de acuerdo a la situación que le dio origen.
α_2	Comunica y representa las ideas matemáticas (CR)	Es la capacidad de comprender el significado de las ideas matemáticas, y expresarlas en forma oral y escrita usando el lenguaje matemático y diversas formas de representación con material concreto, gráfico, tablas, símbolos y recursos TIC, y transitando de una representación a otra.
α_3	Elabora y usa estrategias (EE)	Es la capacidad de planificar, ejecutar y valorar una secuencia organizada de estrategias y diversos recursos, entre ellos las tecnologías de información y comunicación, empleándolas de manera flexible y eficaz en el planteamiento y resolución de problemas, incluidos los matemáticos. Esto implica ser capaz de elaborar un plan de solución, monitorear su ejecución, pudiendo incluso reformular el plan en el mismo proceso con la finalidad de llegar a la meta. Asimismo, revisar todo el proceso de resolución, reconociendo si las estrategias y herramientas fueron usados de manera apropiada y óptima.
α_4	Razona y argumenta generando ideas matemáticas (RA)	Es la capacidad de plantear supuestos, conjeturas e hipótesis de implicancia matemática mediante diversas formas de razonamiento (deductivo, inductivo y abductivo), así como el verificarlos y validarlos usando argumentos. Esto implica partir de la exploración de situaciones vinculadas a la matemática para establecer relaciones entre ideas, establecer conclusiones a partir de inferencias y deducciones que permitan generar nuevas conexiones e ideas matemáticas.

Después de ello, los especialistas en matemática de la UMC, definen la matriz \mathbf{Q} , la cual tiene una dimensión $J \times K$, donde J es la cantidad total de ítems ($J = 48$) y K la cantidad de habilidades ($K = 4$). Una porción de esta matriz, donde 1 significa que el j -ésimo ítem requiere de aquella habilidad y 0 en caso contrario, se presenta en el cuadro 5.4:

Cuadro 5.4: Porción de la matriz Q para la prueba de matemática

Ítem	MA	CR	EE	RA
1	1	0	0	0
2	1	0	0	0
3	0	0	1	0
4	0	0	0	1
5	0	0	0	1
⋮	⋮	⋮	⋮	⋮
J = 48	0	0	1	0

5.3.1. Estimación de los parámetros mediante el muestreador de Gibbs

Se realizó la estimación de los parámetros descritos para el modelo en estudio mediante el muestreador de Gibbs. Esta aplicación se realizó con 50000 iteraciones quemando las primeras 500 y con saltos de 5.

Los parámetros fueron estimados mediante el paquete **dina** (Culpepper, 2015b) del software R (R Core Team, 2016) y la convergencia del modelo se evaluó observando el comportamiento de la cadena para cada parámetro de los ítems, los cuales junto a las autocorrelaciones, se muestran en la sección D.2 del apéndice D.

Debido a que no hay evidencia para decir que la convergencia tuvo problemas, se presenta el resumen de los parámetros estimados para los ítems \hat{g} y $1 - \hat{s}$, los cuáles se interpretarían como la probabilidad de acertar el ítem a pesar de no dominar las habilidades requeridas para este, y la probabilidad de acertarlo cuando efectivamente debería hacerlo; respectivamente. Asimismo, debido a que existen $K = 4$ habilidades, la cantidad de clases posibles será $2^4 = 16$, las cuales son todas las posibles combinaciones que se pueden dar con esta cantidad de habilidades.

A continuación en el cuadro 5.5 se presentan los valores de las estimaciones de los parámetros mencionados de los ítems de la prueba de matemática:

Cuadro 5.5: Estimaciones de los parámetros g y s de los ítems de la prueba de matemática de 2do de secundaria

Ítem	Matriz Q				Estimación mediante el Muestreador de Gibbs			
	MA	CR	EE	RA	Adivinación (\hat{g})		1 - Desliz (\hat{s})	
	α_1	α_2	α_3	α_4	estimado	se	estimado	se
1	1	0	0	0	0.216	0.009	0.553	0.018
2	1	0	0	0	0.170	0.009	0.433	0.018
3	0	0	1	0	0.345	0.012	0.758	0.015
4	0	0	0	1	0.242	0.010	0.404	0.016
5	0	0	0	1	0.146	0.008	0.361	0.016
6	0	1	0	0	0.096	0.007	0.413	0.016
7	1	0	0	0	0.255	0.010	0.484	0.018
8	0	0	1	0	0.126	0.008	0.190	0.013
9	1	0	0	0	0.079	0.006	0.314	0.016
10	1	0	0	0	0.466	0.012	0.836	0.014
11	1	0	0	0	0.281	0.010	0.328	0.016
12	1	0	0	0	0.199	0.009	0.389	0.017
13	1	0	0	0	0.205	0.009	0.335	0.016
14	0	0	0	1	0.279	0.011	0.542	0.016
15	0	1	0	0	0.434	0.012	0.850	0.012
16	0	1	0	0	0.307	0.011	0.518	0.015
17	1	0	0	0	0.369	0.011	0.718	0.017
18	0	1	0	0	0.553	0.012	0.771	0.013
19	1	0	0	0	0.201	0.010	0.621	0.018
20	0	0	0	1	0.167	0.009	0.225	0.013
21	0	0	1	0	0.236	0.010	0.380	0.016
22	0	0	1	0	0.300	0.012	0.798	0.014
23	1	0	0	0	0.262	0.010	0.558	0.018
24	1	0	0	0	0.111	0.007	0.502	0.019
25	0	0	0	1	0.175	0.009	0.383	0.016
26	0	0	0	1	0.180	0.008	0.197	0.010
27	1	0	0	0	0.124	0.008	0.487	0.019
28	0	0	0	1	0.322	0.011	0.571	0.016
29	0	1	0	0	0.391	0.012	0.850	0.012
30	0	0	1	0	0.386	0.012	0.773	0.015
31	0	1	0	0	0.772	0.010	0.966	0.006
32	0	1	0	0	0.446	0.012	0.728	0.014
33	1	0	0	0	0.643	0.011	0.977	0.006
34	1	0	0	0	0.535	0.012	0.900	0.012
35	0	1	0	0	0.164	0.010	0.740	0.015
36	0	1	0	0	0.385	0.012	0.563	0.015
37	0	0	0	1	0.426	0.012	0.828	0.013
38	0	1	0	0	0.555	0.012	0.917	0.009
39	1	0	0	0	0.571	0.012	0.914	0.010
40	0	1	0	0	0.562	0.012	0.936	0.008
41	0	0	1	0	0.493	0.013	0.939	0.009
42	0	0	0	1	0.567	0.013	0.975	0.006
43	0	1	0	0	0.606	0.012	0.960	0.006
44	0	0	0	1	0.394	0.012	0.853	0.013
45	0	0	0	1	0.392	0.013	0.897	0.012
46	0	1	0	0	0.155	0.009	0.709	0.016
47	1	0	0	0	0.286	0.010	0.683	0.017
48	0	0	1	0	0.339	0.012	0.764	0.015

Los ítems con estimaciones del parámetro de adivinación (g) que superan el 0.6 son los 31, 33 y 43, estos son los ítems que tienen más del 60% de probabilidades de que el individuo lo adivine o que lo acierte a pesar de no tener las capacidades necesarias. Por otro lado, el parámetro 1-Desliz ($1 - s$) se interpreta como la probabilidad de acertar el ítem cuando si se tienen todas las habilidades necesarias, según la matriz Q , en este caso es interesante observar a aquellos con estimaciones más bajas, como son los ítems 20, 26 y 8.

En el cuadro 5.6 se presenta un resumen descriptivo de los valores estimados para cada uno:

Cuadro 5.6: Descriptivos de la estimación de los parámetros de los ítems, del modelo DINA, en la muestra de alumnos evaluados en 2do de secundaria

Estadístico	\hat{g}	$1 - \hat{s}$
Media	0.332	0.641
Desv. Estándar	0.167	0.237
Mínimo	0.079	0.190
Máximo	0.772	0.977
Primer cuartil	0.185	0.418
Mediana	0.303	0.696
Tercer cuartil	0.443	0.850
Skew	0.549	-0.244
Curtosis	-0.432	-1.184

De igual manera, en el cuadro 5.7 se presenta la distribución de las probabilidades marginales para cada una de las clases latentes que se generaron:

Cuadro 5.7: Distribución de la probabilidad posterior de cada clase latente para la prueba de matemática de 2do de secundaria

Clases latentes (α_c)				π	
α_1	α_2	α_3	α_4	estimado	se
0	0	0	0	0.567	0.012
1	0	0	0	0.002	0.002
0	1	0	0	0.013	0.005
0	0	1	0	0.016	0.006
0	0	0	1	0.005	0.004
1	1	0	0	0.009	0.003
1	0	1	0	0.004	0.002
1	0	0	1	0.002	0.002
0	1	1	0	0.011	0.005
0	1	0	1	0.034	0.006
0	0	1	1	0.014	0.005
1	1	1	0	0.004	0.003
1	1	0	1	0.003	0.003
1	0	1	1	0.002	0.002
0	1	1	1	0.024	0.007
1	1	1	1	0.289	0.011

El 28.9% de los alumnos tienen un patrón de habilidades completo; es decir, dominan las 4 habilidades que requieren los ítems. Aunque existe un 56.7% que no domina ninguna de las cuatro habilidades, siendo la mayoría en toda la muestra evaluada.

5.4. Análisis de los datos mediante Rasch

Uno de los objetivos del presente trabajo es realizar un análisis de correspondencia entre los cuartiles del rendimiento en matemática de la muestra, estimado mediante el modelo Rasch, y los perfiles latentes que se forman mediante el modelo propuesto. De igual manera, se pretende realizar esta correspondencia para los ítems, tomando en cuenta los cuartiles de la escala de dificultad y las habilidades que se requieren, según la matriz \mathbf{Q} , para tener una mayor probabilidad de responderlos correctamente.

A continuación se presentan los resultados de los parámetros estimados a partir del modelo Rasch; es decir, la dificultad del ítem y la habilidad latente del alumno.

En el cuadro 5.8, se presentan las estimaciones de las medidas de dificultad ($\hat{\beta}$) de los ítems desde el punto de vista bayesiano junto a la capacidad que se necesita dominar, según matriz **Q**, para generar los posibles perfiles latentes.

Cuadro 5.8: Estimación de la dificultad de los ítems mediante el modelo de Rasch

Ítem	Habilidad	Dificultad ($\hat{\beta}$)	se
1	MA	0.854	0.045
2	MA	1.253	0.049
3	EE	0.002	0.043
4	RA	0.958	0.046
5	RA	1.42	0.05
6	CR	1.47	0.051
7	MA	0.825	0.045
8	EE	2.005	0.057
9	MA	1.972	0.057
10	MA	-0.412	0.043
11	MA	0.994	0.047
12	MA	1.209	0.047
13	MA	1.292	0.048
14	RA	0.565	0.044
15	CR	-0.471	0.043
16	CR	0.51	0.045
17	MA	0.075	0.043
18	CR	-0.679	0.044
19	MA	0.79	0.045
20	RA	1.683	0.052
21	EE	1.037	0.047
22	EE	0.068	0.044
23	MA	0.677	0.044
24	MA	1.368	0.05
25	RA	1.248	0.048
26	RA	1.705	0.054
27	MA	1.343	0.05
28	RA	0.378	0.044
29	CR	-0.347	0.043
30	EE	-0.148	0.044
31	CR	-1.964	0.056
32	CR	-0.283	0.043
33	MA	-1.276	0.048
34	MA	-0.742	0.045
35	CR	0.517	0.045
36	CR	0.192	0.043
37	RA	-0.382	0.044
38	CR	-0.978	0.046
39	MA	-0.892	0.045
40	CR	-1.042	0.046
41	EE	-0.766	0.045
42	RA	-1.109	0.046
43	CR	-1.244	0.048
44	RA	-0.333	0.044
45	RA	-0.405	0.043
46	CR	0.605	0.045
47	MA	0.401	0.044
48	EE	0.01	0.043

De acuerdo a lo mostrado, los 5 ítems más fáciles son: 31, 33, 43, 42 y 40, los cuales responden a las habilidades: CR, MA, CR, RA y CR, respectivamente; mientras que los 4 más difíciles son: 8, 9, 26 y 20; que responden a las siguientes habilidades: EE, MA, RA y RA, respectivamente.

De igual forma, se presenta el cuadro resumen 5.9 donde se muestra que la prueba en promedio tuvo 0.2907 logits de dificultad, donde el ítem más fácil obtuvo una medida de -1.964 y el más difícil 2.005.

Cuadro 5.9: Descriptivos de la medida de dificultad de la muestra de alumnos evaluados en 2do de secundaria

Estadístico	$\hat{\beta}$
Media	0.291
Desv. Estándar	0.973
Mínimo	-1.964
Máximo	2.005
Primer cuartil	-0.410
Mediana	0.390
Tercer cuartil	1.166
Skew	-0.173
Curtosis	-0.807

Gráficamente, se presenta en la figura 5.1 la distribución de la medida de habilidad mencionada:

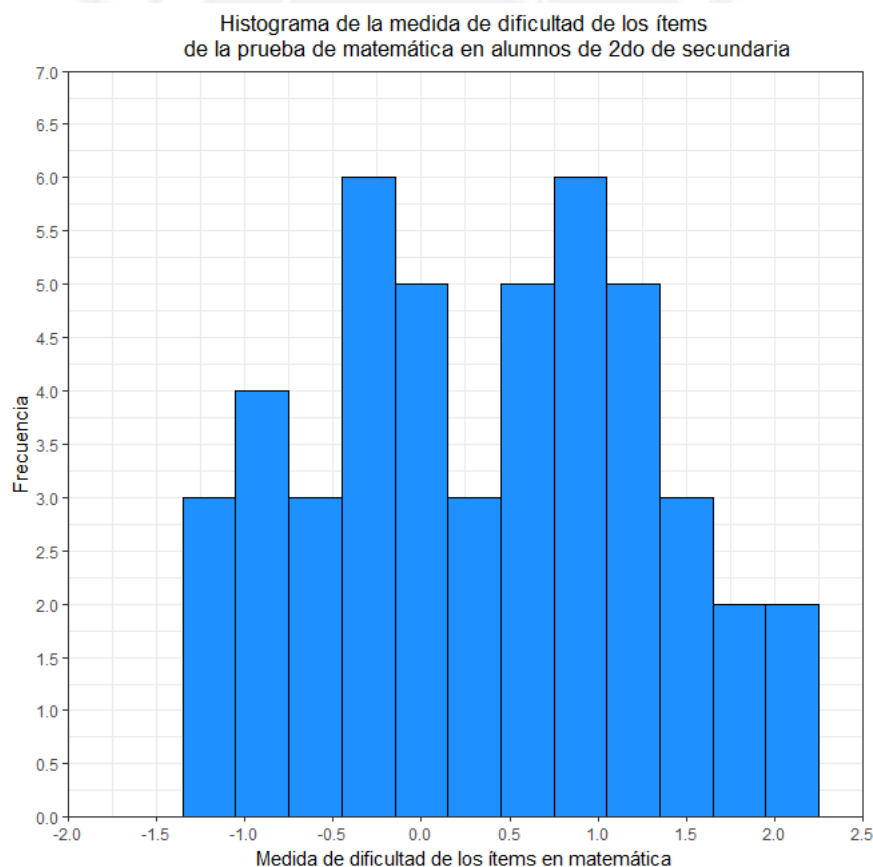


Figura 5.1: Histograma de las medidas de dificultad de los ítems de la prueba de matemática estimadas para la muestra de alumnos evaluados en 2do de secundaria

En la tabla 5.10 se presenta un resumen del análisis descriptivo de las habilidades ($\hat{\theta}$) estimadas de los alumnos evaluados:

Cuadro 5.10: Descriptivos de la medida de habilidad de la muestra de alumnos evaluados en 2do de secundaria

Estadístico	$\hat{\theta}$
Media	0.000
Desv. Estándar	0.850
Mínimo	-2.226
Máximo	2.882
Primer cuartil	-0.598
Mediana	-0.103
Tercer cuartil	0.460
Skew	0.630
Curtosis	0.360

Gráficamente, se presenta en la figura 5.2 la distribución de la medida de habilidad mencionada:

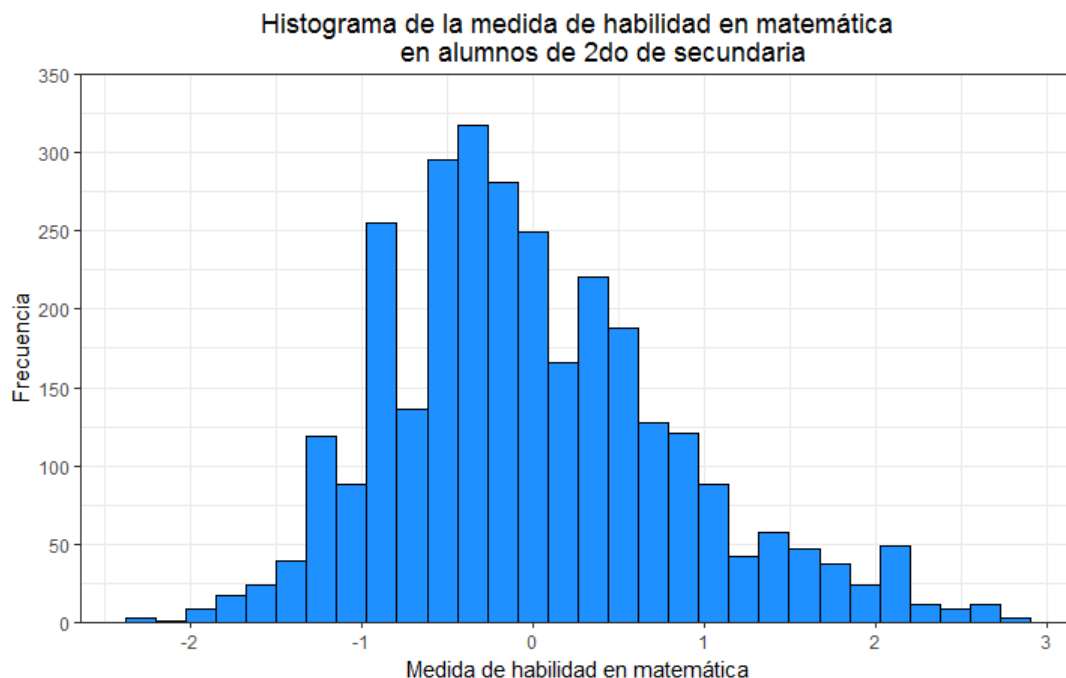


Figura 5.2: Histograma de las medidas de habilidad en matemática estimadas para la muestra de alumnos evaluados en 2do de secundaria

La distribución de las habilidades estimadas de los alumnos muestra una ligera asimetría positiva, la cual corresponde a las evidencias encontradas, las cuales muestran que la prueba resultó difícil para los alumnos.

5.4.1. Análisis de correspondencia entre Rasch y DINA

Después de obtener las estimaciones de las medidas de habilidad de los alumnos y las medidas de dificultad de los ítems de la prueba de 2do de secundaria en matemática mediante el modelo Rasch, se usará esta información para un análisis de correspondencia con las probabilidades de pertenencia a las clases o perfiles latentes y las habilidades requeridas para cada ítem, respectivamente. Para este análisis se generaron dos indicadores, los cuales ayudarán a realizar el cruce de resultados obtenidos por ambos modelos.

El primer indicador es una variable de cuatro categorías, donde cada una representa el cuartil al cual pertenece la unidad de análisis, según sea el caso. Es decir, la categoría 1 representa el cuartil inferior, el 2 el cuartil inferior medio, el 3 el cuartil superior medio, finalmente el 4 el cuartil superior, de la escala correspondiente.

El segundo indicador representará los resultados del modelo DINA y se detallará para las dos unidades de análisis (ítems y alumnos) que se describen en las siguientes secciones.

