

*Modelo LSMIRT para múltiples poblaciones.*

GUALBERTO SEGUNDO AGÁMEZ MONTALVO

ESTADÍSTICO

CÓDIGO: 832501



UNIVERSIDAD NACIONAL DE COLOMBIA

FACULTAD DE CIENCIAS

DEPARTAMENTO DE ESTADÍSTICA

BOGOTÁ, D.C.

MARZO DE 2014

*Modelo LSMIRT para múltiples poblaciones.*

GUALBERTO SEGUNDO AGÁMEZ MONTALVO  
ESTADÍSTICO  
CÓDIGO: 832501

TRABAJO DE TESIS PARA OPTAR AL TÍTULO DE  
MAGÍSTER EN CIENCIAS ESTADÍSTICA

DIRECTOR  
ALVARO MAURICIO MONTENEGRO DÍAZ, PH.D.  
DOCTOR EN ESTADÍSTICA



UNIVERSIDAD NACIONAL DE COLOMBIA  
FACULTAD DE CIENCIAS  
DEPARTAMENTO DE ESTADÍSTICA  
BOGOTÁ, D.C.  
MARZO DE 2014

## **Título en español**

Modelo LSMIRT para múltiples poblaciones.

## **Title in English**

Model LSMIRT for multiple populations.

**Resumen:** Se propone un modelo multidimensional de teoría de respuesta al ítem con una estructura lineal latente para  $g$  grupos o subgrupos poblaciones. El modelo fue diseñado para ajustar datos provenientes de pruebas binarias o dicotomizadas, las cuales están divididas en  $m$  subpruebas y son aplicadas a  $g$  grupos o subgrupos poblaciones. Se asume que cada subprueba está diseñada para medir un trazo latente unidimensional (trazo latente principal o habilidad principal), y medir estos es el objetivo de la prueba. Además, se asume que la prueba completa mide un vector de trazos latentes de los examinados, cuyas componentes no coinciden necesariamente con los trazos latentes principales. En lugar de eso, se asume que los trazos latentes principales son composiciones (combinaciones lineales) de las componentes del vector de trazos latentes. Por lo tanto, los trazos latentes principales tienen una estructura lineal latente. Se supone que cada ítem pertenece exactamente a una subprueba. En el modelo estudiado, la dimensión de la prueba se define como el número de sus subpruebas y no es necesariamente igual a la dimensión del espacio de trazos latentes. Para estimar los parámetros de los modelos propuestos, se desarrolló un algoritmo para implementar un muestreador de Gibbs con datos aumentados (DAGS). El modelo fue utilizado para ajustar los datos del Primer Estudio Internacional Comparativo sobre Lenguaje, Matemática y Factores Asociados en Tercero y Cuarto Grado (PERCE), que fue aplicada por el Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación.

**Abstract:** A multidimensional item response theory model with latent linear structure for several groups is proposed. This model was introduced in order to fit binary tests, which in turn are divided in several subtest and subsequently applied to different groups or populations. It is assumed that each subtest measure a one-dimensional latent trait (main latent trait or main ability). The main aim is to measure these latent traits. Furthermore, it is also assumed that the entire test measures a latent trait vector from tested subjects. This latent trait vector does not necessarily have the same components as the main latent trait. Instead, it is supposed that the main latent traits are linear combinations of latent trait vector components. Therefore, they have a linear latent structure. Each item is assumed to belong to exactly one subtest. In this model, the test dimension is defined as the number of subtest and it may not equal the latent trait space dimension. In order to estimate the parameters, an augmented data Gibbs sampler (DAGS) was implemented and tested in simulations. Besides, the model was used to fit data from the 'First comparative survey on language, math and associated factors for 3rd and 4th year students (PERCE)', which was carried out by the Latinamerican laboratory for assessment of quality of education.

**Palabras clave:** teoría de respuesta al ítem multidimensional, estructura lineal latente, múltiples grupos, subprueba, trazo latente.

**Keywords:** multidimensional item response theory, linear latent structure, multiple groups, subtest, latent trait.

---

---

# Índice general

---

---

<b>Índice general</b>	<b>II</b>
<b>Índice de tablas</b>	<b>IV</b>
<b>Índice de figuras</b>	<b>VI</b>
<b>Introducción</b>	<b>VIII</b>
<b>1. Introducción a los modelos MIRT.</b>	<b>1</b>
1.1. Introducción a los modelos multidimensionales de IRT. . . . .	1
1.2. El modelo MIRT generalizado. . . . .	4
1.3. Funciones de información. . . . .	5
1.4. Componente de referencia de una prueba. . . . .	5
1.5. Trazo o habilidad sintética. . . . .	6
1.6. Introducción a los modelos MIRT para una estructura lineal latente. . . . .	8
1.6.1. Trazos latentes principales y básicos. . . . .	9
1.6.2. Modelo MIRT para una estructura lineal latente. . . . .	10
1.6.3. Identificabilidad del modelo LSMIRT. . . . .	11
1.6.4. Estimación de los parámetros. . . . .	12
<b>2. Modelo LSMIRT para múltiples poblaciones.</b>	<b>13</b>
2.1. Introducción a los modelos LSMIRT para múltiples poblaciones. . . . .	13
2.2. Modelo propuesto. . . . .	13
2.3. Estimación de los parámetros. . . . .	14
2.4. Identificabilidad del modelo. . . . .	15
2.5. Simulación. . . . .	15
2.5.1. Simulación del test. . . . .	15

---

2.5.2. Ajuste de los datos simulados. . . . .	17
2.6. Aplicación a un caso real. . . . .	23
2.6.1. Ajuste del caso de datos reales. . . . .	23
2.6.2. Parámetros estimados. . . . .	23
2.6.3. Bondad de ajuste. . . . .	28
2.6.4. Discusión. . . . .	30
<b>3. Conclusiones</b>	<b>32</b>
<b>4. Trabajos futuros</b>	<b>34</b>
<b>A. Distribuciones posterioris completas</b>	<b>35</b>
A.1. Notación matricial . . . . .	35
A.2. Variables latentes $Z_{ijg}$ . . . . .	36
A.3. Distribución condicional completa a posteriori de $\theta_{ig}$ . . . . .	36
A.4. Distribución condicional completa a posteriori de $\beta_v$ . . . . .	36
A.5. Distribución condicional completa a posteriori de $\zeta_j$ . . . . .	37
A.6. Distribución condicional completa a posteriori de $\mu_g$ . . . . .	38
A.7. Muestrear $\Sigma_g$ . . . . .	38
A.8. Algoritmo DAGS . . . . .	39
<b>B. Figuras de las cadenas de <math>\Sigma_g</math> para los datos simulados.</b>	<b>40</b>
<b>C. Resultados del modelo G LSMIRT para los datos reales.</b>	<b>44</b>
C.1. Figuras de las cadenas de $\Sigma_g$ . . . . .	44
C.2. Algunas figuras de las cadenas de $\alpha$ . . . . .	45
C.3. Parámetros de pendiente e intercepto. . . . .	47
<b>Bibliografía</b>	<b>48</b>

---



---

## Índice de tablas

---



---

2.1.	La columna $a_{vj}$ es el parámetro de discriminación multidimensional en el modelo clásico MIRT; las columnas $\beta_{vj1}$ , $\beta_{vj2}$ , $\beta_{vj3}$ son las componentes de las direcciones originales de los ítems; la columna $\rho_{vj}$ es el producto punto entre la dirección original del ítem y la dirección de la subprueba correspondiente; la columna $\alpha_{vj}$ es el parámetro de pendiente verdadero en el modelo LSMIRT cuyo valor es el producto de $a_{vj}\rho_{vj}  \tilde{\mathbf{B}}^{-1}\boldsymbol{\beta}_v  $ y la columna $\gamma_{vj}$ es el parámetro de intercepto. . . . .	16
2.2.	Las columnas $\hat{a}_{vj}$ y $\hat{\gamma}_{vj}$ son las estimaciones de $a_{vj}$ y $\gamma_{vj}$ respectivamente, así como $sd_{\hat{a}_{vj}}$ y $sd_{\hat{\gamma}_{vj}}$ son las desviaciones estándar posteriori de las estimativas. . . . .	18
2.3.	Comparaciones globales entre $\alpha_{vj}$ y $\gamma_{vj}$ verdaderos y los estimados. . . . .	18
2.4.	Comparaciones globales entre las componentes de referencia verdaderos y los estimados en la prueba simulada. . . . .	20
2.5.	Estimaciones de las componentes de $\boldsymbol{\mu}_g$ , desviaciones estándar posterior, errores absolutos y relativos. . . . .	21
2.6.	Estimaciones de las componentes de $\boldsymbol{\Sigma}_2$ , desviaciones estándar posterior, errores absolutos y relativos. . . . .	21
2.7.	Estimaciones de las componentes de $\boldsymbol{\Sigma}_3$ , desviaciones estándar posterior, errores absolutos y relativos. . . . .	21
2.8.	Estimaciones de las componentes de $\boldsymbol{\mu}_g$ y sus respectivas desviaciones estándar posterior para los datos reales. . . . .	23
2.9.	Estimaciones de las componentes de $\boldsymbol{\Sigma}_2$ y sus respectivas desviaciones estándar posterior. . . . .	24
2.10.	Estimaciones de las componentes de $\boldsymbol{\Sigma}_3$ y sus respectivas desviaciones estándar posterior. . . . .	25
2.11.	Estadísticas resumen para los trazos latentes principales estimados. . . . .	26
2.12.	Correlaciones entre los scores clásicos y los trazos principales estimados. . . . .	28
2.13.	Porcentaje de outliers y valores extremos por país. . . . .	29

- 
- C.1. Las columnas  $\hat{\alpha}_{vj}$  y  $\hat{\gamma}_{vj}$  son las estimaciones de los parámetros de pendiente e intercepto del modelo G-LSMIRT, así como  $sd_{\hat{\alpha}_{vj}}$  y  $sd_{\hat{\gamma}_{vj}}$  son las desviaciones estándar posteriori de las estimativas para los datos reales. . . . 47

---

---

## Índice de figuras

---

---

2.1.	Gráficos de dispersión entre los $a_{vj}$ simulados y los estimados. . . . .	18
2.2.	Gráficos de dispersión entre los $\gamma_{vj}$ simulado y el estimado. . . . .	19
2.3.	Trazos, distribuciones posteriori y autocorrelaciones parciales de las cadenas de Markov para las estimaciones de $\beta_{4d}$ . . . . .	19
2.4.	Diagramas de dispersión entre las componentes de referencia verdaderos y los estimados para cada subprueba de la población 1 . . . . .	20
2.5.	Trazos, distribuciones posteriori y autocorrelaciones parciales de las cadenas de Markov para las estimaciones de $\mu_{2d}$ . . . . .	22
2.6.	Trazos, distribuciones posteriori y autocorrelaciones parciales de las cadenas de Markov para las estimaciones de $\mu_{3d}$ . . . . .	22
2.7.	Trazos, distribuciones posteriori y autocorrelaciones parciales de las cadenas de Markov para las estimaciones de $\mu_{2d}$ . . . . .	24
2.8.	Trazos, distribuciones posteriori y autocorrelaciones parciales de las cadenas de Markov para las estimaciones de $\mu_{3d}$ . . . . .	24
2.9.	Densidades para la primera componente del vector de trazos latentes estimados por población . . . . .	25
2.10.	Densidades para la segunda componente del vector de trazos latentes estimados por población . . . . .	25
2.11.	Densidades de los trazos latentes unidimensionales por población para la subprueba de matemáticas . . . . .	26
2.12.	Densidades de los trazos latentes unidimensionales por población para la subprueba de lenguaje . . . . .	26
2.13.	Comparación entre las distribuciones empíricas de los datos reales y los trazos latentes estimados. . . . .	27
2.14.	Comparación entre las densidades empíricas de los datos reales y los trazos latentes estimados. . . . .	28
2.15.	Diagramas de dispersión de los scores clásicos contra los trazos principales estimados. . . . .	28



---

B.1. Trazos de las cadenas de Markov para las estimaciones de las componentes de $\Sigma_2$ para los datos simulados. . . . .	40
B.2. Distribuciones posteriori de las cadenas de Markov para las estimaciones de las componentes de $\Sigma_2$ para los datos simulados. . . . .	41
B.3. Autocorrelaciones parciales de las cadenas de Markov para las estimaciones de las componentes de $\Sigma_2$ para los datos simulados. . . . .	41
B.4. Trazos de las cadenas de Markov para las estimaciones de las componentes de $\Sigma_3$ para los datos simulados. . . . .	42
B.5. Distribuciones posteriori de las cadenas de Markov para las estimaciones de las componentes de $\Sigma_3$ para los datos simulados. . . . .	42
B.6. Autocorrelaciones parciales de las cadenas de Markov para las estimaciones de las componentes de $\Sigma_3$ para los datos simulados. . . . .	43
C.1. Trazos, distribuciones posteriori y autocorrelaciones parciales de las cadenas de Markov para las estimaciones de las componentes de $\Sigma_2$ para los datos reales. . . . .	44
C.2. Trazos, distribuciones posteriori y autocorrelaciones parciales de las cadenas de Markov para las estimaciones de las componentes de $\Sigma_3$ para los datos reales. . . . .	45
C.3. Trazos de algunas de las cadenas de Markov para las estimaciones de $\alpha$ para los datos reales. . . . .	45
C.4. Distribuciones posteriori de algunas de las cadenas de Markov para las estimaciones de $\alpha$ para los datos reales. . . . .	46
C.5. Autocorrelaciones parciales de algunas de las cadenas de Markov para las estimaciones de $\alpha$ para los datos reales. . . . .	46

---

---

## Introducción

---

---

Esta disertación está basada en el interés de desenvolver un modelo multidimensional de teoría de respuesta al ítem para una estructura lineal latente (LSMIRT), cuando se desea aplicar una prueba que se encuentra dividida en subpruebas aproximadamente unidimensionales<sup>1</sup> a  $g$  grupos o subgrupos poblacionales. Cada una de estas subpruebas es diseñada esencialmente para medir una única habilidad o trazo latente principal de los examinados a cada uno de los  $g$  grupos, y así poder determinar los cambios que estos tuvieron en los diferentes ítems a los cuales fueron sometidos. Por lo tanto, se propone un modelo LSMIRT para  $g$  grupos poblaciones (G-LSMIRT).

El modelo desarrollado en esta disertación puede aplicarse a pruebas (con las características mencionadas encima) en las cuales existen varios grupos de respondientes. Es usual que los grupos tengan diferencias entre ellos, por razones culturales, de idioma, sistemas educativos, entre otras. La prueba PERCE, que se aplicó en 1997 a estudiantes de 13 países latinoamericanos, con el objetivo de percibir los logros educativos de los países a través de enfoques comparados, que permitieran generar indicadores comunes y las pruebas PISA (Programa Internacional para la Evaluación de Estudiantes o por sus siglas en inglés: Program for International Student Assessment) que son aplicadas cada tres años, son ejemplos de este tipo de pruebas.

Se utilizaron procedimientos Bayesianos para estimar los parámetros del modelo propuesto. La estrategia de datos aumentada se utiliza en el algoritmo de estimación. Por experimentos de simulación, se muestra la calidad de los parámetros recuperados por el algoritmo DAGS.

En el capítulo 1 se presenta una breve revisión literaria de los principales aspectos de la teoría de respuesta al ítem, enfocándose en la parte multidimensional y principalmente en el modelo LSMIRT. En el capítulo 2 se presenta el modelo propuesto, así como las condiciones para que este sea identificable. Adicionalmente, se presentan los resultados obtenidos al modelar datos simulados y datos reales bajo el modelo G-LSMIRT. El conjunto de datos reales fue tomado del Primer Estudio Internacional Comparativo sobre Lenguaje, Matemática y Factores Asociados en Tercero y Cuarto Grado (PERCE), que fue aplicada por el Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación; empleando la misma muestra utilizada en Antonio (2013).

---

<sup>1</sup>Ninguna prueba es estrictamente unidimensional, por tanto, se habla de pruebas aproximadamente unidimensionales. Cabe notar que existen pruebas de unidimensionalidad para determinar esto.

---

La prueba fue aplicada a niños de tercer y cuarto grado entre los meses de Junio y Noviembre de 1997 en trece países de América Latina. La prueba tiene dos componentes (subpruebas): matemática con  $K_1 = 32$  ítems, y lenguaje con  $K_2 = 19$  ítems. Se diseñaron dos cuestionarios:  $A$  y  $B$ , que estaban compuestos por diferentes ítems, pero el número de estos en cada cuestionario es el mismo. A cada estudiante le correspondió responder uno de los dos cuestionarios al azar. La muestra aleatoria tomada fue de respuestas del cuestionario  $B$ , de tres de los países evaluados. El tamaño de la muestra fue  $N = 602$ , conformada por  $N_1 = 252$  datos de Bolivia,  $N_2 = 194$  datos de Colombia y  $N_3 = 156$  datos de Cuba. Ver Antonio (2013, pág 12).

En el capítulo 3 se presentan las conclusiones obtenidas en esta disertación, mientras que en el capítulo 4 los posibles trabajos futuros. Finalmente, en los apéndices se presentan las distribuciones condicionales completas, el algoritmo de datos aumentados implementado, y algunos resultados para los datos simulados y reales.

# CAPÍTULO 1

---

---

## Introducción a los modelos MIRT.

---

---

En este capítulo se presenta una revisión de la literatura sobre los principales aspectos de la teoría de respuesta al ítem (IRT) que se estudia en esta disertación, enfocándose en la parte multidimensional y principalmente en el modelo multidimensional de teoría de respuesta al ítem para una estructura lineal latente (LSMIRT) propuesto por Montenegro (2011).

### 1.1. Introducción a los modelos multidimensionales de IRT.

El principal objetivo de la IRT es estimar un vector de trazo latente o también conocido como un vector de habilidades para cada uno de los examinados ( $\theta$ ), por lo cual, una prueba es diseñada como una herramienta para estimar este vector. Reckase (2009) establece que hay dos tipos principales de modelos de teoría de respuesta al ítem multidimensional (MIRT): los modelos compensatorios y los modelos no compensatorios o parcialmente compensatorios.

**Definición 1.** *Modelo compensatorio.* Es un tipo de modelo que está basado en una combinación lineal de coordenadas  $\theta$ . Esa combinación lineal es usada con una ojiva normal o una forma logística para especificar la probabilidad de una respuesta correcta. La combinación lineal de coordenadas  $\theta$  pueden producir la misma suma con varias combinaciones de valores  $\theta$ . Si una coordenada  $\theta$  es baja, la suma será la misma si otra coordenada  $\theta$  es suficientemente alta.

**Definición 2.** *Modelo no compensatorio.* Este tipo de modelo separa las tareas cognitivas en una prueba de ítems en partes y usa un modelo unidimensional para cada una. La probabilidad de respuesta correcta para el ítem es el producto de las probabilidades para cada parte. El uso del producto de probabilidades resulta en características no lineales para esta clase de modelos. Además, el hecho que la probabilidad de respuesta correcta no pueda ser superior a la mayor de las probabilidades en el producto reduce la compensación de una coordenada  $\theta$  alta para una coordenada  $\theta$  baja.

**Definición 3.** La probabilidad de una respuesta correcta para el caso unidimensional, Baker & Seok-Ho (2004), Bock (1972), en el modelo logística de dos parámetros está dada por:

$$p_j(\theta_i) = P(Y_{ij} = 1 | \theta_i, a_j, b_j) = \frac{1}{1 + e^{-a_j(\theta_i - b_j)}} \quad (1.1)$$

donde,  $Y_{ij}$  es la respuesta del examinado  $i$  al ítem  $j$ :  $Y_{ij} = 1$  si el examinado  $i$  responde correctamente el ítem  $j$ ,  $Y_{ij} = 0$  en otro caso;  $\theta_i$  es el parámetro unidimensional de habilidad para el examinado  $i$ , el parámetro de escala  $a_i$  es llamado el parámetro de discriminación del ítem  $j$ , y  $b_j$  es el parámetro de dificultad o posición del ítem  $j$ .

La función  $f_j(\theta) = p_j(\theta)$ , es llamada la función de respuesta al ítem (IRF) y su gráfica es la curva de respuesta al ítem (IRC). Se puede ver que  $f_j(b_j) = \frac{1}{2}$  y  $f'(b_j) = \frac{1}{4}a_j$ , por lo tanto, excepto por el término  $\frac{1}{4}$ ,  $a_j$  representa la pendiente de la IRC en el punto  $b_j$ .

**Definición 4.** La probabilidad que un examinado con vector de habilidad  $\boldsymbol{\theta}_i$  responda correctamente al ítem  $j$ , en el modelo clásico MIRT compensatorio, donde se asume que hay más de una habilidad que es medida por la prueba, está dada por

$$P(Y_{ij} = 1 | \boldsymbol{\theta}_i, \mathbf{a}_j, \gamma_j) = \frac{1}{1 + e^{-(\mathbf{a}_j^t \boldsymbol{\theta}_i - \gamma_j)}} \quad (1.2)$$

donde,  $\boldsymbol{\theta}_i$  es un vector de  $R^d$  que representa el vector de habilidad del examinado  $i$ ,  $\mathbf{a}_j$  es un vector de  $R^d$  relacionado con la discriminación del ítem y  $\gamma_j$  es una escalar relacionado con la dificultad del ítem. El componente  $\theta_{ik}$  de  $\boldsymbol{\theta}_i$  representa la habilidad del examinado  $i$  en la  $k$ -th dimensión, Reckase (2009).

