

*Estimación Bayesiana de un Modelo TRI Logístico en el
que el Trazo Sigue una Distribución Normal Truncada*

JUAN PABLO MOJICA MACÍAS
MATEMÁTICO



UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE CIENCIAS
DEPARTAMENTO DE ESTADÍSTICA
BOGOTÁ, D.C.
OCTUBRE DE 2017

*Estimación Bayesiana de un Modelo TRI Logístico en el
que el Trazo Sigue una Distribución Normal Truncada*

JUAN PABLO MOJICA MACÍAS
MATEMÁTICO

DISERTACIÓN PRESENTADA PARA OPTAR AL TÍTULO DE
Ms.C. ESTADÍSTICA

DIRECTOR
ÁLVARO MAURICIO MONTENEGRO DÍAZ, Ph.D.
DOCTOR EN ESTADÍSTICA

LÍNEA DE INVESTIGACIÓN
TEORÍA DE RESPUESTA AL ÍTEM



UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE CIENCIAS
DEPARTAMENTO DE ESTADÍSTICA
BOGOTÁ, D.C.
OCTUBRE DE 2017

Título en español

Estimación Bayesiana de un Modelo TRI Logístico en el que el Trazo Sigue una Distribución Normal Truncada.

Title in English

Bayesian Estimation of a Logistic IRT Model in which the Trait follows a Truncated Normal Distribution.

Resumen: En el marco de los modelos logísticos de la TRI, los procesos de estimación de parámetros desde el enfoque clásico o Bayesiano, suponen que el trazo presenta un comportamiento simétrico, de hecho se asume que el sigue una distribución normal o Log-normal. Un método de estimación Bayesiana es propuesto en este trabajo, en el que se supone que el comportamiento del trazo latente y los scores del test no es simétrico. Este algoritmo establece que la distribución del trazo latente θ es Normal Trucada (*NT*). Los resultados obtenidos en la simulación y en los datos reales muestran que el algoritmo está recuperando y estimando de manera efectiva tanto los parámetros de los ítems como el trazo latente, es decir, ajusta adecuadamente los datos. La aplicación a datos reales se realizó en el constructo Sinceridad del Inventario PIHEMA R-2 presentado por los aspirantes a la Facultad de Humanidades y Educación de la Universidad Central de Venezuela, en el año 2006.

Abstract: In the framework of the IRT logistic models, the processes of estimating parameters from the classical or Bayesian approach, assume that the trait has a symmetrical behavior, in fact it's assumed that follows a normal or Log-normal distribution. A Bayesian estimation method is proposed here, we assumed that the behavior of the latent trait and the test scores is not symmetrical. This algorithm states that the distribution of the latent trait θ is Truncated Normal (*NT*). The results obtained in the simulation and real data show that the algorithm is effectively recovering and estimating both the parameters of the items as the latent trait, that is, it adequately adjusts the data. The application to real data was made in the construct Sincerity of the Inventory PIHEMA R-2 presented by the applicants to the Faculty of Humanities and Education of the Central University of Venezuela, in 2006.

Palabras clave: Modelo 2PL, algoritmo Metrópolis-Hastings (M-H), bondad y ajuste, distribución Normal Truncada, estimación Bayesiana.

Keywords: 2PL Model, Metropolis-Hastings algorithm (M-H), GoF, Truncated Normal distribution, Bayesian estimation.

Dedicado a

Leidy Natalia León, mis padres Ana Macías y Pablo Mojica por su apoyo incondicional en la realización de este nuevo logro.

Agradecimientos

Una nueva etapa en mi vida académica culmina, es el momento de agradecer a todas las personas que fueron parte fundamental para alcanzar este objetivo. A pesar de las dificultades siempre he contado con el apoyo de mis padres, novia, amigos y maestros.

Después de tanto esfuerzo, ser Magister en Estadística es ahora una realidad. Agradezco a mis padres Ana Macías y Pablo Mojica por su apoyo, motivación, amor y consejos, que me han traído hasta aquí y también a aquellos maestros que hicieron parte de este camino.

Finalmente, agradezco al profesor Álvaro Mauricio Montenegro por su colaboración para el desarrollo de esta tesis y al grupo de investigación **SICS Research Group** del Departamento de Estadística en el marco del proyecto “*Multidimensional item response theory models for practical application in large tests designed to measure multiple constructs*”, código Colciencias 110156933802.

Índice general

Índice general	I
Índice de tablas	III
Índice de figuras	V
Introducción	VII
1. Marco Teórico	1
1.1. Teoría de Respuesta al Ítem	1
1.1.1. Trazos Latentes en la TRI	2
1.2. Modelos de la TRI	6
1.2.1. Modelo Logístico de Un Parámetro (Modelo de Rasch)	6
1.2.2. Modelo Logístico de Dos Parámetros (Birnbaum o Modelo 2PL)	7
1.2.3. Modelo Logístico de Tres Parámetros (Modelo 3PL)	7
1.3. Estadística Bayesiana	8
1.3.1. Teorema de Bayes	9
1.4. Métodos de Muestreo	10
1.4.1. Muestreo de Gibbs	10
1.4.2. Metrópolis–Hastings	11
1.5. Estimación Bayesiana en TRI	12
2. Modelo 2PL Donde el Trazo Latente Sigue una Distribución Normal Truncada	14
2.1. Modelo Usual de Estimación Bayesiana en TRI	14
2.2. Modelo Propuesto	16
2.3. Método de Estimación	16

2.4. Pruebas de Bondad y Ajuste	18
2.4.1. Test de Heidelberger–Welch	19
3. Simulación	20
3.1. Generación del Test	20
3.2. Recuperación de Parámetros	21
3.3. Comparación entre los valores simulados y recuperados	25
3.4. Pruebas de Bondad y Ajuste	27
4. Aplicación a Datos Reales	29
4.1. Descripción de Datos Reales	29
4.2. Selección del Constructo y Análisis Preliminar	30
4.3. Análisis Preliminar	31
4.4. Estimación de Parámetros	32
4.5. Pruebas de Bondad y Ajuste	35
A. Implementación código en R	37
A.1. Generación de valores aleatorios con Distribución Normal Truncada	37
A.2. Algoritmo de Metropolis–Hasting	39
A.3. Generación de Cadenas	42
A.4. Estimación Conjunta de Parámetros y GoF	42
B. Análisis de Dimensionalidad Datos Reales	46
C. Simulación Otro Test	48
C.1. Generación del Test	48
C.2. Recuperación de Parámetros	49
C.3. Comparación entre los valores simulados y recuperados	51
C.4. Pruebas de Bondad y Ajuste	54
Discusión y Conclusiones	56
Trabajo futuro	58
Bibliografía	59

Índice de tablas

3.1. Medidas Descriptivas de los parámetros a_j , b_j y θ_i Simulados	21
3.2. Valores del trazo estimado y el score estándar para los primeros 10 registros del test	22
3.3. Estadísticas descriptivas para el Error Absoluto del parámetro de dificultad	25
3.4. Estadísticas descriptivas para el Error Absoluto del parámetro de discriminación	25
3.5. Estadísticas descriptivas para el Error Absoluto del Trazo	26
3.6. Comparación entre los valores simulados y recuperados de los parámetros de los ítems	27
3.7. Estadísticas descriptivas para el p -valor de Heidelbergger–Welch para el parámetro de dificultad	28
3.8. Estadísticas descriptivas para el p -valor de Heidelbergger–Welch para el parámetro de discriminación	28
3.9. Estadísticas descriptivas Bondad y Ajuste de los parámetros de los ítems recuperados	28
4.1. Resultados del Análisis Paralelo de Horn para Sinceridad	31
4.2. Valores del trazo estimado y el score estándar para los primeros 10 registros del test	33
4.3. Valores estimados de los parámetros de los ítems	34
4.4. Estadísticas descriptivas para el p -valor de Heidelbergger–Welch para el parámetro de dificultad	36
4.5. Estadísticas descriptivas para el p -valor de Heidelbergger–Welch para el parámetro de discriminación	36
4.6. Estadísticas descriptivas Bondad y Ajuste de los parámetros de los ítems estimados	36
4.7. Estadísticas descriptivas Bondad y Ajuste para el trazo estimado	36
B.1. Resultados del Análisis Paralelo de Horn	46

C.1. Medidas Descriptivas de los parámetros a_j , b_j y θ_i Simulados	49
C.2. Valores del trazo estimado y el score estándar para los primeros 10 registros del test	50
C.3. Estadísticas descriptivas para el Error Absoluto del parámetro de dificultad	52
C.4. Estadísticas descriptivas para el Error Absoluto del parámetro de discriminación	52
C.5. Estadísticas descriptivas para el Error Absoluto del Trazo	53
C.6. Comparación entre los valores simulados y recuperados de los parámetros de los ítems	54
C.7. Estadísticas descriptivas para el p -valor de Heidelbergger–Welch para el parámetro de dificultad	54
C.8. Estadísticas descriptivas para el p -valor de Heidelbergger–Welch para el parámetro de discriminación	55
C.9. Estadísticas descriptivas Bondad y Ajuste de los parámetros de los ítems recuperados	55

Índice de figuras

1.1. Curva Característica de un Ítem	2
1.2. Interpretación del Parámetro de Dificultad	4
1.3. (a) CCI de un ítem bajo los tres modelos logísticos. (b) ICC del ítem j en un modelo 2PL para $b_j = 0.3$, $a_j = 0, 0.5, 1, 2$ y (c) CCI del ítem j en un modelo 2PL para $a_j = 1$, $b_j = -3, -0.5, 0, 2$	7
1.4. (a) ICC con $b_j = 0.3, c_j = 0.1$ y $a_j = 0, 0.5, 1, 2$. (b) ICC con $a_j = 2, c_j = 0.1$ y $b_j = -3, -0.5, 0, 2$. (d) ICC con $a_j = 2, b_j = 0.3$ y $c_j = 0.1, 0.2, 0.3, 0.4$. . .	8
3.1. Función de Densidad $NT(\mu = 0, \sigma = 1, l = 0, u = \infty)$	21
3.2. Cadenas para los parámetros de los ítems y el trazo latente	22
3.3. Gráficos Cadenas, densidad y autocorrelograma para el trazo latente θ	23
3.4. Gráficos Cadenas, densidad y autocorrelograma para el parámetro de dificultad	24
3.5. Gráficos Cadenas, densidad y autocorrelograma para el parámetro de discriminación	24
3.6. Comparación entre los valores simulados y recuperados para el parámetro de dificultad	25
3.7. Comparación entre los valores simulados y estimados para el parámetro de discriminación	26
3.8. Comparación entre los valores simulados y recuperados del Trazo	26
4.1. Histograma de los Scores por Constructo	30
4.2. Análisis Paralelo de Horn para Sinceridad	31
4.3. (a) Diagrama de Barras Autovalores Datos Sinceridad. (b) Curva de Cronbach–Mesbah Datos Sinceridad	32
4.4. Cadenas para los parámetros de los ítems y el trazo latente	33
4.5. Trazo Estimado Vs. Score Estándar	33
4.6. Gráficos cadenas, densidades y autocorrelogramas para el trazo estimado . .	34

4.7. Gráficos bayesianos para el parámetro de dificultad	35
4.8. Gráficos bayesianos para el parámetro de discriminación	35
B.1. Análisis Paralelo de Horn	47
B.2. (a) Diagrama de Barras Autovalores Datos Simulados. (b) Curva de Cronbach–Mesbah Datos Simulados	47
C.1. Cadenas para los parámetros de los ítems y el trazo latente	49
C.2. Gráficos Cadenas, densidad y autocorrelograma para el trazo latente θ	50
C.3. Gráficos Cadenas, densidad y autocorrelograma para el parámetro de dificultad	51
C.4. Gráficos Cadenas, densidad y autocorrelograma para el parámetro de discriminación	51
C.5. Comparación entre los valores simulados y recuperados para el parámetro de dificultad	52
C.6. Comparación entre los valores simulados y estimados para el parámetro de discriminación	53
C.7. Comparación entre los valores simulados y recuperados del Trazo	53

Introducción

En la Teoría de Respuesta al Ítem (TRI), se asume usualmente que los trazos latentes de los individuos en un test dicotómico siguen una distribución simétrica, sin embargo, como resultado de diversas investigaciones se ha logrado establecer que el comportamiento de estos trazos no siempre sigue esa clase de distribuciones. Dentro de la TRI es frecuente encontrar datos en los que el score de los individuos presenta un comportamiento asimétrico, como por ejemplo en pruebas donde se miden rasgos de personalidad. Por esta razón la TRI busca implementar herramientas que permitan el análisis y la estimación de parámetros para esta clase de datos. Recientemente, la investigación sobre esta clase de modelos ha tenido un desarrollo importante y sus implicaciones son cada vez más diversas, por ejemplo en ciencias humanas, en psicometría, entre otros.

Aunque en la mayoría de investigaciones realizadas con modelos TRI en rasgos de personalidad son comúnmente explicados con los modelos logísticos y asumiendo distribuciones simétricas en sus parámetros, se pretende implementar una metodología Bayesiana para la estimación de los mismos suponiendo que siguen una distribución asimétrica.

Los métodos de estimación MCMC permiten la estimación de manera más eficaz y aproximada, entre ellos el algoritmo de Metrópolis–Hastings (M–H) y el muestreo de Gibbs, representan una ventaja dado que la distribución propuesta para los parámetros puede conducir a la simulación de los valores de la distribución posterior. Sin embargo, una de las limitaciones de la metodología surge al seleccionar la distribución propuesta, ya que puede conducir a problemas de convergencia. Tanto en la teoría como en las aplicaciones es común asumir que las distribuciones propuestas son simétricas, como es el caso de la Normal y la distribución Logística. Fox (2010) define un esquema de datos unidimensionales que siguen un modelo de respuesta logístico a través de un método de muestreo M–H.

Las pruebas Inventario PIHEMA fueron realizadas por la Universidad Central de Venezuela en el año 2006 para las admisiones de la Facultad de Humanidades y Educación, con la finalidad de evaluar características relacionadas con el comportamiento de los individuos frente a diversas situaciones (neuroticismo, psicopatía, paranoia, etc.), rasgos que como se afirma en Woods (2008) y Páyer (2013), no siguen una distribución Normal Estándar. Es sobre este conjunto de datos que se implementará la metodología de estimación propuesta en este trabajo.

El objetivo de este documento es presentar los resultados correspondientes a la Estimación Bayesiana y sus implicaciones teóricas y prácticas. Se propone un algoritmo Bayesiano usando un modelo logístico 2PL, en el que la distribución a priori del trazo θ_i no es simétri-

ca. Específicamente se asumirá que la distribución del trazo es Normal Truncada (NT), teniendo en cuenta que el comportamiento del score de los individuos es asimétrico.

En el primer Capítulo del documento se presentarán de manera general los principales conceptos y supuestos de la TRI, los modelos logísticos para test dicotómicos, así como las distintas metodologías de estimación. En el segundo Capítulo se especificará el modelo Logístico de dos parámetros en el que el trazo sigue una distribución Normal Truncada y se describirá el procedimiento de estimación Bayesiana para los parámetros de los ítems y el trazo latente. Posteriormente, se realizará una simulación para evaluar la calidad de recuperación de los parámetros con la metodología propuesta en el Capítulo 3. Finalmente, a lo largo del Capítulo 4 se realizará el ajuste del modelo propuesto en un conjunto de datos reales

CAPÍTULO 1

Marco Teórico

1.1. Teoría de Respuesta al Ítem

La Teoría de Respuesta al Ítem (TRI) se define como una área estadística que relaciona el rendimiento de un respondiente con el trazo latente y las características de los ítems del test (Hambleton & Jones (1993)). En la TRI pueden obtenerse rasgos de los ítems que no dependen del grupo de individuos y mediciones de trazos latentes que no dependen de los test.

Los modelos de TRI a considerar en este documento son aquellos en los que se asume que un solo trazo latente subyace al test, son aplicables a datos dicotómicos y establecen una relación matemática entre el ítem y el trazo latente (función logística de uno, dos o tres parámetros). Los principales supuestos de la TRI son

- Unidimensionalidad el test, hace referencia a que los ítems del test aportan información sobre un único trazo latente.
- Independencia local o condicional, las respuestas de un individuo a ítems diferentes en un test son estadísticamente independientes (Hambleton & Swaminathan (1985)).
- Independencia entre respondientes, las respuestas de individuos distintos al test son independientes.

Por medio de la función logística se establece una relación entre el ítem (que es una característica observable) y el trazo latente θ (característica no observable), de esta relación puede obtenerse la probabilidad $P(\theta)$ de responder correctamente a un ítem en función del trazo latente cuya representación gráfica se conoce como *curva característica* del ítem y se presenta en la Figura 1.1.

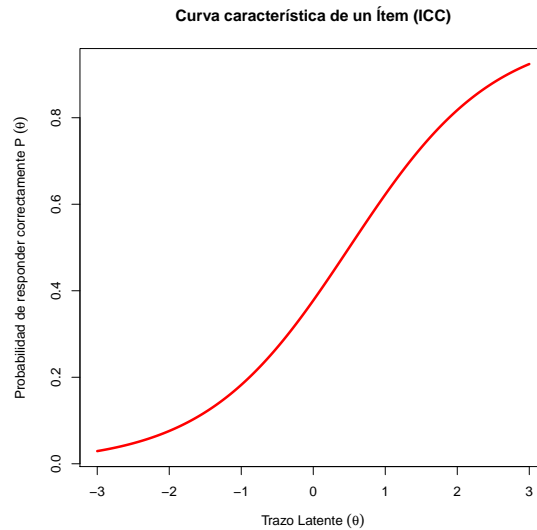


FIGURA 1.1. Curva Característica de un Ítem

La curva característica de un ítem está determinada por los parámetros empleados en la función logística y caracterizan al ítem mismo en términos de su dificultad y capacidad de discriminación. En apartados siguientes se especificarán las principales características de los modelos de la TRI y de los parámetros que en ellos intervienen. La función de información también varía de acuerdo al nivel del trazo latente, por lo que puede determinarse el error estándar de las estimaciones.