Análisis de correspondencia para los ítems

Se evalúa entonces la relación entre el primer indicador descrito y el segundo indicador que, en este caso, será la habilidad requerida para obtener mayor probabilidad de acertar al ítem, las cuales han sido detalladas en el cuadro 5.3, por este motivo tendrá también 4 categorías ya que se han descrito esa cantidad de habilidades.

Se presenta el cuadro 5.11 donde se muestra la lista de los 48 ítems, junto a la matriz Q , los estimados de los parámetros g y $1 - s$, las medidas de dificultad estimadas a partir del modelo Rasch, y el indicador de cuartiles.



Cuadro 5.11: Distribución de los ítems de acuerdo a los parámetros estimados por el modelo DINA y el modelo Rasch, junto a la matriz Q y el indicador de cuartiles

Ítem	Matriz Q				Modelo DINA				Modelo Rasch		Indicador de Cuartiles
	MA	CR	EE	RA	Adivinación ($\hat{\theta}$)		1 - Desliz ($\hat{\delta}$)		Dificultad		
	α_1	α_2	α_3	α_4	estimado	se	estimado	se	estimado	se	
1	1	0	0	0	0.216	0.009	0.553	0.018	0.854	0.045	3
2	1	0	0	0	0.170	0.009	0.433	0.018	1.253	0.049	4
3	0	0	1	0	0.345	0.012	0.757	0.015	0.002	0.043	2
4	0	0	0	1	0.242	0.010	0.403	0.016	0.958	0.046	3
5	0	0	0	1	0.146	0.008	0.361	0.016	1.42	0.05	4
6	0	1	0	0	0.096	0.007	0.413	0.016	1.47	0.051	4
7	1	0	0	0	0.255	0.010	0.484	0.018	0.825	0.045	3
8	0	0	1	0	0.126	0.008	0.190	0.013	2.005	0.057	4
9	1	0	0	0	0.079	0.006	0.314	0.016	1.972	0.057	4
10	1	0	0	0	0.466	0.012	0.836	0.014	-0.412	0.043	1
11	1	0	0	0	0.281	0.010	0.328	0.016	0.994	0.047	3
12	1	0	0	0	0.199	0.009	0.389	0.017	1.209	0.047	4
13	1	0	0	0	0.205	0.009	0.335	0.016	1.292	0.048	4
14	0	0	0	1	0.279	0.011	0.541	0.016	0.565	0.044	3
15	0	1	0	0	0.433	0.012	0.850	0.012	-0.471	0.043	1
16	0	1	0	0	0.307	0.011	0.518	0.015	0.51	0.045	3
17	1	0	0	0	0.369	0.011	0.718	0.017	0.075	0.043	2
18	0	1	0	0	0.553	0.012	0.771	0.013	-0.679	0.044	1
19	1	0	0	0	0.201	0.010	0.621	0.018	0.79	0.045	3
20	0	0	0	1	0.168	0.009	0.225	0.013	1.683	0.052	4
21	0	0	1	0	0.236	0.010	0.380	0.016	1.037	0.047	3
22	0	0	1	0	0.300	0.012	0.798	0.014	0.068	0.044	2
23	1	0	0	0	0.262	0.010	0.558	0.018	0.677	0.044	3
24	1	0	0	0	0.111	0.007	0.502	0.019	1.368	0.05	4
25	0	0	0	1	0.175	0.009	0.383	0.016	1.248	0.048	4
26	0	0	0	1	0.180	0.008	0.197	0.010	1.705	0.054	4
27	1	0	0	0	0.124	0.008	0.487	0.019	1.343	0.05	4
28	0	0	0	1	0.323	0.012	0.570	0.016	0.378	0.044	2
29	0	1	0	0	0.391	0.012	0.850	0.012	-0.347	0.043	2
30	0	0	1	0	0.386	0.012	0.773	0.015	-0.148	0.044	2
31	0	1	0	0	0.772	0.010	0.966	0.006	-1.964	0.056	1
32	0	1	0	0	0.446	0.012	0.728	0.014	-0.283	0.043	2
33	1	0	0	0	0.643	0.011	0.977	0.006	-1.276	0.048	1
34	1	0	0	0	0.535	0.011	0.900	0.012	-0.742	0.045	1
35	0	1	0	0	0.165	0.010	0.740	0.015	0.517	0.045	3
36	0	1	0	0	0.385	0.012	0.563	0.015	0.192	0.043	2
37	0	0	0	1	0.426	0.012	0.827	0.013	-0.382	0.044	2
38	0	1	0	0	0.555	0.012	0.917	0.009	-0.978	0.046	1
39	1	0	0	0	0.571	0.012	0.914	0.010	-0.892	0.045	1
40	0	1	0	0	0.563	0.012	0.936	0.008	-1.042	0.046	1
41	0	0	1	0	0.493	0.013	0.939	0.009	-0.766	0.045	1
42	0	0	0	1	0.566	0.013	0.975	0.006	-1.109	0.046	1
43	0	1	0	0	0.606	0.012	0.960	0.006	-1.244	0.048	1
44	0	0	0	1	0.394	0.013	0.853	0.013	-0.333	0.044	2
45	0	0	0	1	0.392	0.013	0.896	0.012	-0.405	0.043	2
46	0	1	0	0	0.156	0.009	0.709	0.016	0.605	0.045	3
47	1	0	0	0	0.285	0.010	0.683	0.017	0.401	0.044	3
48	0	0	1	0	0.339	0.012	0.764	0.015	0.01	0.043	2

A partir de lo anterior, se presenta el cuadro 5.12 donde se cruzan ambos indicadores:

Cuadro 5.12: Tabla de contingencia entre las habilidades presentadas por la matriz Q y el indicador de cuartiles de la escala de dificultad estimada a partir del modelo Rasch

Habilidades	Cuartil inferior		Cuartil medio inferior		Cuartil medio superior		Cuartil superior	
	n	%	n	%	n	%	n	%
	CR	6	46.2 %	3	23.1 %	3	23.1 %	1
EE	1	14.3 %	4	57.1 %	1	14.3 %	1	14.3 %
MA	4	23.5 %	1	5.9 %	6	35.3 %	6	35.3 %
RA	1	9.1 %	4	36.4 %	2	18.2 %	4	36.4 %

En líneas generales, 12 de los 18 ítems cuya habilidad necesaria es “Matematiza situaciones”, son los que se ubican en el 50 % superior de la escala de dificultad. Esto puede deberse a que esta habilidad es menos mecánica algorítmicamente y más de razonamiento lógico. A su vez, 5 de los 7 ítems que requieren de la habilidad “Elabora y usa estrategias” se encuentran en la parte más baja de la escala de dificultad, esto puede ser debido a que esta habilidad está más relacionada a la ejecución de estrategias matemáticas aprendidas.

Otra de las habilidades requeridas para los ítems que en su mayoría se encuentran en la escala más baja es “Comunica y Representa”, 9 de 13 preguntas, la totalidad de los ítems con esta característica representan el 27 % de la prueba (13 de 48).

La distribución de los ítems cuya habilidad “Razona y argumenta generando ideas matemáticas” es requerida, presenta un comportamiento más disperso entre la clasificación por cuartiles.

En la figura 5.3 se muestra el análisis de las correspondencias simple entre los indicadores:

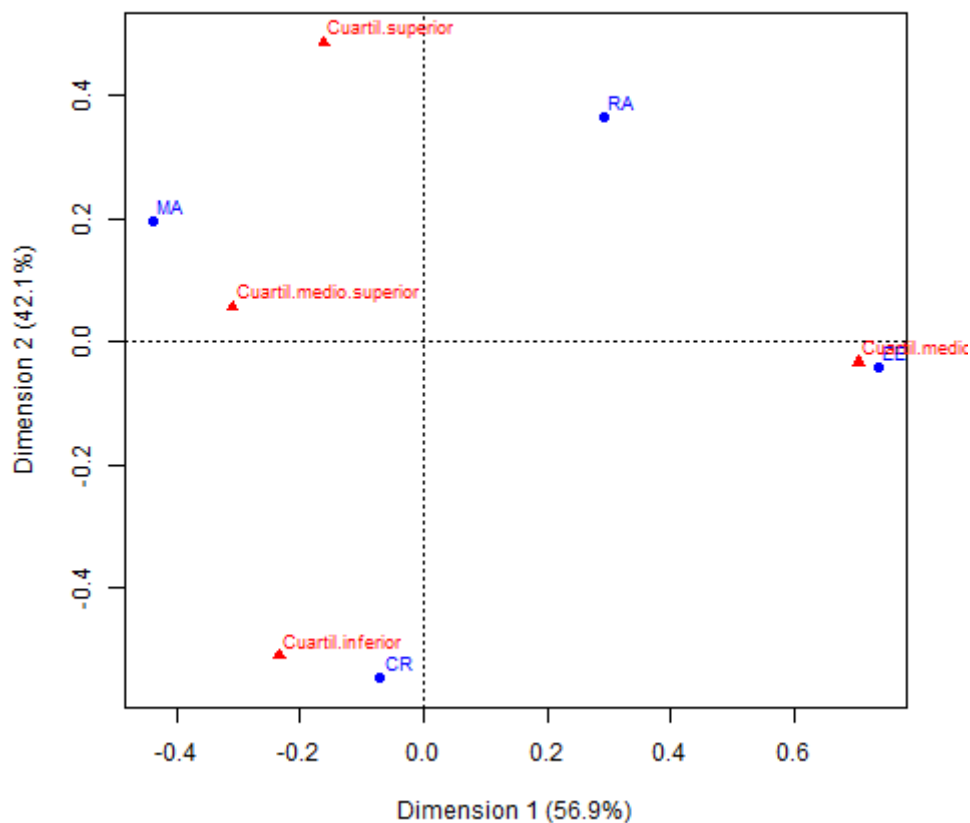


Figura 5.3: Biplot del análisis de correspondencias simple para los indicadores de los ítems

Finalmente, para estudiar con más detalle los parámetros de ruido del ítem que son obtenidos por el modelo en estudio: Adivinación y Desliz. Se presentan las relaciones de estos con indicadores cercanos a su interpretación, como son: la dificultad del ítem y la correlación ítem-medida. Esta correlación es la de Pearson, entre la respuesta observada al ítem y la medida estimada para cada persona (Minedu, 2016).

En la figura 5.4 se muestra la relación entre las medidas de dificultad de los ítems halladas por el modelo de Rasch y las estimaciones del parámetro de adivinación.

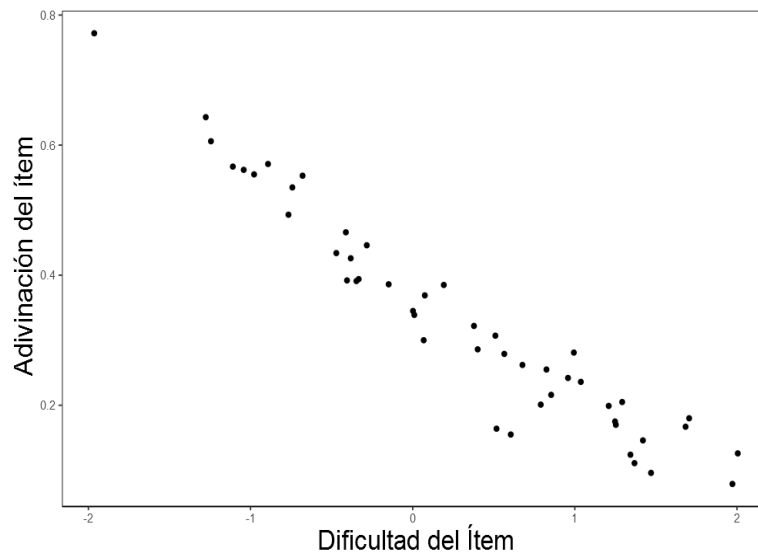


Figura 5.4: Gráfico de dispersión de la relación entre las estimaciones de dificultad y adivinación de los ítems

Se observa que la relación es fuertemente negativa (-0.96); es decir, a mayores medidas de dificultad del ítem, menor es la probabilidad de adivinar.

Asimismo, el modelo Rasch no estima el parámetro de discriminación del ítem; sin embargo, según la teoría clásica se puede tener ciertas evidencias de ello mediante la correlación ítem-medida. Debido a esto, utilizaremos este indicador para estudiar la relación que pueda tener con el parámetro desliz que es estimado por el modelo DINA. En la figura 5.5 se muestra la relación mencionada:

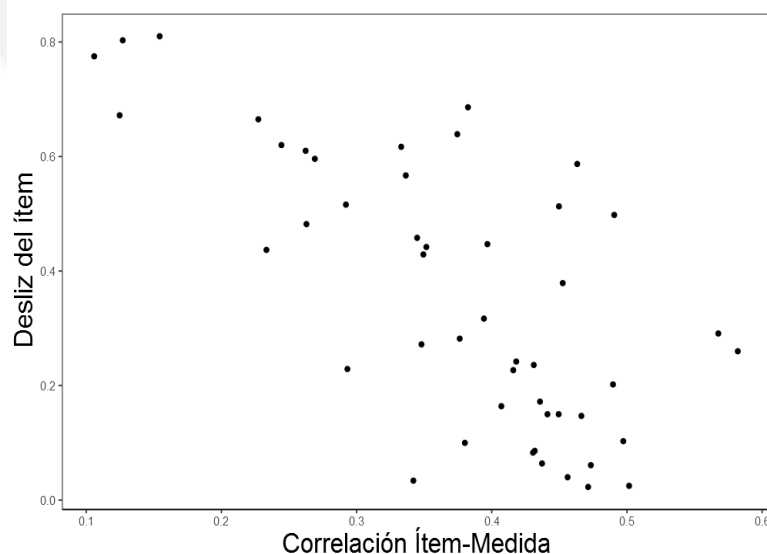


Figura 5.5: Gráfico de dispersión de la relación entre la correlación ítem-medida y las estimaciones del desliz de los ítems

Se observa también una relación negativa entre ambos indicadores, aunque no es tan fuerte (-0.67) como la de la figura anterior. Esto quiere decir, que a mayor discriminación del ítem, menor es la probabilidad de fallarlo a pesar de dominar todas las habilidades requeridas.

Análisis de correspondencia para los alumnos

En el caso del análisis para los alumnos, también se evaluará la relación entre dos indicadores: el primero, es una variable de cuatro categorías, donde cada una representa el cuartil al cual pertenece el alumno evaluado en la escala de habilidad; mientras que el segundo indicador tendrá tantas categorías como clases posibles se estimen. En este caso, tiene 16 categorías. Es así que el alumno pertenecerá a una de esas categorías dependiendo del valor más alto de sus probabilidades de pertenencia a cada uno de los perfiles latentes.

La correspondencia entre ambas categorías se muestra en el cuadro 5.13:

Cuadro 5.13: Tabla de contingencia entre los cuartiles de la escala de habilidad latente estimada a partir del modelo Rasch y el perfil latente estimado a partir del modelo DINA

Clases latentes (α_c)				Cuartil inferior		Cuartil medio inferior		Cuartil medio superior		Cuartil superior	
α_1	α_2	α_3	α_4	n	%	n	%	n	%	n	%
0	0	0	0	760	40.9%	741	39.9%	353	19.0%	2	0.1%
1	0	0	0	0	0.0%	0	0.0%	1	100.0%	0	0.0%
0	1	0	0	0	0.0%	5	27.8%	13	72.2%	0	0.0%
0	0	1	0	0	0.0%	3	33.3%	6	66.7%	0	0.0%
0	0	0	1	0	0.0%	0	0.0%	1	100.0%	0	0.0%
1	1	0	0	0	0.0%	1	4.8%	17	81.0%	3	14.3%
1	0	1	0	0	0.0%	0	0.0%	4	80.0%	1	20.0%
1	0	0	1	0	0.0%	0	0.0%	0	0.0%	0	0.0%
0	1	1	0	0	0.0%	0	0.0%	13	92.9%	1	7.1%
0	1	0	1	0	0.0%	7	7.6%	76	82.6%	9	9.8%
0	0	1	1	0	0.0%	3	9.4%	28	87.5%	1	3.1%
1	1	1	0	0	0.0%	0	0.0%	0	0.0%	0	0.0%
1	1	0	1	0	0.0%	0	0.0%	0	0.0%	0	0.0%
1	0	1	1	0	0.0%	0	0.0%	0	0.0%	0	0.0%
0	1	1	1	0	0.0%	0	0.0%	28	93.3%	2	6.7%
1	1	1	1	0	0.0%	0	0.0%	220	22.9%	741	77.1%

En general, la mayor parte de la muestra de alumnos está ubicada en los extremos de los perfiles latentes; es decir, tienen un perfil completo o no dominan ninguna habilidad.

De esta manera podemos observar que, los alumnos más hábiles obtienen el perfil latente que contiene el dominio de las cuatro habilidades a la vez (1111). El 77.1% de alumnos con el perfil mencionado, se encuentra en el cuartil superior de rendimiento en matemática ($\hat{\theta}$); mientras que el 22.9% restante se encuentran en el cuartil medio superior. Esta clasificación es bastante congruente.

Sin embargo, analizando a los alumnos con bajo rendimiento, se observa que el 40.9% de estudiantes con el perfil latente menos favorable; es decir, aquel en el cual no domina ninguna de las habilidades propuestas (0000); se encuentra en el cuartil inferior del rendimiento estimado a partir del modelo Rasch. El otro 39.9% se encuentra en el cuartil medio inferior. No obstante, existe un 19% de alumnos que a pesar de tener el perfil latente 0000 se ubica en los cuartiles más altos de la escala, lo cual no guarda mucha congruencia. Específicamente, estos son dos alumnos, los cuales tienen el siguiente patrón de respuestas:

- Alumno 1: 11010010011101101101000000110111101011011111000, el cual representa el 54% de acierto de la prueba, un valor estimado de habilidad de 0.462 logits y una probabilidad de 0.42 de pertenecer al perfil latente más desventajado. Además, este alumno pertenece a un colegio público urbano.

- Alumno 2: 00100011011000011100110010011011111011111000011, el cual representa el 54% de acierto de la prueba, un valor estimado de habilidad de 0.465 logits y una probabilidad de 0.33 de pertenecer al perfil latente más desventajado. Además, este alumno pertenece a un colegio público rural.

Adicionalmente, en el cuadro 5.13 es interesante observar que los alumnos con un perfil latente 0101 (“comunica y representa las ideas matemáticas” juntamente con “razona y argumenta generando ideas matemáticas”) y 0011 (“Elabora y usa estrategias” juntamente con “razona y argumenta ideas matemáticas”); a pesar de no dominar las cuatro habilidades, el 92.4% y el 90.6% de alumnos se ubican dentro de los cuartiles superiores de la escala; respectivamente. Esto podría ser debido a que estos perfiles latentes agrupan una habilidad que es requerida para los ítems más fáciles (CR y EE) y otra para los más difíciles (RA).