Reckase (1985; 1997; 2007), afirma que el modelo MIRT no provee una interpretación directa de los parámetros  $\mathbf{a}_j$  y  $\gamma_j$ . En este caso, la función de respuesta al ítem (IRF)  $f_j(\boldsymbol{\theta}) = p_j(\boldsymbol{\theta})$  es una función multivariante, y su grafico el cual representa la curva de respuesta al ítem (IRC) es una hipersuperficie.

Por otro lado,  $\alpha_j$  es la norma del vector  $\mathbf{a}_j$ , es decir,

$$\alpha_j = \sqrt{\sum_{k=1}^d a_{jk}^2}$$

donde  $a_{jk}$  son las componentes del vector  $\mathbf{a}_j$ . Entonces, el vector  $\mathbf{a}_j$  puede ser reescrito como

$$\mathbf{a}_j = \alpha_j \boldsymbol{\beta}_j$$

donde  $\boldsymbol{\beta}_j = (\beta_{j1}, \beta_{j2}, \dots, \beta_{jd})^t$ ,  $\beta_{jk} = a_{jk}/\alpha_j$  y claramente se puede ver que  $\boldsymbol{\beta}_j$  es un vector unitario de  $R^d$ . Esta nomenclatura fue propuesta en Montenegro & Cepeda (2010).

Por el resultado anterior, el modelo de la ecuación 1.2 puede reescribirse como

$$P(Y_{ij} = 1 | \boldsymbol{\theta}_i, \alpha_j, \boldsymbol{\beta}_j, b_j) = \frac{1}{1 + \exp -\alpha_j(\boldsymbol{\beta}_j^t \boldsymbol{\theta}_i - b_j)} \quad (1.3)$$

donde  $b_j = \gamma_j/\alpha_j$ .

Reckase (1985) define el valor  $\alpha_j$  como el parámetro de discriminación multidimensional (MDISC) y el valor  $b_j$  como el parámetro de dificultad multidimensional (MDIFF).

En los modelos de teoría de respuesta al ítem unidimensional (UIRT), la función de respuesta al ítem y la curva de respuesta son casi intercambiables, sin embargo, en el caso multidimensional el asunto no es tan simple.

En Montenegro & Cepeda (2010) y Montenegro (2011), se muestra que cualquier ítem de un modelo MIRT compensatorio es esencialmente unidimensional, además se tiene que la hipótesis de respuesta al ítem de un ítem en un modelo MIRT es monótona a lo largo de cualquier dirección. Esta propiedad permite intercambiar la función de respuesta del ítem y la hipótesis de respuesta del ítem (IRSH) como en el caso unidimensional, pero también nos permite determinar que un ítem realmente mide en un modelo de MIRT.

**Teorema 1.** *La IRF multidimensional  $f_j(\boldsymbol{\theta})$  de un modelo MIRT es una extensión trivial del modelo clásico UIRF.*

*Demostración.* Sea  $\boldsymbol{\theta}$  un vector en  $R^d$ , y sea  $\{\boldsymbol{\beta}_j, \mathbf{v}_1, \dots, \mathbf{v}_{d-1}\}$  una base ortogonal normalda de  $R^d$  que contiene el vector  $\boldsymbol{\beta}_j$ . Entonces, existen números reales  $t, t_1, \dots, t_{d-1}$  tales que

$$\boldsymbol{\theta} = t\boldsymbol{\beta}_j + t_1\mathbf{v}_1 + \dots + t_{d-1}\mathbf{v}_{d-1},$$

entonces,

$$\boldsymbol{\beta}_j^t \boldsymbol{\theta} = (\boldsymbol{\beta}_j^t \boldsymbol{\beta}_j) = t.$$

$$f_j(\boldsymbol{\theta}) = \frac{1}{1 + \exp -\alpha_j \boldsymbol{\beta}_j^t \boldsymbol{\theta} + \gamma} = \frac{1}{1 + \exp -\alpha_j t + \gamma} = \frac{1}{1 + \exp -\alpha_j(t - b_j)} = p_{\boldsymbol{\beta}_j}(t).$$

□

El Teorema 1 muestra que cada ítem  $j$  en un modelo MIRT es esencialmente unidimensional a lo largo de la dirección dada por el vector  $\beta_j$ . El ítem  $j$  mide el compuesto  $\beta_j^t \theta_i$ . Entonces, cada ítem mide una combinación lineal diferente de la  $\theta_i$ , a menos que toda la  $\beta_j$  vectores tienen la misma dirección.

**Definición 5.** Sea  $p_j(t)$  la UIRF definida por

$$p_j(t) = \frac{1}{1 + \exp -\alpha_j(t - b_j)}. \quad (1.4)$$

La función  $p_j(t)$  puede ser trivialmente extendida a una función multivariada por  $p_j(t_1, \dots, t_d) = p_j(t_1)$ . La hipótesis original es obtenida por una rotación rígida de la hipótesis definida por  $p_j(t_1, \dots, t_d)$  sobre el hiperplano definido por los vectores canónicos  $e_1, e_2, \dots, e_d$ , que alinea el vector  $e_1$  con el vector  $\beta_j$ .

## 1.2. El modelo MIRT generalizado.

Un modelo MIRT tiene básicamente cuatro componentes, los cuales son similares a los componentes del modelo lineal generalizado propuesto por Nelder & Wedderburn (1972), excepto por el primer componente. Los componentes de un modelo MIRT son:

1. *El vector de trazo latente.* El vector de trazo latente del examinado  $i$  se denota como  $\theta_i = (\theta_{i1}, \theta_{i2}, \dots, \theta_{id})^t$ . Se supone que  $\theta_i$  es un valor de una variable latente  $\Theta = N_d(\mathbf{0}, \Sigma)$ . Otras distribuciones simétricas son posibles, como la distribución t-student multivariada, sin embargo, algunos autores han propuesto distribuciones asimétricas como la normal o t-student multivariadas sesgadas, Bazán (2006), Da Silva (2008).
2. *El predictor lineal latente.* El predictor lineal latente asociado con la respuesta del examinado  $i$  al ítem  $j$ , está definido como

$$\eta_{ij} = \alpha_j \beta_j^t \theta_i - \gamma_j, \quad i = 1, 2, \dots, N; \quad j = 1, 2, \dots, p \quad (1.5)$$

3. *El componente estocástico.* Sea  $Y_{ij}$  una variable aleatoria que representa la respuesta del examinado  $i$  al ítem  $j$ . Entonces, se supone que  $Y_{ij}$  es una variable binaria con parámetro  $p_{ij} = P(Y_{ij} = 1)$ .
4. *La función de enlace.* Zhang & Stout (1999), propuso que la función de enlace es una función no decreciente  $H_j(x)$  con  $H_j'(x) \geq 0$  para todo  $x$  y  $H_j'(x)H_l'(y)$  no cero para  $j, l = 1, \dots, p$ .

En general, una función de distribución acumulativa univariada (cdf) que tiene una función de densidad de probabilidad (pdf) satisface la definición anterior, sin embargo, las funciones de densidad de probabilidad logística y la normal estándar

univariada son los vínculos más extendidos.

La función de enlace relaciona el predictor lineal latente con la media de la variable aleatoria binaria  $Y_{ij}$ , esta relación es expresada como

$$P(Y_{ij} = 1 | \boldsymbol{\theta}_i, \alpha_j, \boldsymbol{\beta}_j, \gamma_j) = H(\eta_{ij}) \quad (1.6)$$

### 1.3. Funciones de información.

**Definición 6.** La información del ítem  $j$  del punto  $\boldsymbol{\theta}$  en la dirección  $u$  se denota por  $I_{ju}(\boldsymbol{\theta})$ , y se define como

$$I_{ju}(\boldsymbol{\theta}) = \frac{[\nabla H_j(\boldsymbol{\theta}) \cdot \mathbf{u}]^2}{H_j(\boldsymbol{\theta})[1 - H_j(\boldsymbol{\theta})]}, \quad (1.7)$$

donde  $\nabla H_j(\boldsymbol{\theta})$  representa el gradiente de la función de enlace  $H_j(\cdot)$  en el punto  $\boldsymbol{\theta}$ , Reckase (2009).

**Definición 7.** La función de información de la prueba (TIF) en la dirección  $\mathbf{u}$  es definida como

$$I_u(\boldsymbol{\theta}) = \sum_{j=1}^p \frac{[H'_j(\boldsymbol{\theta})]^2}{H_j(\boldsymbol{\theta})[1 - H_j(\boldsymbol{\theta})]} [\alpha_j \boldsymbol{\beta}_j^t \mathbf{u}]^2. \quad (1.8)$$

La definición anterior coincide con la TIF clásica para el caso unidimensional. En el caso en que todos  $\boldsymbol{\beta}_j$  son paralelas con el vector  $\mathbf{u}$ , el modelo MIRT es una extensión trivial del modelo unidimensional.

**Definición 8.** La esperanza de la función de información del test (ETIF) en la dirección  $\mathbf{u}$  es definida como

$$ETIF = E_{\boldsymbol{\theta}}[I_u(\boldsymbol{\theta})] = \sum_{j=1}^p E_{\boldsymbol{\theta}} \left( \frac{[H'_j(\boldsymbol{\theta})]^2}{H_j(\boldsymbol{\theta})[1 - H_j(\boldsymbol{\theta})]} \right) [\alpha_j \boldsymbol{\beta}_j^t \mathbf{u}]^2 \quad (1.9)$$

### 1.4. Componente de referencia de una prueba.

**Definición 9.** Un componente o compuesto  $\Theta_{\beta}$  del vector de trazo latente completo  $\boldsymbol{\Theta}$  es una combinación lineal de  $\boldsymbol{\Theta}$ , es decir,  $\Theta_{\beta} = \boldsymbol{\beta}^t \boldsymbol{\Theta} = \sum_{k=1}^d \beta_k \Theta_k$ , donde  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_d)^t$  es un vector de constantes llamado la dirección del componente  $\Theta_{\beta}$ . Si  $Var(\Theta_{\beta}) = 1$ ,  $\Theta_{\beta}$  es llamado un componente o compuesto normalizado, Wang (1986).



Montenegro (2011) propuso la siguiente definición de componente de referencia de una prueba basada en el concepto de información esperada de la prueba.

**Definición 10.** La dirección de referencia de una prueba es la dirección  $u$  que maximiza la esperanza de la función de información de la prueba. La componente de referencia de la prueba es la componente determinada por la dirección de referencia de la prueba.

## 1.5. Trazo o habilidad sintética.

**Definición 11.** Sea  $\{X_1, X_2, \dots, X_m\}$  una subprueba,  $Y = \sum_{j=1}^m X_j$  el score del número de correctas de la subprueba y  $\xi(\boldsymbol{\theta}) = \sum_{j=1}^m p_j(\boldsymbol{\theta})$  es el verdadero score de la subprueba. Zhang & Stout (1999) define la dirección del score  $Y$  como un vector  $\boldsymbol{\xi}$  que maximiza la esperanza multidimensional de la razón crítica (EMCR) definida como

$$EMCR(\boldsymbol{\xi}, \boldsymbol{\theta}; Y) = E \left( \frac{\Delta_{\boldsymbol{\xi}} \xi(\boldsymbol{\theta})}{[Var(Y|\boldsymbol{\theta})]^{1/2}} \right)$$

donde  $\Delta_{\boldsymbol{\xi}} \xi(\boldsymbol{\theta})$  es la derivada direccional del verdadero score  $\xi(\boldsymbol{\theta})$  en la dirección  $\boldsymbol{\xi}$ .

Ellos demuestran que el vector  $\boldsymbol{\xi}$  esta dado por

$$\boldsymbol{\xi} = \sum_{j=1}^m w_j \boldsymbol{\beta}_j$$

donde  $w_j = cE \left( \mathbf{H}_i^t(\alpha_j \boldsymbol{\beta}_j^t \boldsymbol{\theta} - \gamma_j) / \sqrt{Var(Y|\boldsymbol{\theta})} \right)$ .  $\mathbf{H}_i(\cdot)$  representa la función de respuesta al ítem.

La dirección  $\boldsymbol{\xi}$  depende de la función de respuesta, y es un promedio del trazo latente poblacional. En este caso,  $\boldsymbol{\xi}^t \boldsymbol{\theta}$  es la componente que es mejor medida por la subprueba. La dirección de referencia  $\boldsymbol{\xi}$  fue llamada la dirección de la subprueba.

Wang (1986) construyó una aproximación unidimensional a una matriz de datos multidimensional que llamó el trazo del componente de referencia. El uso la transformación  $y = \ln[p/(1-p)]$ , el score logístico del ítem y reescribió el modelo MIRT logístico como

$$\mathbf{Y} = \boldsymbol{\theta} \mathbf{A}^t - \mathbf{1}_k \boldsymbol{\gamma}^t,$$

donde  $\boldsymbol{\theta}$  es una matriz de trazos latentes,  $\mathbf{A}_{k \times d}$  es una matriz de parámetros de discriminación en el modelo MIRT,  $k$  el número de ítems,  $\mathbf{1}_k$  es un vector de unos y  $\boldsymbol{\gamma}$  es un vector asociado a la dificultad. La función objetivo en este caso es la traza de  $(\mathbf{Y} - \hat{\mathbf{Y}})^t (\mathbf{Y} - \hat{\mathbf{Y}})$ , donde  $\hat{\mathbf{Y}} = \mathbf{G} \mathbf{H}^t - \mathbf{1}_k \boldsymbol{\gamma}^t$ ,  $\mathbf{G}$  es el trazo latente unidimensional en la aproximación del

modelo y  $\mathbf{H}$  el vector de parámetros de discriminación del ítem en este modelo. Se supone que los parámetros de dificultad no cambian, Wang mostró que en este caso

$$\mathbf{G} = \boldsymbol{\theta}\mathbf{w},$$

donde  $\mathbf{w}$  es el vector propio asociado con el valor propio mas grande de la matriz  $\mathbf{A}^t\mathbf{A}$ .

Por otro lado, Montenegro (2011) deriva una habilidad sintética unidimensional de una manera diferente que las anteriores en la literatura. Este observó que si todos los  $\beta_j$  son los mismos,  $\beta_j = \beta$ ,  $j = 1, 2, \dots, k$ , donde  $k$  es el número de ítems en la prueba, entonces la ecuación 1.3 se reduce a

$$P(Y_{ij} = 1 | \boldsymbol{\theta}_i, \alpha_j, \boldsymbol{\beta}, b_j) = \frac{1}{1 + \exp -\alpha_j(\boldsymbol{\beta}^t\boldsymbol{\theta}_i - b_j)} \quad (1.10)$$

que es una extensión trivial de un modelo UIRT, donde cada uno de los ítems mide el mismo compuesto de la habilidad dado por  $\boldsymbol{\beta}^t\boldsymbol{\theta}_i$ . Debido a que estos vectores tienen longitud 1, ellos están en la hipersfera unitaria de  $R^d$ . Por otra parte, se puede suponer que los componentes de los vectores  $\beta_j$  son todos no negativos, entonces todos los vectores están en el mismo hiper-cuadrante. Por lo tanto, es razonable esperar que el vector que resume todos los  $\beta_j$  es el mismo hiper-cuadrante de la hipersfera unitaria. Esto llevo a buscar un vector  $\boldsymbol{\beta}$  mediante la optimización de la función objetivo dada por

$$h(\beta_1, \beta_2, \dots, \beta_d) = \sum_{l=1}^d \sum_{k=1}^k (\beta_{kl}^2 - \beta_l^2)^2$$

cuya solución es el vector unitario dado por

$$\beta_l = \sqrt{\frac{1}{k} \sum_{j=1}^k \beta_{kl}^2}.$$

Se denotó el vector de la solución en este caso como  $\boldsymbol{\beta}_h$ . Alternativamente, se optimizo la función objetivo

$$g(\beta_1, \beta_2, \dots, \beta_d) = \sum_{l=1}^d \sum_{k=1}^k (\beta_{kl} - \beta_l)^2$$

cuya solución, considerando un vector unitario está dado por

$$\beta_l = \frac{\sum_{j=1}^k \beta_{kl}}{\|\sum_{j=1}^k \beta_{kl}\|}.$$

El vector solución en este caso se denota como  $\boldsymbol{\beta}_g$ .

Es usual asumir que las habilidades de los examinandos son una muestra extraída de una distribución normal  $d$  - dimensional  $N(0, \boldsymbol{\Sigma})$ . La estimación marginal EM se

basa en esta hipótesis, Bock & Aitkin (1981). Para obtener un modelo identificable, la mayoría de programas escritos para estimar modelos MIRT asumen que  $\Sigma = \mathbf{I}_d$ , donde  $\mathbf{I}_d$  es la matriz idéntica de orden  $d$ , sin embargo, esta situación no es muy realista. Bégin & Glass (2001) y De la Torre & Patz (2005) proponen los algoritmos MCMC que simultáneamente estiman los parámetros del ítem, las habilidades latentes y la matriz  $\Sigma$ . En Montenegro (2011), se asume que sólo los elementos de la diagonal son todos 1, este supuesto define una escala común a lo largo del eje canónica del espacio de habilidad. Ackerman (1989) declaró que en el caso donde la matriz  $\Sigma$  no es la identidad, la dificultad y la dimensionalidad pueden ser confundidas.

Por otro lado, Montenegro (2011) asumió que  $\theta$ , la habilidad latente de los examinados, es una muestra de una distribución normal  $N(\mathbf{0}, \Sigma)$ . Entonces  $\Sigma$  tiene la representación estocástica dada por  $\theta = \Sigma^{1/2}\gamma$ , donde  $\gamma$  tiene una distribución normal estándar multivariada, y  $\Sigma^{1/2}$  es la raíz cuadrada de  $\Sigma$ . Entonces, se tiene que

$$\beta^t \theta = \left( \Sigma^{1/2} \beta \right)^t \quad (1.11)$$

Por lo tanto, cuando en el proceso de estimación se asume que la matriz de correlación es la matriz identidad, la dirección de cada ítem es estimada en un espacio transformado determinado por  $\Sigma^{1/2}$ . Además, la ecuación 1.11 se muestra un procedimiento para calcular la dirección de referencia cuando la matriz de correlación se encuentra disponible.

Si  $\theta$  tiene una distribución normal multivariada  $N(\mathbf{0}, \Sigma)$ , entonces, cualquier combinación  $\beta^t \theta$  tiene una escala diferente, ya que  $Var(\beta^t \theta) = \beta^t \Sigma \beta$ . En este caso, la dirección de referencia debe ser calculado a partir de los vectores transformados  $\Sigma^{1/2} \beta$ , y la habilidad sintética debe ser calculada utilizando la habilidad transformada  $\gamma = \Sigma^{1/2} \theta$ .

En síntesis, se tiene que en Wang (1986) Zhang & Stout (1999) fueron revisados. Wang, Zhang y Stout se propusieron dos índices alternativos sintéticos dados por  $(\mathbf{w}^t \theta)$  y  $(\xi^t \theta)$ , y en Montenegro (2011) propuso dos nuevos índices sintéticos:  $(\beta_h^t \theta)$  y  $(\beta_g^t \theta)$ . Estos índices alternativos pueden ser calculados más fáciles que los índices anteriores, son más naturales y fáciles de usar por los expertos.

## 1.6. Introducción a los modelos MIRT para una estructura lineal latente.

Montenegro (2011) introduce estos modelos que han sido pensados para ser utilizados en pruebas de evaluación a gran escala, diseñadas explícitamente para medir más que un trazo latente. Estas pruebas por lo general están divididas en subpruebas, donde cada una de estas subpruebas está diseñada para medir principalmente un único trazo latente unidimensional específicamente.

En las pruebas de admisión de algunas universidades, es común dividir la prueba en varias subpruebas, donde cada una de estas está diseñada para medir una habilidad principal o trazo latente. Para este tipo de pruebas, hay por lo menos tres procedimientos de estimación que se utilizan antes del planteado por Montenegro:

- El primer procedimiento, consiste en la estimación de los parámetros de cada subprueba por separado, utilizando los modelos unidimensionales de teoría de respuesta al ítem, Reckase (1985).
- El segundo procedimiento, se basa en el uso del modelo multidimensional de teoría de respuesta al ítem, Reckase (1985; 1997; 2007; 2009).
- El tercer procedimiento se basa en el uso de un modelo de estructura simple, también llamado modelo multiunidimensional, De la Torre & Patz (2005), Sheng (2007).

### 1.6.1. Trazos latentes principales y básicos.

La teoría de respuesta al ítem usa el supuesto que los trazos latentes de los examinados pueden ser representados por vectores en un espacio Euclidiano llamado espacio de trazos latentes. Desde un punto de vista muy general, los trazos latentes individuales no tienen una interpretación directa. Adicionalmente, la dimensión del espacio de trazo latente es desconocida apriori.

De otra parte, el objetivo de una prueba es medir algún trazo latente de los examinados, que pueden ser interpretados a través de una teoría subyacente. La variable latente  $Z_j$  puede ser escrita como

$$Z_j = \alpha_j \beta_j \boldsymbol{\theta} - \gamma_j + e_j \quad e_j \sim N(0, 1) \quad (1.12)$$

donde  $\beta_j$  es un vector unitario llamado la dirección del ítem  $j$  y es la dirección en la cual el ítem  $j$  discrimina mejor. Esto significa, que el ítem  $j$  discrimina mejor entre los valores del trazo latente sintético dado por la componente normalizada  $\beta_j \boldsymbol{\theta}$ .