1.1.1. Trazos Latentes en la TRI

En la TRI se pretende medir o predecir el rendimiento de un respondiente a un test por un conjunto de *trazos*, *trazos latentes* o *habilidades* que no puede ser medidos directamente. Hambleton et al. (1991) y Hambleton & Swaminathan (1985) presentan una relación entre la respuesta de un individuo a un ítem y el conjunto de trazos latentes, es decir que el comportamiento del individuo frente a un ítem puede ser establecido. Como lo presenta Attorresi et al. (2009), uno de los objetivos de la TRI es construir test como instrumentos de medición que no se vean afectados por las características de las poblaciones, es decir, dos individuos con trazos similares tendrán la misma probabilidad acertar al ítem sin importar la población de la que provengan.

Si el trazo latente es más alto se espera que la probabilidad de dar una respuesta correcta sea también mayor y si el trazo latente es bajo la probabilidad de acertar al ítem será a su vez más baja, este comportamiento se describe gráficamente por una función monótona creciente llamada *curva característica del ítem*, (CCI).

Como se mencionó anteriormente, la probabilidad de responder a un ítem correctamente depende de la habilidad o habilidades del individuo y las características de los ítems. La posición de los respondientes en el r -ésimo trazo se denota por θ_r , si se consideran varios trazos latentes se dirá que el espacio latente es k -dimensional y se representará por un vector de scores de trazos $(\theta_1, \dots, \theta_k)$ llamado *espacio latente completo*. Entonces la relación entre trazos e ítems tiene una forma matemática específica llamada *modelo*, que

son clasificados según la cantidad de trazos latentes observados al momento de responder a un ítem.

Por lo general, se supone que la respuesta para un ítem está en función de un único trazo latente y esto es conocido como un modelo *unidimensional*, estableciendo además una independencia entre las respuestas de cada ítem conociendo el trazo θ , es decir que la respuesta a un ítem no depende de la respuesta a cualquier otro, este supuesto es conocido como independencia local. Como consecuencia, se establece en Hambleton et al. (1991) que:

- θ_r es el espacio latente, $r = 1, 2, \dots, k$, si $r = 1$ tenemos un espacio latente unidimensional.
- Y_{ij} la respuesta del i -ésimo respondiente al ítem j . $i = 1, 2, \dots, N$ y $j = 1, \dots, p$
- $P(Y_{ij} = 1|\theta_i)$ probabilidad de responder correctamente al ítem j . Por lo tanto $P(Y_{ij} = 0|\theta_i) = 1 - P(Y_{ij} = 1|\theta_i)$

Luego, para un individuo se tiene que

$$P(Y_{11}, \dots, Y_{1p}|\theta_i) = P(Y_{11}|\theta_i) \cdots P(Y_{1p}|\theta_i) = \prod_{j=1}^p P(Y_{1j}|\theta_i) \quad (1.1)$$

Por ejemplo, supongamos que las respuestas de un test están dadas por,

$$Y_{11} = 1, Y_{12} = 1, Y_{13} = 0, Y_{14} = 0,$$

$$\begin{aligned} P(Y_{11} = 1, Y_{12} = 1, Y_{13} = 0, Y_{14} = 0|\theta_i) &= P(Y_{11}|\theta_i)P(Y_{12}|\theta_i)P(Y_{13}|\theta_i)P(Y_{14}|\theta_i) \\ &= P_1P_2Q_3Q_4 \end{aligned} \quad (1.2)$$

Donde $P_j = P(Y_{ij} = 1|\theta_i)$ y $Q_j = 1 - P(Y_{ij} = 1|\theta_i)$, con $j = 1, \dots, p$.

Como se mencionó anteriormente la CCI representa la probabilidad de responder correctamente a un ítem en función de los trazos latentes θ . Hambleton & Swaminathan (1985) establece que al seleccionar un modelo para ajustar los datos las mismas curvas características no dependen de la distribución del trazo latente de los respondientes usada para estimar los parámetros de los ítems (dificultad, discriminación y adivinación), estos parámetros dependen del modelo en particular (*Rasch, Birnbaum y 3PL*).

Hambleton et al. (1991) y Attorresi et al. (2009) muestran que el parámetro de dificultad b_j para el ítem j , es el punto sobre el eje del trazo latente θ , donde la probabilidad de responder correctamente es 0.5, es decir el mayor trazo latente que requiere un individuo para tener el 50% de probabilidad de responder correctamente un ítem. Por esta razón, también se puede ver como un parámetro de localización, indicando la posición de la ICC en relación a la escala del trazo latente, por lo que, si un ítem i tiene mayor valor en b que el ítem j se requiere un valor mayor en el trazo latente. Los ítems con mayor dificultad son ubicados al lado derecho de la escala de habilidad mientras que los ítems fáciles serán ubicados al lado izquierdo de esta escala. Algunas ICC para un modelo de un parámetro son mostradas (Ver Figura 1.2).

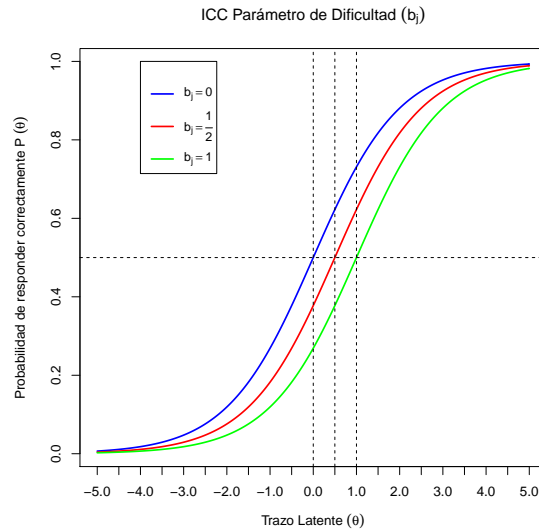


FIGURA 1.2. Interpretación del Parámetro de Dificultad

El parámetro a_j es conocido como el parámetro de discriminación para el ítem j . Este parámetro se puede interpretar como la pendiente de la curva característica en el punto b_j , la funcionalidad de este parámetro es separar los respondientes en diferentes niveles del trazo latente en el punto de corte con la escala del trazo latente, es decir, discrimina respondientes con valor de trazo latente menor igual que θ y respondientes con valor de trazo latente mayor que θ . Por lo que se supone que para los valores menores de θ el respondiente tendrá menor habilidad para responder correctamente al ítem y viceversa, para lo valores mayores de θ el respondiente tendrá mayor habilidad para responder correctamente el ítem. Cabe resaltar que entre mayor sea el valor de a_j mejor es la discriminación del ítem.

Finalmente, Attorresi et al. (2009) presenta el parámetro de adivinación c_j para el ítem j , como el valor de probabilidad en los valores más bajos del trazo latente, es decir, que los respondientes con trazo latente muy bajo reflejan una probabilidad de responder correctamente al ítem, posiblemente por “suerte”.

Para la estimación de los trazos latentes bajo el supuesto de unidimensionalidad e independencia local se considera las siguientes afirmaciones. Sea Y_j la respuesta de un individuo al ítem $j = 1, 2, \dots, p$, donde

$$Y_j = \begin{cases} 1, & \text{si la respuesta es correcta} \\ 0, & \text{si la respuesta es incorrecta} \end{cases}$$

Entonces

- θ espacio latente unidimensional.
- $P(Y_j|\theta)$ es la probabilidad de que un respondiente de respuesta al ítem j dado θ
- $P_j = P(Y_j = 1|\theta)$ es la probabilidad de respuesta correcta.
- $Q_j = P(Y_j = 0|\theta) = 1 - P_j$ es la probabilidad de respuesta incorrecta.

Note que, la variable Y_j tiene forma de una variable de Bernoulli, por lo que la probabilidad de respuesta se puede ver como

$$P(Y_j|\theta) = P(Y_j = 1|\theta)^{Y_j} P(Y_j = 0|\theta)^{1-Y_j} = P_j^{Y_j} (1 - P_j)^{1-Y_j} = P_j^{Y_j} Q_j^{1-Y_j} \quad (1.3)$$

Luego, por la independencia local si un individuo responde a un test con p ítems, la probabilidad conjunta de las Y_j respuestas con $j = 1, \dots, p$ se puede ver como

$$\begin{aligned} P(Y_1, \dots, Y_p|\theta) &= P(Y_1|\theta) \cdots P(Y_p|\theta) = \prod_{j=1}^p P(Y_j|\theta) \\ &= \prod_{j=1}^p P_j^{Y_j} (1 - P_j)^{1-Y_j} \\ &= \prod_{j=1}^p P_j^{Y_j} Q_j^{1-Y_j} \end{aligned} \quad (1.4)$$

Hambleton & Swaminathan (1985) muestra que cuando la variables aleatorias Y_j toman un valor específico y_j , con $y_j = 1$ ó $y_j = 0$, la expresión anterior deja de ser una probabilidad, para ser una función donde θ es conocido, más específicamente una *función de verosimilitud*, es decir

$$L(y_1, \dots, y_p|\theta) = \prod_{j=1}^p P_j^{y_j} Q_j^{1-y_j} \quad (1.5)$$

Entonces el valor de θ que maximiza la función de verosimilitud, puede ser tomado como el estimador de θ , o más conocido como el estimador de máxima verosimilitud, denotado por $\hat{\theta}$.

Además, el logaritmo de la función máximo verosímil (log-Verosimilitud) está dada por

$$\begin{aligned} \ell(y_j|\theta) &= \ln L(y_1, \dots, y_p|\theta) = \ln \left(\prod_{j=1}^p P_j^{Y_j} Q_j^{1-Y_j} \right) \\ &= \sum_{j=1}^p (y_j \ln P_j + (1 - y_j) \ln Q_j) \end{aligned} \quad (1.6)$$

vale la pena resaltar que tanto P_j como Q_j dependen de θ como se observa en la expresión (1.3). Entonces el estimador máximo verosímil de θ se obtiene al resolver la siguiente ecuación

$$\frac{d}{d\theta} \ell(y_j|\theta) = 0.$$

Ahora supongamos que un grupo de N respondientes es examinado con un test de p ítems y además con los parámetros de los ítems conocidos, luego el proceso común en TRI para

la estimación del trazo latente θ es la función de verosimilitud para los N individuos.

$$L(y_{ij}|\theta_i) = \prod_{i=1}^N \prod_{j=1}^p P_{ij}^{y_{ij}} (1 - P_{ij})^{1-y_{ij}} \quad (1.7)$$

Sea $Q_{ij} = 1 - P_{ij}$, entonces la log-verosimilitud se puede ver como:

$$\ln L(y_{ij}|\theta_i) = \ell(y_{ij}|\theta_i) = \sum_{i=1}^N \sum_{j=1}^p [y_{ij} \ln(P_{ij}) + (1 - y_{ij}) \ln(Q_{ij})] \quad (1.8)$$

Entonces, la derivada parcial $\left(\frac{\partial \ell}{\partial \theta_i}\right)$, queda determinada como sigue

$$\frac{\partial \ell}{\partial \theta_i} = \sum_{j=1}^p \left[\frac{y_{ij} - P_{ij}}{P_{ij} Q_{ij}} \right] \frac{\partial P_{ij}}{\partial \theta_i} \quad (1.9)$$

con $i = 1, \dots, N$.

Cabe resaltar que el estimador máximo verosímil $\hat{\theta}$ es asintóticamente normal con media θ y varianza $[I(\theta)]^{-1}$, donde $I(\theta)$ es la matriz de información de θ .

1.2. Modelos de la TRI

Los modelos logísticos son utilizados en esta teoría así como los supuestos que fundamentan la TRI con relación a los parámetros a lo que se ha hecho alusión. Formalmente, los modelos logísticos de uno, dos y tres parámetros se basan en la función de distribución acumulativa logística que está determinada por

$$P(\eta) = \frac{e^\eta}{1 + e^\eta} = \frac{1}{1 + e^{-\eta}} \quad (1.10)$$

Ya se había establecido la importancia de este tipo de modelos, pues permiten relacionar el trazo latente con las características de los ítems. Así pues, se considera un grupo de N respondientes cada uno con un trazo latente θ_i ($i = 1, \dots, N$), a los que se les aplica un test con p ítems, bajo los supuestos de unidimensionalidad e independencia local, el interés recae en determinar la probabilidad de que el individuo i responda acertadamente al ítem j bajo el trazo latente correspondiente.

1.2.1. Modelo Logístico de Un Parámetro (Modelo de Rasch)

También conocido como modelo de Rasch¹ o modelo 1PL, está determinado por

$$P_{ij}(\theta_i) = \frac{1}{1 + e^{-(\theta_i - b_j)}} \quad (1.11)$$

donde b_j ($j = 1, \dots, p$) es el parámetro de dificultad que corresponde al punto sobre la escala del trazo latente en el que el respondiente tiene un 50% de probabilidad de responder

¹El modelo fue propuesto por el matemático danés Georg Rasch en 1960.

correctamente al ítem j ; en este modelo se asume que la discriminación para los j ítems es $a_j = 1$, en la Figura 1.2 se habían presentado algunas gráficas de curvas características variando el valor de b_j .

1.2.2. Modelo Logístico de Dos Parámetros (Birnbbaum o Modelo 2PL)

El modelo logístico 2PL, en este modelo, las curvas características de los j ítems del test están dadas por

$$P_{ij}(\theta_i) = \frac{1}{1 + e^{-a_j(\theta_i - b_j)}} = \frac{1}{1 + e^{-a_j\theta_i + d_j}} \quad (1.12)$$

donde a_j es el parámetro de discriminación del ítem j y b_j el parámetro de dificultad del mismo.

En la Figura 1.3(b) y 1.3(c) se presentan las curvas características del ítem j para variaciones de los parámetros del modelo.

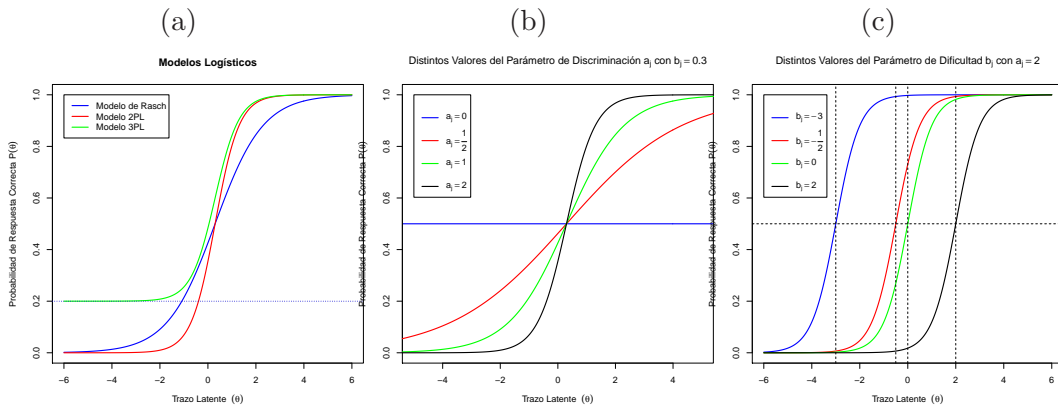


FIGURA 1.3. (a) CCI de un ítem bajo los tres modelos logísticos. (b) ICC del ítem j en un modelo 2PL para $b_j = 0.3$, $a_j = 0, 0.5, 1, 2$ y (c) CCI del ítem j en un modelo 2PL para $a_j = 1$, $b_j = -3, -0.5, 0, 2$

1.2.3. Modelo Logístico de Tres Parámetros (Modelo 3PL)

El modelo de tres parámetros o modelo 3PL, incluye un tercer parámetro para la caracterización del rendimiento de los ítems y está determinado por

$$P_{ij}(\theta_i) = c_j + \frac{(1 - c_j)}{1 + e^{-a_j(\theta_i - b_j)}}$$

Este tercer parámetro c_j ($j = 1, \dots, p$), constituye una asíntota horizontal inferior para las curvas características del ítem j y se interpreta como un parámetro de “adivinación” en el modelo, es decir, admite que un respondiente con bajo trazo latente tenga una buena probabilidad de responder correctamente a un ítem de alta dificultad. Generalmente toma valores en $(0, 1)$.

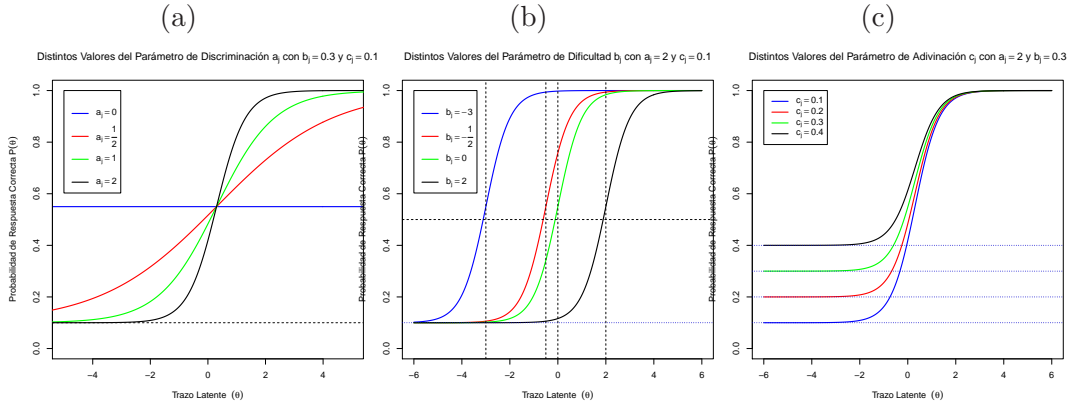


FIGURA 1.4. (a) ICC con $b_j = 0.3, c_j = 0.1$ y $a_j = 0, 0.5, 1, 2$. (b) ICC con $a_j = 2, c_j = 0.1$ y $b_j = -3, -0.5, 0, 2$. (c) ICC con $a_j = 2, b_j = 0.3$ y $c_j = 0.1, 0.2, 0.3, 0.4$.