Por otro lado, se describieron tres estratos en la descripción de la muestra, a continuación se presenta en el cuadro 5.14 la distribución de los perfiles latentes de acuerdo a los estratos:

Cuadro 5.14: Distribución de los 16 perfiles latentes de acuerdo al estrato al que pertenece la institución educativa del alumno

Perfil latente	Urbano Público		Urbano Privado		Rural	
	n	%	n	%	n	%
1 0000	651	62.0%	384	37.7%	821	84.6%
2 1000	0	0.0%	0	0.0%	1	0.1%
3 0100	8	0.8%	7	0.7%	3	0.3%
4 0010	3	0.3%	3	0.3%	3	0.3%
5 0001	0	0.0%	1	0.1%	0	0.0%
6 1100	6	0.6%	13	1.3%	2	0.2%
7 1010	3	0.3%	1	0.1%	1	0.1%
8 1001	0	0.0%	0	0.0%	0	0.0%
9 0110	5	0.5%	5	0.5%	4	0.4%
10 0101	40	3.8%	44	4.3%	8	0.8%
11 0011	13	1.2%	13	1.3%	6	0.6%
12 1110	0	0.0%	0	0.0%	0	0.0%
13 1101	0	0.0%	0	0.0%	0	0.0%
14 1011	0	0.0%	0	0.0%	0	0.0%
15 0111	16	1.5%	8	0.8%	6	0.6%
16 1111	305	29.0%	540	53.0%	116	11.9%

De la misma manera, en el cuadro 5.15 se muestra la distribución de los cuartiles donde se ubican los alumnos según el estrato al que pertenecen:

Cuadro 5.15: Distribución de los cuartiles de rendimiento de acuerdo al estrato al que pertenece la institución educativa del alumno

Estrato	Cuartil inferior		Cuartil medio inferior		Cuartil medio superior		Cuartil superior	
	n	%	n	%	n	%	n	%
Urbano Público	224	21.3%	284	27.0%	319	30.4%	223	21.2%
Urbano Privado	104	10.2%	185	18.2%	274	26.9%	456	44.7%
Rural	432	44.5%	291	30.0%	167	17.2%	81	8.3%

A partir de los dos cuadros presentados, se puede ver claramente que los alumnos del estrato

rural son los que presentan los resultados más bajos con cualquiera de los indicadores; es decir, el 84.6 % de los alumnos que estudian en una IE rural tienen una alta probabilidad de pertenecer al perfil latente que no domina ni una de las 4 habilidades, así también el 44.5 % de ellos se ubican en el cuartil de rendimiento inferior. Por otro lado, el 53.0 % de los alumnos que pertenecen a colegios privados en una zona urbana tienen una alta probabilidad de pertenencia a la clase latente 1111, a su vez el 44.7 % de ellos se encuentra en el cuartil de rendimiento en matemática más alto.

Finalmente, así como para los ítems, en la figura 5.6 se muestra el análisis de correspondencias simple entre los indicadores:

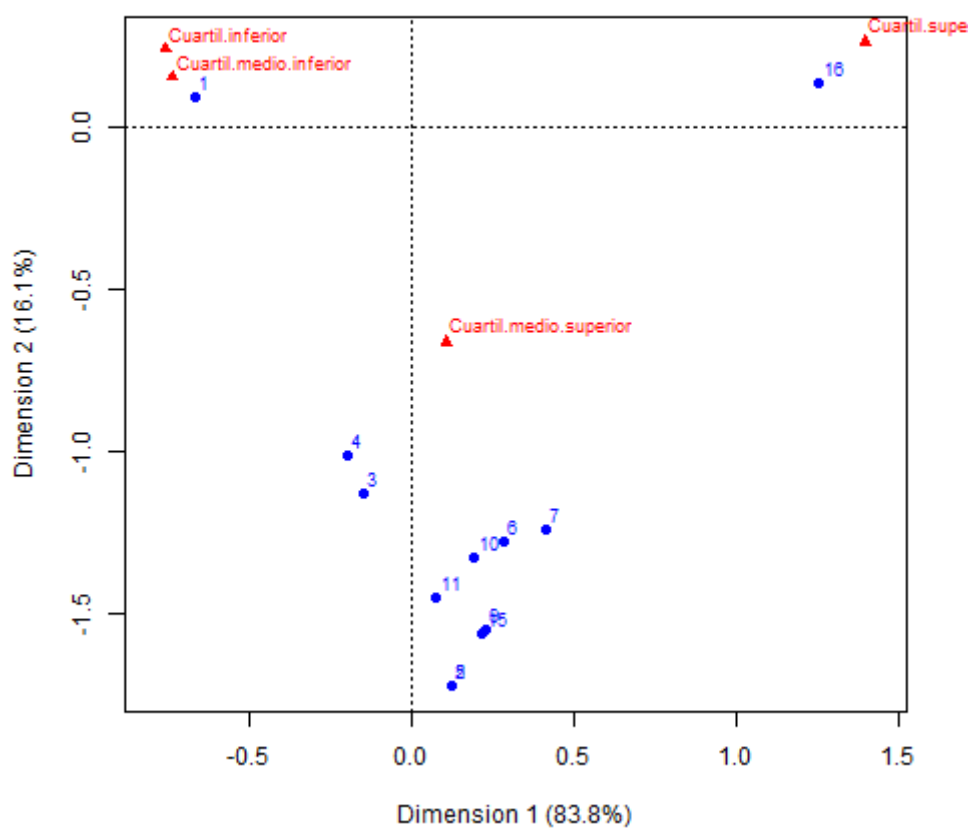


Figura 5.6: Biplot del análisis de correspondencia simple para los indicadores de los alumnos

Capítulo 6

Conclusiones

6.1. Conclusiones

- En este documento de tesis se desarrolló el modelo DINA bajo principalmente el enfoque bayesiano con el muestreador de Gibbs, buscando evaluar la recuperación de parámetros en comparación al enfoque clásico y aplicándolo luego al campo educacional en una muestra de 3040 alumnos que rindieron la prueba de matemática de la ECE del 2do de secundaria 2015 realizada por la UMC.
- Se realizaron además 8 estudios de simulación, donde se hicieron combinaciones de dos características por cada parámetro en estudio: s , g y π ; considerando tres tamaños de población: $N = 100$, $N = 300$ y $N = 500$.
 - En el primer estudio de simulación se ensayaron escenarios con tres tamaños de población, en el cual los parámetros de los ítems fueron fijados con el fin de que se presente una alta o fuerte relación con las habilidades de los individuos. Además, se fijaron las probabilidades de cada clase latente en 2^{-K} . Los resultados encontrados en los escenarios muestran que el muestreador de Gibbs estima ligeramente mejor los valores del parámetro π especialmente cuando el tamaño de la población es pequeña, esta afirmación se da observando los valores del RMSE. Sin embargo, el algoritmo EM muestra mejores estimaciones para los parámetros de los ítems considerando los valores del sesgo, esto es más marcado en el escenario 2, cuando el tamaño de la población es de 300.
 - En el segundo estudio de simulación se ensayaron escenarios con tres tamaños de población, en el cual los parámetros de los ítems fueron fijados con el fin de que se presente una baja o débil relación con las habilidades de los individuos. Además, se fijaron las probabilidades de cada clase latente en 2^{-K} . Los resultados encontrados en los escenarios muestran que, al igual que el estudio anterior, el muestreador de Gibbs obtiene mejores estimaciones, considerando los valores del RMSE y del sesgo, para el parámetro π cuando $N = 100$. Sin embargo, el algoritmo EM en el escenario con un tamaño de población de 300 muestra una clara ventaja en la estimación de los parámetros de los ítems. En el escenario con una población más grande, ambos enfoques brindan muy similares estimaciones para los parámetros en estudio.
 - En el tercer estudio de simulación se ensayaron escenarios con tres tamaños de población, en el cual los parámetros de los ítems fueron fijados con el fin de que el

parámetro s muestre una alta relación y el g un bajo impacto con las habilidades de los individuos. Manteniendo constantes las probabilidades de cada clase latente en 2^{-K} . Los resultados encontrados en los escenarios siguen sosteniendo que el muestreador de Gibbs tiene ligeras ventajas en la estimación de los π con tamaños de población pequeños. Aunque, en este mismo escenario, el algoritmo EM brinda mejores estimaciones para los parámetros del ítem, especialmente para el de adivinación. Después de ello, no hay mayores diferencias.

- En el cuarto estudio de simulación se ensayaron escenarios con tres tamaños de población, en el cual los parámetros de los ítems fueron fijados con el fin de que el parámetro s muestre una alta relación y el g un bajo impacto con las habilidades de los individuos. Manteniendo constantes las probabilidades de cada clase latente en 2^{-K} . Los resultados muestran que las mayores diferencias se encontraron, considerando los valores del sesgo, en la adivinación para los escenarios 1 y 2 a favor del algoritmo EM; mientras que, considerando los valores del RMSE, el muestreador de Gibbs obtiene mejores estimaciones para π en el escenario 1 y ligeramente en el escenario 3.
- En el quinto estudio de simulación se ensayaron escenarios con tres tamaños de población, en el cual los parámetros de los ítems fueron fijados con el fin de que se presente una alta o fuerte relación con las habilidades de los individuos. Aunque, para el parámetro π se consideran dos valores de acuerdo a la cantidad de habilidades que conforman la clase latente. Esta diferenciación se definió como un efecto alto del parámetro respecto a las habilidades de los alumnos. Dicho esto, es interesante evaluar este cambio por lo que los resultados muestran que aún con valores diferenciados para π el muestreador de Gibbs continúa mostrando ligeras ventajas cuando la población es pequeña; y en los demás escenarios los resultados son similares con ambos enfoques. Aunque, es necesario mencionar que con respecto a los parámetros de los ítems y considerando ambos indicadores (RMSE y el sesgo), el algoritmo EM es ampliamente superior al otro método en los escenarios 2 y 3.
- En el sexto estudio de simulación se ensayaron escenarios con tres tamaños de población, en el cual los parámetros de los ítems fueron fijados con el fin de que se presente una baja o débil relación con las habilidades de los individuos; y considerando también una alta influencia del parámetro π con estas. Los resultados encontrados en los escenarios muestran que las estimaciones son insesgadas y con valores pequeños para el RMSE en ambos enfoques. No obstante, las diferencias suelen ser en los parámetros de los ítems, aunque en esta ocasión son mínimas. De la misma forma, las estimaciones para π son similares para todos los escenarios y considerando ambos indicadores. No existen claras ventajas de un enfoque respecto al otro.
- En el séptimo estudio de simulación se ensayaron escenarios con tres tamaños de población, en el cual los parámetros de los ítems fueron fijados con el fin de que el parámetro s muestre una alta relación y el g un bajo impacto con las habilidades de los individuos. Asimismo, se considera una alta influencia del parámetro π con estas habilidades. Los resultados encontrados en los escenarios muestran que las estimacio-

nes se mantienen con buenos indicadores; es decir, son incesgados y con valores bajos del RMSE. Sin embargo, las diferencias entre ambos enfoques son más notorias en los escenarios 2 y 3 cuando analizamos los parámetros de los ítems. Adicionalmente, las estimaciones para el π son similares utilizando cualquiera de los dos enfoques, aunque todavía este muestra a veces ligeras ventajas en tamaños de población pequeños a favor del muestreador de Gibbs.

- En el octavo estudio de simulación se ensayaron escenarios con tres tamaños de población, en el cual los parámetros de los ítems fueron fijados con el fin de que el parámetro s muestre un alto impacto y el g un bajo impacto con las habilidades de los individuos. De igual forma que en los tres estudios previos, se estudió un alto impacto del parámetro π . Los resultados encontrados en los escenarios muestran que no hay diferencias notorias entre ambos enfoques. Este y el estudio anterior podrían considerarse lo más cercano al comportamiento de una prueba regular ya que se ensayaron características de los distintos parámetros.
- Si bien es cierto que en varios de los análisis presentados en los estudios de simulación, el algoritmo EM obtuvo resultados con sesgos más pequeños; se debería considerar la restricción del cálculo de errores estándar (Paek y Cai, 2014), el cual dicta que es necesario que la cantidad de estudiantes evaluados sea necesariamente mucho más grande que la cantidad de ítems en la prueba. De esta forma, si se desea aplicar este modelo en aulas de clase con pocos alumnos cabe la posibilidad que existan problemas en esta estimación. Lo último no ocurre con el enfoque bayesiano.
- Todos los estudios de simulación fueron realizados con el software R con los paquetes DINA y CDM. El tiempo de demora aproximado por estudio fue de 3 horas, considerando que cada uno de ellos presenta 3 escenarios.
- La aplicación fue realizada con datos del Ministerio de Educación, específicamente para una muestra de alumnos que rindieron la prueba de 2do de secundaria del 2015 a nivel nacional, en la competencia matemática. De los resultados obtenidos se concluye que:
 - La correspondencia entre los indicadores de los ítems son bastante congruentes, ya que se observa que los ítems que requieren la habilidad “Comunica y representa las ideas matemáticas” resultan ser los más fáciles según el modelo Rasch. Esto posiblemente ocurre debido a que es una capacidad donde el alumno realiza más aplicaciones de su conocimiento algorítmico en las matemáticas y no tanto utiliza el razonamiento lógico.
 - Por otro lado, los ítems que requieren el dominio de las capacidades “Razona y argumenta generando ideas matemáticas” y “Matematiza situaciones” resultan ubicarse, en su mayoría, en los cuartiles superiores en la escala de dificultad obtenida por el modelo Rasch. Esto ocurre, según especialistas en el área, porque el alumno debe desarrollar problemas o situaciones de la vida real mediante modelos matemáticos; es decir, razonar antes que aplicar algoritmos.

- La relación entre las estimaciones del parámetro adivinación y la dificultad de los ítems resulta ser fuertemente negativa (-0.96); es decir, conforme se incrementa la medida de dificultad del ítem, se reduce la probabilidad de adivinarlo. En el DINA el parámetro de adivinación de un ítem j podría interpretarse como un parámetro de dificultad de este.
- La relación entre las estimaciones del parámetro desliz y la correlación ítem medida, el cual consideraremos un indicador para estudiar la discriminación del ítem; resulta ser negativa (-0.67). Conforme aumenta la discriminación de un ítem, la probabilidad de fallarlo se reduce a pesar de dominar las habilidades requeridas. En el DINA el parámetro de desliz de un ítem j podría interpretarse como un parámetro de discriminación de este.
- La correspondencia entre los indicadores de los alumnos muestran congruencia en la parte alta; en otras palabras, los alumnos con un patrón de habilidades perfecto, que dominan las cuatro capacidades, se ubican en el cuartil superior de la escala.
- No obstante, en la parte baja de la escala encontramos una ligera incongruencia, debido a que algunos alumnos tienen un perfil latente, el cual describe que no dominan ninguna de las cuatro habilidades; sin embargo, estas fueron ubicadas en los cuartiles más altos, según el modelo Rasch.
- Ambos modelos apuntan a obtener evidencias educativas confiables para una adecuada toma de decisiones en políticas públicas; es por ello, que este trabajo no pretende decir que un modelo es mejor que otro sino busca complementar la información que puedan brindar ambos para obtener más especificidad. Por ejemplo, si priorizamos una escala continua para el rendimiento de un área educativa, la información que reportaríamos del alumno sería un puntaje o medida, la ubicación dentro de un percentil o dentro de un punto de corte propuesto por los especialistas; por otra parte, con los MDC podríamos reportar el dominio o no de este estudiante en las diferentes habilidades matemáticas, o describir un perfil de ellos, incluso controlando por niveles socioeconómicos o estratos de gestión educativa como pertenecientes a un colegio público o privado. Lo último sería de vital importancia para profesores, ya que les permitirá tener mejores descripciones de los atributos que dominan o no cada uno de sus estudiantes en la competencia evaluada.

6.2. Sugerencias para futuras investigaciones

- A partir de la metodología, se sugiere realizar una extensión a modelos que consideren ítems o preguntas abiertas; es decir con créditos parciales. Un ejemplo de ello, es el modelo GDINA de la misma familia de modelos de diagnóstico cognitivo (De La Torre, 2011).
- Estudiar nuevos escenarios de simulación se evalúen pruebas con diferentes cantidades de ítems aplicados a los alumnos. Incluyendo escenarios donde la cantidad de ítems es mayor que la cantidad de alumnos evaluados.
- Estudiar nuevos escenarios de simulación con distintas estructuras para la matriz Q , donde se consideren más de una habilidad requerida por ítem.

- Se sugiere realizar comparaciones entre los modelos conjuntivos, como el presentado, y los no conjuntivos; es decir aquellos que no tienen como principal requerimiento que el alumno domine todas las habilidades para un ítem sino al menos una.
- En vista que en la última década están en auge las pruebas adaptativas, por ejemplo la prueba internacional PISA, se sugiere estudiar la efectividad de estos modelos para este tipo de prueba, considerando que en el presente trabajo solo se tomo en cuenta la estructura de prueba única. Una muestra de ello se puede ver en el trabajo de [Huebner \(2010\)](#) y [McGlohen \(2004\)](#).
- Buscar alternativas computacionales, más allá de los paquetes presentados, para mejorar el proceso de estimación; así como generar las rutinas de código para los modelos sugeridos como el GDINA desde el enfoque bayesiano.



Apéndice A

Código en R del estudio de simulación

```
rm(list=ls())
#####
# Estudio de simulación: Modelo DINA
#####
dir <- "D:/Yuriko Sosa/Otros/PUCP/Tesis/3 Análisis del modelo"
setwd(dir)
dir.result <- paste0(dir, "/Simulaciones/Items fijos", sep="")
dir.images <- paste0(dir, "/Simulaciones/Items fijos/Convergencia", sep="")

# Cargar paquetes
library(dina)
library(foreign)
library(CDM)

# Matriz Q fija
Q <- read.csv("Matriz_Q_Mate_2S_C1_C6.csv", sep=",", T)
Q <- as.matrix(Q[,-(1:2)])
# Matriz de posibles combinaciones según K
A <- read.csv("matriz_CxK.csv", T, sep=",")
A <- as.matrix(A[,-1])

# Variables fijas
K <- ncol(Q) # número de habilidades
J <- nrow(Q) # número de ítems
chainLength <- 50000 # número de iteraciones
burnin <- 1000 # Quemar las primeras 1000
thin <- 5

# Definiendo los escenarios
nrep <- 50 # número de repeticiones
escenarios <- list(c("100", "rep(.1,J)", "rep(.1,J)", "rep(1/(2^K),2^K)"), # 1
c("300", "rep(.1,J)", "rep(.1,J)", "rep(1/(2^K),2^K)"), # 2
c("500", "rep(.1,J)", "rep(.1,J)", "rep(1/(2^K),2^K)"), # 3
c("100", "rep(.2,J)", "rep(.2,J)", "rep(1/(2^K),2^K)"), # 4
c("300", "rep(.2,J)", "rep(.2,J)", "rep(1/(2^K),2^K)"), # 5
c("500", "rep(.2,J)", "rep(.2,J)", "rep(1/(2^K),2^K)"), # 6
c("100", "rep(.1,J)", "rep(.2,J)", "rep(1/(2^K),2^K)"), # 7
c("300", "rep(.1,J)", "rep(.2,J)", "rep(1/(2^K),2^K)"), # 8
c("500", "rep(.1,J)", "rep(.2,J)", "rep(1/(2^K),2^K)"), # 9
c("100", "rep(.2,J)", "rep(.1,J)", "rep(1/(2^K),2^K)"), # 10
```

```

c("300", "rep(.2,J)", "rep(.1,J)", "rep(1/(2^K),2^K)", # 11
c("500", "rep(.2,J)", "rep(.1,J)", "rep(1/(2^K),2^K)", # 12
c("100", "rep(.1,J)", "rep(.1,J)", "c(rep(0.021,11),rep(0.042,5))", # 13
c("300", "rep(.1,J)", "rep(.1,J)", "c(rep(0.021,11),rep(0.042,5))", # 14
c("500", "rep(.1,J)", "rep(.1,J)", "c(rep(0.021,11),rep(0.042,5))", # 15
c("100", "rep(.2,J)", "rep(.2,J)", "c(rep(0.021,11),rep(0.042,5))", # 16
c("300", "rep(.2,J)", "rep(.2,J)", "c(rep(0.021,11),rep(0.042,5))", # 17
c("500", "rep(.2,J)", "rep(.2,J)", "c(rep(0.021,11),rep(0.042,5))", # 18
c("100", "rep(.1,J)", "rep(.2,J)", "c(rep(0.021,11),rep(0.042,5))", # 19
c("300", "rep(.1,J)", "rep(.2,J)", "c(rep(0.021,11),rep(0.042,5))", # 20
c("500", "rep(.1,J)", "rep(.2,J)", "c(rep(0.021,11),rep(0.042,5))", # 21
c("100", "rep(.2,J)", "rep(.1,J)", "c(rep(0.021,11),rep(0.042,5))", # 22
c("300", "rep(.2,J)", "rep(.1,J)", "c(rep(0.021,11),rep(0.042,5))", # 23
c("500", "rep(.2,J)", "rep(.1,J)", "c(rep(0.021,11),rep(0.042,5))", # 24