Teóricamente, una prueba es unidimensional si está diseñado para medir principalmente un componente en el espacio de trazo latente. Es este caso, por lo general se puede dar una interpretación directa a la componente de referencia de la prueba. Sin embargo, el espacio de trazo latente pueden ser multidimensional, y cada componente del vector de trazo latente no puede tener una interpretación directa.

En general, el significado de un trazo latente depende del diseño de la prueba. Montenegro (2011) propone los nombres de trazos latentes básicos a las componentes de los vectores de trazos latentes y trazos latentes principales a los trazos latentes que la prueba intenta medir. En consecuencia, los trazos latentes principales se definen de acuerdo al diseño de la prueba.

En otras palabras, en el modelo LSMIRT hay dos tipos de trazos latentes que se consideran: los trazos latentes principales y los trazos latentes básicos. Como los trazos latentes principales corresponden a aquellas habilidades que la prueba intenta medir; son definidos según el diseño de la prueba, y pueden ser interpretados directamente por la teoría subyacente que conduce al diseño de la prueba. Por otro lado, los trazos latentes principales son combinaciones lineales de los trazos latentes básicos.

### 1.6.2. Modelo MIRT para una estructura lineal latente.

Montenegro (2011) introduce la nomenclatura y las suposiciones del modelo MIRT para una estructura lineal latente. Las suposiciones del modelo LSMIRT son las siguientes:

- La prueba está dividida en  $m$  subpruebas. Se supone que cada subprueba es esencialmente unidimensional. Por lo tanto, cada subprueba trata de medir sólo un trazo latente principal. Cada subprueba tiene  $K_v$  ítems, por lo que toda la prueba tiene  $K = K_1 + K_2 + \dots + K_m$  ítems.
- Se asume que los trazos latentes básicos de los examinandos son una muestra aleatoria extraída de una distribución normal multivariada  $N_d(\mathbf{0}, \mathbf{\Sigma})$ , donde  $\mathbf{\Sigma}$  es una matriz de correlación, y  $d \leq m$ .
- Los trazos latentes principales de los examinandos siendo medidos por cada subprueba son las combinaciones lineales de los vectores de trazos latentes básicos.
- La función de enlace es la ojiva de la normal estándar, denotada  $\Phi(\cdot)$ .
- Parámetros de respuesta al azar no son incluidos.

El ítem  $j$ -ésimo de la subprueba  $v$  será llamado  $vj$ . El modelo LSMIRT es especificado por la probabilidad del suceso del examinado  $i$  para el ítem  $vj$  dada por

$$P(Y_{vij} = 1 | \alpha_{vj}, \gamma_{vj}, \beta_v, \theta_i) = \Phi(\alpha_{vj} \beta_v^t \theta_i - \gamma_{vj}) \quad (1.13)$$

donde  $\alpha_{vj}$  y  $\gamma_{vj}$  se llamarán respectivamente el parámetro de pendiente (la discriminación) y el parámetro de intercepto del ítem  $vj$ . El vector  $\beta_v = (\beta_{v1}, \dots, \beta_{vd})^t$  es un vector unitario en el espacio de trazo latente, se llamará la dirección de la subprueba  $v$  y  $\theta_i = (\theta_{i1}, \dots, \theta_{id})^t$  representa el vector de trazos latentes básicos del examinado  $i$ . Por otro lado se tiene que

- El parámetro de dificultad clásico esta dado por  $b_{vj} = \gamma_{vj} / \alpha_{vj}$ . Montenegro denota los parámetros del ítem  $vj$  como  $\xi_{vj} = (\alpha_{vj}, \gamma_{vj})^t$ . El trazo latente principal medido por la prueba  $v$  esta dado por el compuesto  $\beta_v^t \theta$ , para  $v = 1, 2, \dots, m$ .
- El predictor lineal latente del ítem  $vj$ , esta dado por la expresión

$$\eta_{vj} = \alpha_{vj} \beta_v^t \theta - \gamma_{vj}.$$

- Para la distribución de los vectores de trazos latentes, otras distribuciones simétricas son posibles como la distribución t-student multivariada, Bazán (2006), Da Silva (2008). Recientemente, algunos autores han propuesto distribuciones asimétricas, sin embargo, Montenegro sólo trabajó con la distribución normal multivariada.
- Cuando la dimensión del trazo latente coincide con la dimensión de la prueba, la prueba es llamada: prueba de estructura simple. En este caso, la probabilidad de éxito del examinado  $i$  al ítem  $j$  se reducir a

$$P(Y_{vij} = 1 | \boldsymbol{\xi}_{vj}, \boldsymbol{\theta}_i) = \Phi(\alpha_{vj}\boldsymbol{\theta}_i - \gamma_{vj}), \quad (1.14)$$

porque en ese caso todas las direcciones de las subpruebas se pueden identificar con los vectores de base canónica del espacio euclidiano  $R^d$ .

Además, Montenegro encontró que

$$\hat{\boldsymbol{a}}_{vj} = \frac{\alpha_{vj}\boldsymbol{\beta}_v^t\boldsymbol{\beta}_{vj}}{\sqrt{1 + \alpha_{vj}^2 - \alpha_{vj}^2(\boldsymbol{\beta}_v^t\boldsymbol{\beta}_{vj})^2}} \quad (1.15)$$

y

$$\hat{\gamma}_j = \frac{\gamma_{vj}}{\sqrt{1 + \alpha_{vj}^2 - \alpha_{vj}^2(\boldsymbol{\beta}_v^t\boldsymbol{\beta}_{vj})^2}} \quad (1.16)$$

- La función de respuesta del ítem  $vj$  está dado por

$$f_{vj}(\boldsymbol{\theta}) = \Phi(\eta_{vj}). \quad (1.17)$$

### 1.6.3. Identificabilidad del modelo LSMIRT.

En Montenegro (2011) se establece que el modelo LSMIRT no es identificable y para obtener un modelo identificable, en primer lugar los vectores de las direcciones de referencia de la subprueba  $\boldsymbol{\beta}_v$ , son vectores unitarios. Sin embargo, esta restricción no es suficiente para tener un modelo identificable. Por lo tanto, propuso dos parametrizaciones, una de las cuales, permite una interpretación fácil de los parámetros, incluyendo los trazos latentes básicos. Las parametrizaciones se basan en la relación entre el trazo latente principal y el básico. Sea  $\tilde{\boldsymbol{\Theta}}$  un vector aleatorio de  $m \times 1$  que representa los trazos latentes principales de la prueba. Entonces,

$$\tilde{\boldsymbol{\Theta}}_{m \times 1} = \boldsymbol{B}_{m \times d}\boldsymbol{\Theta}_{d \times 1} \quad (1.18)$$

donde  $\boldsymbol{B}$  es una matriz cuyas filas son los vectores  $\boldsymbol{\beta}_v$ . La matriz de covarianza del vector  $\tilde{\boldsymbol{\Theta}}$  está dada por

$$Cov(\tilde{\boldsymbol{\Theta}}) = \boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}^t. \quad (1.19)$$

La ecuación 1.18 representa la relación entre los trazos latentes básicos y los trazos latentes principales. De esta ecuación, dos parametrizaciones son consideradas. Sin pérdida de generalidad, Montenegro supuso que la matriz  $\boldsymbol{\Sigma}$  es definida positiva y que las primeras  $d$  filas de la matriz  $\boldsymbol{B}$  son linealmente independientes. Sea  $\boldsymbol{\Sigma}^{1/2}$  la raíz cuadrada de la matriz  $\boldsymbol{\Sigma}$ . Para la primera parametrización, se reescribe  $\tilde{\boldsymbol{\Theta}}$  como

$$\tilde{\boldsymbol{\Theta}} = \left(\boldsymbol{B}\boldsymbol{\Sigma}^{1/2}\right) \left(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Theta}\right) \quad (1.20)$$

La ecuación 1.20 implica que se puede asumir que  $\Theta \sim N_d(\mathbf{0}, \mathbf{I}_d)$ . En ese caso, los trazos latentes básicos no están correlacionados. En esta parametrización, la matriz  $\Sigma$  es la idéntica y consecuentemente es necesario estimar las  $m \times d$  componentes de las direcciones  $\beta_v$ .

Para la segunda parametrización, sea  $\tilde{\mathbf{B}}$  la submatriz de  $\mathbf{B}$  que contiene las primeras  $d$  filas. Entonces,  $\tilde{\Theta}$  puede ser escrita como

$$\tilde{\Theta} = \left( \mathbf{B} \tilde{\mathbf{B}}^{-1} \right) \left( \tilde{\mathbf{B}} \Theta \right) \quad (1.21)$$

La ecuación 1.21 implica que las primeras  $d$  direcciones de referencia están alineados con los ejes de coordenadas. Si la varianza del trazo latente básico es fijada en 1, como es usual, sólo hay  $(m - d) \times d + d \times (d - 1)/2$  parámetros para estimar, que corresponden a las componentes de las direcciones de referencia que no están alineadas con los ejes coordenados y los elementos no diagonales de la matriz de correlación.

La segunda parametrización tiene algunas ventajas. Primero, hay menos parámetros a ser estimados; segundo, los trazos latentes básicos son directamente el compuesto de referencia de las primeras  $d$  subpruebas; tercero, los ejes de coordenadas se establecen de antemano, así, problemas de identificabilidad causados por transformaciones ortogonales del espacio de trazo latente son imposibles. Por otro lado, es importante notar que en la primera parametrización las direcciones del ítem tienen una mejor proyección a lo largo del compuesto de referencia correspondiente. De acuerdo a las ecuaciones 1.15 y 1.16, esto implica que los valores de los parámetros de ítem son más cercanos a los valores de los parámetros del ítem del modelo MIRT, del cual puede ser derivado el modelo LSMIRT. Cabe notar que es posible cambiar de una parametrización a otra.

#### 1.6.4. Estimación de los parámetros.

Sea  $p_{vij} = P(Y_{vij} = 1 | \theta_i, \beta_v, \xi_{vj})$ . Sea  $\theta_{N \times d}$  la matriz de trazos latentes de examinados en la muestra. Sea  $\xi$  el vector de todos los parámetros de ítems de la prueba. Sea  $\beta_{m \times d}$  la matriz de las  $m$  direcciones de referencia de las subpruebas. Entonces, bajo el supuesto de independencia local, es decir, que la probabilidad de responder correctamente a un ítem no está afectada por las respuestas dadas a otros ítems, la función de verosimilitud está dada por

$$f(\mathbf{y} | \theta, \beta, \xi) = \prod_{i=1}^n \prod_{v=1}^m \prod_{j=1}^{k_v} p_{vij}^{y_{vij}} (1 - p_{vij})^{1 - y_{vij}}. \quad (1.22)$$

donde  $y_{vij}$  es la respuesta observada del examinado  $i$  al ítem  $vj$ , Montenegro (2011).

Montenegro utilizo un algoritmo de datos aumentados del muestreador de Gibbs (DAGS) para estimar conjuntamente los parámetros del ítem y el trazo latente.

---

---

### Modelo LSMIRT para múltiples poblaciones.

---

---

En este capítulo se propone un modelo de teoría de respuesta al ítem multidimensional con una estructura lineal latente para  $g$  grupos (poblaciones) o subgrupos (G-LSMIRT), por lo tanto, este modelo incluye un indicador que identifica cada conjunto de parámetros correspondientes a cada una de las  $g$  poblaciones respecto al modelo LSMIRT propuesto por Montenegro. El modelo propuesto en esta sección es más parsimonioso que los modelos disponibles para el tipo de pruebas mencionadas en la sección anterior.

#### 2.1. Introducción a los modelos LSMIRT para múltiples poblaciones.

La teoría de respuesta al ítem es un campo en donde se han desarrollado diversos modelos para la estimación de las habilidades o trazos latentes de los examinados en diferentes tipos de pruebas y escenarios de las mismas. No obstante, el modelo LSMIRT propuesto por Montenegro (2011), no presenta una especificación cuando estos son aplicados a múltiples poblaciones o una población dividida en subgrupos.

Por lo tanto, esta sección se base en el interés de establecer un modelo LSMIRT cuando se desea aplicar una prueba que se encuentran dividida en subpruebas aproximadamente unidimensionales, donde cada una de estas es diseñada esencialmente para medir una habilidad o trazo latente principal de los examinados a  $g$  poblaciones o subgrupos poblacionales, para examinar los cambios que los examinados tuvieron en los diferentes ítems.

#### 2.2. Modelo propuesto.

Siguiendo la metodología utilizada en Bégin & Glass (2001), donde extienden los modelos de teoría de respuesta al ítem multidimensionales al caso donde se tienen varias poblaciones y teniendo en cuenta el modelo LSMIRT propuesto por Montenegro (2011), un posible modelo especificado por la probabilidad de suceso (éxito) del examinado  $i$  para el ítem  $v_j$  de la población  $g$  podría ser:



$$P(Y_{vij} = 1 | \alpha_{vj}, \gamma_{vj}, \beta_v, \theta_{ig}) = \Phi(\alpha_{vj} \beta_v^t \theta_{ig} - \gamma_{vj}) \quad (2.1)$$

donde:

- $\alpha_{vj}$  y  $\gamma_{vj}$  se llamarán respectivamente el parámetro de pendiente (la discriminación) y el parámetro de intercepto del ítem  $vj$ .
- El vector  $\beta_v = (\beta_{v1}, \dots, \beta_{vd})^t$  es un vector unitario en el espacio de trazo latente, se llamará la dirección de la subprueba  $v$ .
- $\theta_{ig} = (\theta_{i1g}, \dots, \theta_{idg})^t$  representa el vector de trazos latentes básicos del examinado  $i$  de la población  $g$ .

Las suposiciones del modelo LSMIRT para  $g$  grupos son las siguientes:

- La prueba está dividida en  $m$  subpruebas. Se supone que cada subprueba es esencialmente unidimensional. Por lo tanto, cada subprueba trata de medir sólo un trazo latente principal. Cada subprueba tiene  $K_v$  ítems, por lo que toda la prueba tiene  $K = K_1 + K_2 + \dots + K_m$  ítems.
- Se asume que los trazos latentes básicos de los examinandos de cada grupo o subgrupo son una muestra aleatoria extraída de una distribución normal multivariada  $N_d(\mathbf{0}, \Sigma_g)$ , donde  $\Sigma_g$  denota una matriz de correlación en cada población, y  $d \leq m$ .
- Los trazos latentes principales de los examinandos siendo medidos por cada subprueba son las combinaciones lineales de los vectores de trazos latentes básicos en cada una de las  $g$  poblaciones.
- La función de enlace es la ojiva de la normal estándar, denotada  $\Phi(\cdot)$ .
- Parámetros de respuesta al azar no son incluidos.

Por otro lado se tiene que

- El parámetro de dificultad clásico esta dado por  $b_{vj} = \gamma_{vj}/\alpha_{vj}$ . Se denotan los parámetros del ítem  $vj$  como  $\xi_{vj} = (\alpha_{vj}, \gamma_{vj})^t$ . El trazo latente principal medido por la subprueba  $v$  de la población o subgrupo poblacional  $g$  esta dado por el compuesto  $\beta_v^t \theta_g$ , para  $v = 1, 2, \dots, m$  y  $g = 1, 2, \dots, G$ .
- El predictor lineal latente del ítem  $vj$  para el examinado  $i$  de la población  $g$ , esta dado por la expresión

$$\eta_{vij} = \alpha_{vj} \beta_v^t \theta_{ig} - \gamma_{vj}.$$

### 2.3. Estimación de los parámetros.

Sea  $p_{vij} = P(Y_{vij} = 1 | \theta_{ig}, \beta_v, \xi_{vj})$ . Sea  $\theta_g$  de orden  $(N_g \times d)$  la matriz de trazos latentes de examinados en la muestra para la población  $g$  ésima. Sea  $\xi$  el vector de todos los

parámetros de ítems de la prueba. Sea  $\beta$  de orden  $(m \times d)$  la matriz de las  $m$  direcciones de referencia de las subpruebas. Entonces, bajo el supuesto de independencia local la función de verosimilitud está dada por

$$f(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\xi}) = \prod_{i=1}^n \prod_{v=1}^m \prod_{j=1}^{k_v} \prod_{g=1}^G p_{vijg}^{y_{vijg}} (1 - p_{vijg})^{1-y_{vijg}}. \quad (2.2)$$

donde  $y_{vijg}$  es la respuesta observada del examinado  $i$  al ítem  $vj$ , para la población  $g$  ésima.

Se utilizo un algoritmo de datos aumentados del muestreador de Gibbs (DAGS) para estimar conjuntamente los parámetros del ítem y el trazo latente, siguiendo la estrategia propuesta por Albert (1992). Se introdujo variables  $Z_{vijg} \sim N(\eta_{vijg}, 1)$ , donde  $\eta_{vijg} = \alpha_{vj} \boldsymbol{\beta}_v^t \boldsymbol{\theta}_g - \gamma_{vj}$ . Cabe notar, que enfoques similares para estimar los parámetros de modelos MIRT y modelos IRT multiunidimensional fueron utilizados por Montenegro (2011), Bégin & Glass (2001), Lee (1995), Sheng (2010) y otros autores.

Las distribuciones condicionales completas para implementar el algoritmo de DAGS son derivadas en el apéndice. El algoritmo de DAGS fue implementado en el software R Core Team (2014). Para evaluar el algoritmo varios escenarios de simulación fueron realizados, en cada uno de estos un conjunto de datos respuesta de una prueba fue creado, los parámetros recuperados y comparados con los simulados.

## 2.4. Identificabilidad del modelo.

Para obtener un modelo identificable es necesario restringir la suma de los  $\gamma_{vj}$  a cero, como en Fox (2010). Los  $\boldsymbol{\beta}_v$  como vectores unitarios como en Montenegro (2011). Tomar los trazos latentes principales de la población 1 con vector de media cero y matriz de covarianzas la identidad según Bégin & Glass (2001). Finalmente, se utilizo la primera parametrización presentada en Montenegro (2011) descrita en la subsección 1.6.3 de esta disertación.

## 2.5. Simulación.

El modelo propuesto fue aplicado a varios conjuntos de datos simulados conformados por diferentes poblaciones y conjuntos de parámetros. Dado que en todos los escenarios simulados se obtuvieron resultados similares, en esta sección se muestran los parámetros recuperados de uno de ellos. Los datos de la respuesta fueron generados usando un modelo MIRT clásico para cada una de las poblaciones y fueron ajustados con un modelo LSMIRT para múltiples poblaciones.

### 2.5.1. Simulación del test.

El escenario seleccionado, constaba de una prueba de tamaño  $K = 100$ . Los 100 ítems fueron divididos en 4 grupos (subpruebas), cada una de 25 ítems. Es decir,

$K_1 = K_2 = K_3 = K_4 = 25$ . Los parámetros de pendiente en el modelo MIRT son denotados  $a_{vj}$ , y fueron generados de una distribución lognormal con parámetros  $(0, 0.25)$ ; y los parámetros  $\gamma_{vj}$  fueron generados de una distribución normal con parámetros  $(0, 1)$ .

Se definieron inicialmente 4 direcciones en  $R^3$  para generar la dirección de los ítems, estas direcciones fueron  $(1.00, 0.20, 0.10)$ ,  $(0.30, 1.00, 0.10)$ ,  $(0.25, 0.25, 1.00)$  y  $(0.25, 0.25, 0.50)$  respectivamente. A partir de estas direcciones de los ítems normalizadas, las direcciones de referencia de cada grupo se obtuvieron, mediante el primer vector propio de  $\mathbf{B}_v^t \mathbf{B}_v$  donde  $\mathbf{B}_v$  es la matriz cuyas filas contienen las direcciones en cada grupo  $v$ ,  $v = 1, 2, 3, 4$ .  $\beta$  es la matriz cuyas filas son las verdaderas direcciones de referencia de la prueba simulada, la cual está dada por

$$\beta = \begin{pmatrix} 0.9754 & 0.1179 & 0.1861 \\ 0.2069 & 0.9740 & 0.0927 \\ 0.1228 & 0.1098 & 0.9863 \\ 0.8216 & 0.1883 & 0.5381 \end{pmatrix} \quad (2.3)$$

Para generar los datos de la respuesta simulada para cada una de las poblaciones se utilizaron las direcciones de los ítems  $\beta_{vj}$ , los parámetros de pendiente  $a_{vj}$  y los parámetros de intercepto  $\gamma_{vj}$  originales, de los cuales se muestran algunos en la Tabla 2.1. Por otra parte, dado que los datos se ajustaron a un modelo LSMIRT para múltiples poblaciones, los parámetros recuperados fueron el conjunto  $a_{vj}$ ,  $\gamma_{vj}$  y  $\beta$ . Las relaciones presentas en la tabla 2.1 se pueden ver en Montenegro (2011, pág 50).