Algunas consideraciones importantes:

- En los tres modelos aparece el parámetro b_j que en el modelo 3PL representa el punto en el que la probabilidad de responder correctamente al ítem j es $(1 + c_j)/2$.
- En las Figuras 1.3 y 1.4 puede observarse que a mayores valores de a_j el valor de la pendiente es mayor en la CCI, lo que se traduce en que el ítem discrimina más a los respondientes.
- Cuando $c_j = 0$ en el modelo 3PL se obtiene naturalmente el modelo 2PL, lo que significa que un respondiente con trazo latente bajo tiene aproximadamente una probabilidad nula de responder correctamente a ítems de alta dificultad.
- El modelo de Rasch tiene como ventaja sobre los otros modelos la estimación de los parámetros, dado que el score total del test constituye una estadística suficiente para la estimación del trazo latente, en tanto que, el número de respondientes que contestaron correctamente un ítem es una estadística suficiente para la estimación del parámetro b_j (Harris (1989)). Sin embargo, el modelo de Rasch tiene como desventaja el hecho de no considerar los diferentes aspectos que proporcionan los datos reales, por ejemplo, el suponer que todos ítems tienen la misma discriminación y que todos sus respondientes no “adivinarán” ninguna respuesta.
- El modelo 2PL tiene como ventajas la viabilidad en las estimaciones de sus parámetros, así como el hecho de que permite analizar de manera simultánea la dificultad y discriminación de un ítem.

1.3. Estadística Bayesiana

La estadística Bayesiana tiene gran importancia en el desarrollo de nuevos problemas estadísticos al buscar valores de probabilidad aplicando el Teorema de Bayes de acuerdo al contexto de investigación. Según Hoff (2009) el uso de probabilidades para expresar información se puede hacer a través de un sentido matemático, y que se puede desarrollar por medio del Teorema de Bayes o comúnmente denominado método Bayesiano.

En general, los métodos Bayesianos se desprenden de la inferencia Bayesiana, cuyo objetivo a grandes rasgos es,

- Estimar parámetros con buenas propiedades estadísticas,
- Describir datos simples a través de sus valores de probabilidad conjunta,
- Predicción de datos ausentes.

Entonces el uso de la estadística Bayesiana proporciona métodos que caracterizan poblaciones por medio de sus parámetros.

Según Hoff (2009) y Gelman et al. (2003), las características numéricas de una población pueden ser explicadas en términos de un parámetro θ y un conjunto de datos y cuyas propiedades estadísticas son desconocidas. Por lo que, la mejor forma de establecer un método Bayesiano comienza con una formulación del conjunto de datos y y el parámetro θ por medio de probabilidades sobre el espacio muestral Y y el espacio de parámetros Θ . Así, para cada valor de $\theta \in \Theta$, la distribución *a priori*, $p(\theta)$, que se supone de manera subjetiva, describe de manera aproximada el valor verdadero de θ en la población. En tanto, para cada $\theta \in \Theta$ y $y \in Y$, la *distribución muestral*, $p(y|\theta)$, describe la suposición de lo que sería el resultado si se conociera el valor verdadero de θ y para cada valor de $\theta \in \Theta$ la distribución *posteriori*, $p(\theta|y)$, describe de manera aproximada el valor verdadero de θ teniendo los valores observados de y . Entonces, bajo estos supuestos de probabilidad, el Teorema de Bayes es aplicado para encontrar o estimar el valor de θ dado y

1.3.1. Teorema de Bayes

Como se mencionó anteriormente y bajo los supuestos de probabilidad que se estableció sobre el parámetro θ y los datos observados y , el modelo se puede ver como una distribución conjunta para θ y y . Esta función de densidad conjunta se puede escribir como el producto de dos probabilidades, $p(\theta)$ (*a priori*) y $p(y|\theta)$ (*distribución muestral*), es decir

$$p(\theta, y) = p(\theta)p(y|\theta)$$

Luego usando las propiedades de probabilidad condicional y el Teorema de Bayes, se puede encontrar la función de densidad *posteriori*

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y|\theta)}{p(y)} \quad (1.13)$$

Donde

$$p(y) = \sum_{\theta} p(\theta)p(y|\theta)$$

y para el caso continuo se puede ver como

$$p(y) = \int_{\theta} p(\theta)p(y|\theta)$$

Note que el factor $p(y)$ no depende de θ y para valores fijos de y puede ser considerada como una constante de normalización para la densidad posteriori, por lo que la Ecuación

(1.14) toma la siguiente forma

$$p(\theta|y) \propto p(\theta)p(y|\theta). \quad (1.14)$$

Para los datos y , que pueden ser observados, pero que a su vez son desconocidos, se puede analizar a través de la distribución marginal de y , que está dada por

$$p(y) = \int p(y, \theta)d\theta = \int p(\theta)p(y|\theta)d\theta \quad (1.15)$$

Según Fox (2010) la Ecuación (1.13) puede ser factorizada como se indica en la Ecuación (1.14), dado que se puede realizar el producto de la distribución a priori y la verosimilitud, esto se debe a que las dos muestras contienen la misma información acerca del parámetro θ cuando las verosimilitudes son proporcionales (ver Casella & Berger (2002)).

Fox (2010), indica que en los modelos de TRI, el objetivo está en las distribuciones posteriori de los parámetros de los respondientes y los ítems, entonces de acuerdo al Teorema de Bayes, la densidad posteriori conjunta de los parámetros pueden ser establecidos como

$$p(\theta, \xi|y) = \frac{p(y|\theta, \xi)p(\theta, \xi)}{p(y)} = \frac{p(y|\theta, \xi)p(\theta)p(\xi)}{p(y)}$$

donde θ es el trazo latente y ξ es el vector de parámetros de los ítems (a_j, b_j, c_j) , que varía de acuerdo al modelo TRI aplicado y además se asume que las probabilidades a priores son independientes.

1.4. Métodos de Muestreo

El interés recae en la simulación de los valores posteriores, aunque la obtención de estos valores resulte bastante dispendiosa. Por lo que se necesitan métodos que permitan encontrar estos resultados. Por ejemplo, Fox (2010) presenta algunos métodos de simulación para la distribución posterior. Los métodos Monte-Carlo de cadenas de Markov (MCMC) es una de ellas, que permite construir secuencias que convergen en distribución a la distribución posteriori.

1.4.1. Muestreo de Gibbs

Es uno de los métodos de muestreo más importantes, de acuerdo a Fox (2010) este muestreo comienza por la partición de parámetros o vectores de interés en subvectores $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \dots, \theta_Q)$. Luego la distribución conjunta posteriori de los vectores aleatorios $p(\theta|y)$ es lo que se conoce como la distribución objetivo. Por lo que define un proceso de iteraciones de $\boldsymbol{\theta}^{(m)}$ a $\boldsymbol{\theta}^{(m+1)}$ de la función de probabilidad condicional asociada a cada subvector de la partición, es decir,

$$\begin{aligned} \boldsymbol{\theta}_1^{(m+1)} &\sim p(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2^{(m)}, \dots, \boldsymbol{\theta}_Q^{(m)}, \mathbf{y}) \\ \boldsymbol{\theta}_2^{(m+1)} &\sim p(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1^{(m+1)}, \dots, \boldsymbol{\theta}_Q^{(m)}, \mathbf{y}) \\ &\vdots \\ \boldsymbol{\theta}_Q^{(m+1)} &\sim p(\boldsymbol{\theta}_Q|\boldsymbol{\theta}_1^{(m+1)}, \dots, \boldsymbol{\theta}_{Q-1}^{(m+1)}, \mathbf{y}) \end{aligned}$$

Por lo que bajo condiciones de regularidad $\theta^{(m)}$ converge en distribución a θ cuando $m \rightarrow \infty$.

1.4.2. Metrópolis–Hastings

El algoritmo de Metrópolis–Hastings (M–H) es un algoritmo más general que el muestreo de Gibbs. El algoritmo M–H considera valores de $\theta^1, \dots, \theta^m$ no necesariamente son independientes e idénticamente distribuidos de las muestras obtenidas de $p(\theta|y)$ (Robert & Casella (2009) y Chib & Greenberg (1995)). Luego, se pretende agregar un nuevo valor $\theta^{(m+1)}$. Este proceso se realiza generando un valor θ^* que se aproxima al valor de $\theta^{(m)}$. Hoff (2009) enuncia que si el valor de la probabilidad conjunta $p(\theta^*|y) > p(\theta^{(m)}|y)$, es mejor agregar más valores de θ^* en el conjunto que valores $\theta^{(m)}$, de esta manera podemos realizar la comparación de $p(\theta^*|y)$ a $p(\theta^{(m)}|y)$. Pero no siempre es posible incluir los valores θ^* , por lo que la siguiente ecuación muestra la regla de decisión para la inclusión de estos valores en el conjunto de valores de $\theta^{(m)}$.

$$r = \frac{p(\theta^*|y)}{p(\theta^{(m)}|y)} = \frac{p(y|\theta^*)p(\theta^*)}{p(y|\theta^{(m)})p(\theta^{(m)})} \quad (1.16)$$

De acuerdo a la ecuación anterior si el valor de $r > 1$, indica que el valor de $p(\theta^*|y)$ es mayor que $p(\theta^{(m)}|y)$. Por lo tanto la regla de decisión es incluir el valor de θ^* en el conjunto. Tomando $\theta^{(m+1)} = \theta^*$. De lo contrario, si $r < 1$, indica que se tiene un valor para un instante del valor θ^* , entonces el valor $\theta^{(m+1)} = \theta^*$ con probabilidad r o en su defecto $\theta^{(m+1)} = \theta^{(m)}$ con probabilidad $1 - r$.

Hoff (2009) y Fox (2010) muestran que el proceso del algoritmo (M–H) consiste en muestrear un valor propuesto para θ^* haciendo uso de una distribución propuesta $q(\theta^*|\theta^{(m)})$, preferiblemente simétrica. Entonces si se toman dos valores diferentes θ_a y θ_b se tiene que

$$q(\theta_b|\theta_a) = q(\theta_a|\theta_b)$$

(*Simetría*). Por lo tanto, el proceso (M–H) se puede escribir aplicando los siguientes pasos

1. Muestrear $\theta^* \sim q(\theta|\theta^{(m)})$,
2. Calcular el radio de aceptación

$$r = \frac{p(\theta^*|y)}{p(\theta^{(m)}|y)} = \frac{p(y|\theta^*)p(\theta^*)}{p(y|\theta^{(m)})p(\theta^{(m)})} = \frac{p(\theta_i^*|y)/q(\theta_i^*|\theta_i^{(m)})}{p(\theta_i^{(m)}|y)/q(\theta_i^{(m)}|\theta_i^*)}$$

3. Si

$$\theta^{(m+1)} = \begin{cases} \theta^* & \text{con probabilidad } \min(r,1) \\ \theta^{(m)} & \text{con probabilidad } 1-\min(r,1) \end{cases}$$

El proceso del paso 1 se puede realizar a través de un muestreo de Gibbs según Robert & Casella (2009).

1.5. Estimación Bayesiana en TRI

La estimación Bayesiana en la TRI consiste en la estimación de los parámetros de los ítems y la habilidad del individuo sobre un modelo de probabilidad de respuesta correcta, que en general está basada en un modelo logístico. La idea es la incorporación de toda la información a priori que se tiene de los parámetros desconocidos para estimar los parámetros sobre los resultados de los valores posteriores.

Hay varios métodos de estimación Bayesiana en TRI, uno de ellos es el propuesto por Mislevy (1985) que generaliza la marginalización propuesta en el artículo de Bock & Aitkin (1981), esta estimación vía marginalización parte de un modelo de estimación Bayesiana lineal jerárquico, Fox (2010) define un esquema de datos unidimensionales que siguen un modelo de respuesta logístico a través de un método de muestreo M-H. Este método es explicado con más detalle en Andrade et al. (2000), pero a continuación se presenta una breve explicación de lo que se muestra en este texto.

La idea fundamental de la estimación Bayesiana es establecer distribuciones a priori para los parámetros de interés, y de esta llegar a una distribución posteriori, que, como se mencionó en las secciones anteriores permite la estimación de los parámetros deseados. El *método de máxima verosimilitud marginal* de acuerdo a Bock & Aitkin (1981) y descrito por Andrade et al. (2000) es el siguiente:

- Se establece una distribución para el trazo latente θ en función de un vector de parámetros η con densidades $p(\theta|\eta)$.
- Se establece una distribución ζ_i para los individuos, con $i = 1, 2, \dots, I$, en función de un vector de parámetros τ , con densidad $p(\zeta|\tau)$.
- Finalmente, se establece distribuciones a priori $f(\tau)$ y $g(\eta)$ para η y τ .

Luego, de acuerdo a lo anterior la distribución conjunta para θ y los parámetros de los individuos ζ es

$$\begin{aligned} f(\theta, \zeta, \eta, \tau) &= f(\zeta|\tau)g(\theta|\eta)f(\tau)g(\eta) \\ &= \left(\prod_{i=1}^I f(\zeta_i|\tau) \right) \left(\prod_{j=1}^n g(\theta_j|\eta) \right) f(\tau)g(\eta) \end{aligned}$$

Por lo tanto, la distribución posteriori queda dada por

$$f(\theta, \zeta, \eta, \tau|y) \propto L(y, \theta, \zeta)f(\zeta|\tau)g(\theta|\eta)f(\tau)g(\eta). \quad (1.17)$$

Andrade et al. (2000) dice, que las inferencias para los parámetro de los ítems está dada por la marginalización de la integral posteriori con relación a θ y τ obteniendo la distribución posteriori de ζ y η de la siguiente manera

$$f^*(\zeta, \eta|y) \propto L(\zeta, \eta)f(\zeta)g(\eta) \quad (1.18)$$

Luego el valor de la moda posteriori vista como un estimador de ζ , es aquella que maximiza la distribución posteriori marginal. Luego se tiene que

$$\log f^*(\zeta, \eta|y) = K + \log L(\zeta, \eta) + \log f(\zeta) + \log g(\eta)$$

Donde el valor de K es un valor constante. Luego basta con realizar las derivadas parciales con respecto a ζ para encontrar el valor que maximiza la distribución posteriori.

Ahora con relación a los parámetro de los ítems, en la TRI en general se supone que el parámetro de discriminación a_i sigue una distribución *log-normal* o en su defecto una distribución chi-cuadrado y el parámetro de dificultad b_i una distribución normal. Entonces, de acuerdo a esto, las estimaciones para estos parámetros están dadas por:

$$\frac{\partial \log f(a_i|\mu_a, \sigma_a^2)}{\partial a_i} = -\frac{1}{a_i} \left[1 + \frac{\log a_i - \mu_a}{\sigma_a^2} \right] \quad (1.19)$$

y,

$$\frac{\partial \log f(b_i|\mu_b, \sigma_b^2)}{\partial b_i} = -\frac{b_i - \mu_b}{\sigma_b^2} \quad (1.20)$$

Para mayores detalle ver (Andrade et al. (2000)).

CAPÍTULO 2

Modelo 2PL Donde el Trazo Latente Sigue una Distribución Normal Truncada

En este capítulo se presentará el modelo propuesto y el método de estimación Bayesiana teniendo en cuenta algunas definiciones y métodos de estimación descritos en el Capítulo 1. En particular, se emplearán métodos Bayesianos en los que la distribución del trazo es asimétrica. En la mayoría de literatura sobre TRI es común suponer que los trazos latentes siguen una distribución simétrica, por ejemplo, en Fox (2010) se asume que $\theta_i \sim N(0, 1)$ o una distribución log-Normal estándar $LN(0, 1)$.

Sin embargo, en algunas investigaciones se ha podido evidenciar que el trazo no siempre sigue una distribución normal estándar (ver Woods (2008)). Por esta razón, se propone una metodología de estimación en la que los scores de los individuos presentan o tienen una tendencia asimétrica, esto se presentará con más detalle por medio de una aplicación a datos reales en el Capítulo 4.

2.1. Modelo Usual de Estimación Bayesiana en TRI

En modelo usual de estimación Bayesiana descrito en Fox (2010) para el muestreo de parámetros condicionados, es el uso de variables auxiliares que permiten muestrear distribuciones complicadas desde la simulación. Estas variables auxiliares o datos aumentados, denotado como Z , están directamente vinculados con los datos observados, por lo tanto las inferencias sobre los parámetros θ se pueden realizar sobre los datos observados $p(y|\theta)$ o usando el modelo de datos aumentados $p(z|\theta)$.

Sea Z_{ij} la variable que determina la habilidad del individuo i para el ítem j . Entonces si $Z_{ij} > 0$ significa que el individuo respondió correctamente al ítem j y si $Z_{ij} \leq 0$ significará respuesta incorrecta. Por lo tanto, la probabilidad que el individuo i responda al ítem j

correctamente se define como

$$\begin{aligned}
 P(Y_{ij} = 1 | \theta_i, \zeta_j) &= P(Z_{ij} > 0 | \theta_i, \zeta_j) \\
 &= \int_0^{\infty} \phi(z; a_j \theta_i - b_j) dz = \Phi(a_j \theta_i - b_j) \\
 &= \int_0^{\infty} \psi(z; d(a_j \theta_i - b_j)) dz = \Psi(d(a_j \theta_i - b_j))
 \end{aligned} \tag{2.1}$$

donde $\Phi(\cdot)$ y $\Psi(\cdot)$, denotan las funciones de densidad normal y logística, respectivamente, y d el factor de escala logístico.

Sea \mathcal{L} y \mathcal{N} denotan la logística estándar y la función de densidad normal estándar respectivamente. Dos diferentes enfoque que pueden tomar los datos aumentados. Por lo tanto, sea $Z_{ij} = a_j \theta_i - b_j$. Luego distribución condicional se define como

$$Z_{ij} | Y_{ij}, \theta_i, \zeta_j \sim \begin{cases} \mathcal{L}(d(a_j \theta_i - b_j), 1) \\ \mathcal{N}(a_j \theta_i - b_j, 1) \end{cases} \tag{2.2}$$

donde Y_{ij} toma el valor de uno si $Z_{ij} > 0$ y $d = 1.7$. El primer caso de la ecuación (2.2) es utilizado en un proceso de estimación usual, mientras que el segundo caso es cuando asumimos que los datos aumentados son normales estándar.