```

```
#####
```

```
# Definir objetos
```

```

mG_bayes <- matrix(0,J,nrep)
mS_bayes <- matrix(0,J,nrep)
PIoutput_bayes <- matrix(0,2^K,nrep)
mG_EM <- matrix(0,J,nrep)
mS_EM <- matrix(0,J,nrep)
PIoutput_EM <- matrix(0,2^K,nrep)

```

```

resbg <- matrix(0,J,nrep); resemg<- matrix(0,J,nrep)
resbs<- matrix(0,J,nrep); resemS <- matrix(0,J,nrep)
rmse_bayes_g<-matrix(0,J); rmse_em_g<-matrix(0,J)
rmse_bayes_s<-matrix(0,J); rmse_em_s<-matrix(0,J)
errorG1<-matrix(0,J); errorG2<-matrix(0,J)
errorS1 <-matrix(0,J); errorS2<-matrix(0,J)
resbpi <- matrix(0,2^K,nrep); resempi<- matrix(0,2^K,nrep)
rmse_bayes_PI <- matrix(0,2^K); rmse_em_PI <- matrix(0,2^K)
errorPI1 <-matrix(0,2^K);errorPI2<-matrix(0,2^K)
errorPI1 <-NA;errorPI2<-NA

```

```

resumen.guessing <- matrix(0,J,4)
resumen.slipping <- matrix(0,J,4)
resumen.pi <- matrix(0,2^K,4)

```

```
#####
```

```
ptm <- proc.time()
```

```

for (i in 1:length(escenarios)){
ptm <- proc.time()
N <- eval(parse(text=escenarios[[i]][1]))
s <- eval(parse(text=escenarios[[i]][2]))
g <- eval(parse(text=escenarios[[i]][3]))
PI <- eval(parse(text=escenarios[[i]][4]))

```

```
#####
```

```
# Simulación
```

```
#####
```

```

for (f in 1:nrep){
CL <- c((1:(2^K))%% rmultinom(n=N,size=1,prob=PI))
Alphas <- A[CL,] # Matriz de alphas A_{NxK}
gen <- DINAsim(Alphas,Q,s,g); Y_sim <- gen$Y # Data simulada
# Estimación
### Muestreador de Gibbs
outchain <- DINA_Gibbs(Y_sim, Amat=A, Q, chain_length=chainLength)
### Algoritmo EM
outdin <- din(Y_sim, q.matr = Q, progress=FALSE)
# Guessing parameter
mG_bayes[,f] <- apply(outchain$GamS[,seq(burnin, chainLength, by = thin)],1,mean)
# Slipping parameter
mS_bayes[,f] <- apply(outchain$SigS[,seq(burnin, chainLength, by = thin)],1,mean)
# Guessing parameter
mG_EM[,f] <- outdin$item[-1]$guess
# Slipping parameter
mS_EM[,f] <- outdin$item[-1]$slip
# Proporción de alumnos en clases latentes
PIoutput_bayes[,f] <- apply(outchain$PIs[,seq(burnin, chainLength, by = thin)],1,mean)
# Proporción de alumnos en clases latentes
PIoutput_EM[,f] <- outdin$attribute.patt$class.prob
} # cierre de repetición

#####
# Plots de convergencia para parámetros de los ítems G y S
for(j in 1:J){
png(filename=paste0(dir.images,"/Guessing/cadena_item",j,"_esc",i,".png", sep=""))
ts.plot(outchain$GamS[j,seq(burnin, chainLength, by = thin)], col="dodgerblue",
xlab="Iteraciones" , ylab = "Adivinación", main="")
dev.off()
png(filename=paste0(dir.images,"/Guessing/acf_item",j,"_esc",i,".png", sep=""))
acf(outchain$GamS[j,seq(burnin, chainLength, by = thin)], col= "red",
xlab="Lag", ylab="Autocorrelación", main="")
dev.off()

png(filename=paste0(dir.images,"/Slipping/cadena_item",j,"_esc",i,".png", sep=""))
ts.plot(outchain$SigS[j,seq(burnin, chainLength, by = thin)], col="dodgerblue",
xlab="Iteraciones" , ylab = "Adivinación", main="")
dev.off()
png(filename=paste0(dir.images,"/Slipping/acf_item",j,"_esc",i,".png", sep=""))
acf(outchain$SigS[j,seq(burnin, chainLength, by = thin)], col= "red",
xlab="Lag", ylab="Autocorrelación", main="")
dev.off()}

# Plots de convergencia para PIs
for(k in 1:2^K){
png(filename=paste0(dir.images,"/Pi/cadena_hab",k,"_esc",i,".png", sep=""))
ts.plot(outchain$PIs[k,seq(burnin, chainLength, by = thin)], col="dodgerblue",
xlab="Iteraciones" , ylab = "Probabilidad", main="")
dev.off()
png(filename=paste0(dir.images,"/Pi/acf_hab",k,"_esc",i,".png", sep=""))
acf(outchain$PIs[k,seq(burnin, chainLength, by = thin)], col= "red",

```

```

xlab="Lag", ylab="Autocorrelación", main="")
dev.off()}

# Resultados de los escenarios
for(h in 1:J){
  for(cont in 1:nrep){
    resbg[h,cont] <- (mG_bayes[h,cont] - g[h])^2
    resemg[h,cont] <- (mG_EM[h,cont] - g[h])^2
    resbs[h,cont] <- (mS_bayes[h,cont] - s[h])^2
    resems[h,cont] <- (mS_EM[h,cont] - s[h])^2

    rmse_bayes_g[h] <- sqrt(sum(resbg[h,])/nrep)
    rmse_em_g[h] <- sqrt(sum(resemg[h,])/nrep)
    rmse_bayes_s[h] <- sqrt(sum(resbs[h,])/nrep)
    rmse_em_s[h] <- sqrt(sum(resems[h,])/nrep)
    errorG1[h] <- mean(mG_bayes[h,]- g[h])
    errorG2[h] <- mean(mG_EM[h,]- g[h])
    errorS1[h] <- mean(mS_bayes[h,]- s[h])
    errorS2[h] <- mean(mS_EM[h,]- s[h])}

    resumen.guessing <- cbind(round(rmse_bayes_g,3),round(errorG1,3),
    round(rmse_em_g,3),round(errorG2,3))
    resumen.slipping <- cbind(round(rmse_bayes_s,3),round(errorS1,3),
    round(rmse_em_s,3),round(errorS2,3))

    for(it in 1:2^K){
      for(cont in 1:nrep){
        resbpi[it,cont] <- (PIoutput_bayes[it,cont] - PI[it])^2
        resempi[it,cont] <- (PIoutput_EM[it,cont] - PI[it])^2
        rmse_bayes_PI[it] <- sqrt(sum(resbpi[it,])/nrep)
        rmse_em_PI[it] <- sqrt(sum(resempi[it,])/nrep)
        errorPI1[it] <- mean(PIoutput_bayes[it,] - PI[it])
        errorPI2[it] <- mean(PIoutput_EM[it,] - PI[it])}

      resumen.pi <- cbind(round(rmse_bayes_PI,3),round(errorPI1,3),
      round(rmse_em_PI,3),round(errorPI2,3))

      # Etiquetas de resumen
      colnames(resumen.guessing) <- c("RMSE_Gibbs", "Sesgo_Gibbs",
      "RMSE_EM", "Sesgo_EM")
      colnames(resumen.slipping) <- c("RMSE_Gibbs", "Sesgo_Gibbs",
      "RMSE_EM", "Sesgo_EM")
      colnames(resumen.pi) <- c("RMSE_Gibbs", "Sesgo_Gibbs",
      "RMSE_EM", "Sesgo_EM")

      write.csv(resumen.guessing, file=paste(dir.result, "/Guessing_Escenario", i, ".csv", sep = ""))
      write.csv(resumen.slipping, file=paste(dir.result, "/Slipping_Escenario", i, ".csv", sep = ""))
      write.csv(resumen.pi, file=paste(dir.result, "/PI_Escenario", i, ".csv", sep = ""))
      print(proc.time() - ptm)
    }
  }
  proc.time() - ptm
}

```

Apéndice B

Rutinas de código en R y Winbugs de la Aplicación

B.1. Modelo DINA

```
rm(list=ls())
#####
# Datos
#####
setwd("E:/PUCP/TESIS/Ciclo IV/Investigación estadística_Tesis/3 Análisis del modelo")
dir.images <- "E:/PUCP/TESIS/Ciclo IV/Investigación estadística_Tesis/
3 Análisis del modelo/Aplicación/Dina_Conv"
#####
# Librerías
#####
library(mvtnorm)
library(foreign)
library(dina)

#####
# Matriz de nxj (personas x ítems)
resp <- read.csv("muestra_respuestas.csv", T, sep = ",")

# Matriz Q
matrizq <- read.csv("Matriz_Q_Mate_2S_C1_C6.csv", T, sep = ",")
matrizq <- matrizq[,-1]

# Matriz de posibles  $2^k$  clases.
comb_matriz <- read.csv("matriz_CxK.csv", T, sep=",")
comb_matriz <- comb_matriz[,-1]
#####
# Modelo DINA mediante estimación bayesiana (Muestreador de Gibbs)
#####
set.seed(100)
# Ejecutar MCMC
chainLength = 50000
burnin = 500
thin = 5

ptm <- proc.time()
d2_Gibbs <- DINA_Gibbs(Y = as.matrix(resp[,-1]),
Amat = as.matrix(comb_matriz),
Q = as.matrix(matrizq[,-1]),
```



```

chain_length=chainLength)
proc.time() - ptm

#####
# Resumen de los resultados (posterior) para g y 1-s
#####

## Parámetro de Adivinación
mGs <- apply(d2_Gibbs$GamS[,seq(burnin,chainLength, by = thin)],1,mean)
sGs <- apply(d2_Gibbs$GamS[,seq(burnin,chainLength, by = thin)],1,sd)

## 1 - Parámetro de Desliz
m1mSS <- 1 - apply(d2_Gibbs$SigS[,seq(burnin,chainLength, by = thin)],1,mean)
s1mSS <- apply(d2_Gibbs$SigS[,seq(burnin,chainLength, by = thin)],1,sd)

# Resultados: resumen
output <- cbind(mGs,sGs,m1mSS,s1mSS)
colnames(output) = c('g Est', 'g SE', '1-s Est', '1-s SE')
rownames(output) = paste0('Item ',1:48)
output <- as.data.frame(output)
print(output,digits=3)
write.csv(output, file = "param_items_d2_Gibbs_new.csv")

#####
# Resumen de la distribución marginal de habilidades
# para las proporciones de clases latentes
#####
PIoutput <- cbind(apply(d2_Gibbs$PIs[,seq(burnin,chainLength, by=thin)],1,mean),
apply(d2_Gibbs$PIs[,seq(burnin,chainLength, by=thin)],1,sd))

#####
# PIs son los parámetros de cada clase (patrón de respuesta)
#####
colnames(PIoutput) <- c('pi_est', 'pi_se')
rownames(PIoutput) <- apply(comb_matriz,1,paste0,collapse=' ')
print(PIoutput,digits=3)
write.csv(PIoutput, file = "maginal_skills_d2_Gibbs_new.csv")

#####
# Plots de convergencia para parámetros
#####
## Plots de la cadena para parámetro G
J=ncol(resp[,-1])
for(j in 1:J){
png(filename=paste0(dir.images,"/Guessing/cadena_item",j,".png", sep=""))
ts.plot(d2_Gibbs$GamS[j,seq(burnin,chainLength, by=thin)], col="dodgerblue",
xlab="Iteraciones" , ylab = "Adivinación", main="")
dev.off()
png(filename=paste0(dir.images,"/Guessing/acf_item",j,".png", sep=""))
acf(d2_Gibbs$GamS[j,seq(burnin,chainLength, by=thin)], col= "red",
xlab="Lag", ylab="Autocorrelación", main="")
dev.off()}

```

```
## Plots de la cadena para parámetro S
for(j in 1:J){
png(filename=paste0(dir.images,"/Slipping/cadena_item",j,".png", sep=""))
ts.plot(d2_Gibbs$SigS[j,seq(burnin,chainLength, by=thin)], col="dodgerblue",
xlab="Iteraciones" , ylab = "Desliz", main="")
dev.off()
png(filename=paste0(dir.images,"/Slipping/acf_item",j,".png", sep=""))
acf(d2_Gibbs$SigS[j,seq(burnin,chainLength, by=thin)], col= "red",
xlab="Lag", ylab="Autocorrelación", main="")
dev.off()}

## Plots de la cadena para PIs
K=nrow(comb_matriz)
for(k in 1:K){
png(filename=paste0(dir.images,"/Pi/cadena_hab",k,".png", sep=""))
ts.plot(d2_Gibbs$PIs[k,seq(burnin,chainLength, by=thin)], col="dodgerblue",
xlab="Iteraciones" , ylab = "Probabilidad", main="")
dev.off()
png(filename=paste0(dir.images,"/Pi/acf_hab",k,".png", sep=""))
acf(d2_Gibbs$PIs[k,seq(burnin,chainLength, by=thin)], col= "red",
xlab="Lag", ylab="Autocorrelación", main="")
dev.off()}
```

B.2. Modelo Rasch

B.2.1. Código en R

```
rm(list=ls())
#####
# Datos
#####
setwd("D:/Yuriko Sosa/Otros/PUCP/Tesis/Análisis/Rasch bayesiano")
library(foreign)
# Base de respuestas de la muestra de alumnos
datos <- read.csv("muestra_respuestas.csv", sep = ",", T)
resp_mate <- datos[,-1]
#####
# Análisis Rasch con WinBugs (estimación bayesiana)
#####
install.packages(c("R2WinBUGS","coda","mcmcplots","denstrip", "boot"),
dependencies = c("Depends"))

# Librerías
library("boot")
library("coda")
library("R2WinBUGS")
library("mcmcplots")

bugs.dir <- "D:/Yuriko Sosa/WinBUGS/winbugs14/WinBUGS14"
bugs.file <- file.path("Rasch.bug")

Y <- as.matrix(resp_mate)
n <- nrow(Y)
```

```

p <- ncol(Y)

#Prioris propuestas por Sahu (2002), Patz e Junker (1999)
m.delta <- 0.0
s.delta <- sqrt(2)

data <- list("Y", "n", "p",
            "m.delta", "s.delta")
monitor <- c("delta", "theta")

n.burn <-500
n.thin<-5
n.sim <- 12500

mate.rasch.bayes.time<-system.time(mate.rasch.bayes <- bugs(data=data, inits=NULL,
parameters.to.save=monitor,
model.file=bugs.file,
bugs.directory=bugs.dir,
n.iter=n.sim, n.thin=n.thin,
n.burnin=n.burn,
n.chains=2
))

# Ver directorio con los resultados
tempdir()

# Diagnostico de Convergencia
plot(mate.rasch.bayes) # muestra los Rhats
mcmcplot(mate.rasch.bayes, random=20)
#####
# Resumen de resultados
#####
## Matriz con parámetros estimados de los ítems
### Dificultades
mate.rasch.bayes.fits <- mate.rasch.bayes$mean$delta
mate.rasch.bayes.fits <- as.matrix(mate.rasch.bayes.fits)
mate.rasch.bayes.error.fits <- mate.rasch.bayes$sd$delta
mate.rasch.bayes.error.fits <- as.matrix(mate.rasch.bayes.error.fits)
mate.rasch.bayes.fits <- data.frame(mate.rasch.bayes.fits, mate.rasch.bayes.error.fits)
colnames(mate.rasch.bayes.fits) <- c("beta.bayes","error.bayes")
write.csv(mate.rasch.bayes.fits, file = "mate.rasch.bayes.fits.csv")

## Matriz con parámetros estimados de las personas
### Habilidades
mate.rasch.bayes.theta <- mate.rasch.bayes$mean$theta
mate.rasch.bayes.theta <- as.matrix(mate.rasch.bayes.theta)
mate.rasch.bayes.sd.theta <- mate.rasch.bayes$sd$theta
mate.rasch.bayes.sd.theta <- as.matrix(mate.rasch.bayes.sd.theta)
mate.rasch.bayes.theta <- data.frame(mate.rasch.bayes.theta, mate.rasch.bayes.sd.theta)
colnames(mate.rasch.bayes.theta) <- c("theta.bayes","error.bayes")
summary(mate.rasch.bayes.theta);summary(mate.rasch.bayes.sd.theta)
hist(mate.rasch.bayes.theta$theta.bayes, probability = TRUE)

```

```

mate.rasch.bayes.theta
write.csv(mate.rasch.bayes.theta, file = "mate.rasch.bayes.theta.csv")

## Guardar archivo de resumen
dificultad_mate <- data.frame(dificultad_mate,mate.rasch.bayes.fits)
habilidad_mate <- data.frame(habilidad_mat, mate.rasch.bayes.theta)
write.csv(dificultad_mate, "dificultad_mate.csv", row.names = TRUE )
write.csv(habilidad_mate, "habilidad_mate.csv",row.names=FALSE )

#####
# Gráficos para la distribución (complementarios)
#####
## Habilidades de los alumnos
library(psych)
library(ggplot2)
habilidades <- read.csv("mate.rasch.bayes.theta.csv")
colnames(habilidades) <- c("ID", "theta", "error")
ggplot(habilidades, aes(x=theta))+
#geom_histogram(aes(fill = ..count..))
geom_histogram(colour = "black", fill = "dodgerblue")+
scale_y_continuous(limits = c(0, 350), expand = c(0,0), breaks = seq(0,350,50))+
#ylab("Frecuencia")+ xlab("Medida de habilidad en matemática")+
labs(x= "Medida de habilidad en matemática", y="Frecuencia",
title = "")+
theme_bw()+
theme(plot.title = element_text(hjust = 0.5))

## Dificultades del ítem
library(psych)
library(ggplot2)
dificultades <- read.csv("mate.rasch.bayes.fits.csv")
colnames(dificultades) <- c("Item", "habilidad_K","beta", "error")
ggplot(dificultades, aes(x=beta))+
geom_histogram(colour = "black", fill = "dodgerblue", binwidth=0.3)+
scale_y_continuous(limits = c(0, 7), expand = c(0,0), breaks = seq(0,7,0.5)) +
scale_x_continuous(limits = c(-2, 2.5), expand = c(0,0), breaks = seq(-2,2.5,.5))+
labs(x= "Medida de dificultad de los ítems en matemática", y="Frecuencia",
title = "Histograma de la medida de dificultad de los ítems
de la prueba de matemática en alumnos de 2do de secundaria")+
theme_bw()+
theme(plot.title = element_text(hjust = 0.5))

```

B.2.2. Código Winbugs

```
model{
for (i in 1:n){
for (j in 1:p){
Y[i, j] ~ dbern(prob[i, j])
logit(prob[i, j]) <- theta[i] - delta[j]
}
theta[i] ~ dnorm(0.0, 1.0)
}

## Prioris de parametros de los ítems
for (j in 1:p){
delta[j] ~ dnorm(m.delta, pr.delta)
}
pr.delta <- pow(s.delta, -2)
}
```



Apéndice C

Ejemplos de ítems de la prueba de Matemática

C.1. Capacidad: matematiza situaciones

Una empresa ofrece en alquiler el siguiente modelo de auto. Observa:



EN ALQUILER
S/ 100 por día +
S/ 40 por derecho de
alquiler (pago único)

Según esta información, ¿cuál de las siguientes expresiones representa el pago total "T" que se realizará al alquilar por "d" días este auto?

- a) $T = 40d$
- b) $T = 100d$
- c) $T = 40 + 100d$
- d) $T = 40d + 100$

Figura C.1: Ítem de la prueba modelo de matemática que requiere de la habilidad: Matematiza situaciones

C.2. Capacidad: comunica y representa ideas matemáticas

4

Se entrevistó a 40 jóvenes para conocer cuánto dinero gastan mensualmente en transporte. Estos fueron los resultados.

Monto (soles)	Cantidad de estudiantes
5	14
8	10
10	12
40	4

De la información proporcionada, se puede obtener las siguientes medidas de tendencia central:

Media = 10,75

Mediana = 8

Moda = 5

¿Cuál de las medidas sería un valor representativo del monto de dinero que gastaron en transporte el grupo de jóvenes mensualmente?