ítem	$a_{vj}$	$\beta_{vj1}$	$\beta_{vj2}$	$\beta_{vj3}$	$\rho_{vj}$	$\ \tilde{\mathbf{B}}^{-1} \beta_v\ $	$\alpha_{vj}$	$\gamma_{vj}$
5	0.860	0.975	0.053	0.215	0.998	1.005	0.863	-0.267
15	0.746	0.984	0.023	0.176	0.995	1.005	0.747	-0.670
25	1.476	0.971	0.060	0.230	0.997	1.005	1.480	2.468
35	0.871	0.057	0.991	0.122	0.988	0.987	0.850	-0.196
45	0.823	0.282	0.959	0.000	0.993	0.987	0.807	-0.159
55	0.570	0.151	0.151	0.977	0.999	1.008	0.574	-0.446
65	0.977	0.000	0.244	0.970	0.983	1.008	0.969	2.048
75	0.890	0.228	0.226	0.947	0.987	1.008	0.886	0.418
85	0.872	0.768	0.281	0.576	0.994	0.882	0.764	0.803
95	0.846	0.776	0.397	0.490	0.976	0.882	0.728	2.034

TABLA 2.1. La columna  $a_{vj}$  es el parámetro de discriminación multidimensional en el modelo clásico MIRT; las columnas  $\beta_{vj1}$ ,  $\beta_{vj2}$ ,  $\beta_{vj3}$  son las componentes de las direcciones originales de los ítems; la columna  $\rho_{vj}$  es el producto punto entre la dirección original del ítem y la dirección de la subprueba correspondiente; la columna  $\alpha_{vj}$  es el parámetro de pendiente verdadero en el modelo LSMIRT cuyo valor es el producto de  $a_{vj} \rho_{vj} \|\tilde{\mathbf{B}}^{-1} \beta_v\|$  y la columna  $\gamma_{vj}$  es el parámetro de intercepto.

Tres muestras ( $N_1 = 3500$ ,  $N_2 = 3200$  y  $N_3 = 3700$ ) de vectores de trazo latente fueron generados de una distribución normal multivariada con diferentes  $\mu_g$  y matrices de covarianzas  $\Sigma_g$ , pero los mismos parámetros  $\alpha_{vj}$  y  $\gamma_{vj}$  (en este caso particular  $G = 3$  poblaciones). Siguiendo la metodología planteada en Bégin & Glass (2001), tenemos para la población 1 que  $\mu_1 = (0, 0, 0)$  y  $\Sigma_1 = \mathbf{I}_d$ .

Para las demás poblaciones se han fijado los valores de los parámetros de interés como:

$$\boldsymbol{\mu}_2 = (1.0, 0.5, 1.5) \quad (2.4)$$

$$\boldsymbol{\mu}_3 = (1.0, 2.0, 3.0) \quad (2.5)$$

$$\boldsymbol{\Sigma}_2 = \begin{pmatrix} 0.24 & 0.24 & 0.52 \\ 0.24 & 0.36 & 0.56 \\ 0.52 & 0.56 & 1.16 \end{pmatrix} \quad (2.6)$$

y

$$\boldsymbol{\Sigma}_3 = \begin{pmatrix} 1.50 & 1.17 & 0.78 \\ 1.17 & 1.14 & 0.44 \\ 0.78 & 0.44 & 0.72 \end{pmatrix} \quad (2.7)$$

Finalmente, las  $N \times K \times G$  ( $N = N_1 + N_2 + N_3$ ) respuestas binarias fueron generadas usando un modelo MIRT con un enlace logit de dos parámetros. Para cada vector de trazo latente  $\boldsymbol{\theta}_{ig}$  y el conjunto de parámetros  $(\boldsymbol{\beta}_{vj}, a_{vj}, \gamma_{vj})$  la probabilidad de una respuesta correcta fue calculada usando la ecuación

$$P(Y_{vijg} = 1 | \boldsymbol{\theta}_{ig}, a_{vj}, \gamma_{vj}) = \frac{e^{a_{vj}\boldsymbol{\beta}_{vj}^t\boldsymbol{\theta}_{ig} - \gamma_{vj}}}{1 + e^{a_{vj}\boldsymbol{\beta}_{vj}^t\boldsymbol{\theta}_{ig} - \gamma_{vj}}}, \quad (2.8)$$

donde  $i = 1, \dots, N$ ,  $j = 1, \dots, p$ ,  $v = 1, \dots, m$  y  $g = 1, 2, \dots, G$ .

Entonces, un número aleatorio  $u$  fue obtenido de una distribución uniforme en el rango  $[0, 1]$ . Si la probabilidad de respuesta correcta fue mayor o igual que  $u$ , se le asigna 1 para el respuesta, en otro caso se le asigna 0.

### 2.5.2. Ajuste de los datos simulados.

Para las simulaciones se utilizó un periodo de arranque de 2.000 iteraciones y posteriormente se realizaron 5.000 iteraciones para calcular las estimaciones bayesianas. En todos los casos, el estimador Bayesiano utilizado es la media muestral, ya que la media y la mediana son muy similares en casi todos los casos.

La Tabla 2.2 muestra las estimaciones de los parámetros de pendiente e intercepción presentados en la Tabla 2.1; se puede observar que la rutina implementada en R recupera bien estos parámetros. En estudios anteriores con los modelos IRT, los autores han observado que los parámetros de intercepción se recuperan muy bien. Sin embargo, las

estimaciones de los parámetros de pendiente son menos estables. Ver, por ejemplo, Sheng (2008) y Montenegro (2011).

item	$a_{vj}$	$\hat{a}_{vj}$	$sd_{\hat{a}}$	$\gamma_{vj}$	$\hat{\gamma}_{vj}$	$sd_{\hat{\gamma}}$
5	0.860	0.849	0.020	-0.283	-0.291	0.017
15	0.746	0.716	0.019	-0.686	-0.651	0.018
25	1.476	1.512	0.036	2.452	2.463	0.049
35	0.871	0.851	0.019	-0.212	-0.206	0.018
45	0.823	0.736	0.018	-0.175	-0.150	0.017
55	0.570	0.567	0.015	-0.462	-0.440	0.019
65	0.977	1.014	0.020	2.031	2.052	0.039
75	0.890	0.936	0.020	0.402	0.415	0.022
85	0.872	0.907	0.018	0.787	0.793	0.023
95	0.846	0.922	0.019	2.018	2.080	0.039

TABLA 2.2. Las columnas  $\hat{a}_{vj}$  y  $\hat{\gamma}_{vj}$  son las estimaciones de  $a_{vj}$  y  $\gamma_{vj}$  respectivamente, así como  $sd_{\hat{a}_{vj}}$  y  $sd_{\hat{\gamma}_{vj}}$  son las desviaciones estándar posteriori de las estimativas.

Las figuras 2.1 y 2.2 muestran los diagramas de dispersión de  $\alpha_{vj}$  y  $\gamma_{vj}$  simulados contra sus correspondientes estimaciones, respectivamente. En la tabla 2.3 se muestran estadísticas resumen para comparar los parámetros de inclinación e intercepto con sus respectivas estimaciones, y se puede notar que se obtuvieron buenas estimaciones de estos parámetros.

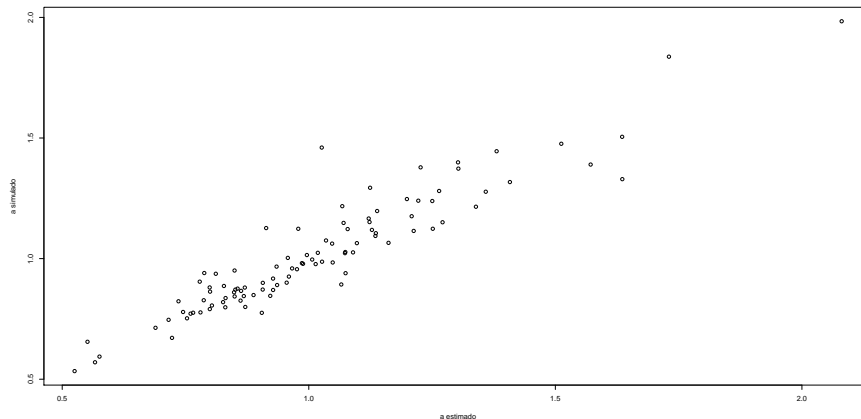
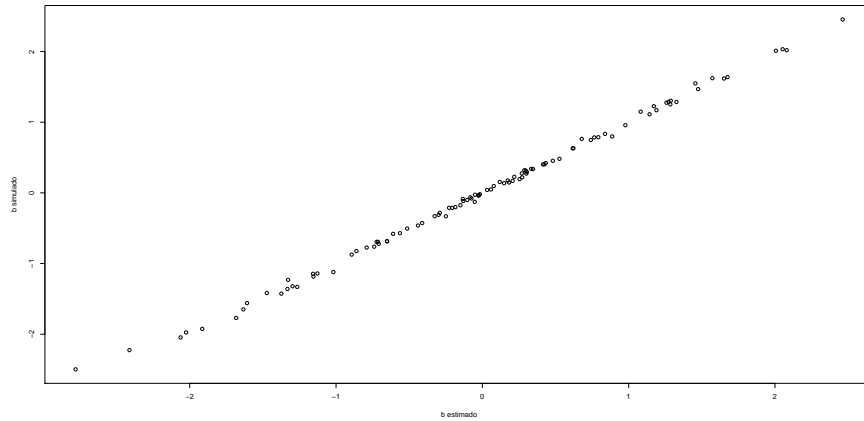


FIGURA 2.1. Gráficos de dispersión entre los  $a_{vj}$  simulados y los estimados.

Parámetro	$cor(\zeta\hat{\zeta})$	$media \zeta - \hat{\zeta} $	$mediana \zeta - \hat{\zeta} $	$\sqrt{media(\zeta - \hat{\zeta})^2}$
$\alpha_{vj}$	0.9354	0.0628	0.0407	0.0913
$\gamma_{vj}$	0.9988	0.0323	0.0204	0.0499

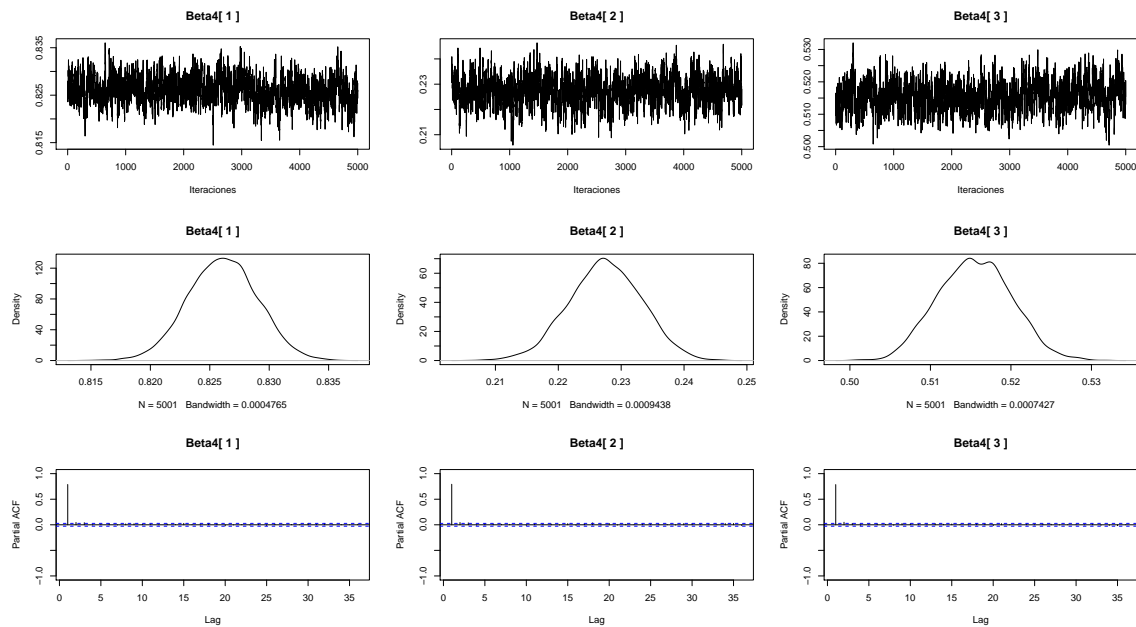
TABLA 2.3. Comparaciones globales entre  $\alpha_{vj}$  y  $\gamma_{vj}$  verdaderos y los estimados.

FIGURA 2.2. Gráficos de dispersión entre los  $\gamma_{vj}$  simulado y el estimado.

Por otro lado, las direcciones de referencias estimadas obtenidas fueron

$$\hat{\beta} = \begin{pmatrix} 0.9706 & 0.1808 & 0.1590 \\ 0.1808 & 0.9768 & 0.1146 \\ 0.1590 & 0.1147 & 0.9806 \\ 0.8261 & 0.2275 & 0.5155 \end{pmatrix}, \quad (2.9)$$

los  $\hat{\beta}$  se recuperaron utilizando la misma parametrización que en Montenegro (2011). En la figura 2.3 se muestran los trazos, las distribuciones posteriori y las autocorrelaciones parciales para las estimaciones de las componentes de  $\beta_4$ . Cabe resaltar que las demas direcciones de referencias estimadas presentaron el mismo comportamiento.

FIGURA 2.3. Trazos, distribuciones posteriori y autocorrelaciones parciales de las cadenas de Markov para las estimaciones de  $\beta_{4d}$

La tabla 2.4 muestra una evaluación global de las estimaciones de los componentes de referencia para cada una de las poblaciones. Note que casi todas las correlaciones entre las componentes de referencia verdaderas y las estimadas están por encima de 0.93 aproximadamente. La figura 2.4 muestra los diagramas de dispersión entre las componentes de referencia verdaderos y los estimados para la población 1.

Parámetro	$cor(\zeta \hat{\zeta})$	$media \zeta - \hat{\zeta} $	$mediana \zeta - \hat{\zeta} $	$\sqrt{media(\zeta - \hat{\zeta})^2}$
Población 1				
$\Theta_{\beta_1}$	0.9642	0.2436	0.2028	0.3105
$\Theta_{\beta_2}$	0.9357	0.2843	0.2454	0.3558
$\Theta_{\beta_3}$	0.9465	0.3086	0.2549	0.3924
$\Theta_{\beta_4}$	0.9585	0.2275	0.1846	0.2893
Población 2				
$\Theta_{\beta_1}$	0.9665	0.2143	0.1726	0.2752
$\Theta_{\beta_2}$	0.9507	0.2366	0.1969	0.2983
$\Theta_{\beta_3}$	0.9686	0.3577	0.2790	0.4657
$\Theta_{\beta_4}$	0.9693	0.2053	0.1658	0.2680
Población 3				
$\Theta_{\beta_1}$	0.9471	0.3521	0.2387	0.5190
$\Theta_{\beta_2}$	0.9275	0.4250	0.3577	0.5413
$\Theta_{\beta_3}$	0.8739	0.8422	0.7937	0.9740
$\Theta_{\beta_4}$	0.9436	0.4201	0.3465	0.5525

TABLA 2.4. Comparaciones globales entre las componentes de referencia verdaderos y los estimados en la prueba simulada.

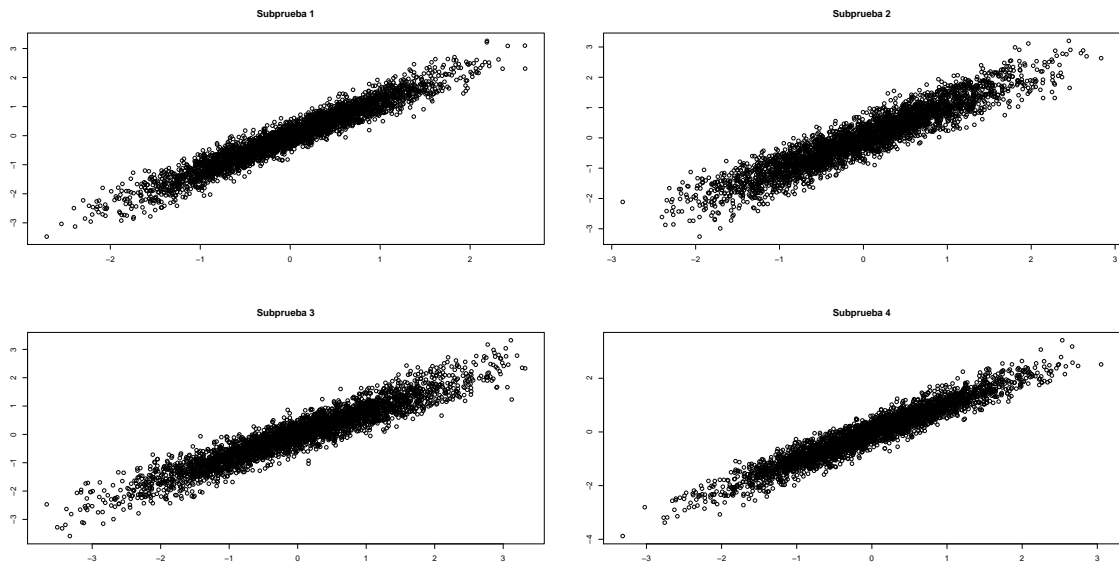


FIGURA 2.4. Diagramas de dispersión entre las componentes de referencia verdaderos y los estimados para cada subprueba de la población 1

En la tabla 2.5 se muestran las estimaciones de las componentes de  $\mu_2$  y  $\mu_3$  (los verdaderos valores para estos parámetros son especificados en las expresiones 2.4 y 2.5, respectivamente), en esta se puede observar que el algoritmo propuesto recupera de forma adecuada las medias de las poblaciones.

Parámetro	Media	Sd	Error A	Error R
$\hat{\mu}_{21}$	1.0357	0.0172	0.0357	0.0357
$\hat{\mu}_{22}$	0.5326	0.0165	0.0326	0.0652
$\hat{\mu}_{23}$	1.4761	0.0252	0.0239	0.0159
$\hat{\mu}_{31}$	0.9850	0.0324	0.0150	0.0150
$\hat{\mu}_{32}$	1.9700	0.0343	0.0300	0.0150
$\hat{\mu}_{33}$	2.9869	0.0378	0.0131	0.0044

TABLA 2.5. Estimaciones de las componentes de  $\mu_g$ , desviaciones estándar posterior, errores absolutos y relativos.

Las tablas 2.6 y 2.7 muestran las estimaciones de los componentes de  $\Sigma_2$  y  $\Sigma_3$ , respectivamente (los verdaderos valores para estos parámetros son especificados en las expresiones 2.6 y 2.7). En estas se puede observar que el algoritmo propuesto recupera de forma adecuada las matrices de covarianzas de las poblaciones.

Parámetro	Media	Sd	Error A	Error R
$\hat{\sigma}_{11}$	0.2491	0.0148	0.0091	0.0379
$\hat{\sigma}_{12}$	0.2350	0.0114	0.0050	0.0208
$\hat{\sigma}_{13}$	0.5085	0.0176	0.0115	0.0221
$\hat{\sigma}_{22}$	0.3567	0.0161	0.0033	0.0092
$\hat{\sigma}_{23}$	0.5486	0.0193	0.0114	0.0204
$\hat{\sigma}_{33}$	1.0789	0.0391	0.0811	0.0699

TABLA 2.6. Estimaciones de las componentes de  $\Sigma_2$ , desviaciones estándar posterior, errores absolutos y relativos.

Parámetro	Media	Sd	Error A	Error R
$\hat{\sigma}_{11}$	1.3969	0.0564	0.1031	0.0687
$\hat{\sigma}_{12}$	1.1559	0.0425	0.0141	0.0121
$\hat{\sigma}_{13}$	0.7334	0.0370	0.0466	0.0597
$\hat{\sigma}_{22}$	1.1997	0.0523	0.0597	0.0524
$\hat{\sigma}_{23}$	0.4693	0.0307	0.0296	0.0673
$\hat{\sigma}_{33}$	0.6724	0.0402	0.0476	0.0611

TABLA 2.7. Estimaciones de las componentes de  $\Sigma_3$ , desviaciones estándar posterior, errores absolutos y relativos.

Las figuras 2.5 y 2.6 muestran los trazos, las distribuciones posteriori y las autocorrelaciones parciales de las cadenas de Markov para las estimaciones de las componentes de  $\mu_2$  y  $\mu_3$ , mientras que las figuras correspondientes a los componentes de  $\Sigma_2$  y  $\Sigma_3$  se muestran en el Apéndice B.