Ahora, supongamos que la distribución para los datos observados está dada por

$$\tilde{Z}_{ij} | Y_{ij}, \theta_i, \zeta_j \sim \begin{cases} \mathcal{L}(0, 1) \\ \mathcal{N}(0, 1) \end{cases} \tag{2.3}$$

donde Y_{ij} toma el valor de uno si $\tilde{Z}_{ij} > d(a_j \theta_i - b_j)$ y cero en otro caso. Note que las variables auxiliares Z_{ij} y \tilde{Z}_{ij} de las ecuaciones (2.2) y (2.3) están relacionadas entre sí. Esto indica que $Z_{ij} = \tilde{Z}_{ij} + d(a_j \theta_i - b_j)$, tal que $Z_{ij} \sim \mathcal{N}(d(a_j \theta_i - b_j), 1)$ y $\tilde{Z}_{ij} \sim \mathcal{N}(0, 1)$. Luego, se define

$$\Delta_l = \max_{j|y_{ij}=1} \frac{(db_j - \tilde{z}_{ij})}{da_j} \quad \text{y} \quad \Delta_u = \min_{j|y_{ij}=0} \frac{(db_j - \tilde{z}_{ij})}{da_j} \tag{2.4}$$

Por lo tanto, la distribución posterior puede ser escrita como

$$P(\theta_i | y, \tilde{z}, \zeta, \mu_\theta, \sigma_\theta^2) \propto I_{(\Delta_l, \Delta_u)}(\theta_i) p(\theta_i | \mu_\theta, \sigma_\theta^2) \tag{2.5}$$

Entonces un proceso de estimación M-H para un modelo 2PL cuando asumimos que los datos observados están normalmente distribuidos, está dado por

1. Muestreo de los datos aumentados $\tilde{z}^{(m+1)}$.
2. Se define la distribución a priori para θ_i y se realiza el muestreo para los valores de $\theta_i^{(m+1)}$
3. Se define la distribución a priori para ζ_j (parámetros de los ítems) y se muestrea los valores $\zeta_j^{(m+1)}$
4. Se muestrea los valores a priori para los parámetros $\mu_\theta^{(m+1)}$ y $(\sigma_\theta^2)^{(m+1)}$.

Este es el proceso de estimación cuando se asume que los datos siguen una distribución Normal Estándar (ver Fox (2010)). Un proceso descrito de manera similar también se

encuentra en Kim & Bolt (2007), donde presenta las ventajas y el proceso de estimación para modelos TRI a través del algoritmo M–H.

2.2. Modelo Propuesto

En esta sección y en general en el proceso de este trabajo, el análisis y desarrollo estará basado un test dicotómico o de respuesta binomial, aplicado a un número N de individuos con un número p de ítems. Donde la probabilidad de que el respondiente i responda correctamente al ítem j esta descrito por un modelo logístico de dos parámetros (Modelo 2PL). (Ver Ecuación (1.12)). Dado que el proceso de estimación se efectúa un método Bayesiano basado en un algoritmo de tipo Metrópolis–Hastings con distribución propositiva Normal Truncada $NT(\mu, \sigma^2, l_x, u_x)$ cuya densidad está dada por

$$f(x|\mu, \sigma^2, l_x, u_x) = \frac{\exp(-(x - \mu)^2/2\sigma^2)}{\sqrt{2\pi}\sigma[\Phi((u_x - \mu)/\sigma) - \Phi((l_x - \mu)/\sigma)]} \mathbb{I}_{l_x \leq x \leq u_x} \quad (2.6)$$

donde l_x y u_x son el extremo inferior y superior de acotamiento respectivamente y $\mathbb{I}_{l_x \leq x \leq u_x}$ es la función indicadora. En este trabajo se realizó el proceso de simulación de las variables normal truncadas de la ecuación (2.6) siguiendo el esquema propuesto por Robert (1995).

Entonces, para la aplicación del modelo propuesto supondremos lo siguiente

1. El test trabajado es esencialmente unidimensional.
2. El trazo latente θ sigue una distribución a priori Normal Truncada ($NT(\mu, \sigma^2, l_\theta, u_\theta)$).
3. La distribución de los score de los individuos tiene una tendencia asimétrica.

2.3. Método de Estimación

En la Sección 2.1 se presentó el proceso usual de estimación Bayesiana suponiendo que el trazo latente sigue una distribución normal estándar y además aplicando el método de variables auxiliares. En este trabajo se realizará el proceso de inferencia Bayesiana sobre los datos observados, es decir, sin utilizar tales variables.

El proceso de estimación está basado en un método Bayesiano, más específicamente en una estimación MCMC, por medio del algoritmo de Metrópolis–Hastings (M–H), obteniendo así una estimación conjunta de los valores de las distribuciones posteriores de los parámetros θ , a_j y b_j .

Para establecer la verosimilitud es necesario establecer los siguientes supuestos

- 1) Los patrones de respuesta entre los individuos son independientes
- 2) Dado el trazo latente θ , la respuestas Y_{ij} son independientes para $i = 1, \dots, N$ y $j = 1, \dots, p$, esto se conoce como independencia local.

Por lo tanto, la distribución posterior conjunta está dada por

$$p(\theta_i, \zeta_j | y) \propto \prod_{i=1}^N \prod_{j=1}^p P_{ij}^{y_{ij}} (1 - P_{ij})^{1-y_{ij}} p(\theta_i) p(\zeta_j) \quad (2.7)$$

y las marginales condicionales están dadas por

$$p(\theta_i | y_{ij}, \zeta) \propto \prod_{j=1}^p P_{ij}^{y_{ij}} (1 - P_{ij})^{1-y_{ij}} p(\theta_i) \quad (2.8)$$

$$p(\zeta_j | y_{ij}, \theta) \propto \prod_{i=1}^N P_{ij}^{y_{ij}} (1 - P_{ij})^{1-y_{ij}} p(\zeta_j) \quad (2.9)$$

donde $p(\theta_i)$ representa la densidad a priori para el trazo latente del individuo i y $p(\zeta_j)$ es la densidad a priori para los parámetros de los ítems j . De acuerdo a esto, las densidades a priori para los parámetros en este trabajo se establecen como (ver ecuación (2.6))

- Para el trazo latente, $\theta_i \sim NT(\mu_\theta, \sigma_\theta^2, l_\theta, u_\theta)$
- Para el parámetro de discriminación, $a_j \sim NT(\mu_a, \sigma_a^2, l_a, u_a)$
- Para el parámetro de dificultad, $b_j \sim NT(\mu_b, \sigma_b^2, l_b, u_b)$

Luego, al reemplazar estas distribuciones en las ecuaciones (2.7), (2.8) y (2.9) y siguiendo el modelo logístico de dos parámetros (1.12), no es posible encontrar distribuciones posteriores conocidas o que nos permitan obtener información de los parámetros de manera analítica. Por esta razón, se hace uso del algoritmo de M–H para obtener muestras de la distribución posterior de la ecuación (2.7). Según Gamerman & Lopes (2006) y Hoff (2009) estos métodos son utilizados cuando no se puede obtener esta distribución de manera analítica.

En este trabajo se considerará que los parámetros siguen una distribución $NT(\mu, \sigma^2, l_x, u_x)$ (ver ecuación (2.6)) para el parámetro de discriminación así como para los parámetros θ y de dificultad, es decir,

- $\theta_i \sim NT(\mu_\theta, \sigma_\theta^2, l_\theta, u_\theta)$,
- $a_j \sim NT(\mu_a, \sigma_a^2, l_a, u_a)$,
- $b_j \sim NT(\mu_b, \sigma_b^2, l_b, u_b)$,

Luego el esquema del algoritmo propuesto queda determinado de la siguiente manera

- 1) Paso 1. Obtener los valores iniciales para los parámetros de los ítems ζ^* y de los trazos latentes θ^* . Es decir, $\zeta_j^{(1)} = \zeta_j^*$, para $j = 1, 2, \dots, p$ y $\theta_i^{(1)} = \theta_i^*$, para $i = 1, 2, \dots, N$
- 2) Paso $k + 1$. Obtenga una muestra candidata ζ_j^* extraída de la densidad propositiva $q(\zeta_j | \zeta_j^{(k)})$.

3) Calcule la regla de aceptación que se define como

$$\alpha_{\zeta_j} = \min \left\{ 1, \frac{p(\zeta_j^*|y)/q(\zeta_j^*|\zeta_j^{(k)})}{p(\zeta_j^{(k)}|y)/q(\zeta_j^{(k)}|\zeta_j^*)} \right\} \quad (2.10)$$

4) La aceptación del nuevo valor se hace mediante la generación de un valor aleatorio u con distribución uniforme $(0,1)$, es decir, $u \sim U(0,1)$. Si $\alpha \geq u$ acepte el movimiento de $\zeta_j^{(k)}$ a $\zeta_j^{(k+1)} = \zeta_j^*$. En caso contrario $\zeta_j^{(k+1)} = \zeta_j^{(k)}$. En otras palabras. Si $\alpha_{\zeta_j} \geq 1$, acepte ζ_j^* , es decir $\zeta_j^{k+1} = \zeta_j^*$. Si $\alpha_{\zeta_j} < 1$ acepte ζ_j^* con probabilidad α_{ζ_j} .

5) Paso $k + 1$. Obtenga una muestra candidata θ_i^* extraída de la densidad propositiva $q(\theta_i|\theta_i^{(k)})$.

6) Calcule la regla de aceptación que se define como

$$\alpha_{\theta_i} = \min \left\{ 1, \frac{p(\theta_i^*|y)/q(\theta_i^*|\theta_i^{(k)})}{p(\theta_i^{(k)}|y)/q(\theta_i^{(k)}|\theta_i^*)} \right\} \quad (2.11)$$

7) La aceptación del nuevo valor se hace mediante la generación de un valor aleatorio u con distribución uniforme $(0,1)$, es decir, $u \sim U(0,1)$. Si $\alpha \geq u$ acepte el movimiento de $\theta_i^{(k)}$ a $\theta_i^{(k+1)} = \theta_i^*$. En caso contrario $\theta_i^{(k+1)} = \theta_i^{(k)}$. En otras palabras. Si $\alpha_{\theta_i} \geq 1$, acepte θ_i^* , es decir $\theta_i^{k+1} = \theta_i^*$. Si $\alpha_{\theta_i} < 1$ acepte θ_i^* con probabilidad α_{θ_i} .

Asuma que la distribución propositiva q para el trazo latente $\theta_i \sim NT(\mu_\theta, \sigma_\theta^2, l_\theta, u_\theta)$.

2.4. Pruebas de Bondad y Ajuste

El proceso de bondad y ajuste (GoF) para los datos del modelo propuesto, se centra en la toma de un muestreo aleatorio $k = 1, 2, \dots, I$ de las distribuciones conjuntas condicionales de θ y los datos observados y , estableciendo una función de comprobación (ver Gelfand et al. (1992)). Gelman et al. (1996) basado en el trabajo realizado por Gelfand et al. (1992), establece una función $T(y)$ y un modelo usando distribuciones predictivas basado en métodos de muestreo y la máxima verosimilitud de la discrepancia entre frecuencias esperadas y observadas. Entonces sea

$$P(z|y) = \int P(z|y, \theta, \zeta) d\theta d\zeta \quad (2.12)$$

donde y son los datos observados y z son los datos no observados o réplicas de y . Por lo tanto, se realiza un muestreo Bayesiano de los parámetros $\theta^{(k)}$ y $\zeta^{(k)}$, donde $k = 1, 2, \dots, I$, en un modelo 2PL, es decir,

$$\pi_{ij}^{(k)} = \frac{1}{1 + \exp \left\{ -(a_j^{(k)t} \theta_i^{(k)} + d_j^{(k)}) \right\}} \quad (2.13)$$

Es decir, que de acuerdo a la ecuación (2.13), $z^{(k)}$ es un test generado (o réplicas) de una distribución de Bernoulli de parámetro $\pi_{ij}^{(k)}$.

Por lo tanto, la prueba de Bondad y Ajuste para el individuo i para el ítem j (ver Gelfand et al. (1992), Gelman et al. (1996)) se define respectivamente como

$$T(y)_{\zeta_j} = I^{-1} \sum_{k=1}^I 1_{L_{\zeta_j}(z_j)^k - L_{\zeta_j}(y_j)^k > 0} \quad (2.14)$$

$$T(y)_{\theta_i} = I^{-1} \sum_{k=1}^I 1_{L_{\theta_i}(z_i)^k - L_{\theta_i}(y_i)^k > 0} \quad (2.15)$$

Donde $1_{\{\cdot\}}$ es la función característica y ζ_j es el conjunto de parámetros de los ítems. Cabe resaltar que si los valores de las ecuaciones (2.14) y (2.15) se encuentran alrededor de 0.5 se tiene un buen ajuste, de lo contrario si estos valores están cercanos a 0 o 1 se tiene un mal ajuste.

2.4.1. Test de Heidelberg–Welch

Heidelberg & Welch (1983) estima la convergencia de una cadena desde dos enfoques, el primero se basa en evaluar la estacionariedad y el segundo evalúa la precisión en que el valor medio de la cadena se aproxima al valor medio de la distribución posterior, este proceso se conoce como el Test de Heidelberg–Welch.

Esta prueba contrasta la hipótesis nula (H_0) de que las muestras tomadas provienen de una distribución estacionaria. En primer lugar la prueba Heidelberg–Welch es aplicada a toda la cadena, si se rechaza H_0 , entonces es descartado el 10 % de la cadena, si aún así, H_0 sigue sin ser aceptada se descartará un 20 % de la cadena y así sucesivamente hasta que la hipótesis nula no sea rechazada o el 50 % de una cadena sea descartado. Si el 50 % ha sido descartada constituye el rechazo absoluto de H_0 lo que indica que es necesario una aplicación de muestreo MCMC más larga.

La prueba de amplitud media calcula un intervalo de confianza del 95 % para la media usando la proporción de la cadena que paso la prueba de estacionariedad. La mitad del ancho de este intervalo con la estimación de la media. Si la razón entre dicha mitad y la media es menor que un $\epsilon > 0$, por default en la librería coda de R, $\epsilon = 0.1$. La prueba se considera pasada. De otra manera la cadena se considera demasiado corta para la estimación de la media.

Simulación

Este capítulo tiene como objetivo ilustrar el funcionamiento del modelo Bayesiano propuesto en el capítulo 2 a través de varias simulaciones de datos con un modelo TRI Logístico de dos parámetros (2PL). A continuación se mostrará la recuperación de parámetros para uno de los conjuntos de datos simulados, mientras que en el apéndice C se presenta el procedimiento con la simulación de otro test.

3.1. Generación del Test

En esta sección se mostrará la generación del test, así como la recuperación de los parámetros, teniendo en cuenta los supuestos teóricos del modelo propuesto.

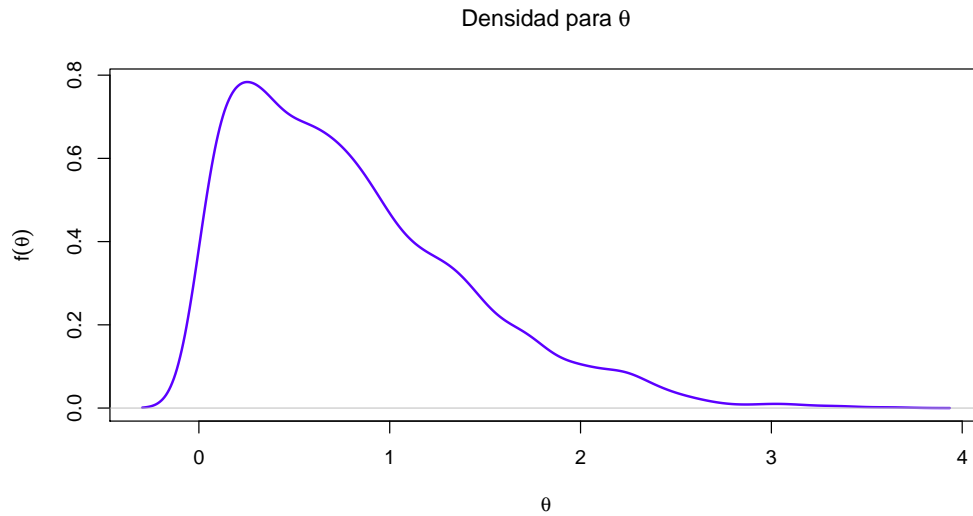
La probabilidad de que el individuo i responda correctamente al ítem j está dada por

$$P_{ij}(Y_{ij}|\theta_i) = \frac{1}{1 + e^{-a_j(\theta_i - b_j)}} = \frac{1}{1 + e^{-a_j\theta_i + d_j}} \quad (3.1)$$

para $i = 1, 2, \dots, N$ y $j = 1, 2, \dots, p$, también conocido como el Modelo 2PL. Entonces se genera un test dicotómico en el software R con la siguiente información y una semilla para reproducir el mismo test varias veces (`set.seed(500)`)

- $p = 20$ ítems y $N = 5000$ individuos
- Modelo 2PL (logit de dos parámetros).
- $a_j \sim LN(0, 1/4)$
- $b_j \sim NT(0, \sigma_b^2 = 1, l_b = 0, u_b = 17)$
- $\theta_i \sim NT(0, \sigma_\theta^2 = 1, l_\theta = 0, u_\theta = 17)$

Donde NT y LN son las distribuciones Normal Truncada y Log–Normal respectivamente. La distribución NT fue simulada de acuerdo a lo establecido por Robert (1995), generando valores aleatorios con un límite inferior de 0 hasta infinito, es decir valores $NT(\mu = 0, \sigma = 1, l = 0, u = \infty)$ (ver Apéndice A.1).

FIGURA 3.1. Función de Densidad $NT(\mu = 0, \sigma = 1, l = 0, u = \infty)$

Se presentan algunas medidas descriptivas de los parámetros simulados con estas distribuciones en la Tabla 3.1

	a_j	b_j	θ_i
Min.	0.66	0.03	0.00
Q_1	0.85	0.16	0.32
\tilde{x}	0.96	0.45	0.67
\bar{x}	1.00	0.55	0.79
Q_3	1.14	0.85	1.13
Max.	1.35	1.70	3.49

TABLA 3.1. Medidas Descriptivas de los parámetros a_j , b_j y θ_i Simulados

De acuerdo a la Tabla anterior y a la Figura 3.1, los valores de la media y la mediana para θ son diferentes, esto nos indica que el comportamiento de la distribución $NT(\mu = 0, \sigma = 1, l = 0, u = \infty)$ para este parámetro es asimétrica.

En la siguiente sección se procederá a la recuperación de los parámetros y el ajuste de los datos con un modelo logístico 2PL empleando una metodología Bayesiana (MCMC algoritmo M-H).