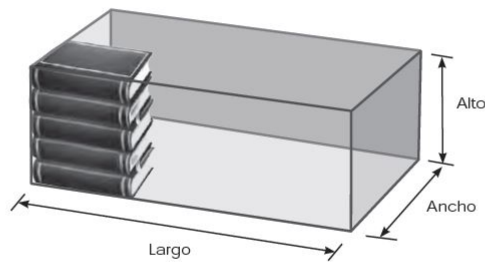
- a) Mediana.
- b) Media.
- c) Moda.
- d) Suma total.

Figura C.2: Ítem de la prueba modelo de matemática que requiere de la habilidad: Comunica y representa ideas matemáticas

C.3. Capacidad: elabora y usa estrategias

22

Se desea llenar la caja mostrada con libros del mismo tamaño. Si se colocan tal como se muestra en la figura, entran 3 libros a lo largo y 2 libros a lo ancho. ¿Cuántos libros como máximo pueden entrar en esta caja?



- a) 50 libros.
- b) 30 libros.
- c) 10 libros.
- d) 6 libros.

Figura C.3: Ítem de la prueba modelo de matemática que requiere de la habilidad: Elabora y usa estrategias

C.4. Capacidad: razona y argumenta generando ideas matemáticas

10

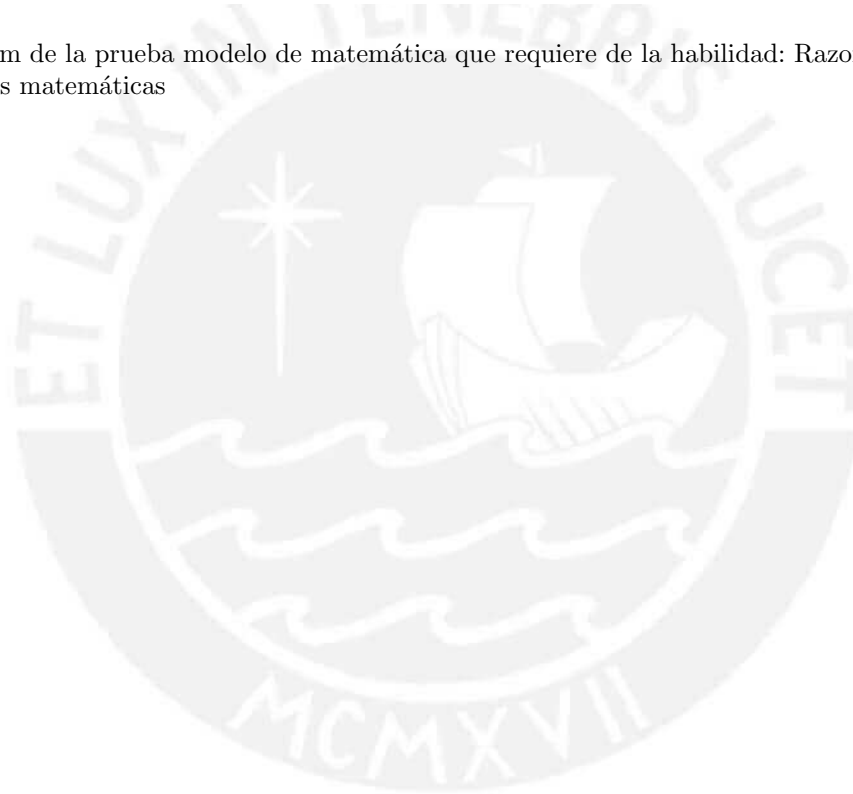
Para saber la cantidad de mayólicas blancas y negras que se utilizarán, el albañil elabora la siguiente tabla:

Cantidad de mayólicas blancas	1	2	3	4	...
Cantidad de mayólicas negras	4	7	10	13	

Si coloca 24 mayólicas blancas, ¿cuántas mayólicas negras utilizará?

- a) 96 mayólicas negras.
- b) 78 mayólicas negras.
- c) 73 mayólicas negras.
- d) 16 mayólicas negras.

Figura C.4: Ítem de la prueba modelo de matemática que requiere de la habilidad: Razona y argumenta generando ideas matemáticas



Apéndice D

Análisis de convergencia

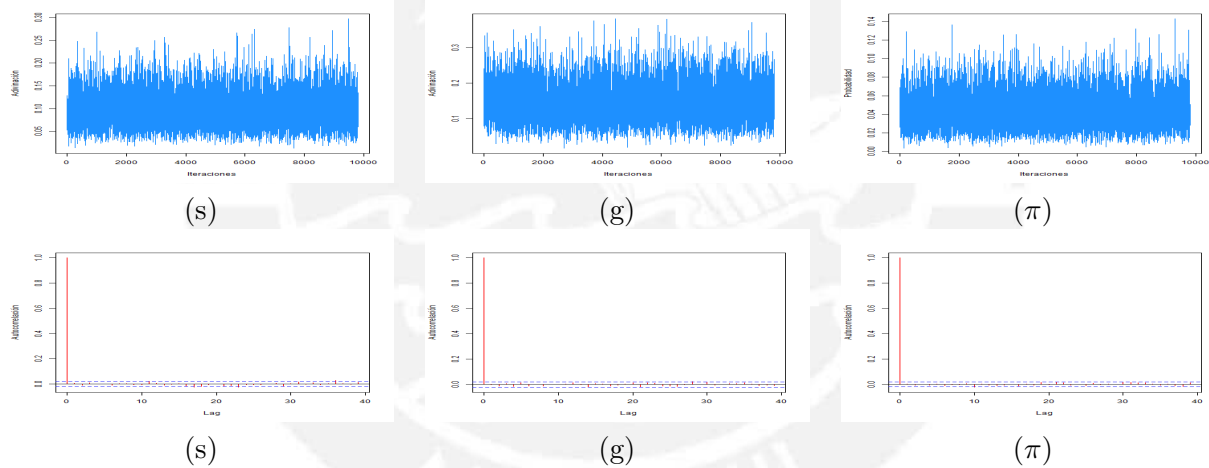
D.1. Convergencia del estudio de Simulación

D.1.1. Estudio de simulación 1

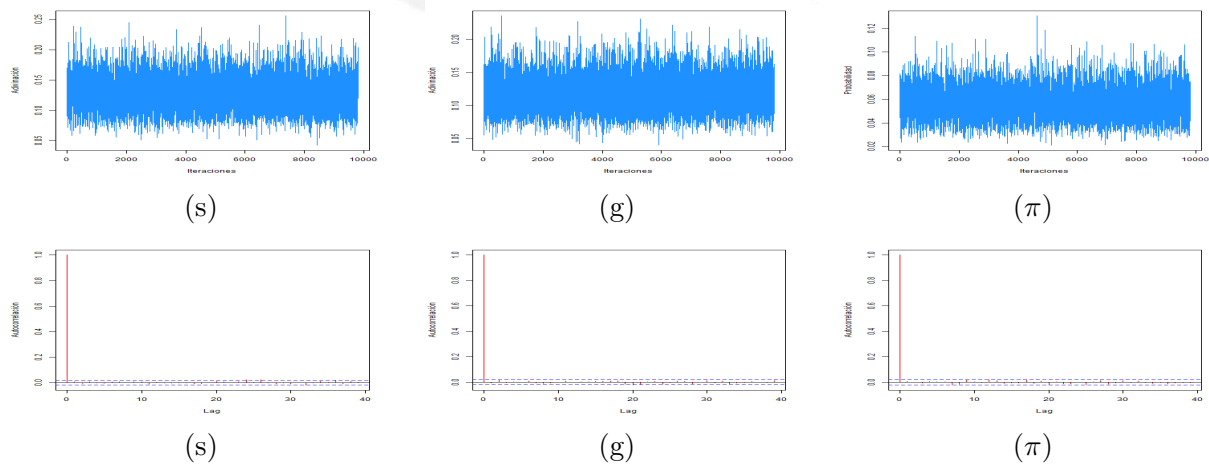
En el cuadro D.1 se presentan los gráficos para las cadenas y autocorrelaciones de las estimaciones de los parámetros de un ítem (s y g) y de la probabilidad de una clase latente (π); para cada uno de los tres escenarios de este estudio.

Cuadro D.1: Gráficos de las cadenas y autocorrelaciones para los parámetros s , g y π de un ítem y una clase, respectivamente, elegido aleatoriamente para el estudio 1 de simulación

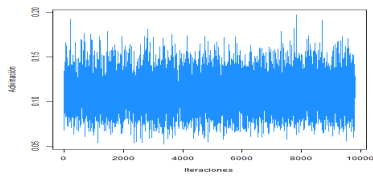
Escenario 1: comportamiento del ítem 5 y la clase 3 (0100).



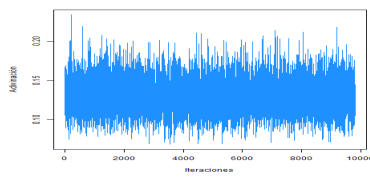
Escenario 2: comportamiento del ítem 8 y la clase 4 (0010)



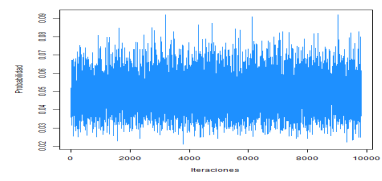
Escenario 3: comportamiento del ítem 25 y la clase 16 (1111)



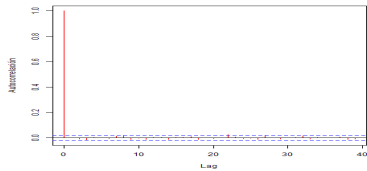
(s)



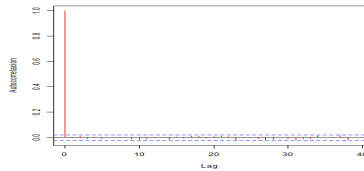
(g)



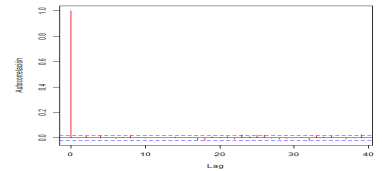
(π)



(s)



(g)



(π)

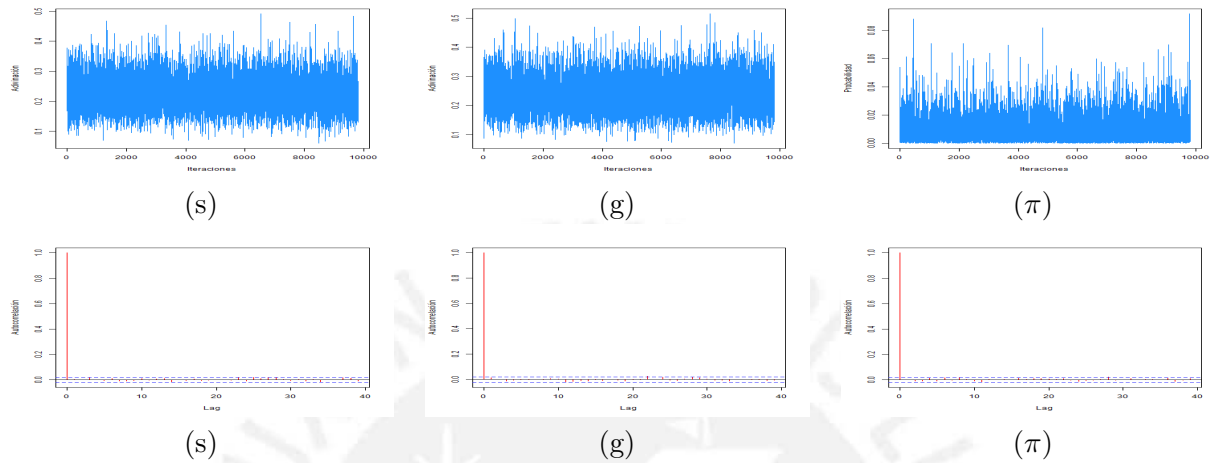


D.1.2. Estudio de simulación 2

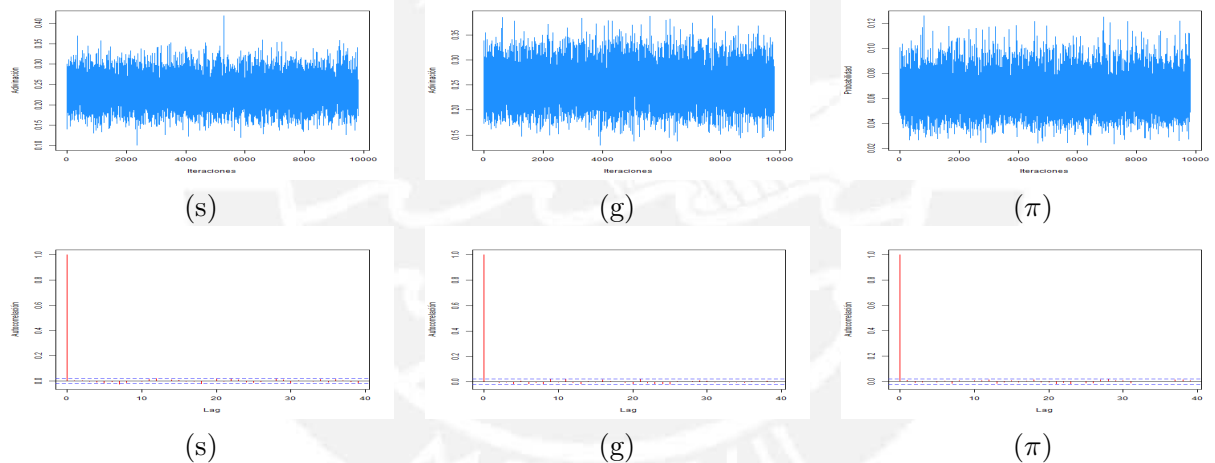
En el cuadro D.2 se presentan los gráficos para las cadenas y autocorrelaciones de las estimaciones de los parámetros de un ítem (s y g) y de la probabilidad de una clase latente (π); para cada uno de los tres escenarios de este estudio.

Cuadro D.2: Gráficos de las cadenas y autocorrelaciones para los parámetros s , g y π de un ítem y una clase, respectivamente, elegido aleatoriamente para el estudio 2 de simulación

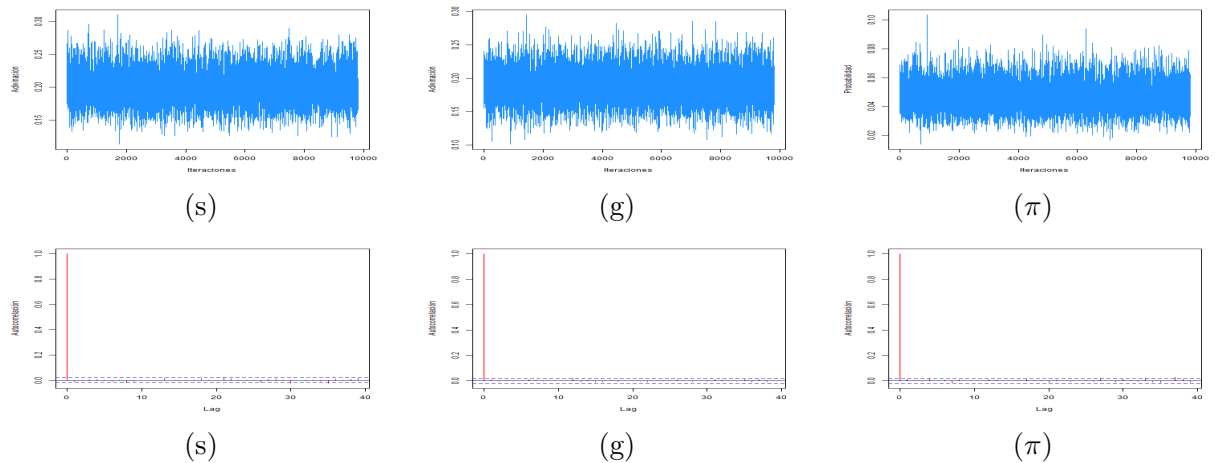
Escenario 1: comportamiento del ítem 35 y la clase 1 (0000).



Escenario 2: comportamiento del ítem 48 y la clase 15 (0111)



Escenario 3: comportamiento del ítem 31 y la clase 2 (1000)

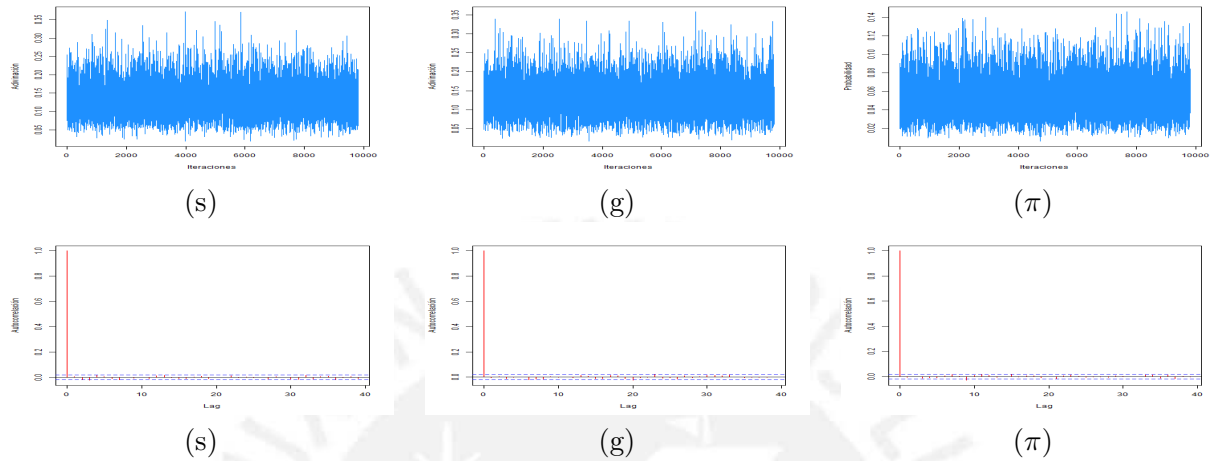


D.1.3. Estudio de simulación 3

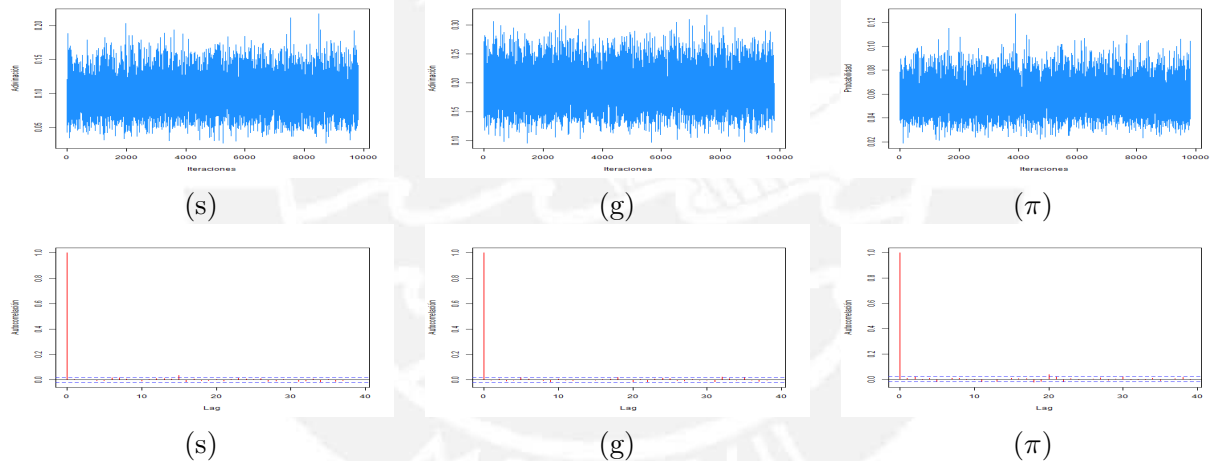
En el cuadro D.3 se presentan los gráficos para las cadenas y autocorrelaciones de las estimaciones de los parámetros de un ítem (s y g) y de la probabilidad de una clase latente (π); para cada uno de los tres escenarios de este estudio.