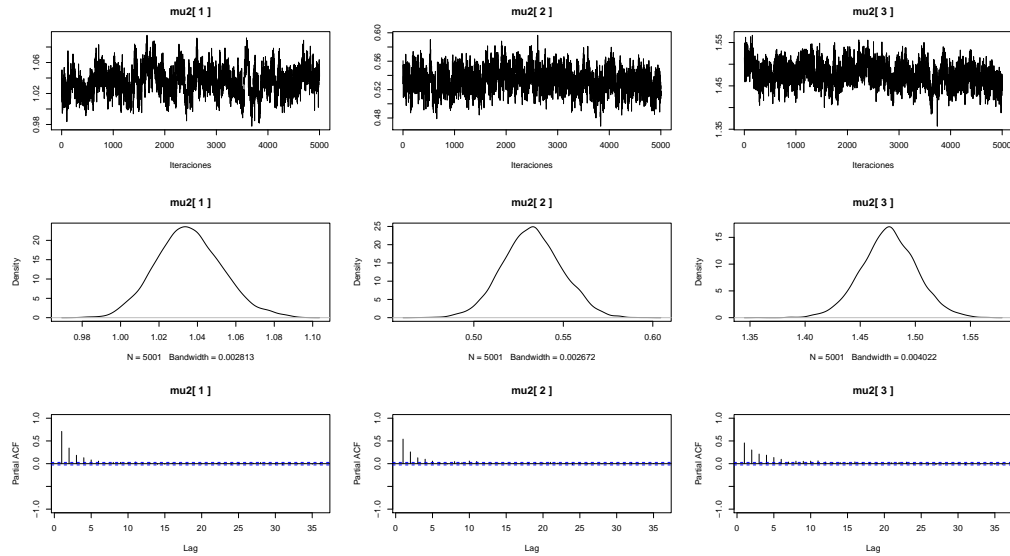


FIGURA 2.5. Trazos, distribuciones posteriori y autocorrelaciones parciales de las cadenas de Markov para las estimaciones de  $\mu_{2d}$

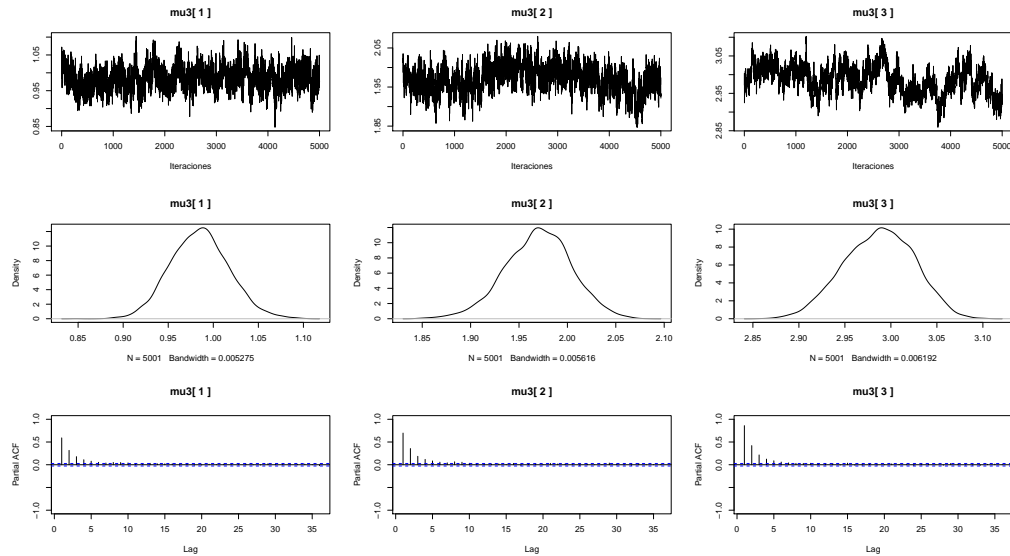


FIGURA 2.6. Trazos, distribuciones posteriori y autocorrelaciones parciales de las cadenas de Markov para las estimaciones de  $\mu_{3d}$

Para todas las cadenas resultantes se realizó el diagnóstico de convergencia. Fue utilizada la prueba de Rubin propuesta en Gelman & Rubin (1992), y para todas las cadenas resultantes se encontró que la varianza entre las diferentes cadenas de cada parámetro generadas desde diferentes puntos de inicio no es más grande que la varianza de cada cadena individual, por tanto, se puede inferir que las cadenas convergen.

Cabe resaltar que en los escenarios de simulación similares a los datos reales descritos en la siguiente sección, se pudo observar que el modelo ajusta bien. Sin embargo, se obtuvieron errores relativos mayores para la estimación de la matriz de covarianza, y una correlación un poco menor entre los parámetros de pendiente recuperados y simulados.

## 2.6. Aplicación a un caso real.

En esta sección el modelo propuesto es aplicado a un conjunto de datos reales, estos fueron tomados del Primer Estudio Internacional Comparativo sobre Lenguaje, Matemática y Factores Asociados en Tercero y Cuarto Grado, facilitados por el ICFES. Fue empleada la misma muestra utilizada en Antonio (2013). El tamaño de la muestra fue  $N = 602$ , conformada por  $N_1 = 252$  datos de Bolivia,  $N_2 = 194$  datos de Colombia y  $N_3 = 156$  datos de Cuba. Las subpruebas fueron: matemática con  $K_1 = 32$  ítems y lenguaje con  $K_2 = 19$  ítems. En Antonio (2013) se establece que la dimensión de los datos es 3, sin embargo, dado que el modelo G-LIMIRT requiere que la dimensión de los datos sea menor o igual que la dimensión de la prueba, trabajaremos con  $d = 2$ . Nuestro principal objetivo es ilustrar el modelo G-LSMIRT.

### 2.6.1. Ajuste del caso de datos reales.

Fue usado un periodo de arranque de 5000 iteraciones. Después, se corrieron 10000 iteraciones para realizar las estimaciones de los parámetros y finalmente, 1000 iteraciones mas fueron utilizadas para examinar la bondad de ajuste del modelo propuesto. El estimador Bayesiano empleado fue la media muestral, ya que la media y la mediana fueron muy similar en casi todos los casos.

El único caso en el que la media muestral y la mediana son un poco diferentes es para la estimación de las matrices de covarianzas, se puede observar en las figuras de las distribuciones posteriori presentadas en el anexo C.1 que estas son asimétricas. Sin embargo, se opto por dejar la media muestral como el estimador Bayesiano.

### 2.6.2. Parámetros estimados.

Era de interés estimar los vectores de medias  $\mu_g$  y las matrices de covarianzas  $\Sigma_g$  de los trazos latentes para cada una de las  $G$  poblaciones. Se tomó a Bolivia como población se referencia y se identificaron los parámetros para Colombia y Cuba como  $g = 2, 3$ , respectivamente. En la tabla 2.8 se muestran las estimaciones de las componentes de  $\mu_2$  y  $\mu_3$  para los datos reales.

Parámetro	Media	Sd
$\hat{\mu}_{21}$	-0.1530	0.0960
$\hat{\mu}_{22}$	0.1626	0.1113
$\hat{\mu}_{31}$	1.5654	0.2713
$\hat{\mu}_{32}$	2.0984	0.2955

TABLA 2.8. Estimaciones de las componentes de  $\mu_g$  y sus respectivas desviaciones estándar posterior para los datos reales.

Las figuras 2.7 y 2.8 muestran los trazos, las distribuciones posteriori y las autocorrelaciones parciales de las cadenas de Markov para las estimaciones de las componentes de  $\mu_2$  y  $\mu_3$ , mientras que las figuras correspondientes a los componentes de  $\Sigma_2$  y  $\Sigma_3$  se muestran en el apéndice C.1 como anteriormente se había mencionado.

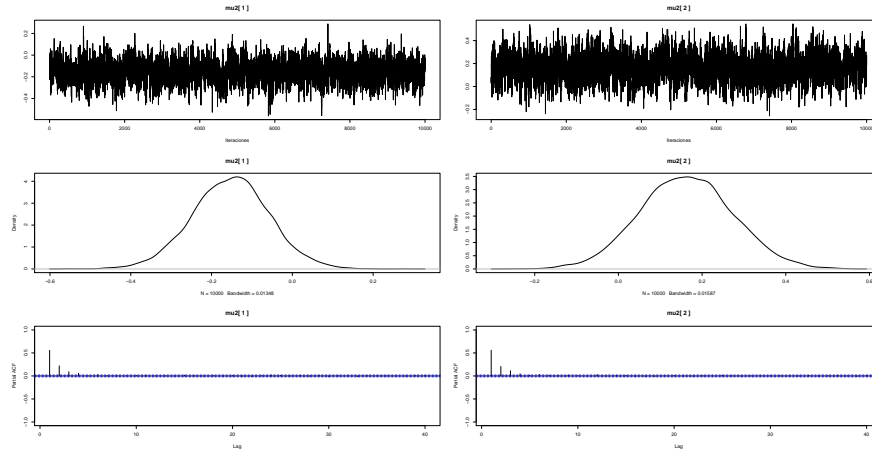


FIGURA 2.7. Trazos, distribuciones posteriori y autocorrelaciones parciales de las cadenas de Markov para las estimaciones de  $\mu_{2d}$

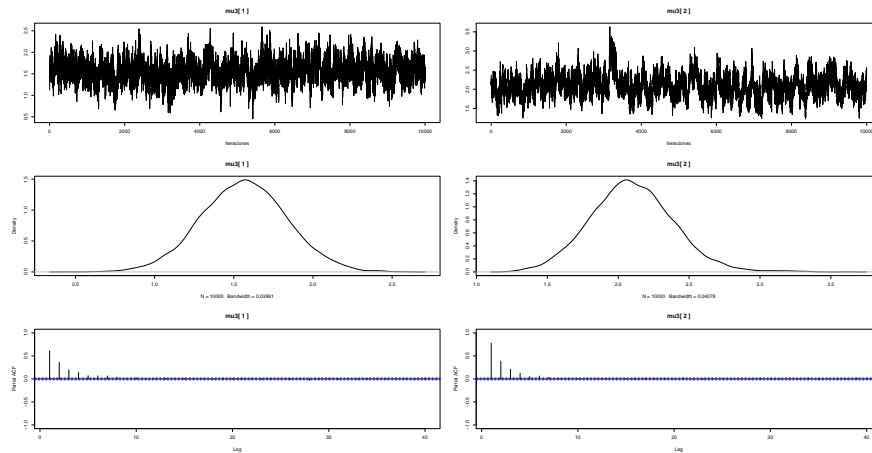


FIGURA 2.8. Trazos, distribuciones posteriori y autocorrelaciones parciales de las cadenas de Markov para las estimaciones de  $\mu_{3d}$

Las tablas 2.9 y 2.10 muestran las estimaciones de los componentes de  $\Sigma_2$  y  $\Sigma_3$ .

Parámetro	Media	Sd
$\hat{\sigma}_{11}$	0.7611	0.1852
$\hat{\sigma}_{12}$	-0.0825	0.1638
$\hat{\sigma}_{22}$	1.0968	0.2743

TABLA 2.9. Estimaciones de las componentes de  $\Sigma_2$  y sus respectivas desviaciones estándar posterior.

La tabla C.1 presentada en el anexo C.3 muestra las estimaciones obtenidas para los parámetros de pendiente e intercepto, así como las correspondientes desviaciones estándar posteriori. En la tabla, los ítems tienen su identificador original. Los ítems de cada subprueba son los siguientes: Matemáticas 1-32 y Lenguaje 33-51. En la figura C.4 presentada en el anexo C.2 se muestran las distribuciones posteriori para algunos parámetros de pendiente.

Parámetro	Media	Sd
$\hat{\sigma}_{11}$	4.4860	1.0307
$\hat{\sigma}_{12}$	-2.2981	0.8259
$\hat{\sigma}_{22}$	3.0963	0.9056

TABLA 2.10. Estimaciones de las componentes de  $\Sigma_3$  y sus respectivas desviaciones estándar posterior.

Otro aspecto de interés era estimar los trazos latentes básicos y principales para cada una de las  $G$  poblaciones; con el objetivo de examinar sus densidades y comparar éstas, con las obtenidas en la disertación de Antonio (2013).

Las figuras 2.9 y 2.10 muestran las densidades de los trazos latentes básicos de las poblaciones para cada una de las componentes que lo conforman. Se puede observar que los trazos latentes básicos estimados con el modelo G-LSMIRT tienen densidades simétricas.

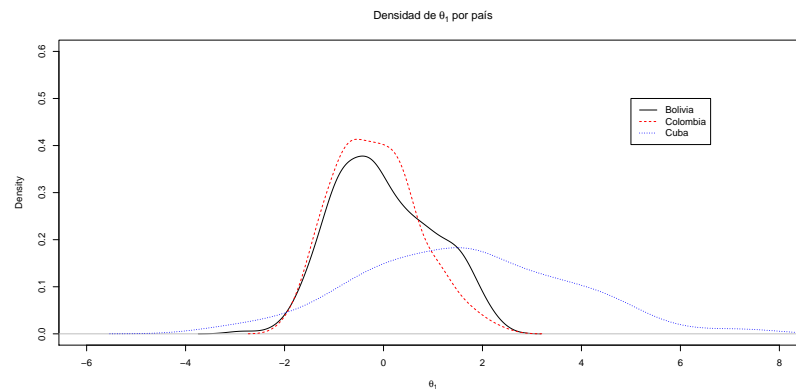


FIGURA 2.9. Densidades para la primera componente del vector de trazos latentes estimados por población

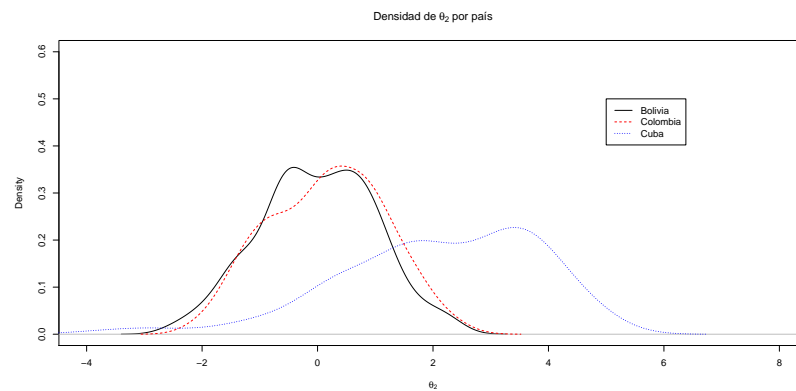


FIGURA 2.10. Densidades para la segunda componente del vector de trazos latentes estimados por población

Por otro lado, en las figuras 2.11 y 2.12 se muestran las densidades de los trazos latentes unidimensionales  $\mathbf{G}_{vg}$  asociados a cada subprueba de la  $g$ -ésima población. Para calcular los trazos latentes unidimensionales se utilizó la dirección principal de cada subprueba y el trazo latente estimado por el modelo LSMIRT para la  $g$ -ésima población, es decir,  $\mathbf{G}_{vg} = \beta_v \theta_g$ .

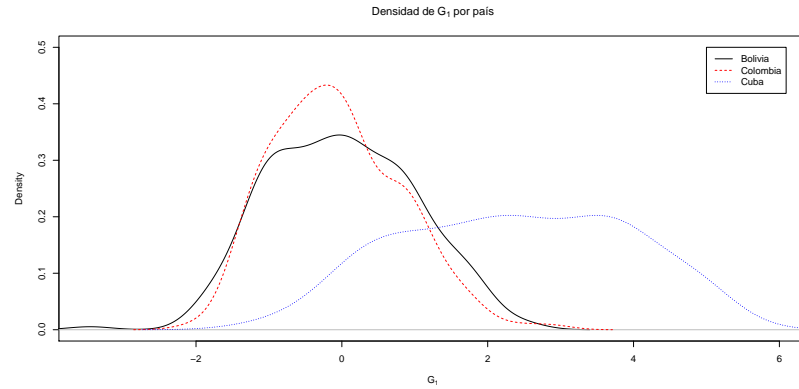


FIGURA 2.11. Densidades de los trazos latentes unidimensionales por población para la subprueba de matemáticas

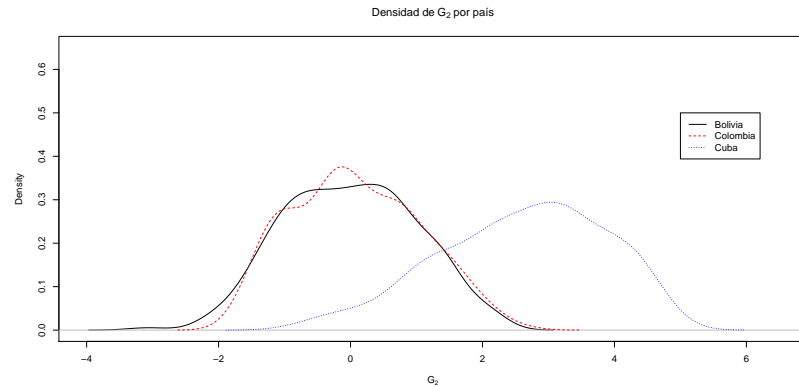


FIGURA 2.12. Densidades de los trazos latentes unidimensionales por población para la subprueba de lenguaje

En la tabla 2.11 se presentan las medias y varianzas para cada uno de los trazos latentes principales estimados en cada una de las poblaciones.

Media (Varianza)	Bolivia	Colombia	Cuba
Matemática	$-2.31 \times 10^{-19}(1)$	$-0.0694(0.7680)$	$2.3769(2.4180)$
Lenguaje	$5.63 \times 10^{-19}(1)$	$0.0808(0.9285)$	$2.6062(1.5504)$

TABLA 2.11. Estadísticas resumen para los trazos latentes principales estimados.

Se puede observar que las densidades de Colombia y Bolivia tienen comportamientos semejantes o similares, mientras que el comportamiento de la densidad de Cuba dista de éstas. Las distribuciones no son simétricas, sin embargo, son considerablemente menos asimétricas que las reportadas por Antonio; lo cual se debe a que cada grupo pertenece a una población diferente, principalmente Cuba que tiene mayor varianza y media respecto a las otras dos poblaciones.

Se construyeron las distribuciones y densidades empíricas de los datos reales, para cada componente de la prueba y cada una de las poblaciones separadamente. En las figuras 2.13 y 2.14 se presentan graficas donde se comparan estas distribuciones y densidades, con las respectivas distribuciones y densidades de los latentes estimados; se puede observar que las densidades son similares para cada componente en Bolivia y Colombia, sin embargo, las distribuciones empíricas de los datos y las correspondientes a los trazos latentes estimados respecto a cada componente de la prueba en Cuba difieren un poco.

Por otro lado, se calcularon las correlaciones de los trazos latentes principales entre las subpruebas en cada una de las poblaciones. Las correlaciones obtenidas fueron 0.7899, 0.7651 y 0.3736 en Bolivia, Colombia y Cuba, respectivamente.

Adicionalmente, en la figura 2.15 se presentan los diagramas de dispersión de los scores clásicos contra los trazos principales estimados, teniendo presente la subprueba y la población; mientras que en la tabla 2.12 se muestran los valores de las correlaciones entre estos. Se puede observar que existe una alta asociación entre los scores clásicos y los trazos principales estimados.

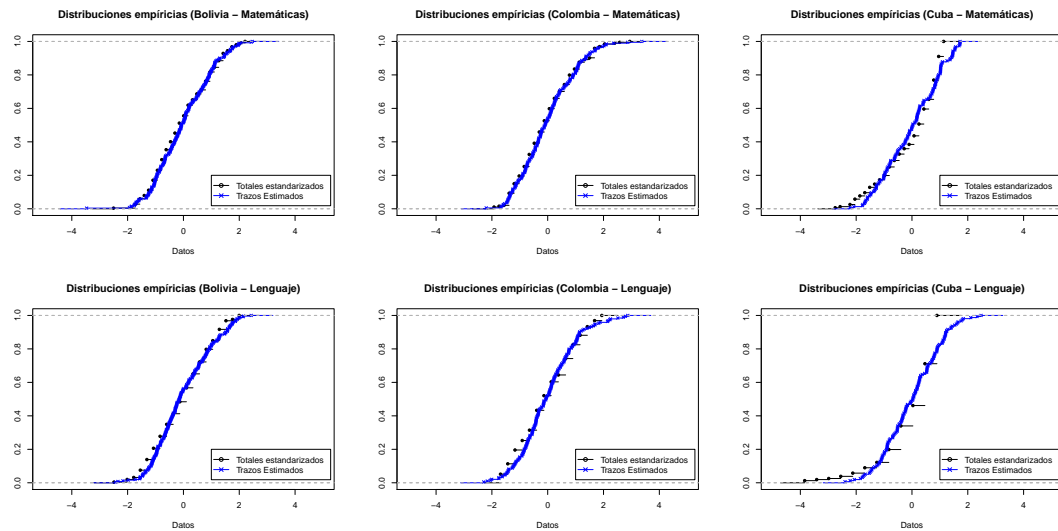


FIGURA 2.13. Comparación entre las distribuciones empíricas de los datos reales y los trazos latentes estimados.

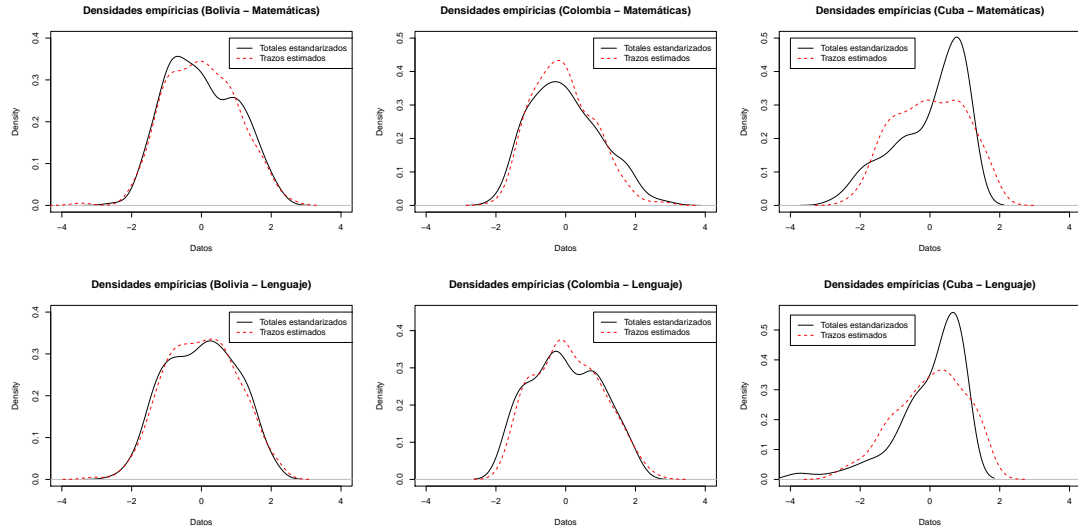


FIGURA 2.14. Comparación entre las densidades empíricas de los datos reales y los trazos latentes estimados.