3.2. Recuperación de Parámetros

Una vez generado el test con los supuestos básicos de la TRI y evidenciando que el test simulado es esencialmente unidimensional, se realiza la depuración del mismo, es decir, se excluirán aquellos individuos que presentan un score perfecto y un score nulo. Por lo tanto, el algoritmo propuesto en la recuperación de los parámetros se aplicará ahora a un Test de $p = 20$ ítems y $N = 4989$ individuos.

Se realizó la recuperación ajustando un modelo logístico de dos parámetros empleando una metodología Bayesiana a través de un muestreo de Metrópolis–Hastings. Para la implementación de la metodología se consideran las siguientes distribuciones a priori para los parámetros de los ítems y para el trazo latente.

- $a_j \sim NT(0, \sigma_a^2 = 1, l_a = 0, u_a = 5)$
- $b_j \sim NT(0, \sigma_b^2 = 1, l_b = 0, u_b = 17)$
- $\theta_i \sim NT(0, \sigma_\theta^2 = 1, l_\theta = 0, u_\theta = 17)$

Adicionalmente, se establecieron las distribuciones propositivas para los parámetros de los ítems y para el trazo latente

- $\theta_i \sim NT(\mu_\theta = \theta^*, \sigma_\theta^2 = \tau_t^2, l_\theta = 0, u_\theta = 17)$
- $b_j \sim NT(\mu_b = b^*, \sigma_b^2 = \tau_b^2, l_b = 0, u_b = 17)$
- $a_j \sim NT(\mu_a = a^*, \sigma_a^2 = \tau_a^2, l_a = 0, u_a = 5)$

Posteriormente se realizaron 10000 muestras con los siguientes parámetros de tuning:

$$\tau_a = 0.06 \quad \tau_b = 0.06 \quad \tau_t = 1.6$$

Las tasas de aceptación para los parámetros de los ítems y del trazo latente fueron 0.349 y 0.383 respectivamente. Tomando 2000 cadenas de calentamiento, las gráficas para el muestreo de los parámetros se presentan en la Figura 3.2.

La recuperación de cada parámetro se realizó utilizando la media. En primer lugar, se presenta los resultados del trazo latente recuperado de los primeros 10 individuos descritos en la Tabla 3.2 en la que, además, se presentan los valores del score (puntaje) estandarizado para los mismos individuos. La correlación entre el trazo estimado y el score estándar es de 0.97.

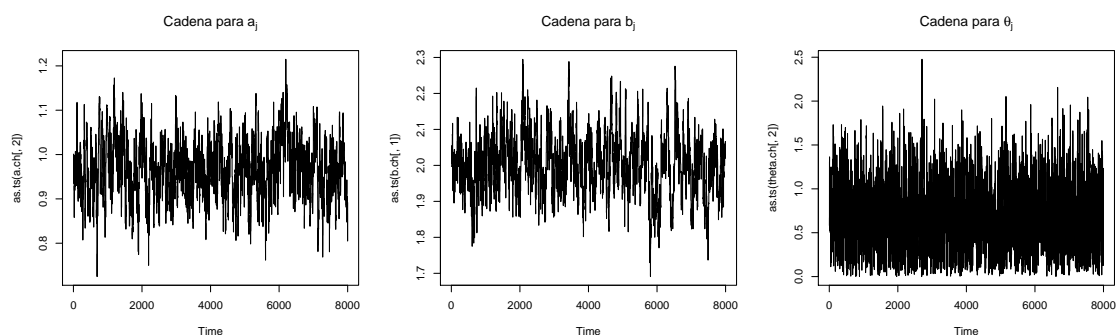


FIGURA 3.2. Cadenas para los parámetros de los ítems y el trazo latente

i	1	2	3	4	5	6	7	8	9	10
$\hat{\theta}_i$	1.19	0.72	0.50	0.27	0.53	0.34	0.38	0.36	1.15	0.80
Score	3.66	2.75	2.44	1.53	2.44	1.84	2.14	1.84	3.66	3.05

TABLA 3.2. Valores del trazo estimado y el score estándar para los primeros 10 registros del test

En la Figura 3.3 se presentan además, las gráficas de las cadenas para algunos valores de θ , las gráficas de las densidades correspondientes, así como los autocorrelogramas, en los que se muestra que para la mayoría de las cadenas de los trazos latentes, la autocorrelación está por debajo de 0.1 entre el Lag 10 y 25.

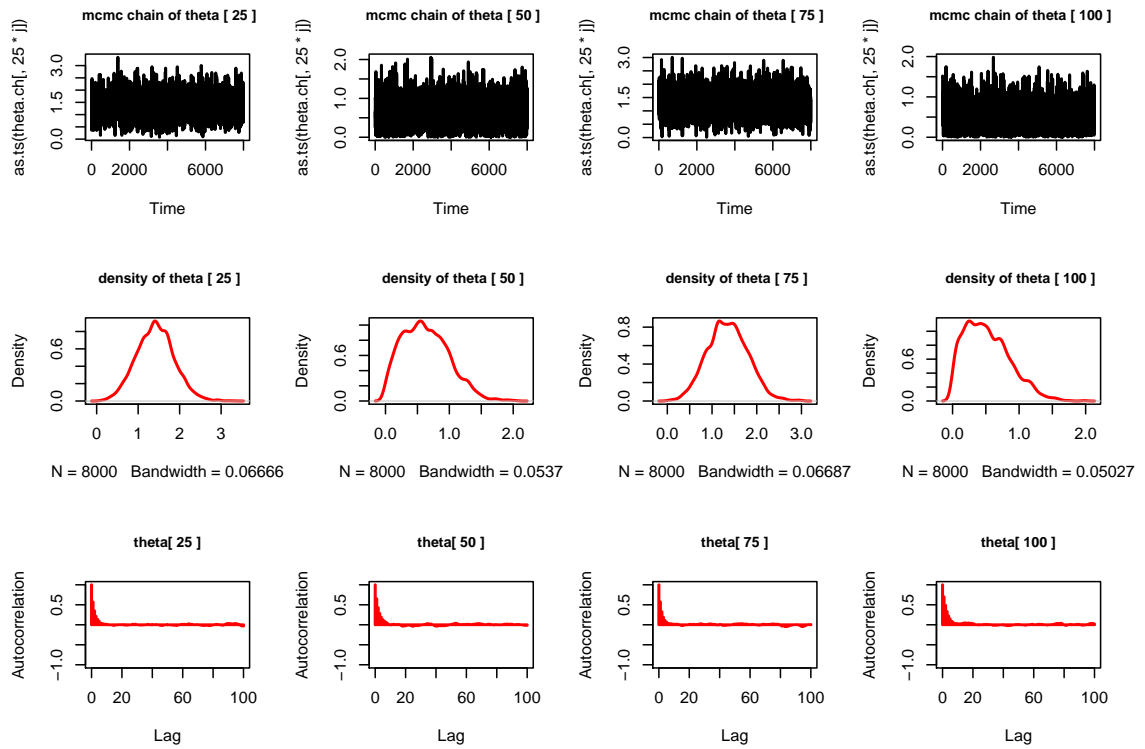


FIGURA 3.3. Gráficos Cadenas, densidad y autocorrelograma para el trazo latente θ

La recuperación de los parámetros de los ítems también se obtuvo a partir de la media. A continuación se presentan los gráficos de las cadenas, autocorrelación y densidades de las mismas para los parámetros de los ítems. Las Figuras 3.4 y 3.5 corresponden a los parámetros de dificultad y discriminación respectivamente.

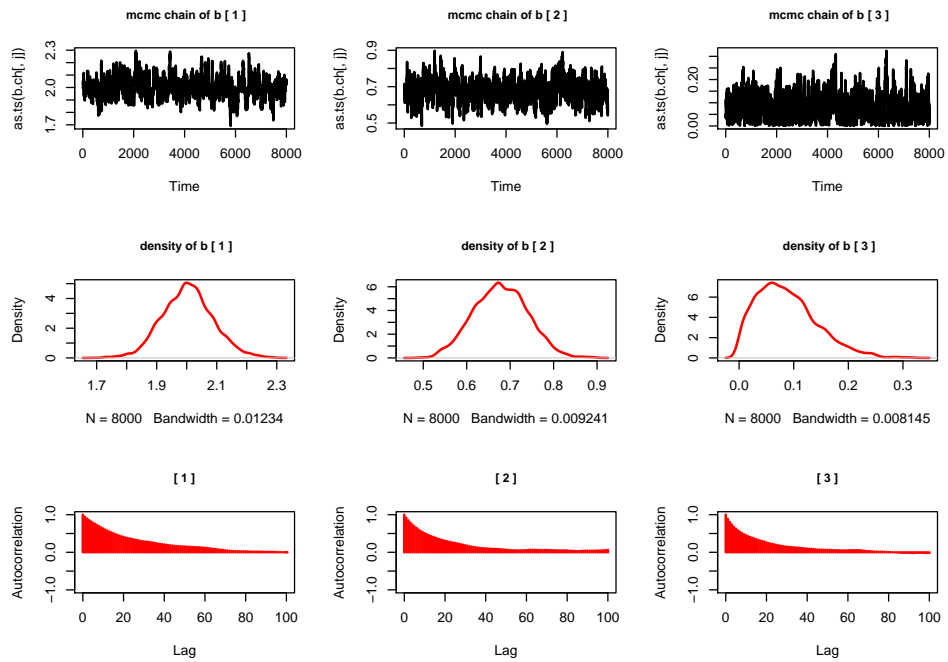


FIGURA 3.4. Gráficos Cadenas, densidad y autocorrelograma para el parámetro de dificultad

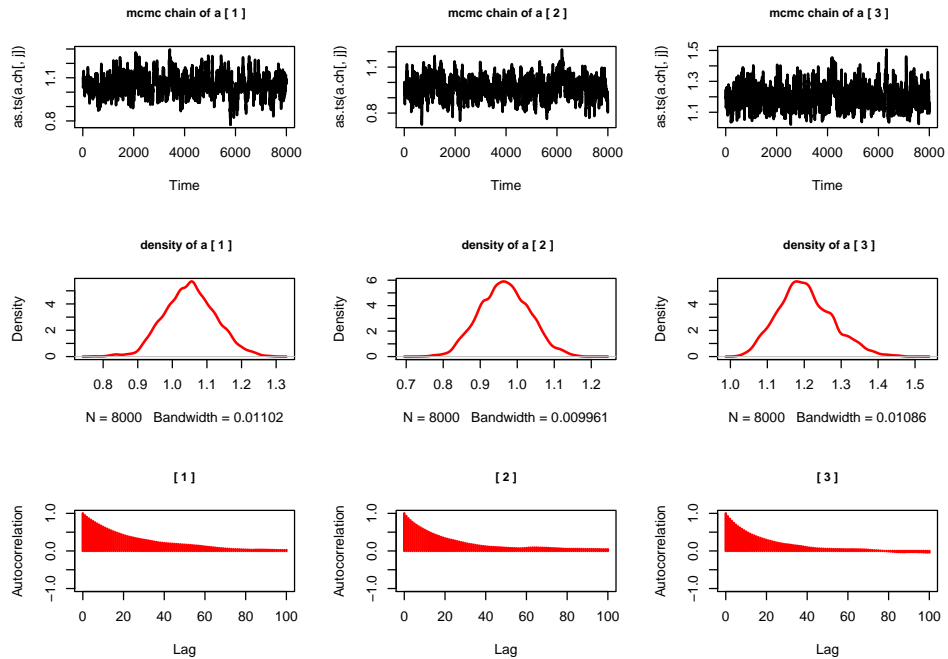


FIGURA 3.5. Gráficos Cadenas, densidad y autocorrelograma para el parámetro de discriminación

En la Figura 3.4 se muestra que para los parámetros de dificultad, la mayoría de autocorrelaciones está por debajo de 0.1 entre los Lags 50 y 80. Para los parámetros de discriminación la Figura 3.5 muestra que la mayoría de autocorrelaciones están por debajo de 0.1 entre los Lags 50 y 80.

3.3. Comparación entre los valores simulados y recuperados

Se procede ahora a comparar los valores de los parámetros estimados con los de la simulación inicial para determinar si la metodología los recupera de manera apropiada.

En primer lugar, se compararon los parámetros de dificultad, la correlación en este caso resultó ser igual a 0.9965246, en tanto que el error absoluto estuvo entre 0.01215 y 0.20950 como se indica en la Tabla 3.3, adicionalmente, en la Figura 3.6 se contrastan los valores simulados y recuperados para el parámetro b_j .

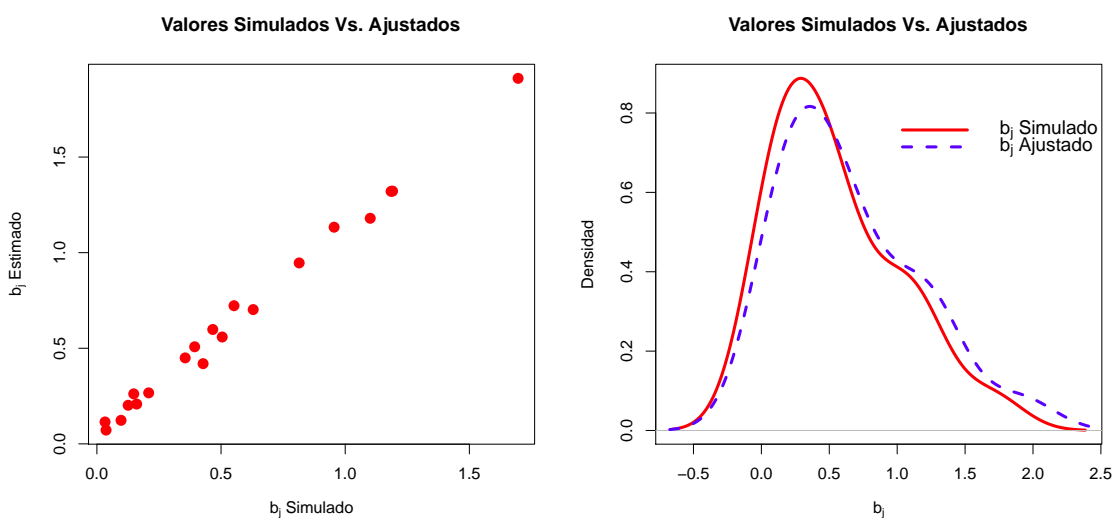


FIGURA 3.6. Comparación entre los valores simulados y recuperados para el parámetro de dificultad

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	sd
0.01215	0.05867	0.08753	0.09700	0.13210	0.20950	0.05306

TABLA 3.3. Estadísticas descriptivas para el Error Absoluto del parámetro de dificultad

Para el parámetro de discriminación se obtuvo una correlación de 0.9423906 entre los valores simulados y los recuperados, además el error absoluto estuvo entre 0.016 y 0.16 con una desviación de 0.04595 como se indica en la Tabla 3.4. Al igual que se hizo con el parámetro de dificultad, en la Figura 3.7 se realiza una comparación entre los valores recuperados y simulados para a_j .

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	sd
0.01614	0.03510	0.05355	0.07090	0.09852	0.16230	0.0459583

TABLA 3.4. Estadísticas descriptivas para el Error Absoluto del parámetro de discriminación

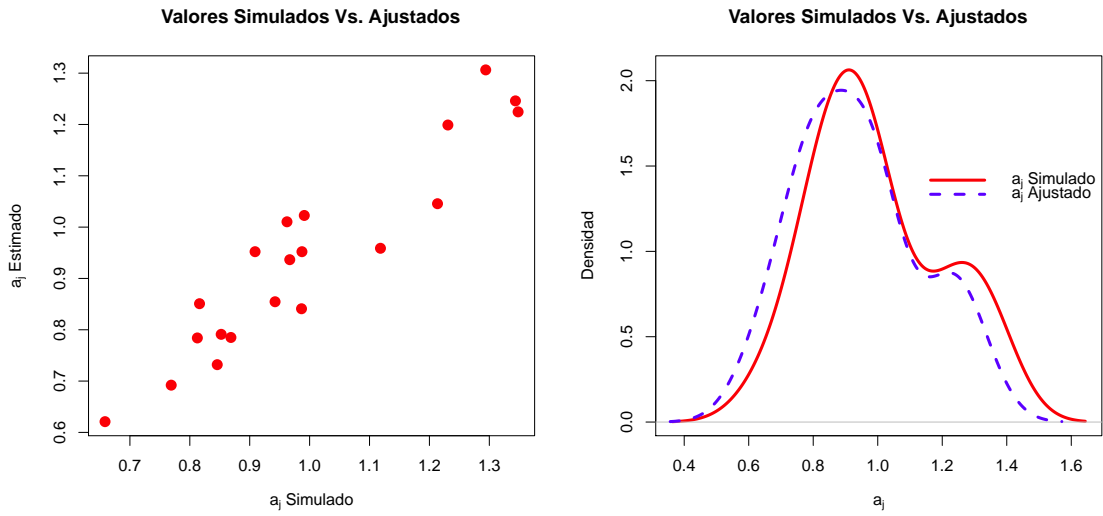


FIGURA 3.7. Comparación entre los valores simulados y estimados para el parámetro de discriminación

Realizando el mismo procedimiento para el trazo, se encontró que el error absoluto fue en promedio de 0.3045 con una desviación estándar igual a 0.235, mientras que la correlación entre el simulado y el recuperado fue de 0.7886151. Como se presenta Tabla 3.5 y la Figura 3.8.