Cuadro D.3: Gráficos de las cadenas y autocorrelaciones para los parámetros s , g y π de un ítem y una clase, respectivamente, elegido aleatoriamente para el estudio 3 de simulación

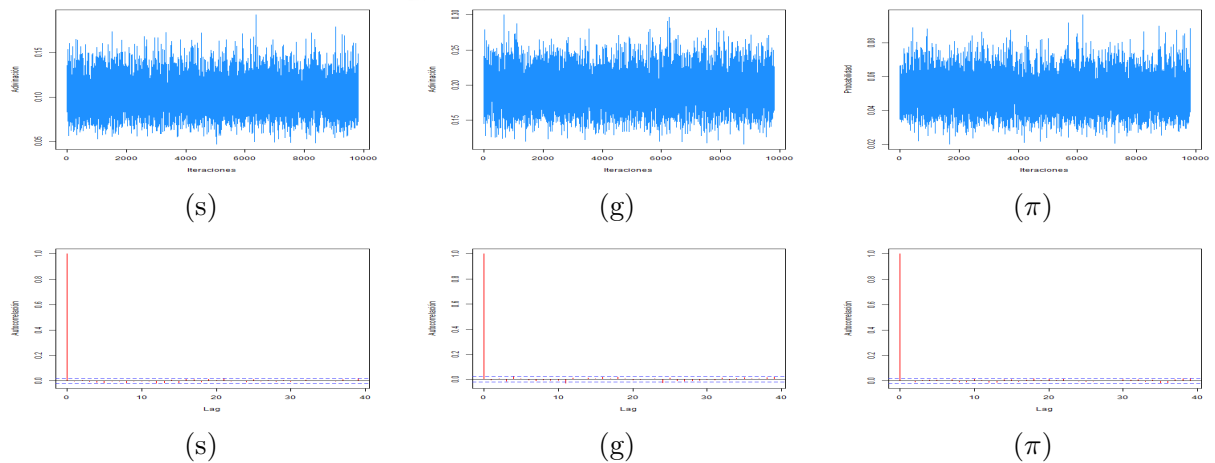
Escenario 1: comportamiento del ítem 18 y la clase 13 (1101).



Escenario 2: comportamiento del ítem 9 y la clase 6 (1100)



Escenario 3: comportamiento del ítem 17 y la clase 12 (1110)

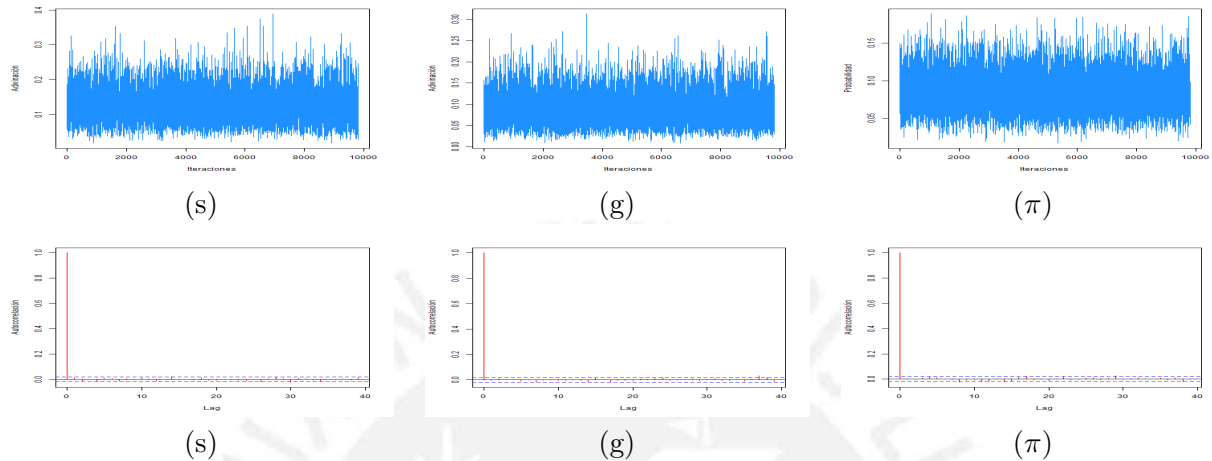


D.1.4. Estudio de simulación 4

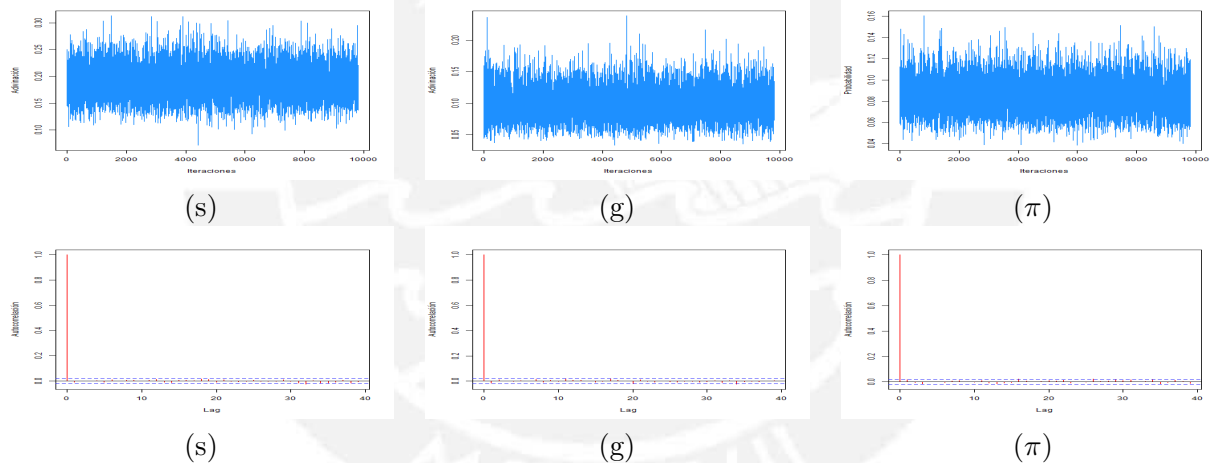
En el cuadro D.4 se presentan los gráficos para las cadenas y autocorrelaciones de las estimaciones de los parámetros de un ítem (s y g) y de la probabilidad de una clase latente (π); para cada uno de los tres escenarios de este estudio.

Cuadro D.4: Gráficos de las cadenas y autocorrelaciones para los parámetros s , g y π de un ítem y una clase, respectivamente, elegido aleatoriamente para el estudio 4 de simulación

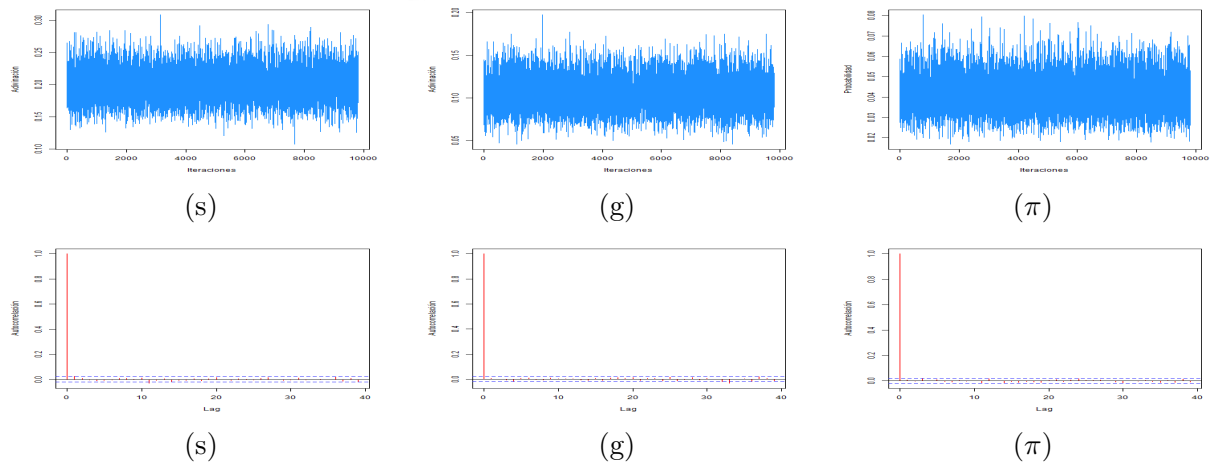
Escenario 1: comportamiento del ítem 46 y la clase 16 (1111).



Escenario 2: comportamiento del ítem 17 y la clase 16 (1111)



Escenario 3: comportamiento del ítem 4 y la clase 2 (1000)

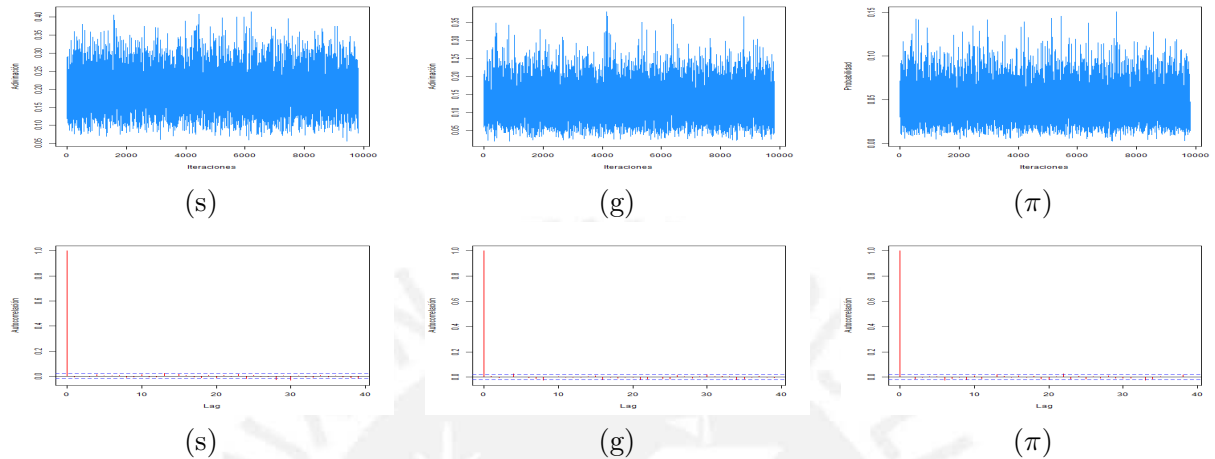


D.1.5. Estudio de simulación 5

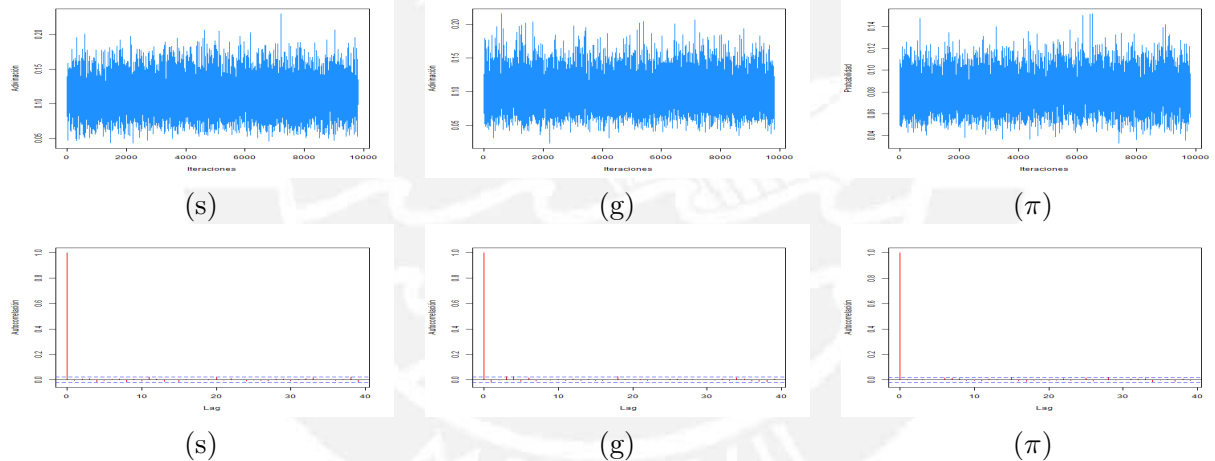
En el cuadro D.5 se presentan los gráficos para las cadenas y autocorrelaciones de las estimaciones de los parámetros de un ítem (s y g) y de la probabilidad de una clase latente (π); para cada uno de los tres escenarios de este estudio.

Cuadro D.5: Gráficos de las cadenas y autocorrelaciones para los parámetros s , g y π de un ítem y una clase, respectivamente, elegido aleatoriamente para el estudio 5 de simulación

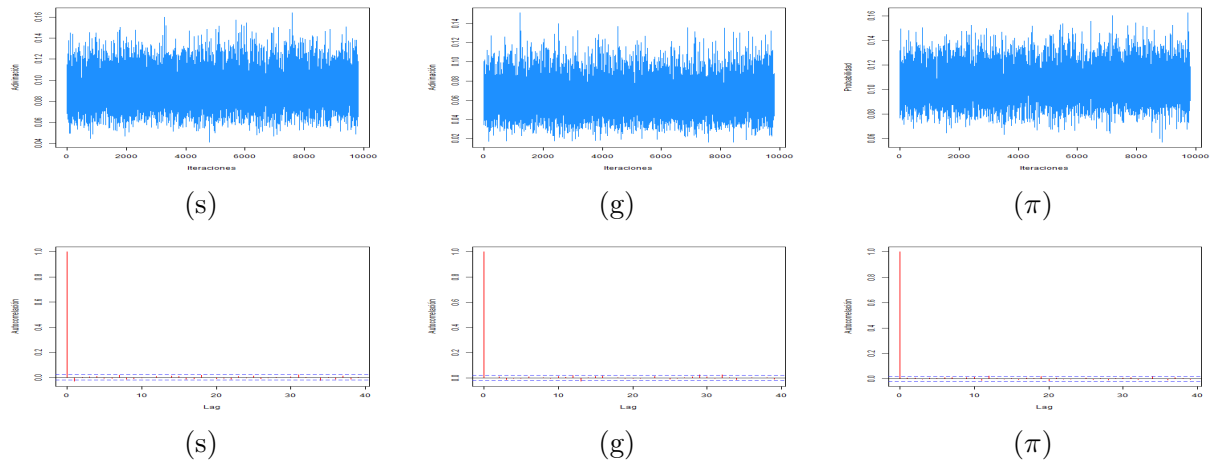
Escenario 1: comportamiento del ítem 23 y la clase 10 (0101).



Escenario 2: comportamiento del ítem 34 y la clase 13 (1101)



Escenario 3: comportamiento del ítem 30 y la clase 15 (0111)

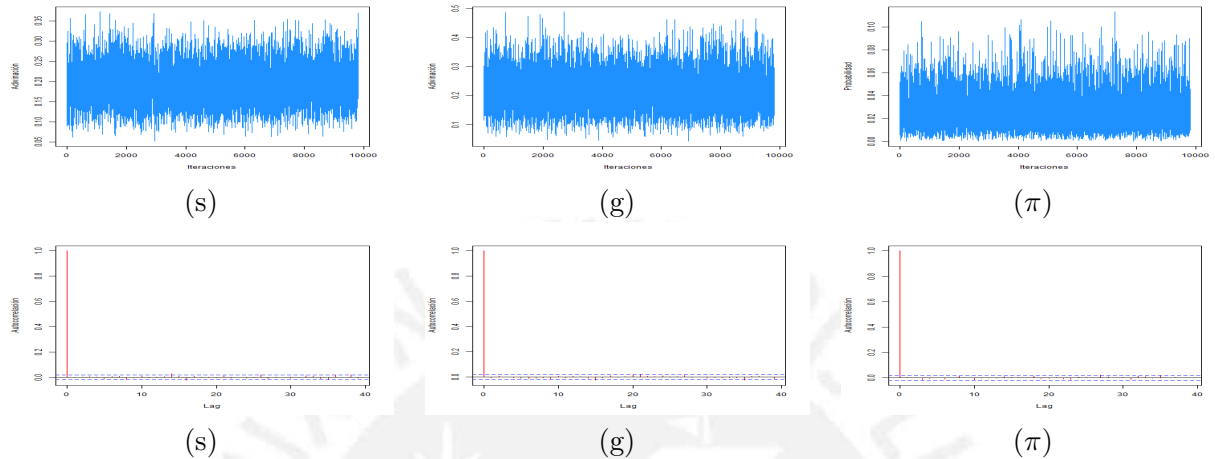


D.1.6. Estudio de simulación 6

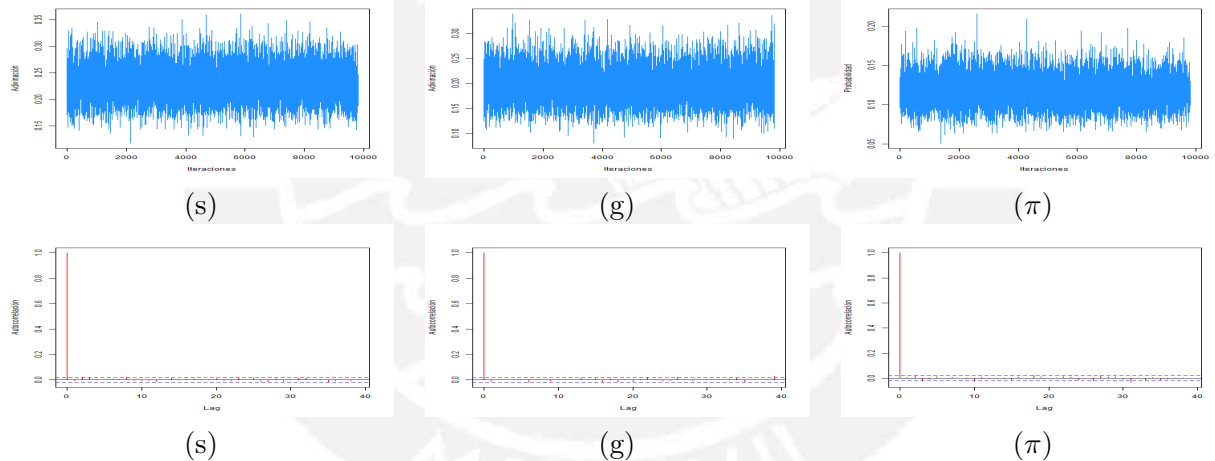
En el cuadro D.6 se presentan los gráficos para las cadenas y autocorrelaciones de las estimaciones de los parámetros de un ítem (s y g) y de la probabilidad de una clase latente (π); para cada uno de los tres escenarios de este estudio.

Cuadro D.6: Gráficos de las cadenas y autocorrelaciones para los parámetros s , g y π de un ítem y una clase, respectivamente, elegido aleatoriamente para el estudio 6 de simulación

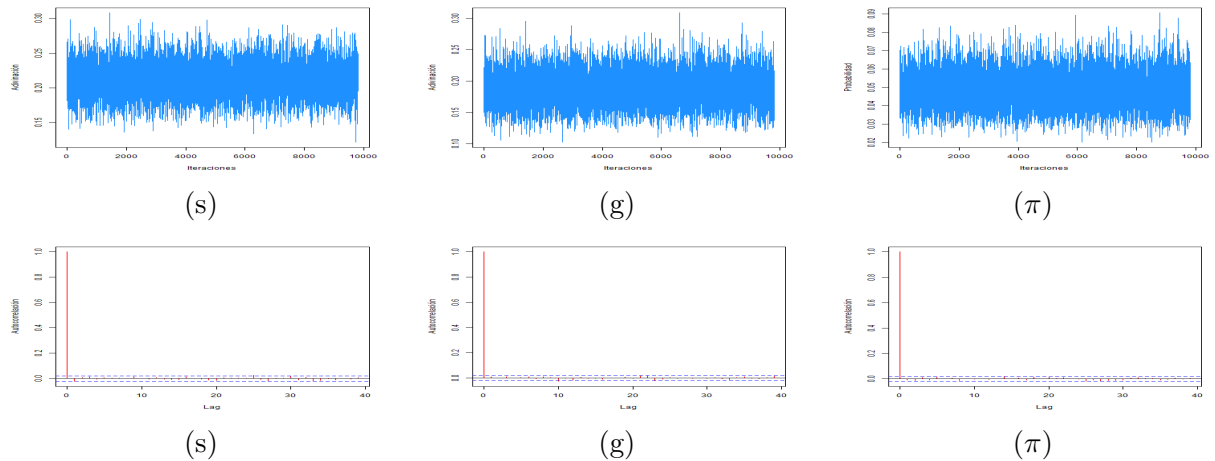
Escenario 1: comportamiento del ítem 34 y la clase 8 (1001).