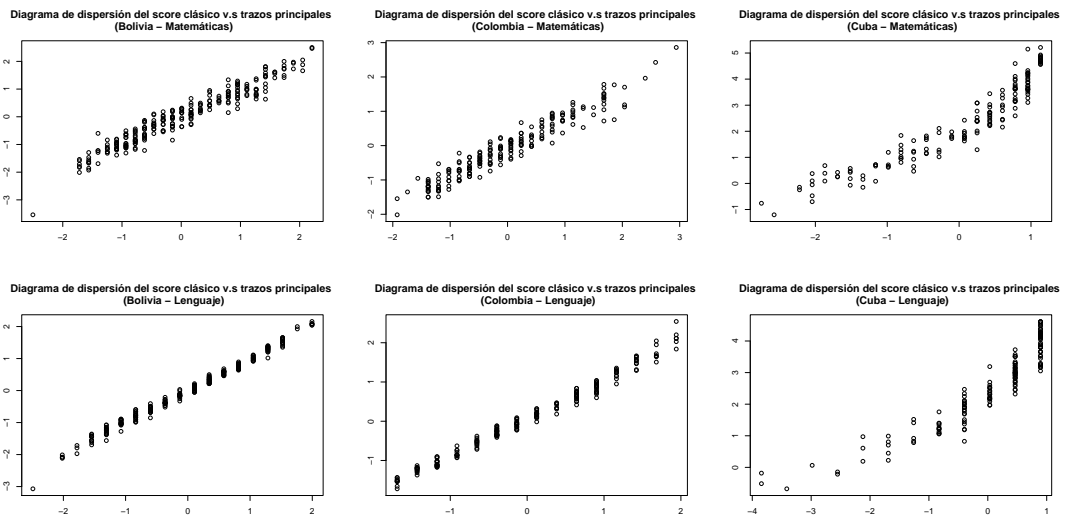


FIGURA 2.15. Diagramas de dispersión de los scores clásicos contra los trazos principales estimados.

Score clásico / Trazos estimados	Bolivia	Colombia	Cuba
Matemática	0.9859	0.9789	0.9415
Lenguaje	0.9945	0.9945	0.9382

TABLA 2.12. Correlaciones entre los scores clásicos y los trazos principales estimados.

### 2.6.3. Bondad de ajuste.

En primera instancia se examinó la presencia de outliers y valores extremos. Según Ntzoufras (2009), valores de CPO inverso mayores de 40 se pueden considerar como posibles outliers y mayores de 70 como valores extremos. Los resultados de los valores de

los CPO inverso se muestran en la tabla 2.13.

	Bolivia	Colombia	Cuba
Outliers	1.45 %	1.41 %	0.73 %
Valores extremos	0.83 %	0.81 %	0.42 %

TABLA 2.13. Porcentaje de outliers y valores extremos por país.

Dado que el porcentaje de outliers es inferior al 5 %, los resultados no son motivo de preocupación según Fox (2010).

Después del análisis de outliers y valores extremos, se examinó el ajuste de los examinados e los ítems tomando como referencia las estadísticas presentes en Fox (2010). Una medida para evaluar el ajuste de un patrón de respuesta del examinado  $i$  bajo el modelo G-LSMIRT basado en los residuales Bayesianos latentes, esta viene dada por

$$X_{p,i,g}^2 = \sum_{v=1}^m \sum_{j=1}^{K_v} (Z_{vijg} - \eta_{vijg})^2 = \sum_{v=1}^m \sum_{j=1}^{K_v} e_{vijg}^2 \quad (2.10)$$

Dado que cada residual Bayesiano latente tiene una distribución normal estándar y bajo el supuesto de independencia condicional, la estadística  $X_{p,i,g}^2$  tiene una distribución chi-cuadrado con  $K$  grados de libertad. El p-valor posterior correspondiente es definido en Fox (2010) como

$$p_0(X_{p,i,g}^2) = \int P(X_K^2 > X_p^2(z_{ig})) p(z_{ig}|y_{ig}) dz_{ig}, \quad (2.11)$$

donde  $y_{ig}$  representa el patrón de respuesta del examinado  $i$  que pertenece a la población  $g$  y  $z_{ig}$  corresponde al patrón de respuesta latente. El p-valor posterior es calculado en cada paso del algoritmo implementado para las últimas 1000 iteraciones, y la media es la estimación del p-valor posterior.

Los p-valores obtenidos para el caso de datos reales según país fueron: Bolivia entre 0.18 y 0.75, Colombia entre 0.30 y 0.71, y para Cuba entre 0.16 y 0.64.

Similarmente, una estadística para examinar el ajuste de los ítems es definida como

$$X_{item,j}^2 = \sum_{g=1}^G \sum_{i=1}^{N_g} (Z_{vijg} - \eta_{vijg})^2 = \sum_{g=1}^G \sum_{i=1}^{N_g} e_{vijg}^2 \quad (2.12)$$

y el correspondiente p-valor posterior es definido en Fox (2010) como

$$p_0(X_{item,vj}^2) = \int P(\chi_N^2 > X_{item}^2(z_j)) p(z_j|y_j) dz_j. \quad (2.13)$$



Los p-valores en el caso de datos reales estuvieron entre 0.41 y 0.64.

Por los resultados mostrados anterior, se puede considerar que el modelo se ajusta adecuadamente a los datos.

#### 2.6.4. Discusión.

En este capítulo se introdujo el modelo G-LSMIRT. El modelo ha sido pensado para ser utilizado en las pruebas de evaluación a gran escala diseñadas expresamente para medir más de un trazo latente, aplicadas a  $g$  grupos poblaciones. Se asumió que las pruebas se dividen en subpruebas y que cada subprueba está diseñado para medir principalmente un rasgo latente unidimensional único.

Un algoritmo del muestreador de Gibbs para datos aumentos (DAGS) fue implementado para ajustar el modelo G-LSMIRT. Los resultados muestran que los parámetros están bien recuperados por el algoritmo DAGS.

Para ilustrar el uso del modelo G-LSMIRT, se utilizaron los datos del Primer Estudio Internacional Comparativo sobre Lenguaje, Matemática y Factores Asociados en Tercero y Cuarto Grado. El tamaño de la muestra fue  $N = 602$ , conformada por  $N_1 = 252$  datos de Bolivia,  $N_2 = 194$  datos de Colombia y  $N_3 = 156$  datos de Cuba. Las subpruebas fueron: matemática con  $K_1 = 32$  ítems y lenguaje con  $K_2 = 19$  ítems. Se utilizó  $d = 2$  para la dimensión de los datos.

En la figura 2.15, específicamente en el diagrama de dispersión del score clásico contra el trazo latente principal correspondiente a la subprueba de lenguaje presentada por los estudiantes de Bolivia por ejemplo, se puede observar que los valores más grandes de los trazos latentes principales corresponden a dos estudiantes que obtuvieron un score clásico inferior a uno en la subprueba. Existen otros estudiantes que obtuvieron un score clásico de uno en la prueba, y tienen un trazo latente principal de menor valor. Lo anterior puede deberse a que esos estudiantes tuvieron mejores resultados en la subprueba de matemáticas. Por otro lado, se encontró que los estudiantes que tuvieron un mayor score clásico en la subprueba de matemáticas, tuvieron un trazo latente más grande en la subprueba de lenguaje.

Las correlaciones de los scores clásicos con sus correspondientes trazos latentes principales estimados para cada una de las poblaciones y cada subprueba separadamente fueron similares (ver tabla 2.12). Aunque se puede observar que las correlaciones obtenidas para Cuba fueron un poco menor, no fueron lo suficientemente marcadas para decir que difieren del comportamiento de las otras poblaciones. Cabe notar, que en la disertación de Antonio se obtuvieron resultados semejantes respecto a éstas correlaciones.

En las figuras 2.13 y 2.14, se puede notar que las densidades empíricas de los datos reales y los trazos latentes estimados son similares para cada componente en Bolivia y Colombia, mientras que en cada componente en Cuba estas densidades difieren un poco. Además se puede ver que las densidades empíricas de los scores clásicos estandarizados para cada subprueba en Cuba son asimétricas hacia la izquierda y esta tendencia es bastante marcada; el modelo propuesto asume simetría, por eso no alcanza a describir

todo el comportamiento de estas.

En Antonio se encontró que los dos valores más grandes del trazo latente calculados en la muestra, tanto en la subprueba de matemáticas como en la de lenguaje, son de estudiantes de Bolivia; sin embargo, en los resultados de esta disertación se encontró que los valores más grandes del trazo latente son de estudiantes de Cuba. Cabe notar que los estudiantes de Cuba tuvieron mayor cantidad de respuestas correctas.

Por otro lado a diferencia de Antonio (2013), el modelo G-LSMIRT supone que los parámetros de las pruebas son invariantes en los  $g$  grupos poblacionales, como se supuso en Bégin & Glass (2001). Se pudo observar que las densidades de Colombia y Bolivia tienen comportamientos semejantes o similares, mientras que el comportamiento de la densidad de Cuba dista de éstas; sin embargo, todas las densidades son menos asimétricas que las presentadas en los resultados encontrados por Antonio. Además, como anteriormente fue mencionado, las densidades de los trazos latentes para Cuba presentan mayor dispersión. Estas observaciones no pueden ser percibidas en la disertación de Antonio.

El modelo propuesto por Antonio presentó un buen ajuste al modelar el mismo conjunto de datos, sin embargo, aparentemente el modelo G-LSMIRT explica mejor el comportamiento los datos.

En general, el modelo G-LSMIRT es más parsimonioso que los modelos existentes para múltiples grupos. Las respuestas en las pruebas simuladas fueron generadas usando modelos MIRT para cada uno de los grupos, manteniendo los mismos parámetros de los ítems en cada grupo. Sin embargo, la estructura de grupos de las pruebas y el hecho que cada subprueba media esencialmente un trazo latente principal fue incorporada a los modelos MIRT. Los resultados obtenidos en este capítulo mostraron que el modelo G-LSMIRT ajusta bien los datos, por lo que el modelo MIRT clásico puede ser substituido por un modelo LSMIRT en estas situaciones.

---

## Conclusiones

---

En esta disertación, se ha propuesto el modelo MIRT con una estructura lineal latente para múltiples grupos o subgrupos poblacionales (G-LSMIRT), una extensión del modelo LSMIRT propuesto por Montenegro (2011). El modelo G-LSMIRT fue escrito como

$$P(Y_{vjig} = 1 | \alpha_{vj}, \gamma_{vj}, \boldsymbol{\beta}_v, \boldsymbol{\theta}_{ig}) = \Phi(\alpha_{vj} \boldsymbol{\beta}_v^t \boldsymbol{\theta}_{ig} - \gamma_{vj})$$

donde  $Y_{vjig}$  es la variable aleatoria binaria que representa la respuesta del individuo  $i$  para la  $j$ -th respuesta de la subprueba  $v$  para el grupo  $g$ . Los  $\alpha_{vj}$  y  $\gamma_{vj}$  son respectivamente el parámetro de pendiente (la discriminación) y el parámetro de intercepto del ítem  $vj$ . El vector  $\boldsymbol{\beta}_v = (\beta_{v1}, \dots, \beta_{vd})^t$  es un vector unitario en el espacio de trazo latente, llamado la dirección de la subprueba  $v$ .  $\boldsymbol{\theta}_{ig} = (\theta_{i1g}, \dots, \theta_{idg})^t$  representa el vector de trazos latentes básicos del examinado  $i$  del grupo o población  $g$  y  $\Phi(\cdot)$  denota la función de enlace es la ojiva de la normal estándar.

Para estimar los parámetros del modelo G-LSMIRT se implementó un algoritmo de datos aumentados, el cual es especificado junto con las distribuciones condicionales posteriori completas en los anexos de esta disertación. En el capítulo 2 se puede observar que el algoritmo recuperó bien todos los parámetros. En general, el modelo G-LSMIRT es más parsimonioso que los modelos existentes para múltiples grupos.

Se aplicó el modelo G-LSMIRT a una muestra de datos reales de la prueba PERCE y el ajuste fue adecuado de acuerdo con los resultados de los criterios de evaluación aplicados. Los componentes de los vectores de medias de los trazos latentes de Bolivia y Colombia son similares para cada subprueba, mientras que los correspondientes a Cuba son considerablemente mayores. Adicionalmente, se pudo ver que para la subprueba de lenguaje el trazo latente de Cuba presentó mayor dispersión que los demás trazos. Por otro lado, se encontró que las densidades de los trazos latentes en cada país son menos asimétricas que las reportadas por Antonio.

Los examinados de la muestra de datos reales que respondieron igual a los ítems de una subprueba, tuvieron diferentes trazos latentes. Lo anterior puede deberse a que en la muestra de datos reales seleccionada, los trazos latentes estimados tuvieron una alta correlación entre las dos subpruebas. Algunos estudiantes que tuvieron mayor cantidad

---

de respuestas correctas en una subprueba, tuvieron un trazo latente más pequeño respecto a otros estudiantes que tuvieron un score clásico mayor en la otra subprueba. Es posible que este resultado se deba también a la correlación entre las dos subpruebas.

Dado que el modelo fue de dos dimensiones que coincide con la dimensión de la prueba es posible proponer para cada subpoblación un modelo de estructura simple. Este procedimiento es equivalente y lleva a los mismos resultados, pero sin pasar por los trazos generales, ni tener que calcular las composiciones.

Para finalizar se puede concluir que los parámetros del modelo G-LSMIRT se pueden interpretar de la misma forma que los del modelo LSMIRT.

## CAPÍTULO 4

---

---

### Trabajos futuros

---

---

- En este trabajo se asumió que la distribución de los trazos latentes era una normal multivariada. Se podría pensar en utilizar otras distribuciones multivariadas, tales como: la t student multivariada o alguna distribución multivariada asimétrica.
- Realizar la generalización a múltiples poblaciones del modelo LSMIRT en presencia de testlets y de igual forma, incluir la presencia de otras distribuciones diferentes a la normal multivariada para los trazos latentes.
- Otro aspecto interesante, sería tener en cuenta pesos de muestreo a los individuos.

---



---

## Distribuciones posterioris completas

---



---

En este Apéndice, son derivadas las distribuciones condicionales completas del modelo LSMIRT para múltiples poblaciones.

### A.1. Notación matricial

Las matrices usadas en el algoritmo DAGS son definidas en esta sección.

$$\boldsymbol{\alpha} = \begin{bmatrix} \boldsymbol{\alpha}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\alpha}_2 & \cdots & \mathbf{0} \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\alpha}_m \end{bmatrix}_{K \times m} \quad (\text{A.1})$$

donde  $\boldsymbol{\alpha}_1 = (\alpha_1, \alpha_2, \dots, \alpha_{K_1})^t, \dots, \boldsymbol{\alpha}_m = (\alpha_{K-K_m+1}, \alpha_{K-K_m+2}, \dots, \alpha_K)^t$ . Sea  $\boldsymbol{\beta}$  la matriz cuyas componentes son los vectores de dirección  $\boldsymbol{\beta}_v, v = 1, 2, \dots, m$ . Sea  $\mathbf{A}$  la matriz definida como  $\mathbf{A} = \boldsymbol{\alpha} \times \boldsymbol{\beta}$ , entonces  $\mathbf{A}$  está dada por

$$\mathbf{A} = \begin{bmatrix} \boldsymbol{\alpha}_1 \beta_{11} & \boldsymbol{\alpha}_1 \beta_{12} & \cdots & \boldsymbol{\alpha}_1 \beta_{1d} \\ \boldsymbol{\alpha}_2 \beta_{21} & \boldsymbol{\alpha}_2 \beta_{22} & \cdots & \boldsymbol{\alpha}_2 \beta_{2d} \\ \cdots & \cdots & \cdots & \cdots \\ \boldsymbol{\alpha}_m \beta_{m1} & \boldsymbol{\alpha}_m \beta_{m2} & \cdots & \boldsymbol{\alpha}_m \beta_{md} \end{bmatrix}_{K \times d} \quad (\text{A.2})$$

La  $j$ -th fila de  $\mathbf{A}$  es denotada por  $\boldsymbol{\alpha}_j^t$ ,  $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_k)^t$  es el vector de interceptos de los ítems,  $\mathbf{Z}_g = [Z_{ijg}]_{N_g \times K}$  la matriz de variables aumentadas de la población  $g$ ,  $\mathbf{1}_{N_1}$  es un vector de unos de tamaño  $N_1$  y  $\boldsymbol{\theta}_g = [\theta_{ijg}]_{N_g \times d}$  es la matriz de todos los vectores de trazo latente de la muestra de examinados de la  $g$ -th población.

Las otras matrices requeridas en el algoritmo DAGS o utilizadas para el cálculo de las distribuciones a posteriori son las siguientes

$$\boldsymbol{\eta}_g = \boldsymbol{\theta}_g \mathbf{A}^t - \mathbf{1}_{N_g}$$

$$\begin{aligned}\mathbf{Z}^t &= [\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_g] \\ \mathbf{D} &= \mathbf{Z} + \mathbf{1}_N \\ \boldsymbol{\theta}^t &= [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_g] \\ \mathbf{D}^t &= [\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_g]\end{aligned}$$

## A.2. Variables latentes $Z_{ijg}$

La distribución condicional de  $Z_{ijg}$  denotada como  $[Z_{ijg}|\cdot]$  es dada por

$$Z_{ijg}|\cdot \sim \begin{cases} N_{(0,\infty)}(\eta_{ijg}, 1), & \text{si } y_{ijg} = 1 \\ N_{(-\infty,0)}(\eta_{ijg}, 1), & \text{si } y_{ijg} = 0 \\ N_{(-\infty,\infty)}(\eta_{ijg}, 1), & \text{en otro caso.} \end{cases} \quad (\text{A.3})$$

## A.3. Distribución condicional completa a posteriori de $\boldsymbol{\theta}_{ig}$

La distribución a priori de los  $\boldsymbol{\theta}_{ig}$ , es por

$$N_d(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g),$$

entonces

$$\begin{aligned}[\boldsymbol{\theta}_{ig}|\cdot] &\propto p(\mathbf{z}_g|\boldsymbol{\beta}, \boldsymbol{\theta}_{ig}, \boldsymbol{\zeta})p(\boldsymbol{\theta}_{ig}|\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g) \\ &\propto \prod_{j=1}^K \exp\left\{-\frac{1}{2}(z_{ijg} - \mathbf{a}_j^t \boldsymbol{\theta}_{ig} + \gamma_j)^2\right\} \times \exp\left\{-\frac{1}{2}((\boldsymbol{\theta}_{ig} - \boldsymbol{\mu}_g)^t \boldsymbol{\Sigma}_g^{-1} (\boldsymbol{\theta}_{ig} - \boldsymbol{\mu}_g))\right\} \\ &= \prod_{j=1}^K \exp\left\{-\frac{1}{2}(\mathbf{a}_j^t \boldsymbol{\theta}_{ig} - D_{ijg})^2\right\} \times \exp\left\{-\frac{1}{2}((\boldsymbol{\theta}_{ig} - \boldsymbol{\mu}_g)^t \boldsymbol{\Sigma}_g^{-1} (\boldsymbol{\theta}_{ig} - \boldsymbol{\mu}_g))\right\} \\ &= \exp\left\{-\frac{1}{2}(\mathbf{A}\boldsymbol{\theta}_{ig} - \mathbf{D}_{ig})^t (\mathbf{A}\boldsymbol{\theta}_{ig} - \mathbf{D}_{ig})\right\} \times \exp\left\{-\frac{1}{2}((\boldsymbol{\theta}_{ig} - \boldsymbol{\mu}_g)^t \boldsymbol{\Sigma}_g^{-1} (\boldsymbol{\theta}_{ig} - \boldsymbol{\mu}_g))\right\} \\ &\propto \exp\left\{-\frac{1}{2}(\boldsymbol{\theta}_{ig}^t \mathbf{A}^t \mathbf{A} \boldsymbol{\theta}_{ig} - 2(\mathbf{A}^t \mathbf{D}_{ig})^t \boldsymbol{\theta}_{ig} + \boldsymbol{\theta}_{ig}^t \boldsymbol{\Sigma}_g^{-1} \boldsymbol{\theta}_{ig} - 2(\boldsymbol{\mu}_g^t \boldsymbol{\Sigma}_g^{-1})^t \boldsymbol{\theta}_{ig})\right\} \\ &= \exp\left\{-\frac{1}{2}(\boldsymbol{\theta}_{ig}^t (\mathbf{A}^t \mathbf{A} + \boldsymbol{\Sigma}_g^{-1}) \boldsymbol{\theta}_{ig} - 2(\mathbf{A}^t \mathbf{D}_{ig} + \boldsymbol{\mu}_g^t \boldsymbol{\Sigma}_g^{-1})^t \boldsymbol{\theta}_{ig})\right\}\end{aligned}$$

Por lo tanto, la distribución condicional completa a posteriori de  $\boldsymbol{\theta}_{ig}$  está dada por

$$\boldsymbol{\theta}_{ig} \sim N_d\left((\mathbf{A}^t \mathbf{A} + \boldsymbol{\Sigma}_g^{-1})^{-1} (\mathbf{A}^t \mathbf{D}_{ig} + \boldsymbol{\mu}_g^t \boldsymbol{\Sigma}_g^{-1}), (\mathbf{A}^t \mathbf{A} + \boldsymbol{\Sigma}_g^{-1})^{-1}\right) \quad (\text{A.4})$$

## A.4. Distribución condicional completa a posteriori de $\boldsymbol{\beta}_v$

La distribución a priori de los  $\boldsymbol{\beta}_v$ , propuesta, es por

$$N_d(\mathbf{b}, \mathbf{T})$$

donde,  $\mathbf{T}$  es una matriz diagonal. Sea  $\mathbf{D}_j$  la columna  $j$ -th de la matriz  $\mathbf{D}$ . Entonces,

$$\begin{aligned} [\boldsymbol{\beta}_v | \cdot] &\propto p(\mathbf{z} | \boldsymbol{\beta}_v, \boldsymbol{\theta}, \boldsymbol{\zeta}) p(\boldsymbol{\beta}_v) \\ &\propto \prod_{g=1}^G \prod_{i=1}^{N_g} \prod_{j \in c_v} \exp\left\{-\frac{1}{2}(\alpha_j \boldsymbol{\theta}_{ig}^t \boldsymbol{\beta}_v - (z_{ijg} + \gamma_j))^2\right\} \times p(\boldsymbol{\beta}_v) \\ &= \prod_{g=1}^G \prod_{i=1}^{N_g} \prod_{j \in c_v} \exp\left\{-\frac{1}{2}(\alpha_j \boldsymbol{\theta}_{ig}^t \boldsymbol{\beta}_v - D_{ijg})^2\right\} \times p(\boldsymbol{\beta}_v) \\ &= \exp\left\{-\frac{1}{2} \sum_{j \in c_v} [\alpha_j \boldsymbol{\theta} \boldsymbol{\beta}_v - \mathbf{D}_j]^t [\alpha_j \boldsymbol{\theta} \boldsymbol{\beta}_v - \mathbf{D}_j]\right\} \times p(\boldsymbol{\beta}_v) \\ &\propto \exp\left\{-\frac{1}{2} \sum_{j \in c_v} [\alpha_j^2 \boldsymbol{\beta}_v^t \boldsymbol{\theta}^t \boldsymbol{\theta} \boldsymbol{\beta}_v - 2\alpha_j (\boldsymbol{\theta} \mathbf{D}_j)]\right\} \times p(\boldsymbol{\beta}_v) \end{aligned}$$

Sea  $\mathbf{W}_v = \sum_{j \in c_v} \alpha_j^2 \boldsymbol{\theta}^t \boldsymbol{\theta}$  y  $\mathbf{D}_v = \sum_{j \in c_v} \alpha_j \mathbf{D}_j$ . Por lo tanto, la distribución condicional completa a posteriori de  $\boldsymbol{\beta}_v$  es dada por

$$\boldsymbol{\beta}_v | \cdot \sim N_d((\mathbf{W}_v + \mathbf{T}^{-1})^{-1}(\boldsymbol{\theta}^t \mathbf{D}_v + \mathbf{T}^{-1} \mathbf{b}), (\mathbf{W}_v + \mathbf{T}^{-1})^{-1}) \quad (\text{A.5})$$

Se propone para esta tesis  $\mathbf{b} = \frac{1}{\sqrt{d}} \mathbf{1}_d$ , donde  $\mathbf{1}_d$  es un vector de unos de dimensión  $d$  y  $\mathbf{T} = \mathbf{I}_d$ .