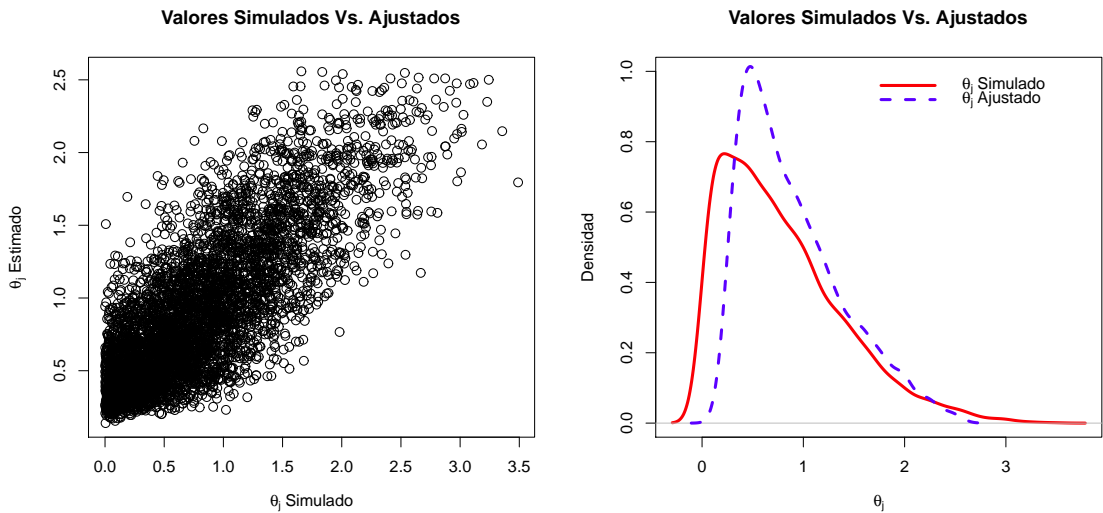


FIGURA 3.8. Comparación entre los valores simulados y recuperados del Trazo

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	sd
0.0000447	0.1197000	0.2553000	0.3045000	0.4351000	1.7160000	0.2357213

TABLA 3.5. Estadísticas descriptivas para el Error Absoluto del Trazo

En la Tabla 3.6 se presentan los valores simulados, recuperados y el error absoluto para cada uno de los parámetros de los ítems.

j	a_j	\hat{a}_j	$ a_j - \hat{a}_j $	b_j	\hat{b}_j	$ b_j - \hat{b}_j $
1	1.214	1.051	0.162	1.696	1.906	0.210
2	1.118	0.966	0.153	0.629	0.702	0.073
3	1.231	1.203	0.028	0.037	0.072	0.036
4	0.658	0.622	0.037	0.428	0.416	0.012
5	0.991	1.021	0.030	0.148	0.260	0.111
6	1.344	1.248	0.095	0.955	1.132	0.176
7	0.909	0.952	0.043	1.101	1.178	0.077
8	1.294	1.310	0.016	0.467	0.599	0.132
9	0.942	0.854	0.088	0.815	0.948	0.134
10	0.769	0.698	0.071	0.126	0.208	0.082
11	0.987	0.957	0.030	1.191	1.317	0.127
12	0.967	0.924	0.043	0.097	0.114	0.017
13	0.986	0.847	0.140	0.356	0.449	0.093
14	0.816	0.854	0.037	0.552	0.722	0.170
15	0.852	0.799	0.053	1.185	1.317	0.132
16	0.813	0.793	0.020	0.505	0.561	0.056
17	1.348	1.228	0.120	0.209	0.268	0.060
18	0.846	0.737	0.109	0.160	0.213	0.053
19	0.869	0.779	0.090	0.033	0.109	0.076
20	0.962	1.016	0.054	0.394	0.508	0.114

TABLA 3.6. Comparación entre los valores simulados y recuperados de los parámetros de los ítems

De acuerdo a lo anterior, el algoritmo parece recuperar de manera apropiada los parámetros simulados. Para corroborar esta información se presentan las pruebas de Bondad y Ajuste para los datos simulados en la siguiente sección.

3.4. Pruebas de Bondad y Ajuste

En esta sección se realizan las pruebas de bondad y ajuste para los parámetros recuperados como se establece en la Sección 2.4 y se presenta consigo la calidad de las cadenas muestreadas, basadas en la prueba estacionaria de Heidelberger & Welch (1983) descrita en la Sección 2.4.1 y que se encuentra en la librería `coda` de R.

La calidad de las cadenas muestreadas son analizadas con un nivel de significancia de $\alpha = 5\%$ y haciendo la comparación con el p -valor, se observa que cantidad de cadenas son rechazadas. Es decir, si el p -valor es menor que α la cadena no pasa el prueba Heidelberger-Welch. A continuación se presenta los resultados de esta prueba para las cadenas de cada uno de los parámetros.

En la Tabla 3.7 y en la Tabla 3.8 se presentan las medidas descriptivas del p -valor para las cadenas de los parámetros de dificultad y discriminación respectivamente. De acuerdo a estos resultados, se puede observar que ninguna cadena es rechazada, es decir, las cadenas presentan una buena calidad de convergencia.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.06678	0.31520	0.46730	0.49470	0.79780	0.92230

TABLA 3.7. Estadísticas descriptivas para el p -valor de Heidelberg–Welch para el parámetro de dificultad

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.08634	0.44700	0.53930	0.55020	0.75170	0.99150

TABLA 3.8. Estadísticas descriptivas para el p -valor de Heidelberg–Welch para el parámetro de discriminación

Para el trazo latente, se encuentra que 43 de las 4989 cadenas o el 0.86% no pasan la prueba con un nivel de 0.05. Dado que el porcentaje es muy bajo, se puede considerar que las autocorrelaciones para el trazo en su mayoría convergen. Por lo que se puede continuar con el método de recuperación.

Ahora, el método de Bondad y Ajuste, está basado de acuerdo al algoritmo descrito en la Sección 2.3. Así, para los parámetros recuperados de los ítems, tenemos que sus valores están entre 0.47 y 0.52 (ver Tabla 3.9) lo que indica que tenemos un buen ajuste. Además, se calculó la diferencia entre las probabilidades observadas y las probabilidades recuperadas de respuesta correcta para cada ítem, y se tiene que sus valores están entre -1.42×10^{-3} y 4.88×10^{-4} , es decir, una diferencia mínima.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.4774	0.4913	0.5026	0.5031	0.5128	0.5290

TABLA 3.9. Estadísticas descriptivas Bondad y Ajuste de los parámetros de los ítems recuperados

Por otro lado, de acuerdo a las pruebas de Bondad y Ajuste, se encontró que 175 de los 4989 trazos latentes recuperados tienen valores por fuera del intervalo (0.05, 0.95), es decir, un 3.5% del 100% de los trazos no presentarían un buen ajuste. Por lo tanto, se puede concluir que el modelo propuesto ajusta de manera adecuada los parámetros de los datos simulados. Es decir, que los parámetros recuperados tienen una buena aproximación a los parámetros poblacionales.

Aplicación a Datos Reales

4.1. Descripción de Datos Reales

En el capítulo anterior se presentó la simulación, recuperación de parámetros y las pruebas de bondad y ajuste para un test dicotómico con un modelo 2PL. En este capítulo se presentará la implementación del modelo propuesto para un conjunto de datos reales. Estos datos corresponden a un conjunto de Pruebas aplicadas como proceso de selección a la Facultad de Humanidades y Educación de la Universidad Central de Venezuela en el año 2006. Esta base de datos consta de 5565 registros en los que se evaluaron aspectos como el rendimiento o éxito escolar y conductas que influyen en el rendimiento académico, entre otras variables como razonamiento verbal y conocimientos generales.

Este conjunto de datos ya fue trabajado con modelos TRI, con el objetivo de determinar si existía alguna relación entre los rasgos que evalúan la personalidad y los puntajes obtenidos en las pruebas de razonamiento, ver detalles en Páyer (2013).

Los constructos fueron contruidos respecto a los rasgos de personalidad de los individuos, en un test psicométrico aplicado a estudiantes mayores de 15 años que aspiraban ingresar a la Universidad Central de Venezuela, más específicamente a la Facultad de Humanidades y Educación en el año 2006. Además, cabe resaltar que estos rasgos de personalidad se basaron en el Inventario PIHEMA R-2 propuesto por Ponce & Granell (2006).

De manera general, este test se aplica de manera escrita donde son evaluados 160 ítems y con un tiempo límite de 30 minutos. Cada respondiente tiene tres categorías de respuesta: *verdadero*, *falso*, *no sabe*. Según Ponce & Granell (2006) este test aporta información en tres dimensiones, que en este caso se tomarán como los constructos.

Las características de cada dimensión o constructos se presentan a continuación:

1. La primera dimensión o constructo hace referencia a los índices de éxito académico, motivación y estabilidad emocional.
2. La segunda dimensión o constructo refleja los rasgos de personalidad que afectan el desarrollo académico, como son: *neuroticismo*, *indicios de psicopatía y paranoia* que son evaluados con el Inventario PIHEMA, enunciado anteriormente.

- La tercera dimensión o constructo observa el comportamiento que tiene el individuo frente a situaciones que son socialmente deseables pero imposibles de cumplir, para hacerse ver de una manera favorable en busca de una aceptación social, (Ponce & Granell (2006)).

4.2. Selección del Constructo y Análisis Preliminar

La selección del constructo fue basada en la siguiente idea. Buscar un comportamiento asimétrico en las respuestas dadas por los individuos a los constructos del test, entonces, para esto se hizo una selección empírica en principio, para luego realizar un análisis estadístico más detallado. El proceso de selección fue basado en la construcción de los histogramas de los score de los individuos en cada constructo evaluado, ver Figura 4.1. Note que la mayoría de los constructos los scores presentan un comportamiento asimétrico. Por forma y comportamiento se decidió la selección de la variable *Sinceridad* que pertenece al tercera dimensión o constructo.

De acuerdo al test, la variable *Sinceridad* consta de $p = 19$ ítems y $N = 5665$ respondientes.

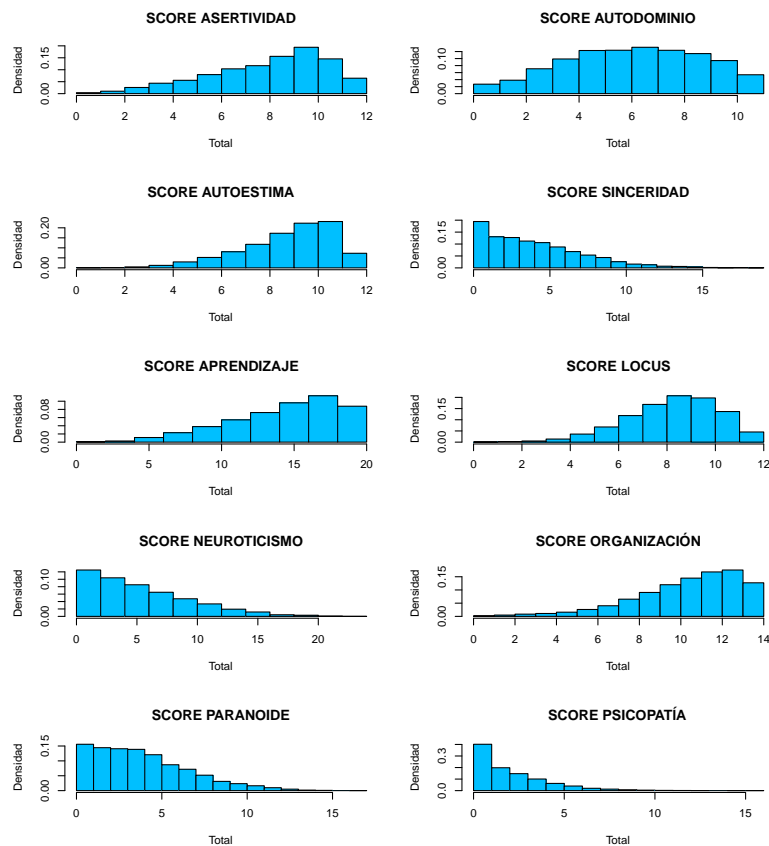


FIGURA 4.1. Histograma de los Scores por Constructo

4.3. Análisis Preliminar

Se realizará un análisis preliminar al conjunto de datos de Sinceridad para determinar la dimensionalidad del test, identificar los ítems problemáticos y los score nulos y perfectos.

Al realizar este análisis de dimensionalidad se encuentra que uno de los ítems de este test parece presentar problemas, por esta razón se decide retirar este ítem (Ver detalles en el Apendice B). Para obtener así un test que consta de $p = 18$ ítems y $N = 5665$ individuos, sobre el cual se realizará la estimación de los parámetros.

Se procede en principio con un análisis de componentes principales (ACP), así como un análisis paralelo de Horn para este nuevo test. Los resultados se presentan en la Tabla 4.1 y en la Figura 4.2

Component	Adjusted Eigenvalue	Unadjusted Eigenvalue	Estimated Bias
1	3.586881	3.684590	0.097709
2	1.607726	1.688065	0.080338
3	1.023151	1.089367	0.066215

TABLA 4.1. Resultados del Análisis Paralelo de Horn para Sinceridad

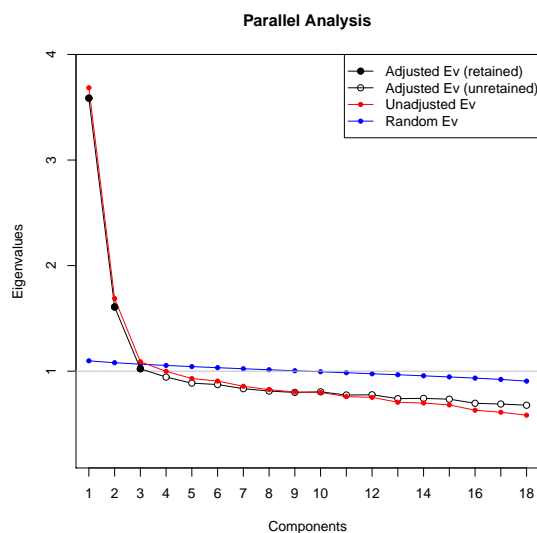


FIGURA 4.2. Análisis Paralelo de Horn para Sinceridad

En este caso, los autovalores ajustados mayores que uno indican que hay 3 componentes a retener. Pero con un ACP se tiene que el mayor porcentaje de varianza acumulada se encuentra en la primera componente, con un 20.47 % lo que podría indicar que el test es unidimensional, pero para confirmar este supuesto se construye la curva de Cronbach–Mesbah mostrando así que es una curva monótona creciente, estos resultados son presentados en la Figura 4.3, Lo que indican que el test simulado es esencialmente unidimensional.

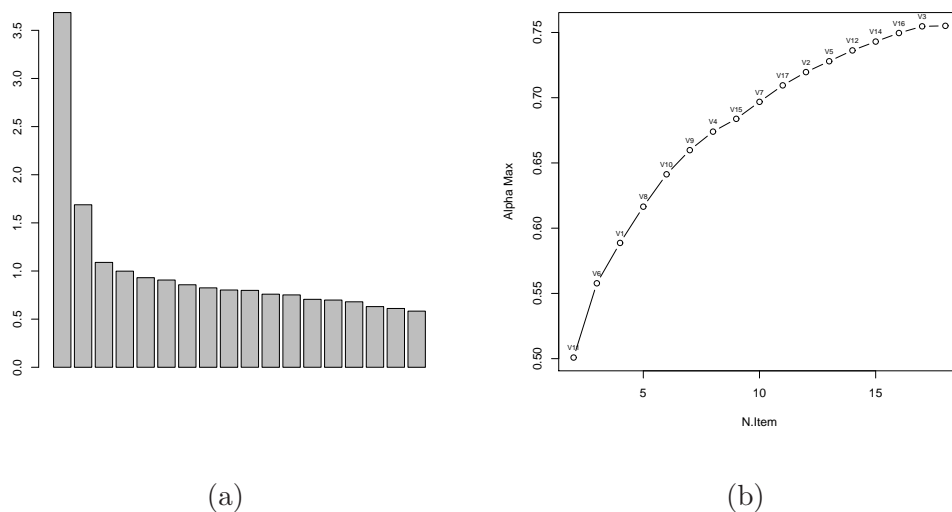


FIGURA 4.3. (a) Diagrama de Barras Autovalores Datos Sinceridad. (b) Curva de Cronbach-Mesbah Datos Sinceridad

Antes de iniciar el procedimiento de estimación se eliminan del test aquellos registros en los que el score es igual a 0 (score nulo) y aquellos en los que es igual a 18 (score perfecto). Por lo tanto, el test resultante consta entonces de $N = 5210$ y $p = 18$.

De acuerdo a los resultados anteriores y con el cumplimiento de los supuestos de la TRI, este test se caracterizará con un Modelo 2PL y de acuerdo a la teoría de estimación Bayesiana la distribución propuestas (proposal) para los parámetros objetivo se suponen de la siguiente manera

- $\theta_i \sim NT(\mu_\theta = \theta^*, \sigma_\theta^2 = \tau_\theta^2, l_\theta = 0, u_\theta = 17)$
- $b_j \sim NT(\mu_b = b^*, \sigma_b^2 = \tau_b^2, l_b = 0, u_b = 17)$
- $a_j \sim NT(\mu_a = a^*, \sigma_a^2 = \tau_a^2, l_a = 0, u_a = 5)$

donde NT es la distribución Normal Truncada.

4.4. Estimación de Parámetros

Una vez establecida la unidimensionalidad del test y el modelo de probabilidad de respuesta correcta. Se procede a la estimación ajustando un modelo logístico de dos parámetros. Se comienza al definir las distribuciones a priori para cada uno de los parámetros de los ítems y el trazo latente

- $a_j \sim NT(0, \sigma_a^2 = 1, l_a = 0, u_a = 5)$
- $b_j \sim NT(0, \sigma_b^2 = 1, l_b = 0, u_b = 17)$
- $\theta_i \sim NT(0, \sigma_\theta^2 = 1, l_\theta = 0, u_\theta = 17)$

Posteriormente se realizaron 10000 muestras con los siguientes parámetros de tuning:

$$\tau_a = 0.06 \quad \tau_b = 0.06 \quad \tau_t = 1.3$$

Las tasas de aceptación para los parámetros de los ítems y del trazo latente fueron 0.396 y 0.384 respectivamente. Tomando 2000 cadenas de calentamiento, las gráficas de las cadenas para el muestreo de los parámetros se presentan en la Figura 4.4.