Escenario 2: comportamiento del ítem 7 y la clase 15 (0111)



Escenario 3: comportamiento del ítem 12 y la clase 4 (0010)

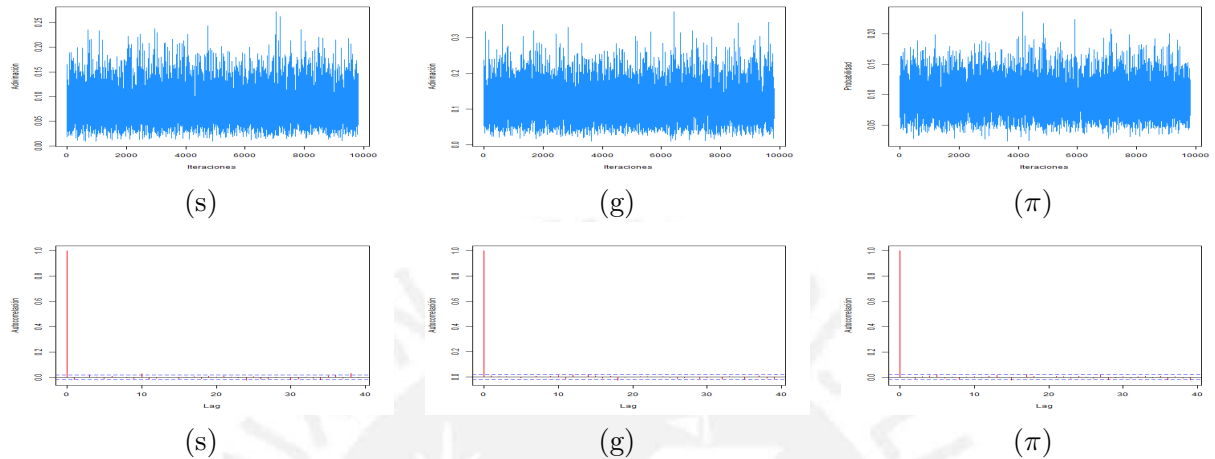


D.1.7. Estudio de simulación 7

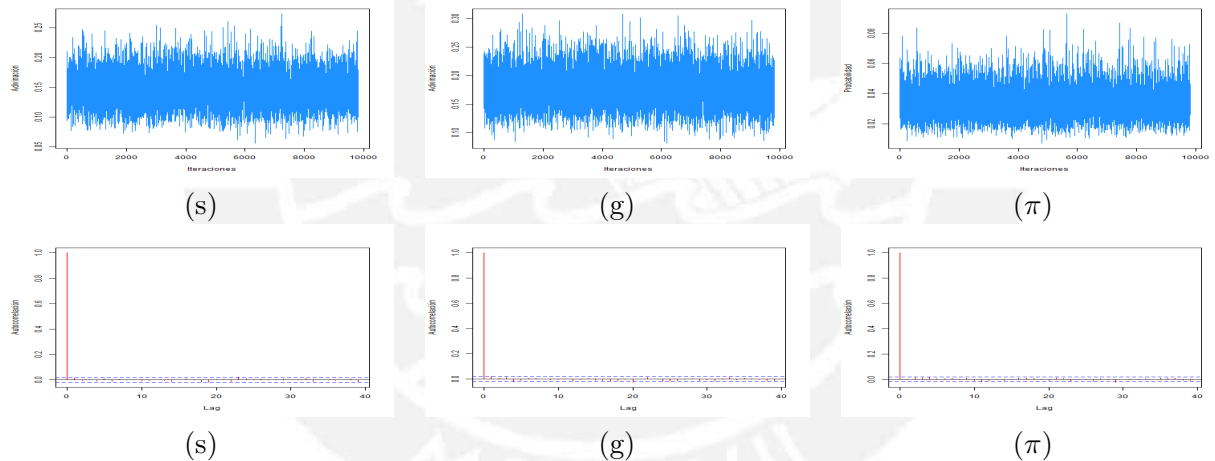
En el cuadro D.7 se presentan los gráficos para las cadenas y autocorrelaciones de las estimaciones de los parámetros de un ítem (s y g) y de la probabilidad de una clase latente (π); para cada uno de los tres escenarios de este estudio.

Cuadro D.7: Gráficos de las cadenas y autocorrelaciones para los parámetros s , g y π de un ítem y una clase, respectivamente, elegido aleatoriamente para el estudio 7 de simulación

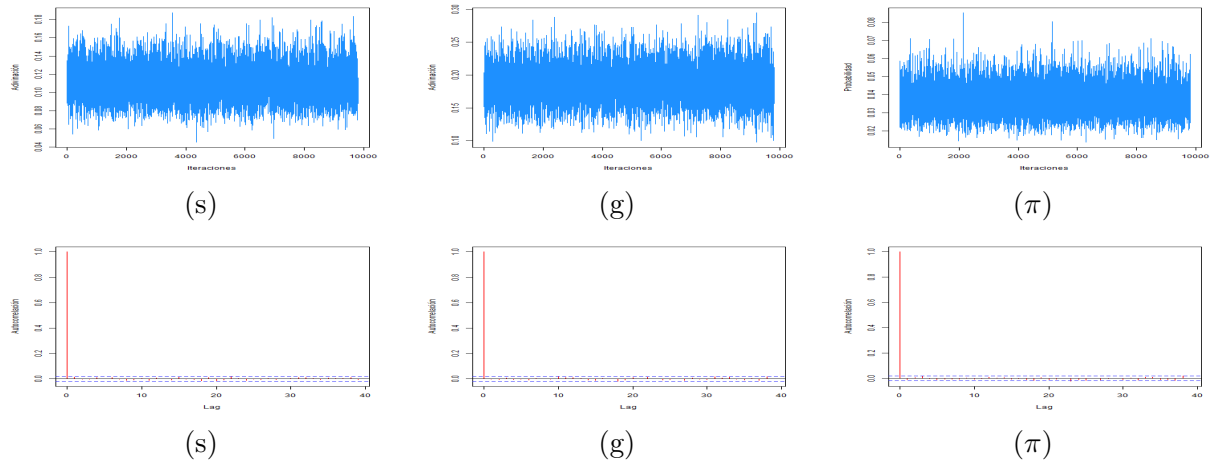
Escenario 1: comportamiento del ítem 37 y la clase 14 (1011).



Escenario 2: comportamiento del ítem 28 y la clase 8 (1001)



Escenario 3: comportamiento del ítem 21 y la clase 10 (0101)

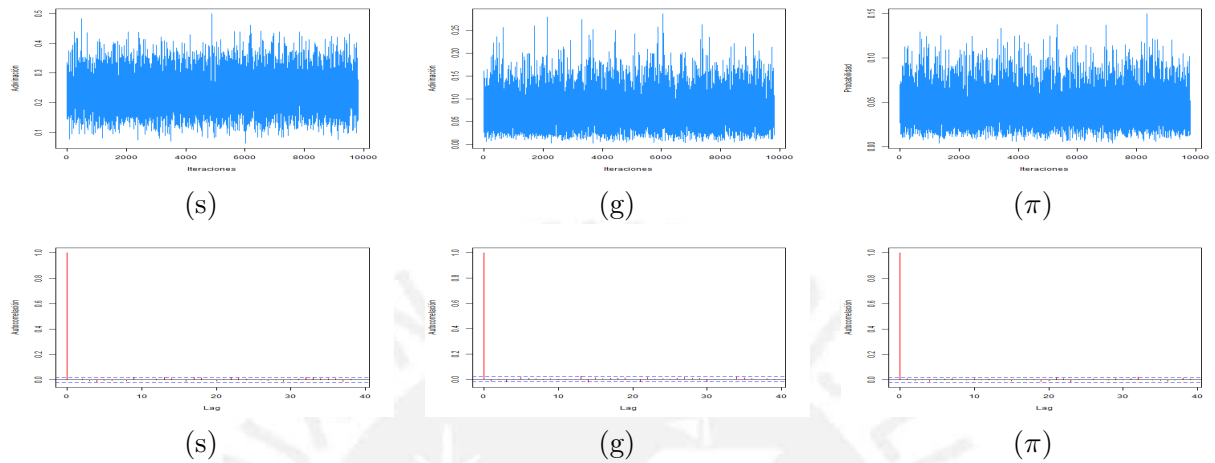


D.1.8. Estudio de simulación 8

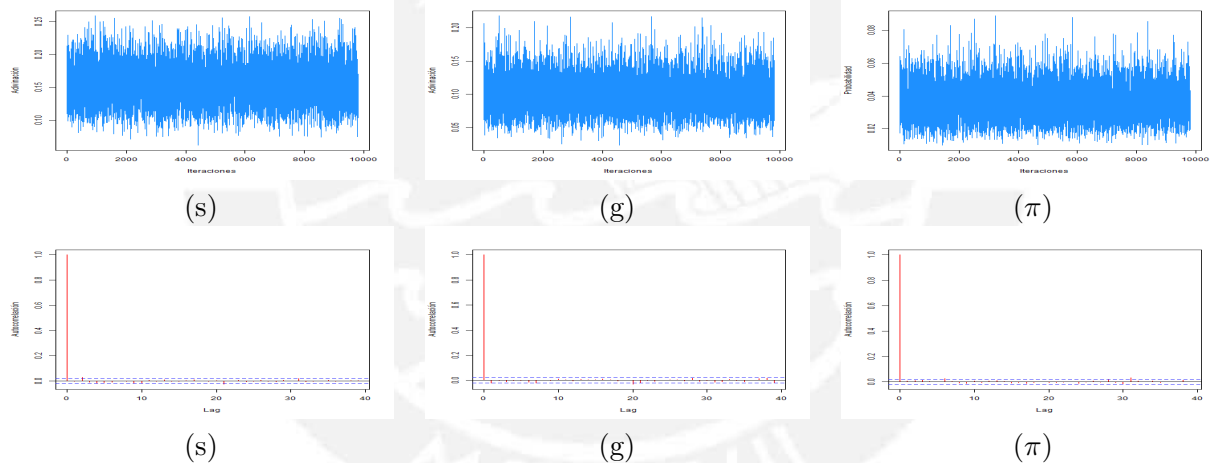
En el cuadro D.8 se presentan los gráficos para las cadenas y autocorrelaciones de las estimaciones de los parámetros de un ítem (s y g) y de la probabilidad de una clase latente (π); para cada uno de los tres escenarios de este estudio.

Cuadro D.8: Gráficos de las cadenas y autocorrelaciones para los parámetros s , g y π de un ítem y una clase, respectivamente, elegido aleatoriamente para el estudio 8 de simulación

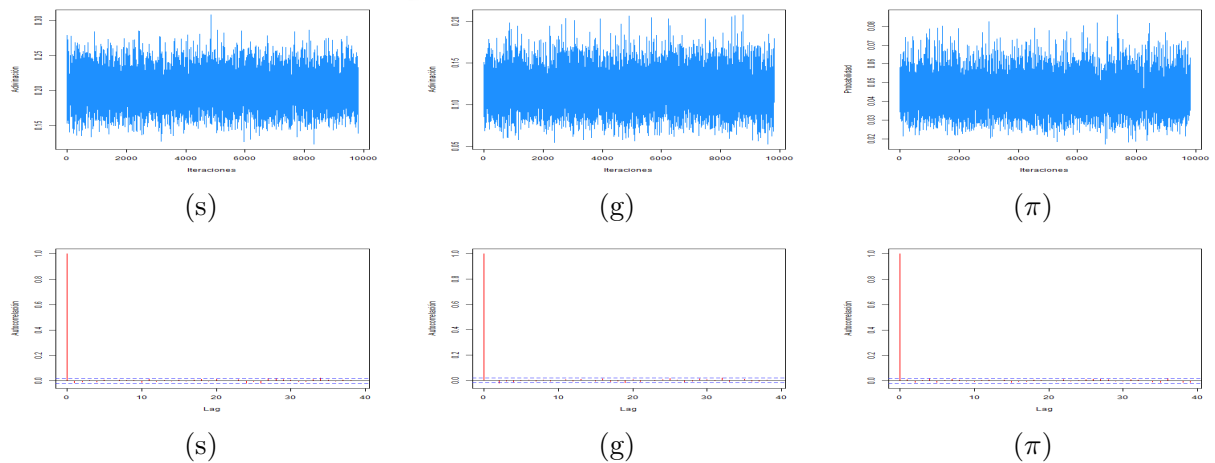
Escenario 1: comportamiento del ítem 42 y la clase 9 (0110).



Escenario 2: comportamiento del ítem 48 y la clase 7 (1010)



Escenario 3: comportamiento del ítem 2 y la clase 1 (0000)

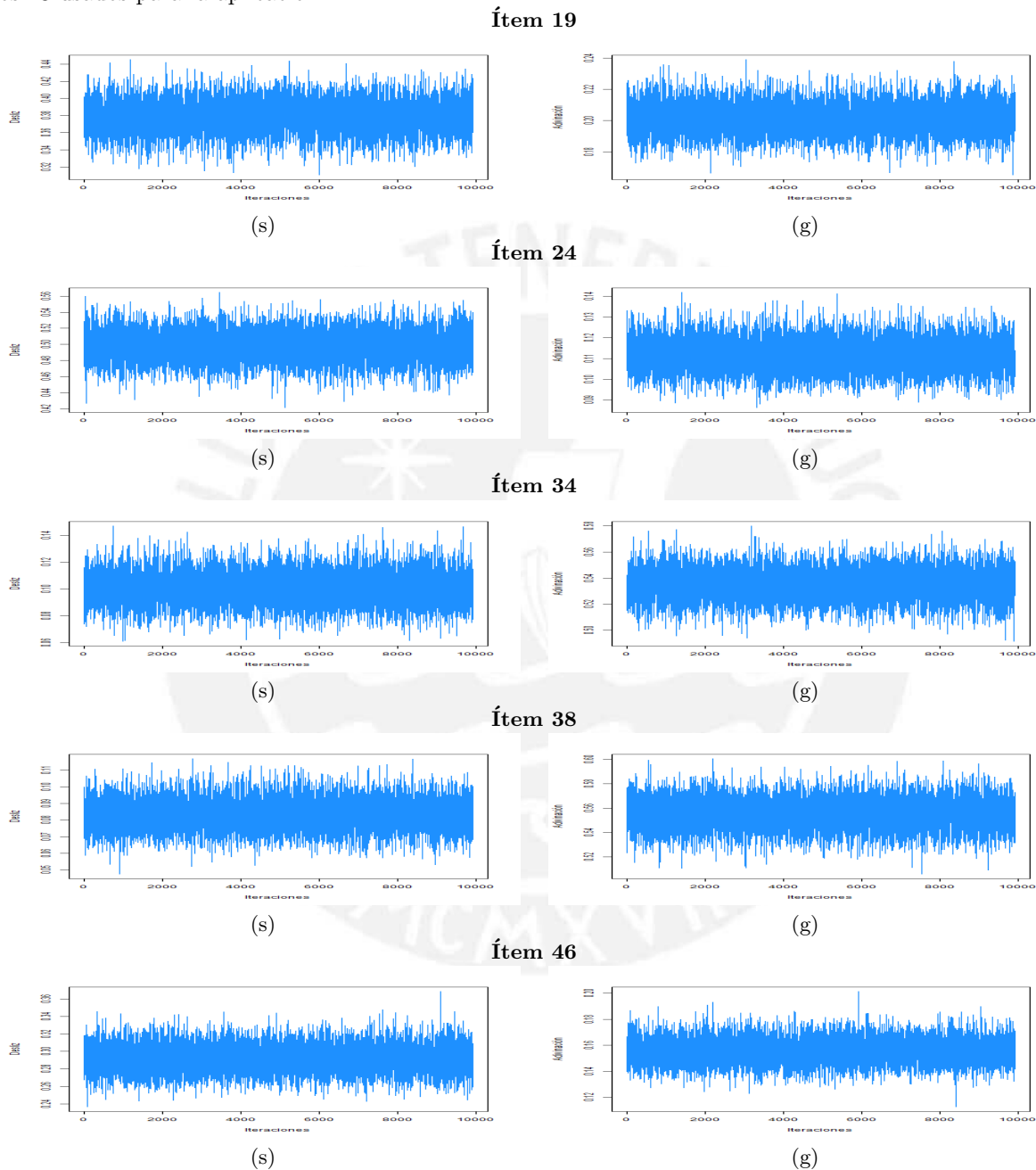


D.2. Convergencia del modelo DINA

D.2.1. Comportamiento de la Cadena y densidad

En el cuadro D.9 se presentan los gráficos para las cadenas de las estimaciones de los parámetros: s y g ; de 5 ítems elegidos aleatoriamente de los 48 utilizados.

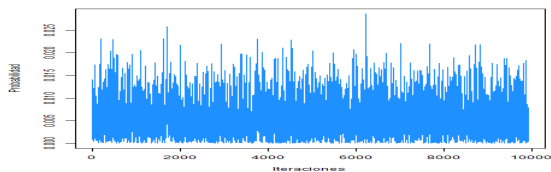
Cuadro D.9: Gráficos de las cadenas para los parámetros s y g de una selección aleatoria de 5 ítems de los 48 usados para la aplicación



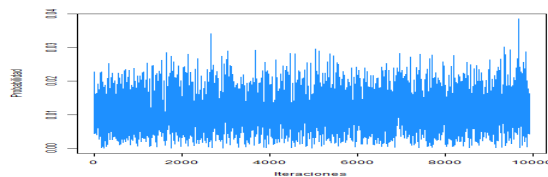
En el cuadro D.10 se presentan los gráficos para las cadenas de las estimaciones del parámetro π ; de 5 clases elegidas aleatoriamente de las 16 posibles.