## A.5. Distribución condicional completa a posteriori de $\boldsymbol{\zeta}_j$

La distribución a priori de los  $\boldsymbol{\zeta}_j$  propuesta, es por

$$p(\boldsymbol{\zeta}_j) \propto \exp\left\{-\frac{1}{2} \boldsymbol{\zeta}_j^t \mathbf{S}^{-1} \boldsymbol{\zeta}_j\right\}$$

donde,  $\mathbf{S}_2$  es una matriz invertible de orden  $2 \times 2$ . Entonces,

$$\begin{aligned} [\boldsymbol{\zeta}_j | \cdot] &\propto p(\mathbf{z} | \boldsymbol{\beta}_v, \boldsymbol{\theta}, \boldsymbol{\zeta}) p(\boldsymbol{\zeta}_j) \\ &\propto \prod_{g=1}^G \prod_{i=1}^{N_g} \exp\left\{-\frac{1}{2}(\alpha_j \boldsymbol{\beta}_v^t \boldsymbol{\theta}_{ig} - \gamma_j - z_{ijg})^2\right\} \times p(\boldsymbol{\zeta}_j) \\ &= \exp\left\{-\frac{1}{2}([\boldsymbol{\theta} \boldsymbol{\beta}_v - \mathbf{1}_N] \boldsymbol{\zeta}_j - \mathbf{Z}_j)^t ([\boldsymbol{\theta} \boldsymbol{\beta}_v - \mathbf{1}_N] \boldsymbol{\zeta}_j - \mathbf{Z}_j)\right\} \times p(\boldsymbol{\zeta}_j) \\ &= \exp\left\{-\frac{1}{2}(\mathbf{X}_v \boldsymbol{\zeta}_j - \mathbf{Z}_j)^t (\mathbf{X}_v \boldsymbol{\zeta}_j - \mathbf{Z}_j)\right\} \times p(\boldsymbol{\zeta}_j) \end{aligned}$$



Por lo tanto, la distribución condicional completa a posteriori de  $\zeta_j$  está dada por

$$\zeta_j \sim N_2 \left( (\mathbf{X}_v^t \mathbf{X}_v + \mathbf{S}^{-1})^{-1} (\mathbf{X}_v^t \mathbf{Z}_j), (\mathbf{X}_v^t \mathbf{X}_v + \mathbf{S}^{-1})^{-1} \right) \mathbf{I}(\alpha_j > 0) \quad (\text{A.6})$$

Tomaremos  $\mathbf{S} = \mathbf{I}_2$ .

## A.6. Distribución condicional completa a posteriori de $\mu_g$

La distribución a priori de los  $\mu_g$  propuesta, es por

$$N_d(\mathbf{0}, \Sigma_{\mu_g})$$

entonces,

$$\begin{aligned} [\mu_g | \cdot] &\propto p(\boldsymbol{\theta}_{ig} | \mu_g, \Sigma_g) p(\mu_g) \\ &\propto \prod_{i=1}^{N_g} \exp\left\{-\frac{1}{2}(\boldsymbol{\theta}_{ig} - \mu_g)^t \Sigma_g^{-1} (\boldsymbol{\theta}_{ig} - \mu_g)\right\} \times \exp\left\{-\frac{1}{2} \mu_g^t \Sigma_{\mu_g}^{-1} \mu_g\right\} \\ &= \prod_{i=1}^{N_g} \exp\left\{-\frac{1}{2}(\boldsymbol{\theta}_{ig}^t \Sigma_g^{-1} \boldsymbol{\theta}_{ig} - 2(\Sigma_g^{-1} \boldsymbol{\theta}_{ig}^t)^t \mu_g + \mu_g^t \Sigma_g^{-1} \mu_g)\right\} \times \exp\left\{-\frac{1}{2} \mu_g^t \Sigma_{\mu_g}^{-1} \mu_g\right\} \\ &\propto \exp\left\{-\frac{1}{2}(\mu_g^t (N_g \Sigma_g^{-1} + \Sigma_{\mu_g}^{-1}) \mu_g - 2 \sum_{i=1}^{N_g} (\Sigma_g^{-1} \boldsymbol{\theta}_{ig}^t)^t \mu_g)\right\} \end{aligned}$$

Por lo tanto, la distribución condicional completa a posteriori de  $\mu_g$  está dada por

$$[\mu_g | \cdot] \sim N_d \left( (N_g \Sigma_g^{-1} + \Sigma_{\mu_g}^{-1})^{-1} \sum_{i=1}^{N_g} \Sigma_g^{-1} \boldsymbol{\theta}_{ig}^t, (N_g \Sigma_g^{-1} + \Sigma_{\mu_g}^{-1})^{-1} \right) \quad (\text{A.7})$$

## A.7. Muestrear $\Sigma_g$

Se propone muestrear la matriz  $\Sigma_g$  como en Bégin & Glass (2001). Además proponemos una a priori no informativa de Jeffreys cuya densidad está dada por

$$p(\mathbf{R}_g) \propto |\mathbf{R}_g|^{-\frac{d+1}{2}}$$

Entonces

$$[\Sigma_g | \cdot] \sim W_d^{-1}(N_g, \mathbf{S}_g), \quad (\text{A.8})$$

donde,  $\mathbf{S}_g = \sum_{i=1}^{N_g} (\boldsymbol{\theta}_{ig} - \mu_g)(\boldsymbol{\theta}_{ig} - \mu_g)^t$ .

## A.8. Algoritmo DAGS

El algoritmo de DAGS utilizado es el siguiente

1. Obtener valores iniciales para los parámetros de los ítems  $\alpha_j^0 = 1$  y  $\gamma_j^0 = 0$  para todo  $j$ ,  $\Sigma_g^0 = \mathbf{I}_d$  para todo  $g$ ,  $\theta_{ig}^0 \sim N_d(\mathbf{0}, \mathbf{I}_d)$  y  $\beta_v^0 = \frac{1}{\sqrt{d}}\mathbf{1}_d$  para cada  $\beta_v$ .
2. Muestrear  $Z$ , para todo  $ijg$ .
3. Muestrear  $\theta_{ig}$ , para todo  $ig$ .
4. Muestrear  $\Sigma_g$  para todo  $g$ .
5. Muestrear  $\mu_g$  para todo  $g$ .
6. Muestrear  $\zeta_j$  para todo  $j$ .
7. Para cada  $v$ , muestrear  $\beta_v$ . Hacer  $s_v = \|\beta_v\|$ . Normalizar  $\beta_v$  y multiplicar  $\alpha_j$  por  $s_v$ , para todo  $j \in c_v$ .

# APÉNDICE B

---

---

Figuras de las cadenas de  $\Sigma_g$  para los datos simulados.

---

---

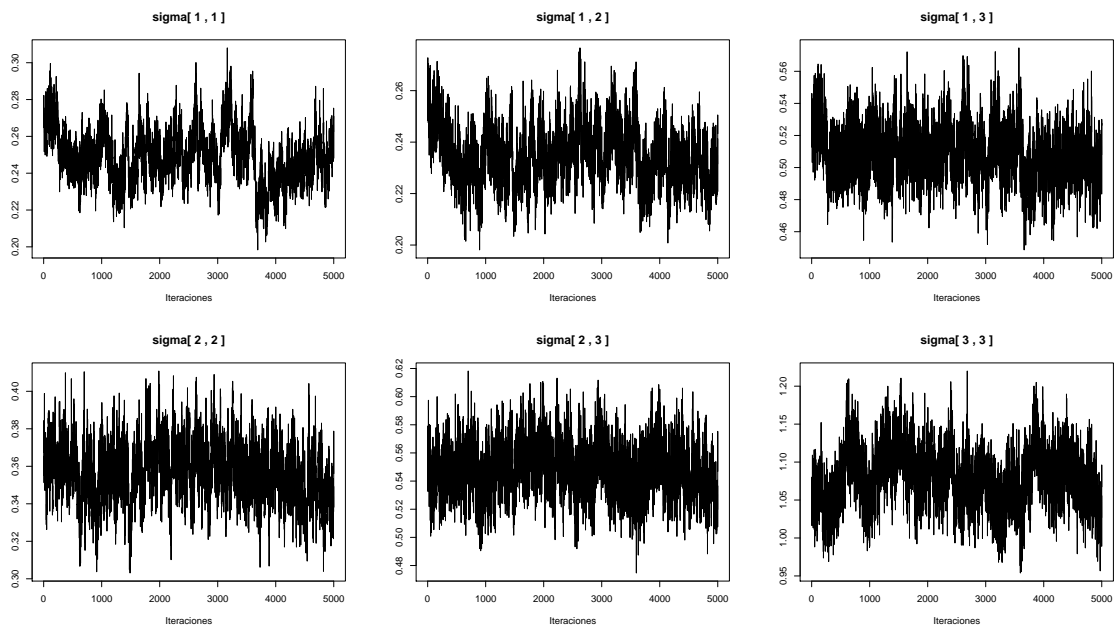


FIGURA B.1. Trazos de las cadenas de Markov para las estimaciones de las componentes de  $\Sigma_2$  para los datos simulados.

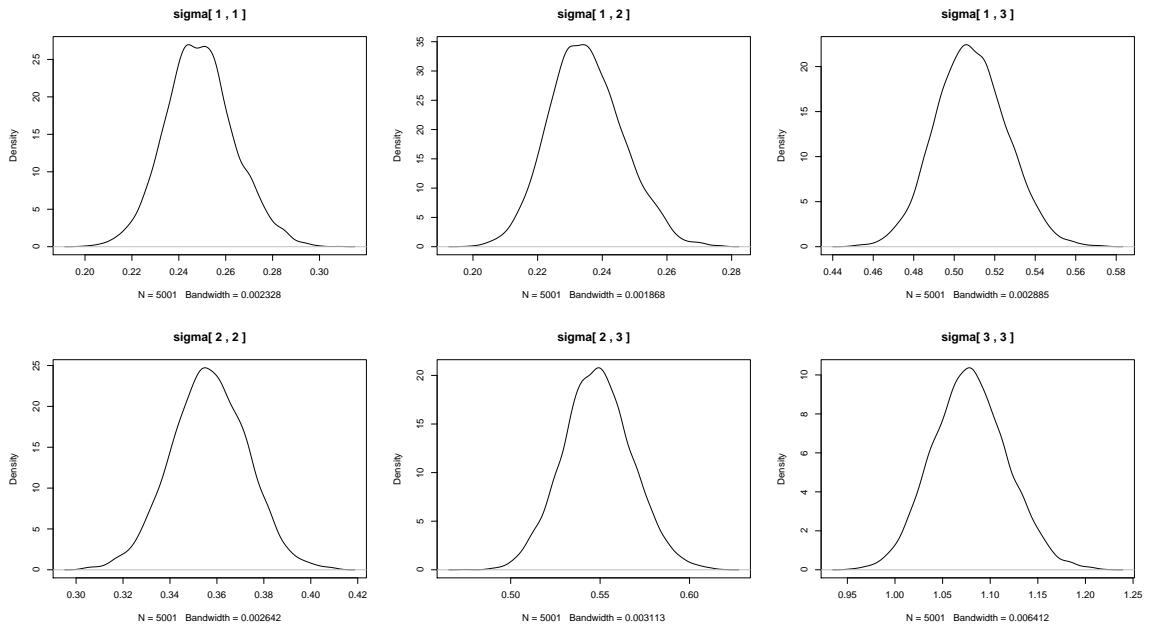


FIGURA B.2. Distribuciones posteriores de las cadenas de Markov para las estimaciones de las componentes de  $\Sigma_2$  para los datos simulados.

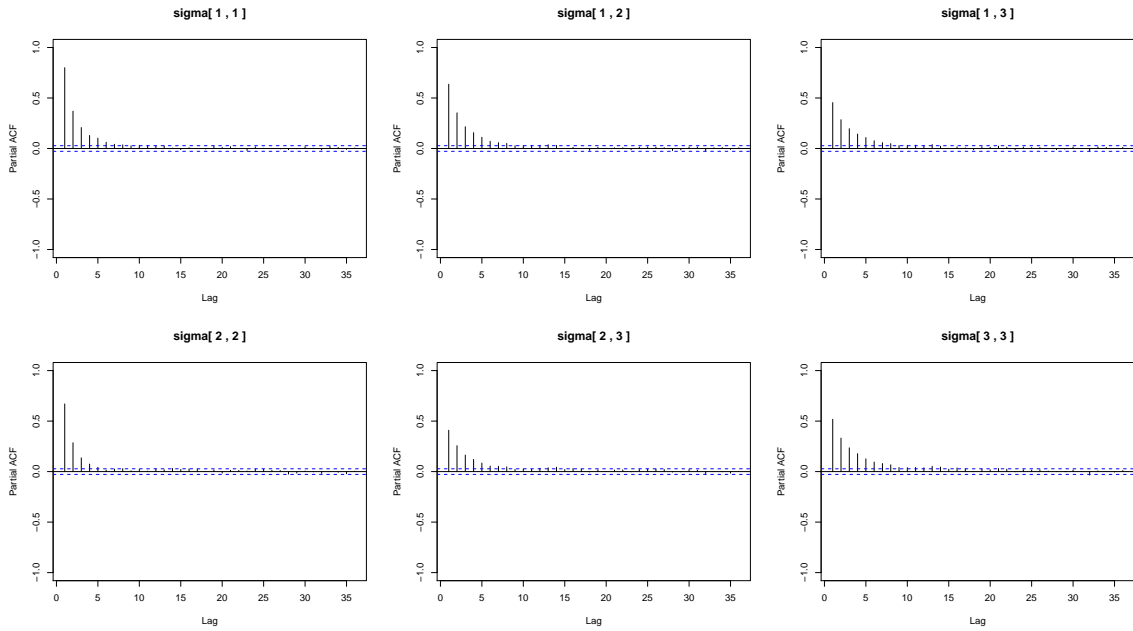


FIGURA B.3. Autocorrelaciones parciales de las cadenas de Markov para las estimaciones de las componentes de  $\Sigma_2$  para los datos simulados.

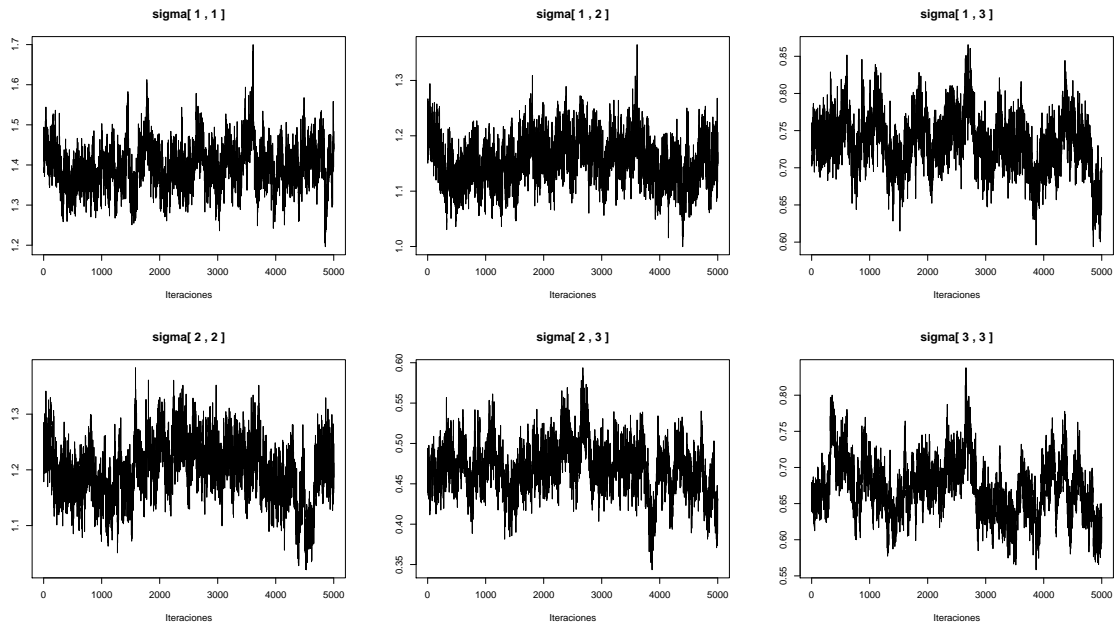


FIGURA B.4. Trazos de las cadenas de Markov para las estimaciones de las componentes de  $\Sigma_3$  para los datos simulados.

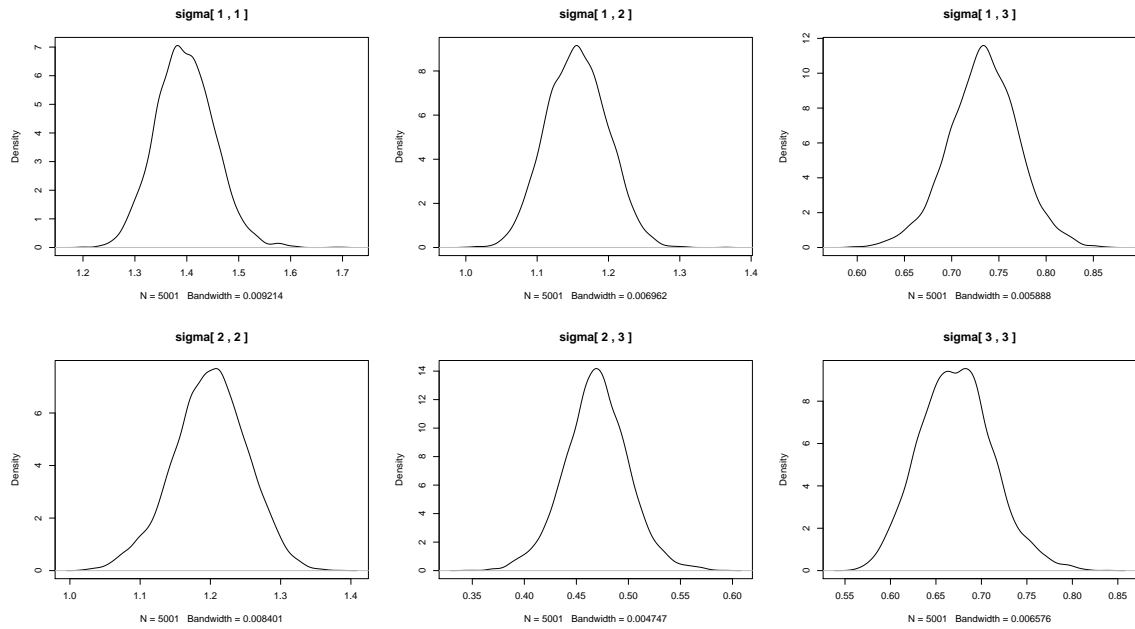


FIGURA B.5. Distribuciones posteriores de las cadenas de Markov para las estimaciones de las componentes de  $\Sigma_3$  para los datos simulados.