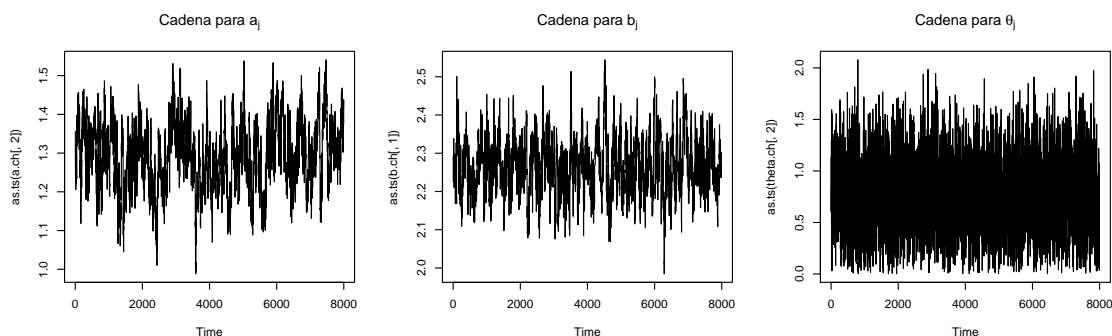


FIGURA 4.4. Cadenas para los parámetros de los ítems y el trazo latente

La estimación de cada parámetro se realizó con los valores de la media a partir del algoritmo Bayesiano. Entonces, en la Tabla 4.2 se presenta el trazo latente estimado de los primeros 10 individuos y además se presentan los valores del score (puntaje) estandarizado para los mismos. La correlación entre el trazo estimado y el score estándar es de 0.9851 y su error medio absoluto es de 0.4348; en la Figura 4.5 se presenta la comparación entre el trazo estimado y el score estándar.

i	1	2	3	4	5	6	7	8	9	10
$\hat{\theta}_i$	0.83	0.84	1.13	0.71	0.73	0.42	2.53	1.92	0.66	0.42
Score	1.33	1.00	1.33	0.67	0.67	0.34	4.29	2.97	1.00	0.34

TABLA 4.2. Valores del trazo estimado y el score estándar para los primeros 10 registros del test

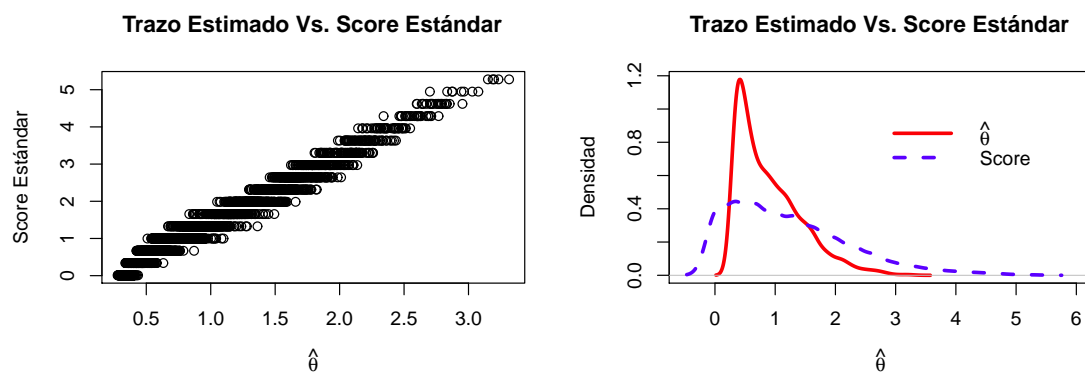


FIGURA 4.5. Trazo Estimado Vs. Score Estándar

En la Figura 4.6 se presentan además, las gráficas de las cadenas para algunos valores muestreados de θ , las gráficas de las densidades correspondientes, así como los autocorrelogramas. Se puede ver que la mayoría de las cadenas de los trazos latentes, la autocorrelación está por debajo de 0.1 entre el Lag 15 y 25.

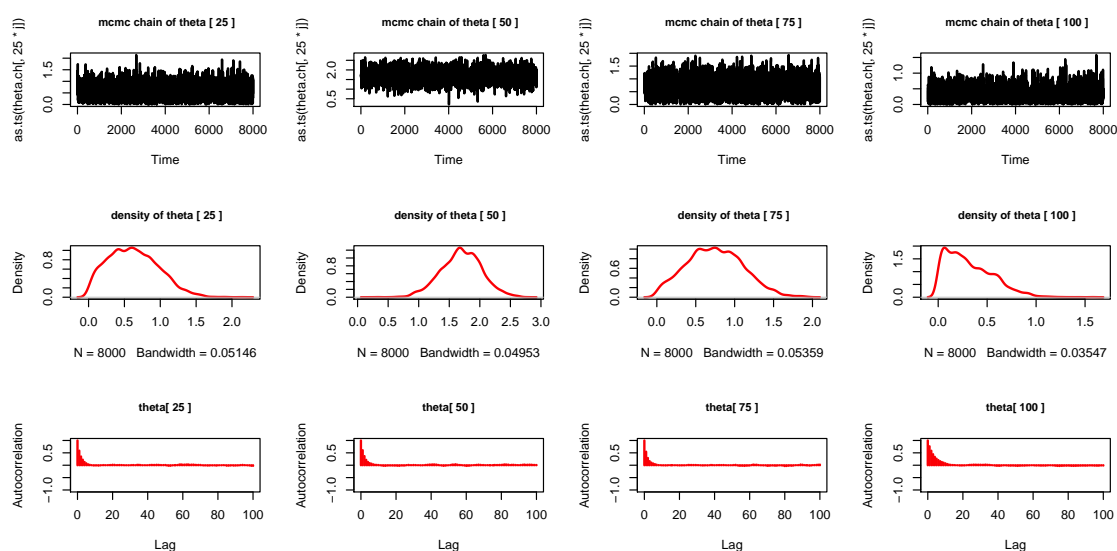


FIGURA 4.6. Gráficos cadenas, densidades y autocorrelogramas para el trazo estimado

Las estimaciones de los parámetros de los ítems que se obtuvieron a partir del algoritmo Bayesiano propuesto. Los resultados se presentan en la Tabla 4.3. Adicionalmente se presentan los gráficos de las cadenas, densidades y autocorrelogramas correspondientes para el parámetro de dificultad y el de discriminación en las Figuras 4.7 y 4.8 respectivamente. En ellas se observa que para los parámetros de dificultad, la mayoría de autocorrelaciones está por debajo de 0.1 entre los lags 50 y 80. Para los parámetros de discriminación se muestra que la mayoría de autocorrelaciones están por debajo de 0.1 entre los lags 45 y 60.

j	1	2	3	4	5	6	7	8	9
\hat{a}_j	1.25	1.28	0.63	2.01	1.16	1.76	1.65	1.38	2.46
\hat{d}_j	2.27	3.47	1.32	4.96	2.96	3.04	4.12	2.85	5.47
\hat{b}_j	1.81	2.71	2.09	2.47	2.55	1.73	2.49	2.07	2.22
j	10	11	12	13	14	15	16	17	18
\hat{a}_j	1.36	1.53	1.01	2.00	0.95	1.98	0.88	1.22	1.63
\hat{d}_j	2.35	1.92	0.70	1.42	1.63	4.86	1.55	2.77	1.73
\hat{b}_j	1.73	1.26	0.70	0.71	1.71	2.45	1.76	2.27	1.06

TABLA 4.3. Valores estimados de los parámetros de los ítems

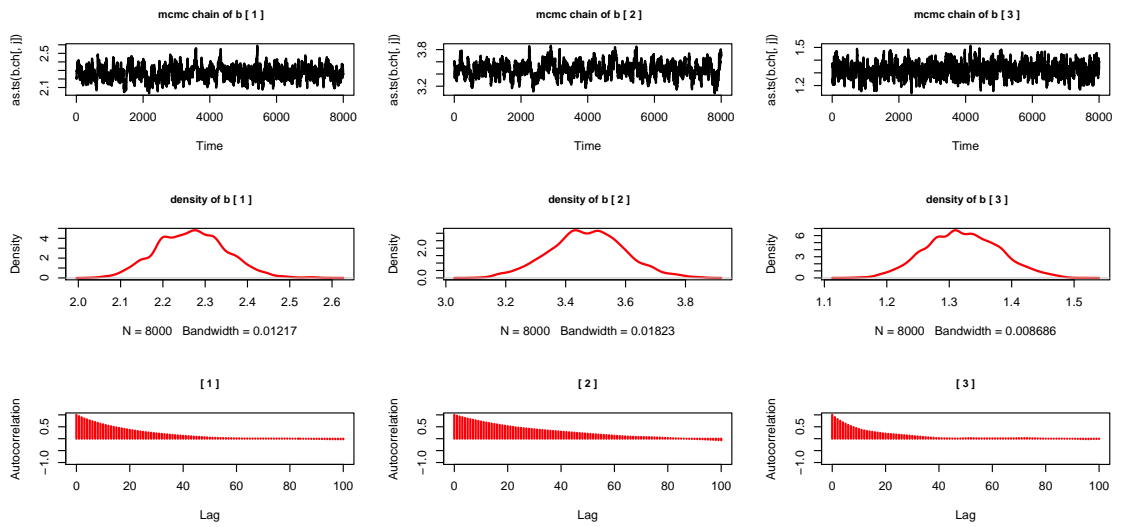


FIGURA 4.7. Gráficos bayesianos para el parámetro de dificultad

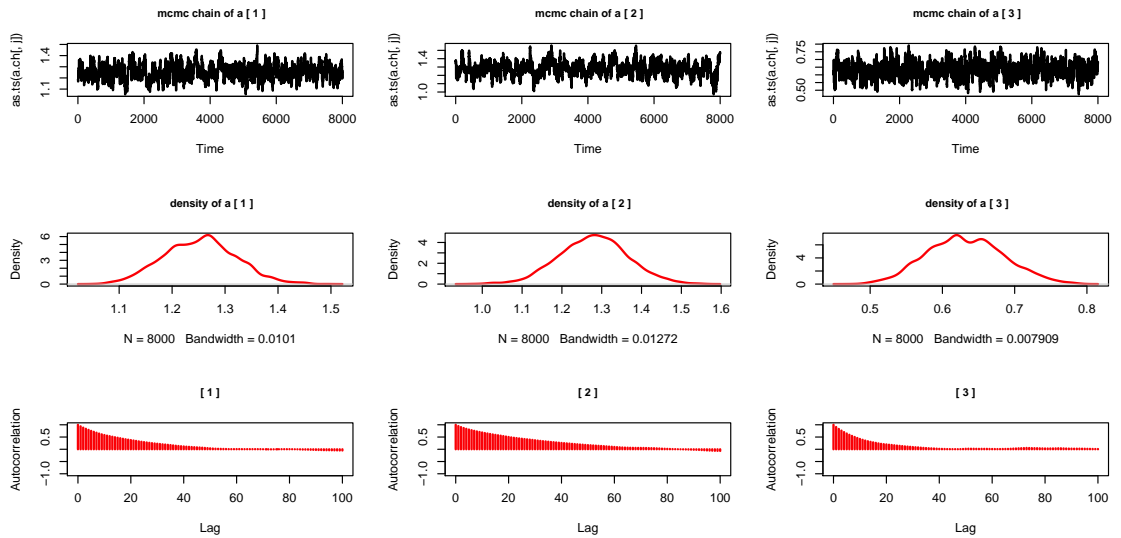


FIGURA 4.8. Gráficos bayesianos para el parámetro de discriminación

4.5. Pruebas de Bondad y Ajuste

En esta sección se presenta las pruebas de bondad y ajuste para los parámetros estimados del test de Sinceridad y la calidad de las cadenas muestreadas, de manera similar que los datos simulados (Ver Sección 3.4). La calidad de las cadenas muestreadas son analizadas con un nivel de significancia de $\alpha = 5\%$. A continuación se presenta la prueba de Heidelberg–Welch para las cadenas de cada uno de los parámetros.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.05164	0.19040	0.49700	0.48920	0.77120	0.93570

TABLA 4.4. Estadísticas descriptivas para el p -valor de Heidelberg–Welch para el parámetro de dificultad

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.1152	0.2411	0.3597	0.4683	0.6684	0.9499

TABLA 4.5. Estadísticas descriptivas para el p -valor de Heidelberg–Welch para el parámetro de discriminación

En la Tabla 4.4 y en la Tabla 4.5 se presentan las medidas descriptivas del p -valor para las cadenas de los parámetros de dificultad y discriminación respectivamente. De acuerdo a estos resultados, no se rechaza ninguna cadena de estos parámetros. Para el trazo latente, se encuentra que 36 de las 5210 cadenas o el 0.07 % no pasan la prueba con un nivel de 0.05. Dado que el porcentaje es muy bajo, se puede considerar que las autocorrelaciones para el trazo en su mayoría convergen.

Ahora, aplicando la prueba de Bondad y Ajuste, para los parámetros estimados de los ítems, se observa que estos tienen valores entre 0.27 y 0.50 (ver Tabla 4.6) lo que indica que tenemos un buen ajuste. Además, la diferencia entre las probabilidades observadas y las probabilidades estimadas de los ítems están entre 7×10^{-5} y 1.03×10^{-3} , es decir, una diferencia mínima.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.2656	0.4257	0.4522	0.4318	0.4828	0.5019

TABLA 4.6. Estadísticas descriptivas Bondad y Ajuste de los parámetros de los ítems estimados

En la Tabla 4.7 se presentan algunas estadísticas descriptivas de las estimaciones del trazo latente por el algoritmo bayesiano

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.2818	0.4663	0.7687	0.9067	1.2180	3.3070

TABLA 4.7. Estadísticas descriptivas Bondad y Ajuste para el trazo estimado

Por otro lado, de acuerdo a las pruebas de Bondad y Ajuste, se encontró que 239 de los 5210 trazos latentes estimados tienen valores por fuera del intervalo (0.05, 0.95), es decir, un 4.5 % del 100 % de los trazos no presentarían un buen ajuste. Por lo tanto, se puede concluir que el modelo Bayesiano propuesto ajusta de manera adecuada los parámetros de los datos de Sinceridad que presentan un comportamiento asimétrico.

Implementación código en R

En este capítulo se presenta la implementación del código en R para el proceso de estimación presentado en este trabajo. En cada una de la secciones se describe la función de cada código

A.1. Generación de valores aleatorios con Distribución Normal Truncada

Se presenta la incorporación del código trabajado para la generación de valores aleatorios que siguen una distribución Normal Truncada con media $\mu = 0$, $\sigma = 1$ y límites inferior y superior 0 e ∞ respectivamente ($NT(\mu = 0, \sigma = 1, l = 0, u = \infty)$).

```
rtnorm <-function (n, mean = 0, sd = 1, lower = 0, upper = Inf)
{
  if (length(n) > 1)
    n <- length(n)
  mean <- rep(mean, length = n)
  sd <- rep(sd, length = n)
  lower <- rep(lower, length = n)
  upper <- rep(upper, length = n)
  lower <- (lower - mean)/sd
  upper <- (upper - mean)/sd
  ind <- seq(length = n)
  ret <- numeric(n)
  alg <- ifelse(lower > upper, -1, ifelse(((lower < 0 & upper ==
  Inf) | (lower == -Inf & upper > 0) | (is.finite(lower) &
  is.finite(upper) & (lower < 0) & (upper > 0) & (upper -
  lower > sqrt(2 * pi))))), 0, ifelse((lower >= 0 & (upper >
  lower + 2 * sqrt(exp(1))/(lower + sqrt(lower^2 + 4)) *
  exp((lower * 2 - lower * sqrt(lower^2 + 4))/4))),
  1, ifelse(upper <= 0 & (-lower > -upper + 2 * sqrt(exp(1))/(-upper+
  sqrt(upper^2 + 4)) * exp((upper * 2 - -upper * sqrt(upper^2 +
  4))/4)), 2, 3))))
```

```

ind.nan <- ind[alg == -1]
ind.no <- ind[alg == 0]
ind.expl <- ind[alg == 1]
ind.expu <- ind[alg == 2]
ind.u <- ind[alg == 3]
ret[ind.nan] <- NaN
while (length(ind.no) > 0) {
  y <- rnorm(length(ind.no))
  done <- which(y >= lower[ind.no] & y <= upper[ind.no])
  ret[ind.no[done]] <- y[done]
  ind.no <- setdiff(ind.no, ind.no[done])
}
stopifnot(length(ind.no) == 0)
while (length(ind.expl) > 0) {
  a <- (lower[ind.expl] + sqrt(lower[ind.expl]^2 + 4))/2
  z <- rexp(length(ind.expl), a) + lower[ind.expl]
  u <- runif(length(ind.expl))
  done <- which((u <= exp(-(z - a)^2/2)) & (z <= upper[ind.expl]))
  ret[ind.expl[done]] <- z[done]
  ind.expl <- setdiff(ind.expl, ind.expl[done])
}
stopifnot(length(ind.expl) == 0)
while (length(ind.expu) > 0) {
  a <- (-upper[ind.expu] + sqrt(upper[ind.expu]^2 + 4))/2
  z <- rexp(length(ind.expu), a) - upper[ind.expu]
  u <- runif(length(ind.expu))
  done <- which((u <= exp(-(z - a)^2/2)) & (z <= -lower[ind.expu]))
  ret[ind.expu[done]] <- -z[done]
  ind.expu <- setdiff(ind.expu, ind.expu[done])
}
stopifnot(length(ind.expu) == 0)
while (length(ind.u) > 0) {
  z <- runif(length(ind.u), lower[ind.u], upper[ind.u])
  rho <- ifelse(lower[ind.u] > 0, exp((lower[ind.u]^2 -
    z^2)/2), ifelse(upper[ind.u] < 0, exp((upper[ind.u]^2 -
    z^2)/2), exp(-z^2/2)))
  u <- runif(length(ind.u))
  done <- which(u <= rho)
  ret[ind.u[done]] <- z[done]
  ind.u <- setdiff(ind.u, ind.u[done])
}
stopifnot(length(ind.u) == 0)
ret * sd + mean
}