Cuadro D.10: Gráficos de las cadenas para el parámetro π de una selección aleatoria de 5 clases de las 16 posibles en la aplicación

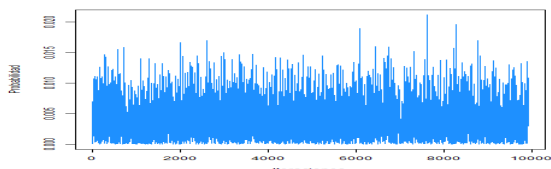
Clase 5: MA=0; CR=0; EE=0; RA=1



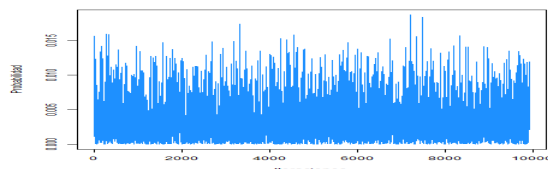
Clase 9: MA=0; CR=1; EE=1; RA=0



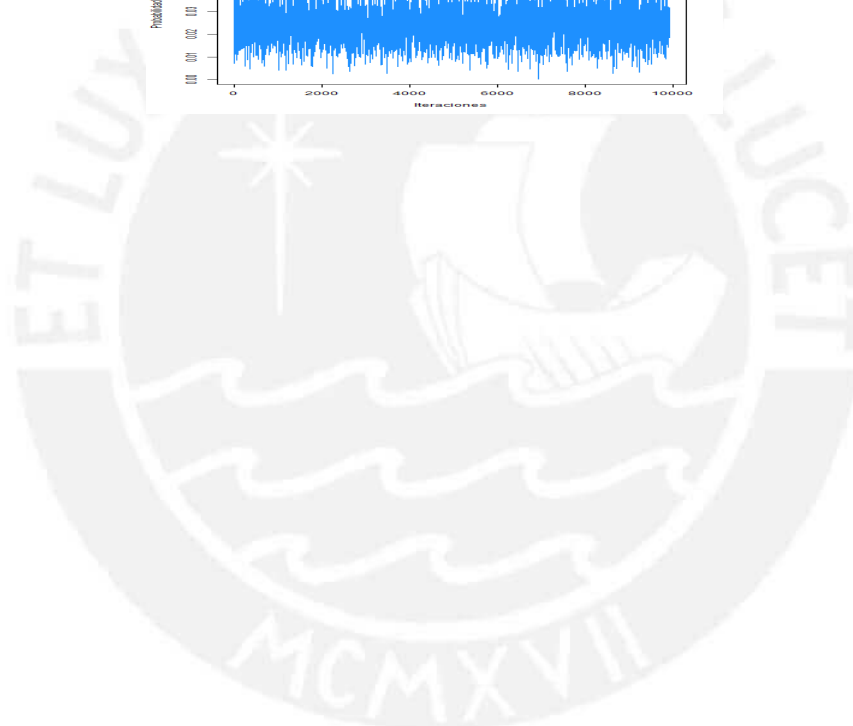
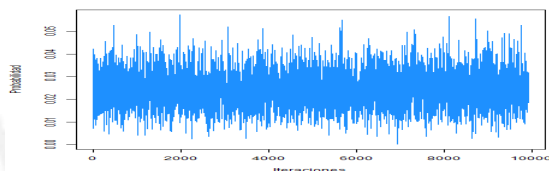
Clase 12: MA=1; CR=1; EE=1; RA=0



Clase 13: MA=1; CR=1; EE=0; RA=1



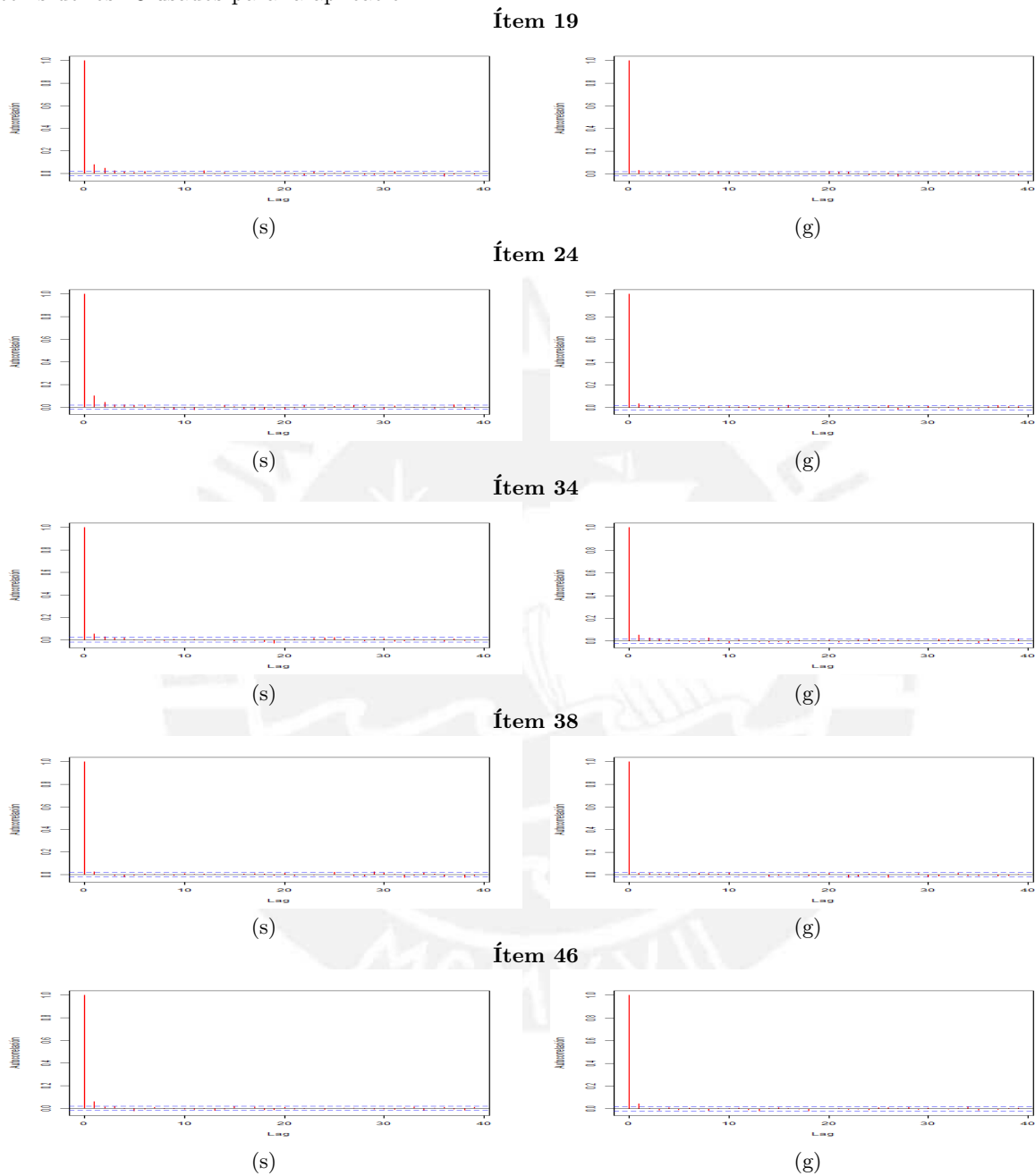
Clase 15: MA=0; CR=1; EE=1; RA=1



D.2.2. Autocorrelaciones

En el cuadro D.11 se presentan los gráficos de autocorrelación de las estimaciones de los parámetros: s y g ; de 5 ítems elegidos aleatoriamente de los 48 utilizados.

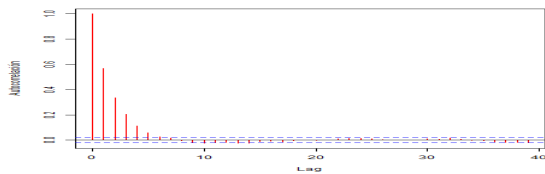
Cuadro D.11: Gráficos de las autocorrelaciones para los parámetros s y g de una selección aleatoria de 5 ítems de los 48 usados para la aplicación



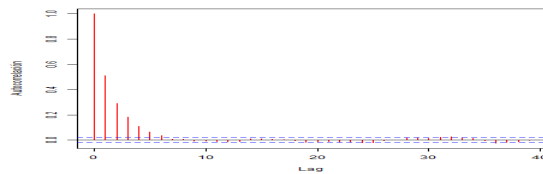
En el cuadro D.12 se presentan los gráficos para las autocorrelaciones de las estimaciones del parámetro π ; de 5 clases elegidas aleatoriamente de las 16 posibles.

Cuadro D.12: Gráficos de las autocorrelaciones para el parámetro π de una selección aleatoria de 5 clases de las 16 posibles en la aplicación

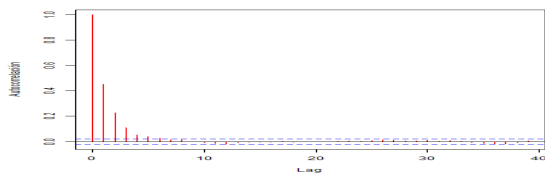
Clase 5: MA=0; CR=0; EE=0; RA=1



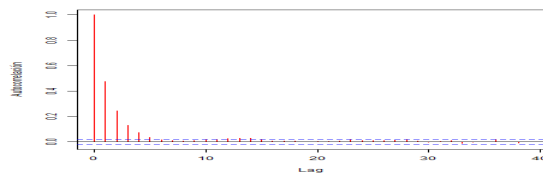
Clase 9: MA=0; CR=1; EE=1; RA=0



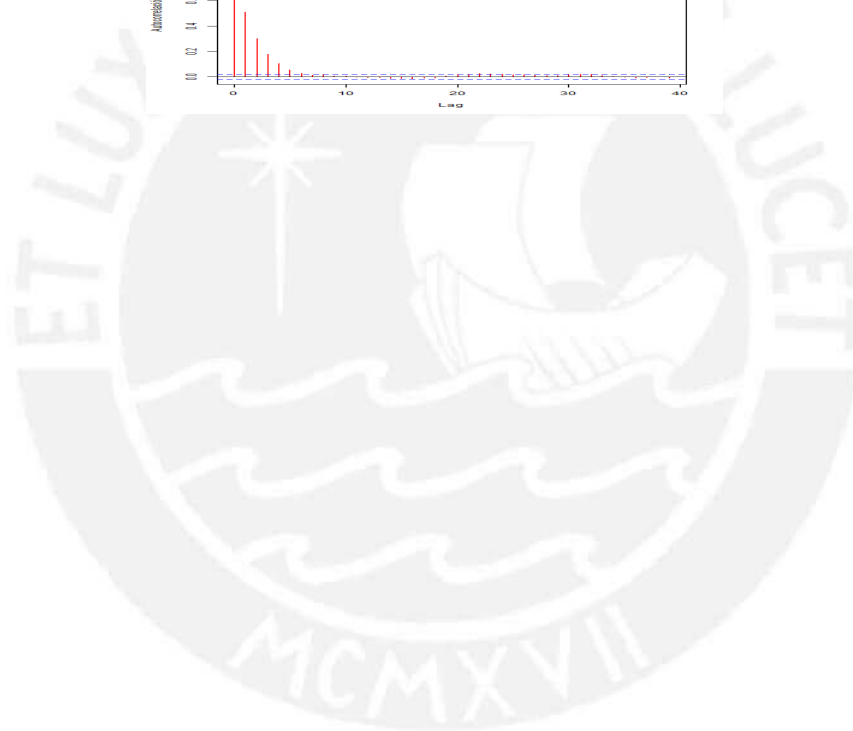
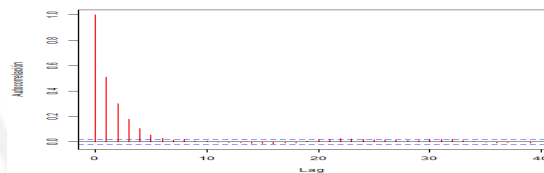
Clase 12: MA=1; CR=1; EE=1; RA=0



Clase 13: MA=1; CR=1; EE=0; RA=1

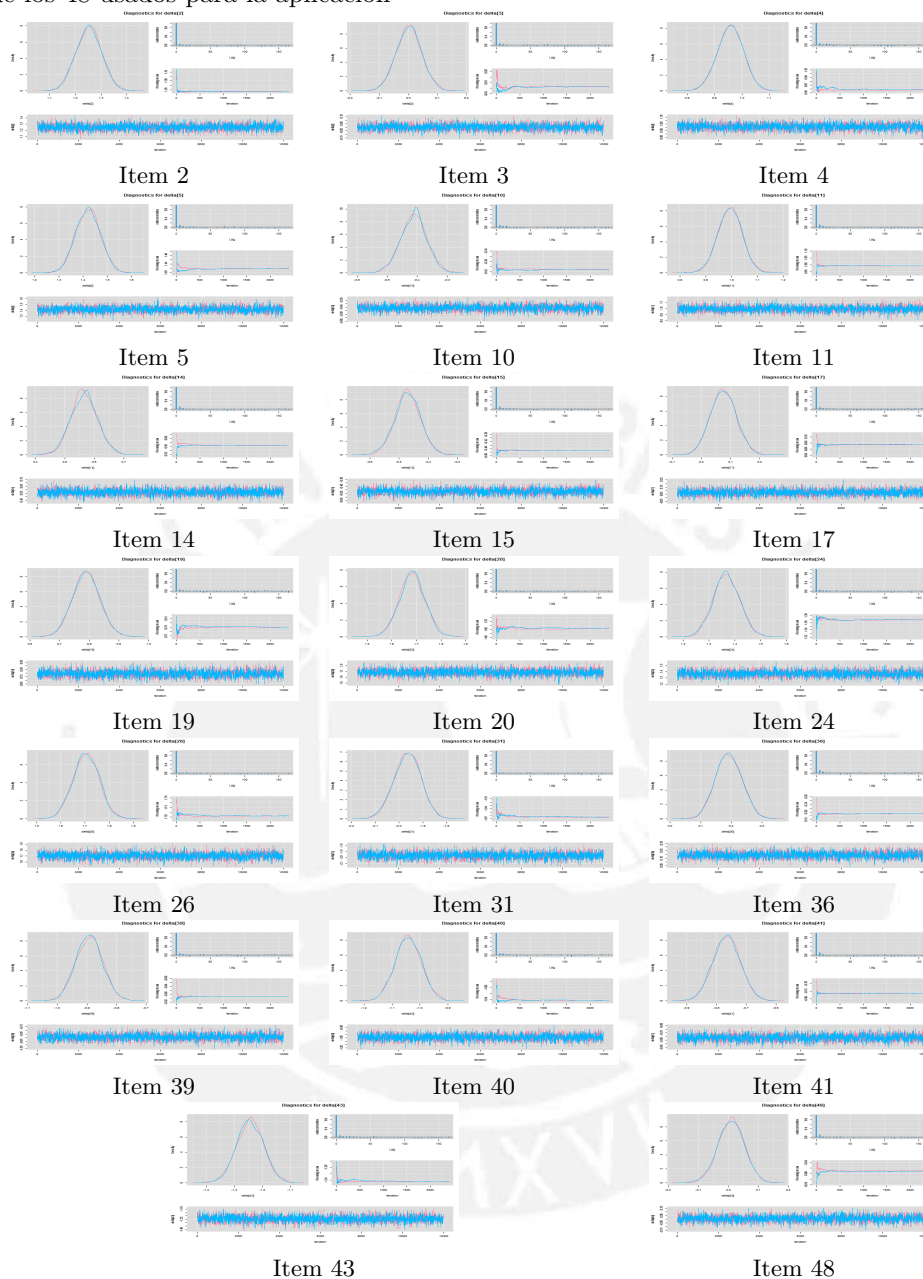


Clase 15: MA=0; CR=1; EE=1; RA=1



D.3. Convergencia del modelo Rasch

Cuadro D.13: Gráficos de densidad, de las cadenas y de autocorrelación de una selección de 20 ítems aleatorios de los 48 usados para la aplicación



Bibliografía

- Chincaro, O. A. (2010). *Modelo de rasch dicotómico con aplicación a la educación*, Master's thesis, Pontificia Universidad Católica del Perú.
- Culpepper, S. (2015a). Bayesian estimation of the dina model with gibbs sampling, *Journal of Educational and Behavioral Statistics* **40**(5): 454–476.
- Culpepper, S. A. (2015b). *dina: Bayesian Estimation of DINA Model*. R package version 1.0.1.
URL: <https://CRAN.R-project.org/package=dina>
- De la Torre, J. (2009a). A Cognitive Diagnosis Model for Cognitively Based Multiple-Choice Options, *Applied Psychological Measurement* **33**(3): 163–183.
URL: <http://apm.sagepub.com/cgi/doi/10.1177/0146621608320523>
- De La Torre, J. (2009b). Dina model and parameter estimation: A didactic, *Journal of Educational and Behavioral Statistics* **34**(1): 115–130.
URL: <http://jeb.sagepub.com/content/34/1/115.abstract>
- De la Torre, J. y Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis, *Psychometrika* **69**(3): 333–353.
- De la Torre, J. y Minchen, N. (2014). Cognitively Diagnostic Assessments and the Cognitive Diagnosis Model Framework, *Psicología Educativa* **20**(2): 89–97.
URL: <http://linkinghub.elsevier.com/retrieve/pii/S1135755X14000116>
- De La Torre, J. (2011). The generalized dina model framework, *Psychometrika* **76**(3): 510–510.
URL: <http://dx.doi.org/10.1007/s11336-011-9214-8>
- DiBello, L., Roussos, L. y Stout, W. (2007). Review of cognitively diagnostic assessment and a summary of psychometric models, in C. R. Rao y S. Sinharay (eds), *Handbook of Statistics, Vol 26*, 1ra edn, Elsevier B.V, Amsterdam, chapter 31A, pp. 979–1030.
- Flores, S. E. (2012). *Modelos testlet logísticos y logísticos de exponente positivo para pruebas de comprensión de textos*, Master's thesis, Pontificia Universidad Católica del Perú.
- Gamerman, D. y Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, Chapman and Hall/CRC.
- George, A., Robitzsch, A., Kiefer, T., Groß, J. y Ünlü, A. (2016). The r package cdm for cognitive diagnosis models, *Journal of Statistical Software* **74**(1): 1–24.
URL: <https://www.jstatsoft.org/index.php/jss/article/view/v074i02>
- Haertel, E. H. (1989). Using Restricted Latent Class Models to Map the Skill Structure of Achievement Items, *Journal of Educational Measurement* **26**(4): 301–321.
URL: <http://doi.wiley.com/10.1111/j.1745-3984.1989.tb00336.x>
- Henson, R. y Douglas, J. (2005). Test construction for cognitive diagnosis, *Applied Psychological Measurement* **29**(4): 262–277.
- Hoff, P. D. (2009). *A First Course in Bayesian Statistical Methods*, 1st edn, Springer Publishing Company, Incorporated.
- Huebner, A. (2010). An overview of recent developments in cognitive diagnostic computer adaptive assessments, *Practical Assessments Research & Evaluation* **15**(3): 1–7.
- Jones, L. y Thissen, D. (2007). A history and overview of psychometrics, in C. R. Rao y S. Sinharay (eds), *Handbook of Statistics, Vol 26*, 1ra edn, Elsevier B.V, Amsterdam, chapter 1, pp. 1 – 27.

- Junker, B. W. y Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory, *Applied Psychological Measurement* **25**(3): 258–272.
- Lim, Y. (2015). *Cognitive Diagnosis Model Comparisons*, PhD thesis, The Academic Faculty, Georgia.
- Lohr, S. L. (1999). *Sampling: Design and Analysis*, 1 edn, Duxbury Press, United States.
- Lunn, D. J., Thomas, A., Best, N. y Spiegelhalter, D. J. (2007). *WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility*, Hingham, MA, USA.
URL: <http://dx.doi.org/10.1023/a:1008929526011>
- McGlohen, M. (2004). *The Application of a Cognitive Diagnosis Model via an Analysis of a Large-Scale Assessment and a Computerized Adaptive Testing Administration*, thesis, The University of Texas.
- Minedu (2015). Rutas del aprendizaje, versión 2015.
URL: <http://www.minedu.gob.pe/rutas-del-aprendizaje/documentos/Secundaria/Matematica-VI.pdf>
- Minedu (2016). Reporte técnico de la evaluación censal de estudiantes (ECE 2015), *Technical report*, Oficina de Medición de la Calidad de los Aprendizajes (UMC), Ministerio de Educación, Perú.
URL: <http://umc.minedu.gob.pe/>
- Paek, I. y Cai, L. (2014). A comparison of item parameter standard error estimation procedures for unidimensional and multidimensional item response theory modeling, *Educational and Psychological Measurement* **74**(1): 58–76.
URL: <http://dx.doi.org/10.1177/0013164413500277>
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: <https://www.R-project.org/>
- Reckase, M. D. (2009). *Multidimensional Item Response Theory*, 1 edn, Springer, New York.
- Robitzsch, A., Kiefer, T., George, A. C. y Uenlue, A. (2016). *CDM: Cognitive Diagnosis Modeling*. R package version 5.1-0.
URL: <https://CRAN.R-project.org/package=CDM>
- Rupp, A. A. (2007). The answer is in the question: A guide for describing and investigating the conceptual foundations and statistical properties of cognitive psychometric models, *International Journal of Testing* **7**(2): 95–125.
- Rupp, A. A. y Templin, J. L. (2008). Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art, *Measurement: Interdisciplinary Research and Perspectives* **6**(4): 219–262.
- Van der Linden, W. J. y Hambleton, R. K. (1997). *Item Response Theory: Brief History, Common Models, and Extensions*, Springer, New York, NY, pp. 1–28.