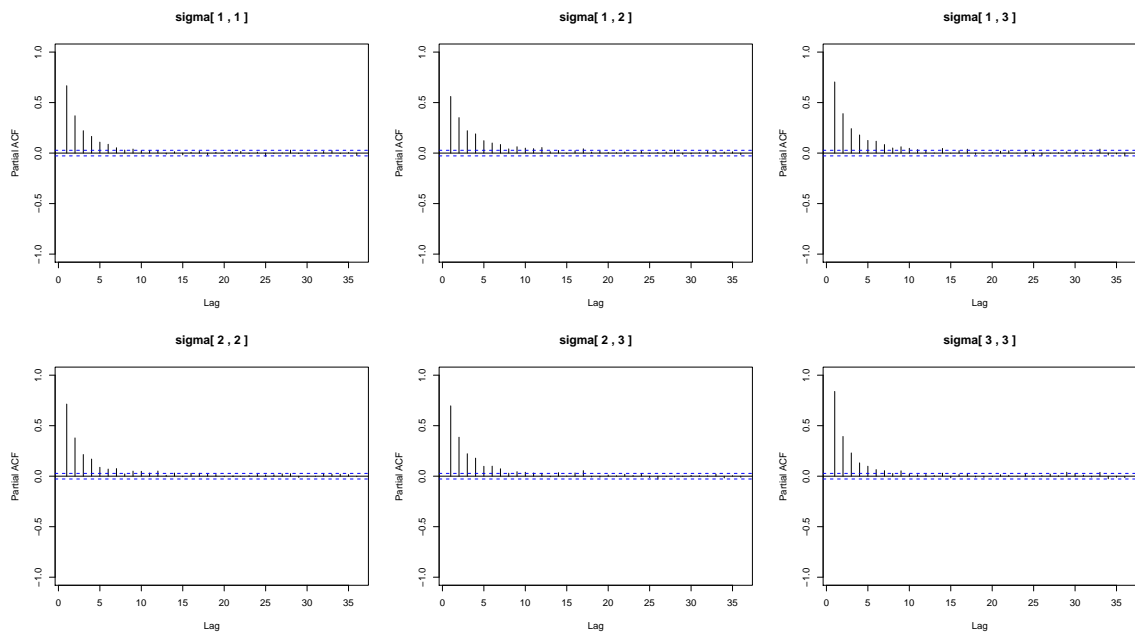


FIGURA B.6. Autocorrelaciones parciales de las cadenas de Markov para las estimaciones de las componentes de  $\Sigma_3$  para los datos simulados.

---



---

## Resultados del modelo G LSMIRT para los datos reales.

---



---

En este Apéndice, se presentan algunos resultados obtenidos por el algoritmo del modelo G-LSMIRT para el conjunto de datos reales.

### C.1. Figuras de las cadenas de $\Sigma_g$ .

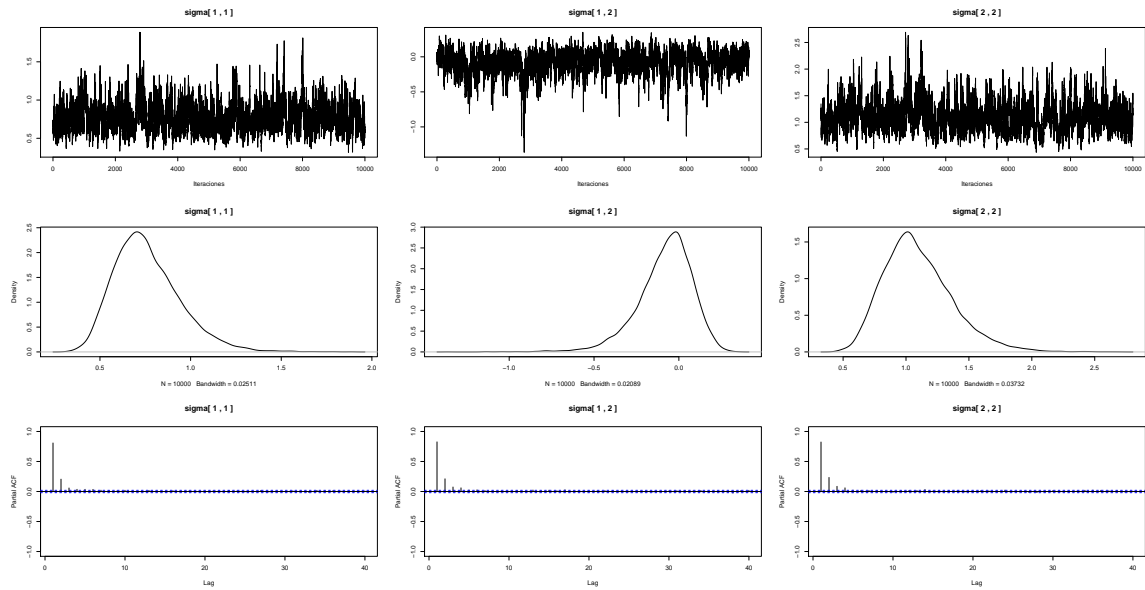


FIGURA C.1. Trazos, distribuciones posteriori y autocorrelaciones parciales de las cadenas de Markov para las estimaciones de las componentes de  $\Sigma_2$  para los datos reales.

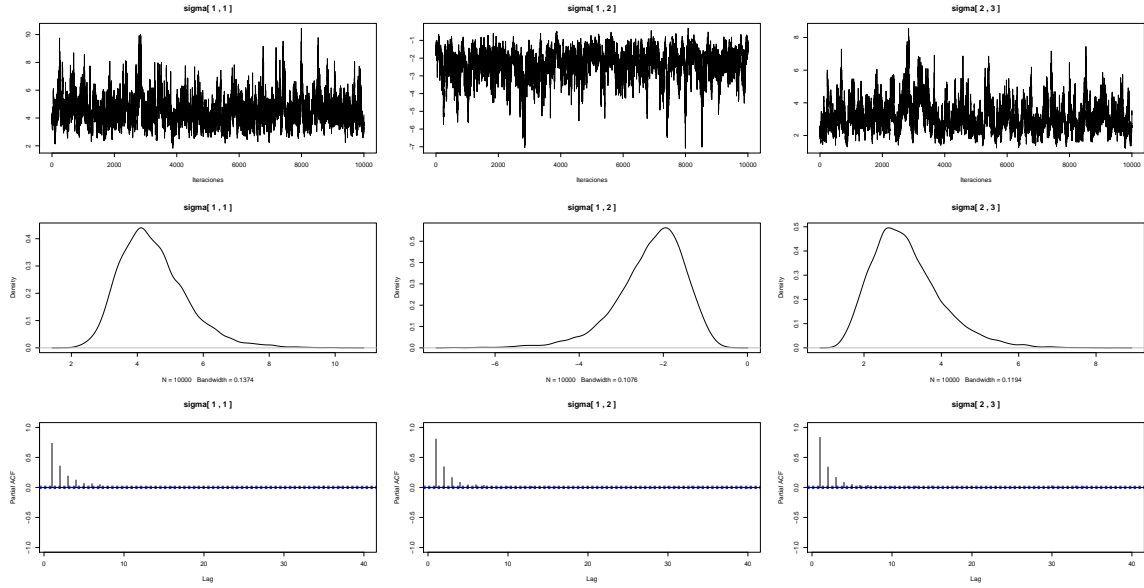


FIGURA C.2. Trazos, distribuciones posteriori y autocorrelaciones parciales de las cadenas de Markov para las estimaciones de las componentes de  $\Sigma_3$  para los datos reales.

### C.2. Algunas figuras de las cadenas de $\alpha$

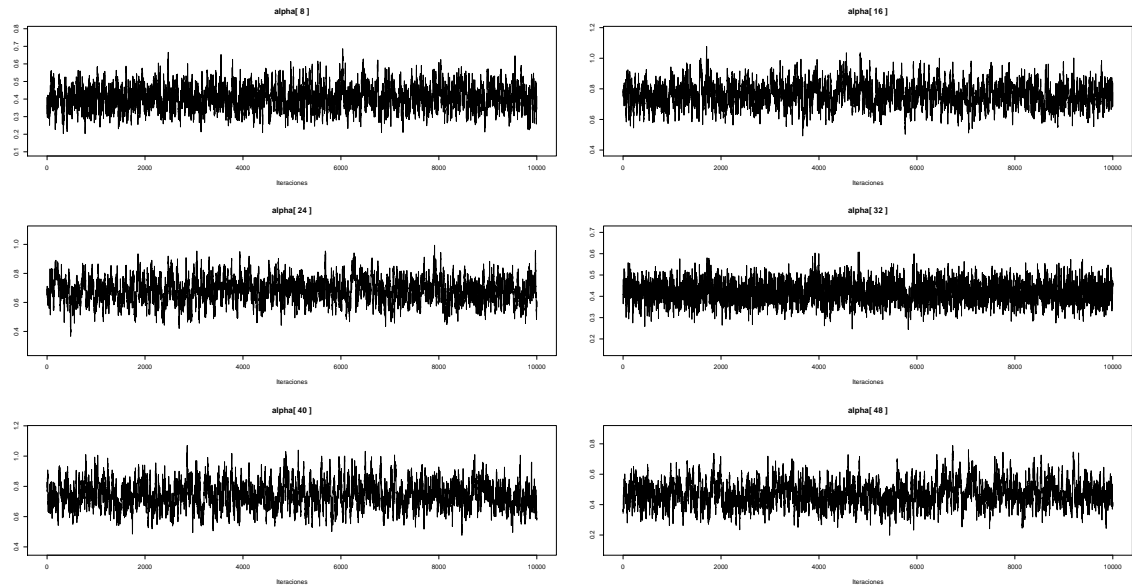


FIGURA C.3. Trazos de algunas de las cadenas de Markov para las estimaciones de  $\alpha$  para los datos reales.



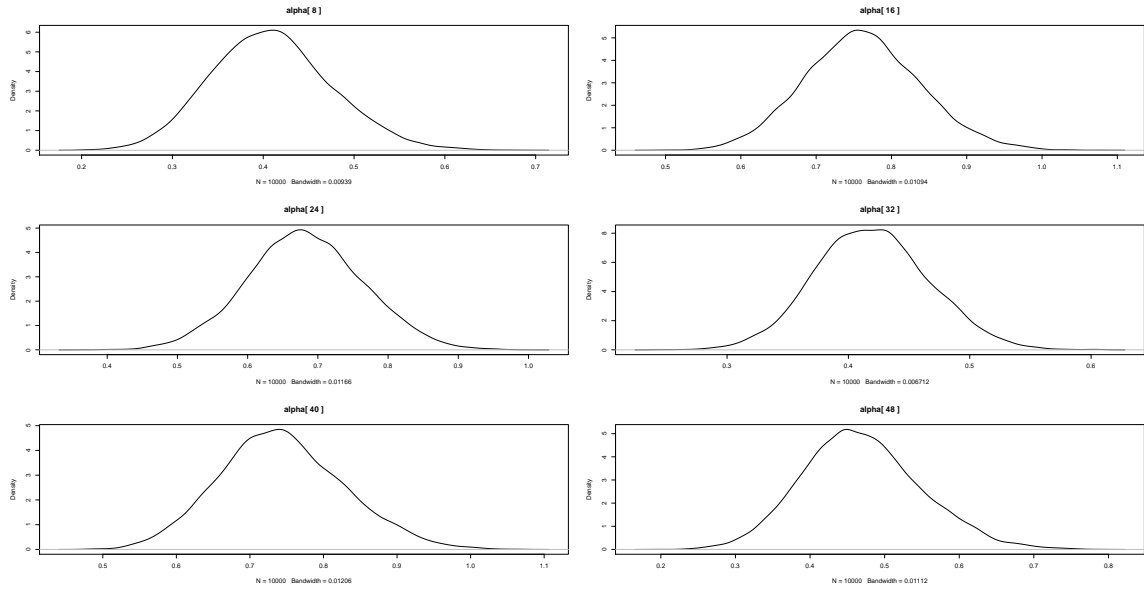


FIGURA C.4. Distribuciones posteriores de algunas de las cadenas de Markov para las estimaciones de  $\alpha$  para los datos reales.

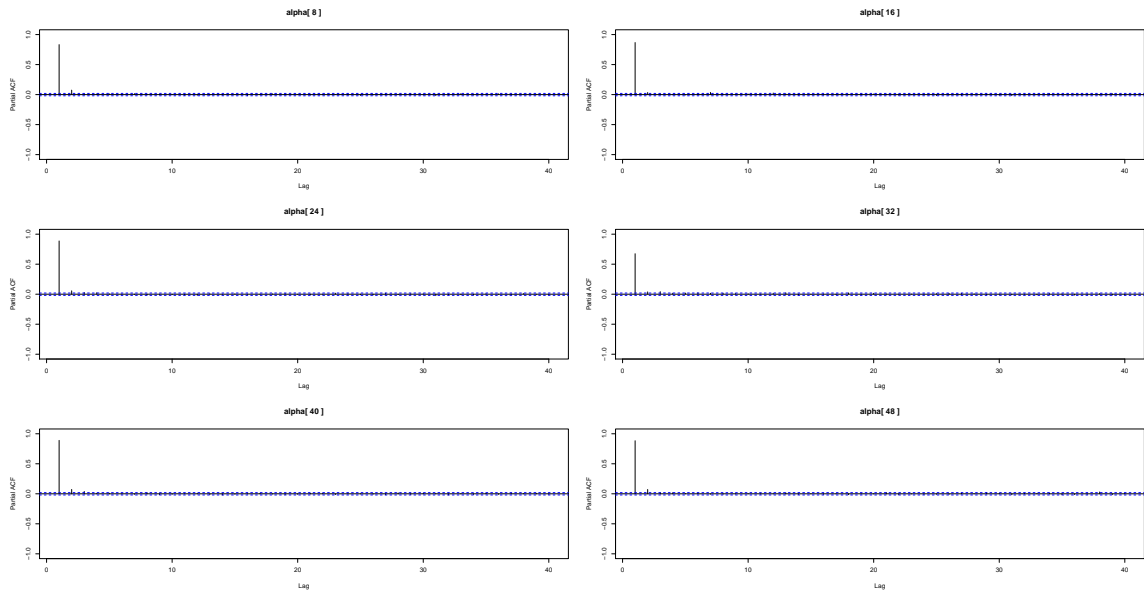


FIGURA C.5. Autocorrelaciones parciales de algunas de las cadenas de Markov para las estimaciones de  $\alpha$  para los datos reales.

### C.3. Parámetros de pendiente e intercepto.

item	$\hat{\alpha}_{vj}$	$sd_{\hat{\alpha}}$	$\hat{\gamma}_{vj}$	$sd_{\hat{\gamma}}$
1	0.76	0.08	-0.29	0.06
2	0.76	0.09	-0.53	0.07
3	0.62	0.06	0.19	0.06
4	0.62	0.10	-1.26	0.08
5	0.49	0.06	-0.01	0.06
6	0.24	0.04	0.55	0.06
7	0.63	0.06	0.97	0.07
8	0.41	0.07	-0.85	0.06
9	0.52	0.05	0.33	0.06
10	0.59	0.08	-0.74	0.07
11	0.50	0.06	-0.26	0.06
12	0.60	0.06	0.06	0.06
13	0.57	0.06	0.08	0.06
14	0.37	0.05	0.33	0.06
15	0.69	0.08	-0.15	0.06
16	0.76	0.08	0.86	0.07
17	0.47	0.05	0.17	0.06
18	0.63	0.07	0.10	0.06
19	0.50	0.05	0.82	0.07
20	0.71	0.07	0.61	0.07
21	0.67	0.07	0.74	0.07
22	0.37	0.05	0.07	0.06
23	0.45	0.05	0.01	0.06
24	0.68	0.08	-0.57	0.06
25	0.52	0.06	-0.09	0.06
26	0.39	0.05	-0.36	0.06
27	0.53	0.06	-0.25	0.06
28	0.42	0.05	0.72	0.06
29	0.62	0.06	0.35	0.06
30	0.68	0.07	-0.11	0.06
31	0.48	0.05	0.48	0.06
32	0.42	0.05	0.38	0.06
33	0.65	0.09	-0.91	0.07
34	0.55	0.06	0.22	0.06
35	0.59	0.06	0.15	0.06
36	0.65	0.09	-0.76	0.07
37	0.63	0.07	0.21	0.06
38	0.57	0.06	0.36	0.06
39	0.77	0.09	-0.43	0.07
40	0.74	0.08	-0.50	0.07
41	0.59	0.06	0.21	0.06
42	0.72	0.08	-0.02	0.06
43	0.43	0.05	0.03	0.06
44	0.56	0.06	0.46	0.06
45	0.49	0.06	0.94	0.07
46	0.46	0.06	-0.36	0.06
47	0.54	0.07	-0.62	0.06
48	0.47	0.08	-1.03	0.07
49	0.78	0.10	-0.49	0.07
50	0.48	0.05	0.10	0.06
51	0.82	0.08	0.10	0.06

TABLA C.1. Las columnas  $\hat{\alpha}_{vj}$  y  $\hat{\gamma}_{vj}$  son las estimaciones de los parámetros de pendiente e intercepto del modelo G-LSMIRT, así como  $sd_{\hat{\alpha}_{vj}}$  y  $sd_{\hat{\gamma}_{vj}}$  son las desviaciones estándar posteriori de las estimativas para los datos reales.

---

---

## Bibliografía

---

---

- Ackerman, T. (1989). Unidimensional irt calibration of compensatory and noncompensatory multidimensional items, *Applied Psychological Measurement* **13**: 113 – 127.
- Albert, J. (1992). Bayesian estimation of normal item response curves using gibbs sampling, *Journal of Educational Statistics* **17**(3): 251 – 269.
- Antonio, D. (2013). *Modelo multidimensional de teoría de respuesta al ítem jerárquico*, Master's thesis, Universidad Nacional de Colombia.
- Baker, F. B. & Seok-Ho, K. (2004). *Item Response Theory*, 2nd edn, Marcel Decker Inc.
- Bazán, J. (2006). A skew item response model, *Bayesian Analysis* **1**(4): 861 – 892.
- Bégin, A. & Glass, C. A. (2001). Mcmc estimation and some model-fit analysis of multidimensional irt models, *Psychometrika* **66**(4): 541 – 562.
- Bock, R. & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: application of an em algorithm, *Psychometrika* **46**: 443 – 459.
- Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories, *Psychometrika* **37**: 29 – 51.
- Carlin, B. (2009). *Bayesian methods for data analysis*, 3 edn, CRC Press: Taylor y Francis Group.
- Da Silva, G. (2008). *Modelos multidimensionais da TRI com distribuições assimétricas para os traços latentes*, PhD thesis, Universidade de São Pablo.
- De la Torre, J. & Patz, R. (2005). Making the most of what we have: a practical application of multidimensional item response theory in test scoring, *Journal of Educational and Behavioral Statistics* **30**(3): 295 – 311.
- Fox, J. (2010). *Bayesian Item Response Modeling. Theory and Methods*, Statistics for Social and Behavior Sciences, Springer.
- Gelman, A. & Rubin, D. (1992). Inference from iterative simulation using multiple sequences, *Statistical Science* **7**(4): 457 – 511.
- Lee, H. (1995). *Markov Chain Monte Carlo Methods for Estimating Multidimensional Ability in Item Response Analysis*, PhD thesis, University of Missouri.

- 
- Montenegro, A. (2011). *Multidimensional Item Response Theory Models where the Ability has a Latent Linear Structure*, PhD thesis, Universidad Nacional de Colombia.
- Montenegro, A. & Cepeda, E. (2010). Synthesizing the ability in multidimensional item response theory models, *Revista Colombiana de Estadística* **33**(1): 127 – 47.
- Nelder, J. & Wedderburn, W. (1972). Generalized linear models, *Journal of the Royal Statistical Society* **135**(3): 370 – 384.
- Ntzoufras, I. (2009). *Bayesian Modelling Using WINBUGS*, John Wiley and Sons, Inc.
- R Core Team (2014). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.  
**URL:** <http://www.R-project.org/>
- Reckase, M. (1985). The difficulty of test items that measure more than one ability, *Applied Psychological Measurement* **9**(9): 401 – 412.
- Reckase, M. (1997). The past and the future of multidimensional item response theory, *Applied Psychological Measurement* **21**(1): 25 – 36.
- Reckase, M. (2007). Multidimensional item response theory, *Handbook of Statistics* **26**: 607 – 642.
- Reckase, M. (2009). *Multidimensional Item Response Theory*, Springer.
- Sheng, Y. (2007). Comparing multiunidimensional and unidimensional item response theory models, *Educational and Psychological Measurement* **67**(6): 899 – 919.
- Sheng, Y. (2008). *Bayesian IRT Models with General and Specific Traits*, VDM Verlag Dr. Müller.
- Sheng, Y. (2010). Bayesian estimation of mirt models with general and specific latent traits in matlab, *Journal of Statistical Software* **34**(3).
- Wang, M. (1986). Fitting a unidimensional model to multidimensional item response data, *Paper presented at the Oce of Naval Research Contractors Meeting*, Garlingburg, TN.
- Zhang, J. & Stout, W. (1999). Conditional covariance structure of generalized compensatory multidimensional items, *Psychometrika* **64**: 129 – 152.