```

A.2. Algoritmo de Metropolis–Hasting

```

library("lattice")
library("coda")
source("rtnorm.r")

# x: Datos del Test
# type.model=2 (Modelo 2PL)
# mcmc_size: Número de Muestras
# file.name: Nombre de Archivo para guardar los resultados
# show.iteration=TRUE (Muestra la iteraciones)
# sigma_t: Desviación del parámetro de theta para la a priori
# sigma_a: Desviación para las a priores de a
# sigma_b: Desviación para las apriores de b
# tau_a : Parámetro de tunning del parámetro a
# tau_b: Parámetro de tunning del parámetro b
# tau_t: Parámetro de tunning para el trazo (proposal)
# l_theta: Límite inferior de theta (proposal)
# u_theta: Límite superior de theta (proposal)
# l_a: Límite inferior de a (proposal)
# u_a: Límite superior de a (proposal)
# l_b: Límite inferior de b (proposal)
# u_b: Límite superior de b (proposal)

irt.Metropolis <- function(x, file.name=NA,type.model=2, mcmc_size=10000,
                           show.iteration=TRUE,sigma_a= 1, sigma_b= 1,sigma_t=1,
                           tau_t=1.6, tau_a = 0.06, tau_b=0.06,l_theta=0,u_theta=17,
                           l_b=0,u_b=17, l_a=0,u_a=5){

  if (missing(x))
    {stop(" array x containig the sample must be specified...")}

# N: Tamaño de la muestra
  N = dim(x)[1]

# K: Tamaño del Test
  K = dim(x)[2]

# Guarda las muestras con valores iniciales

  a.smpl = matrix(1,nrow = mcmc_size, ncol = K)
  b.smpl = matrix(0,nrow = mcmc_size, ncol = K)
  trazo.smpl = matrix(0,nrow = mcmc_size, ncol = N)

# Taza de aceptación para Theta
  accept.rate.t = 0

# Taza de aceptación para los ítems
  accept.rate.i = 0

```

```
#####
#     Valores Iniciales Cadenas     #
#####

trait.smpl[1,] = rtnorm(N,mean=0,sd=sigma_t,lower=l_theta,upper=u_theta)
  b.smpl[1,] = rtnorm(K, mean = 0, sd = sigma_b,lower=l_b,upper=u_b)
  a.smpl[1,] = rtnorm(K, mean = 0,sd = sigma_a,lower = l_a,upper = u_a)

#####
#     Valores Iniciales Proposal     #
#####

theta.c = trazo.smpl[1,]
a.c = a.smpl[1,]
b.c = b.smpl[1,]

times <- Sys.time()

for (i in 2:mcmc_size ){
#####
# Cálculo del numerador para los trazos#
#####
theta.c = rtnorm(N,mean=theta.c,sd=tau_t,lower=l_theta,upper=u_theta)
eta = matrix(theta.c,ncol=1)%*% matrix(a.smpl[i-1,],nrow=1)-
      matrix(b.smpl[i-1,],N,K,byrow=TRUE)
P = 1/(1+exp(-eta))
L.post.num=matrix(apply(iffelse(x=log(P),log(1-P)),1,sum),N,1)-
      matrix((theta.c/sigma_t)^2/2,N,1)
#####
# Cálculo del denominador de los trazos#
#####
eta = matrix(trazo.smpl[i-1,],ncol=1)%*% matrix(a.smpl[i-1,],nrow=1)-
      matrix(b.smpl[i-1,],N,K,byrow=TRUE)
P =1/(1+exp(-eta))
L.post.den = matrix(apply(iffelse(x=log(P),log(1-P)),1,sum),N,1)-
      matrix((trazo.smpl[i-1,]/sigma_t)^2/2,N,1)

alpha = exp(L.post.num - L.post.den)
  alpha = iffelse(alpha>1,1,alpha)

# Prueba de aceptación
  accept = iffelse(alpha>runif(N),TRUE,FALSE)
  accept.rate.t = accept.rate.t + sum(accept)
  theta.c = trazo.smpl[i,] = iffelse(accept, theta.c, trazo.smpl[i-1,])

#####
# Cálculo del numerador para los ítems #
#####
```

```

a.c=rtnorm(K, mean = a.c, sd = tau_a, lower = l_a, upper = u_a)
b.c=rtnorm(K, mean = b.c, sd = tau_b, lower = l_b, upper = u_b)
eta=matrix(trazo.smpl[i,],ncol=1)%*% matrix(a.c,nrow=1)-
      matrix(b.c,N,K,byrow=TRUE)
P =1/(1+exp(-eta))
L.post.num = matrix(apply(ifelse(x,log(P),log(1-P)),2,sum),K,1)-
      matrix((b.c/sigma_b)^2/2,K,1)-
      matrix((a.c/sigma_a)^2/2,K,1)
#####
# Cálculo del denominador para los ítems #
#####
eta = matrix(trazo.smpl[i,],ncol=1)%*% matrix(a.smpl[i-1,],nrow=1)-
      matrix(b.smpl[i-1,],N,K,byrow=TRUE)
P =1/(1+exp(-eta))
L.post.den = matrix(apply(ifelse(x,log(P),log(1-P)),2,sum),K,1)-
      matrix((b.smpl[i-1,]/sigma_b)^2/2,K,1)-
      matrix((a.smpl[i-1,]/sigma_a)^2/2,K,1)

alpha = exp(L.post.num - L.post.den)
alpha = ifelse(alpha>1,1,alpha)

# Prueba de aceptación
accept = ifelse(alpha>runif(K),TRUE,FALSE)
accept.rate.i = accept.rate.i + sum(accept)
a.c = a.smpl[i,] = ifelse(accept, a.c, a.smpl[i-1,])
b.c = b.smpl[i,] = ifelse(accept, b.c, b.smpl[i-1,])

# Muestra iteraciones

if(show.iteration){
  cat( i,"\n " )
  }
}

# Muestra Tiempo
times = Sys.time() - times
cat( "\n mcmc cicle time: ",times,"\n")

#return
accept.rate.t =accept.rate.t/(mcmc_size*N)
accept.rate.i =accept.rate.i/(mcmc_size*K)
l = list(trait.smpl=trait.smpl, a.smpl = a.smpl,
        b.smpl = b.smpl,accept.rate.t= accept.rate.t,
        accept.rate.i=accept.rate.i)
l
}

```

A.3. Generación de Cadenas

Se realiza la generación de cadenas con el algoritmo de Metropolis–Hasting presentado en la sección anterior con los datos del test a trabajar.

```
# Lectura de datos

datos<-read.table("Test_Depurado.txt",header=F)
Y<-as.matrix(datos)

N<-dim(datos)[1]
p<-dim(datos)[2]

#####
# GENERACIÓN DE CADENAS #
#####

samples=irt.Metropolis(Y, file.name=NA,type.model=2, mcmc_size=10000,
                        show.iteration=TRUE,sigma_a= 1, sigma_b= 1,
                        sigma_t=1.0, tau_t=1.3, tau_a = 0.06, tau_b=0.06,
                        l_theta=0,u_theta=17, l_b=0,u_b=17, l_a=0,u_a=5)
accept.t = samples$accept.rate.t
accept.t

accept.i = samples$accept.rate.i
accept.i

#Guarda las cadenas como un .RData

salida = list(samples=samples,accept.t=accept.t,accept.i=accept.i)
save(salida,file ="Ruta destino/Archivo.RData")
```

A.4. Estimación Conjunta de Parámetros y GoF

A continuación se presenta el código para la estimación conjunta de parámetros con el algoritmo Bayesiano propuesto.

```
library("lattice")
library("coda")
library(xtable)

#Lectura del Archivo .RData

load("Ruta Destino/Archivo.RData")
samples=salida$samples
salida$accept.t
```

```

salida$accept.i

# Extracción de muestras (Cadenas)

burning= 1:2000

theta.ch =samples$trazo.smpl[-burning,]
a.ch = samples$a.smpl[-burning,]
b.ch = samples$b.smpl[-burning,]

dim(theta.ch)
dim(a.ch)
dim(b.ch)

# Gráficas de las cadenas para los parámetros
par(mfrow=c(1,3))
plot(as.ts(a.ch[,2]), main=expression(paste('Cadena para ', a[j])))
plot(as.ts( b.ch[,1]), main=expression(paste('Cadena para ', b[j])))
plot(as.ts( theta.ch[,2]), main=expression(paste('Cadena para ', theta[j])))

# Estimación de Parámetros

theta.mean= apply(theta.ch,2,mean)
a.mean= apply(a.ch,2,mean)
d.mean= apply(b.ch,2,mean)
b.mean=d.mean/a.mean

#####
# PRUEBA DE BONDAD Y AJUSTE #
#####
# size.t es tamaño muestreo menos calentamiento (10000-2000)

size.t=8000
prob.item.chain = matrix(NA,size.t,p)

z.model = 0
z.item = array(0,dim=p)
z.person = array(0,dim=N)

# Generación del test replicado
for (i in 1:size.t){
  if(i%10==0){print( i)}
  eta.s=(theta.ch[i,])%*%t(a.ch[i,])-matrix(b.ch[i,],N,p, byrow=TRUE)
  prob.s=1/(1+exp(-eta.s))
  prob.item.chain[i,] = apply(prob.s,2,mean)
  y.sim = ifelse( prob.s>matrix(runif(N*p),N,p),1,0)
  log.like.dat = -2*(ifelse(Y==1,log(prob.s),log(1-prob.s)))
  log.like.sim = -2*(ifelse(y.sim==1,log(prob.s),log(1-prob.s)))
  log.like.dat.item = apply(log.like.dat,2,sum)
}

```

```

log.like.sim.item = apply(log.like.sim,2,sum)
z.item = z.item + ifelse(log.like.dat.item >log.like.sim.item,1,0 )
log.like.dat.person = apply(log.like.dat,1,sum)
log.like.sim.person = apply(log.like.sim,1,sum)
z.person = z.person + ifelse(log.like.dat.person >log.like.sim.person,1,0 )
log.like.dat.model = sum(log.like.dat)
log.like.sim.model = sum(log.like.sim)
z.model = z.model+ ifelse(log.like.dat.model >log.like.sim.model,1,0 )
}

prob.item.fit = apply(prob.item.chain,2,mean)

# Prueba de Bondad y Ajuste

prob.obs=apply(Y,2,mean)
probs=cbind(prob.item.fit,prob.obs)
probs
summary(prob.item.fit-prob.obs)

z.item = z.item/size.t
z.person = z.person/size.t
z.model = z.model/size.t

summary(z.item)
summary(z.person)

#Número de Trazos latentes fuera del intervalo del 95% de confianza

idx.per = which(z.person<0.975 &z.person>0.025)
setdiff(1:N,idx.per)
length(which(z.person<0.05))
length(which(z.person>0.95))

#####
# Convergencia de Heidelberg: #
#Si p-value<0.05, La prueba de estacionaridad No es aprobada#
#####

# parameter b
p.value_heidel.t.b=heidel.diag(mcmc(b.ch/a.ch))
p.value_heidel.t.b=p.value_heidel.t.b[,3]
summary(p.value_heidel.t.b)
which(p.value_heidel.t.b<0.05)

#parameter a
p.value_heidel.t.a=heidel.diag(mcmc(a.ch))
p.value_heidel.t.a=p.value_heidel.t.a[,3]
summary(p.value_heidel.t.a)

```

```
which(p.value_heidel.t.a<0.05)

#latent traits
p.value_heidel.t.t=heidel.diag((theta.ch))
p.value_heidel.t.t=p.value_heidel.t.t[,3]
summary(p.value_heidel.t.t)
which(p.value_heidel.t.t<0.05)
```

APÉNDICE B

Análisis de Dimensionalidad Datos Reales

Se realizará un análisis preliminar al conjunto de datos de Sinceridad para determinar la dimensionalidad del test, identificar los ítems problemáticos y los score nulos y perfectos.

El proceso de comienza con un análisis de componentes principales (ACP) al constructo de sinceridad que tiene $N = 5665$ individuos y $p = 19$ ítems. Luego, se tiene que el mayor porcentaje de varianza acumulada se encuentra en la primera y en la segunda componente con un 19.46 % y 8.9 % respectivamente.

Ahora, haciendo el análisis paralelo de Horn (1965), se obtienen que los autovalores ajustados mayores que uno indican que hay 3 componentes a retener. Los resultados son presentados en la Tabla B.1 y en la Figura B.1.

Component	Adjusted Eigenvalue	Unadjusted Eigenvalue	Estimated Bias
1	3.597877	3.699131	0.101253
2	1.608065	1.691934	0.083868
3	1.019353	1.089561	0.070207

TABLA B.1. Resultados del Análisis Paralelo de Horn

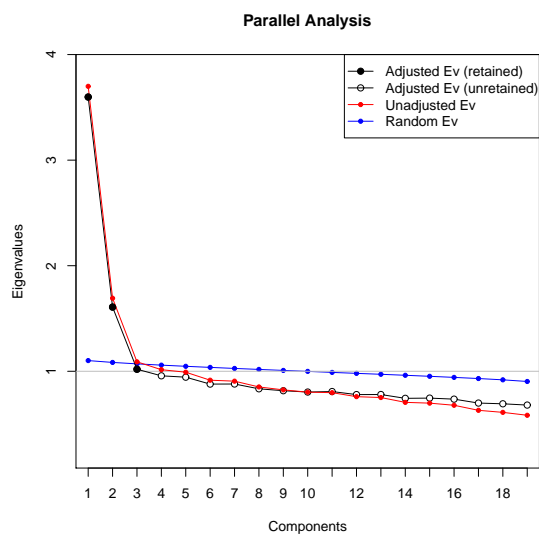


FIGURA B.1. Análisis Paralelo de Horn

Luego, se decide ver y analizar la curva de Cronbach–Mesbah para ver el comportamiento que tiene este test en sus ítems. La curva es monótona creciente, pero en el ítem 12 parece que se presenta un problema, dado que la curva en este punto empieza a decrecer, como se muestra en la Figura B.2. Luego se decide retirar el ítem. Por lo tanto, el Test de Sinceridad a ajustar constará ahora de 18 ítems. La continuación del proceso se puede revisar en el Capítulo 4.

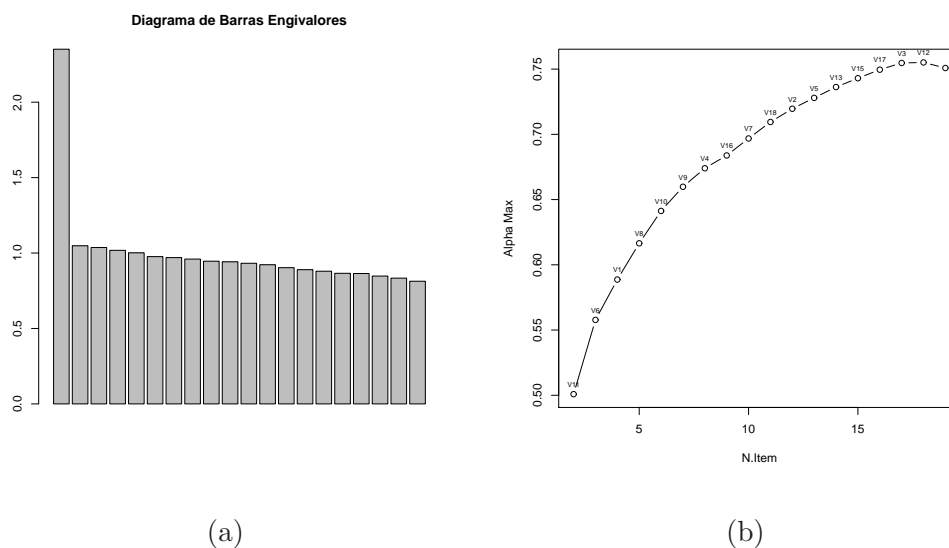


FIGURA B.2. (a) Diagrama de Barras Autovalores Datos Simulados. (b) Curva de Cronbach–Mesbah Datos Simulados

Simulación Otro Test

Se presentará la recuperación de parámetros para otros dos conjuntos de datos simulados, como se presentó en el Capítulo 3.

C.1. Generación del Test

La probabilidad de que el individuo i responda correctamente al ítem j está dada por

$$P_{ij}(Y_{ij}|\theta_i) = \frac{1}{1 + e^{-a_j(\theta_i - b_j)}} = \frac{1}{1 + e^{-a_j\theta_i + d_j}} \quad (\text{C.1})$$

para $i = 1, 2, \dots, N$ y $j = 1, 2, \dots, p$. Nuevamente se genera un test dicotómico en el software R con la siguiente información y una semilla para reproducir el mismo test varias veces (`set.seed(1208)`)

- $p = 20$ ítems y $N = 5000$ individuos
- Modelo 2PL (logit de dos parámetros).
- $a_j \sim LN(0, 1/4)$
- $b_j \sim NT(0, \sigma_b^2 = 1, l_b = 0, u_b = 17)$
- $\theta_i \sim NT(0, \sigma_\theta^2 = 1, l_\theta = 0, u_\theta = 17)$

Donde NT y LN son las distribuciones Normal Truncada y Log–Normal respectivamente.

Se presentan algunas medidas descriptivas de los parámetros simulados con estas distribuciones en la Tabla C.1

	1	2	3
Min.	0.63	0.14	0.00
1st Qu.	0.78	0.34	0.32
Median	0.94	0.80	0.68
Mean	0.95	0.87	0.81
3rd Qu.	1.00	1.23	1.16
Max.	1.53	2.47	4.28

TABLA C.1. Medidas Descriptivas de los parámetros a_j , b_j y θ_i Simulados

C.2. Recuperación de Parámetros

Una vez generado el test con los supuestos básicos de la TRI y evidenciando que el test simulado es esencialmente unidimensional, se realiza la depuración del mismo, es decir, se excluirán aquellos individuos que presentan un score perfecto y un score nulo. Por lo tanto, el algoritmo propuesto en la recuperación de los parámetros se aplicará ahora a un Test de $p = 20$ ítems y $N = 4995$ individuos.

Se realizó la recuperación ajustando un modelo logístico de dos parámetros, de igual manera que en el Capítulo 3. Posteriormente se realizaron 10000 muestras con los siguientes parámetros de tuning:

$$\tau_a = 0.06 \quad \tau_b = 0.06 \quad \tau_t = 1.6$$

Las tasas de aceptación para los parámetros de los ítems y del trazo latente fueron 0.342 y 0.385 respectivamente. Tomando 2000 cadenas de calentamiento, las gráficas para el muestreo de los parámetros se presentan en la Figura C.1.

La recuperación de cada parámetro se realizó utilizando la media. En primer lugar, se presenta los resultados del trazo latente recuperado de los primeros 10 individuos descritos en la Tabla C.2 en la que, además, se presentan los valores del score (puntaje) estandarizado para los mismos individuos. La correlación entre el trazo estimado y el score estándar es de 0.97.

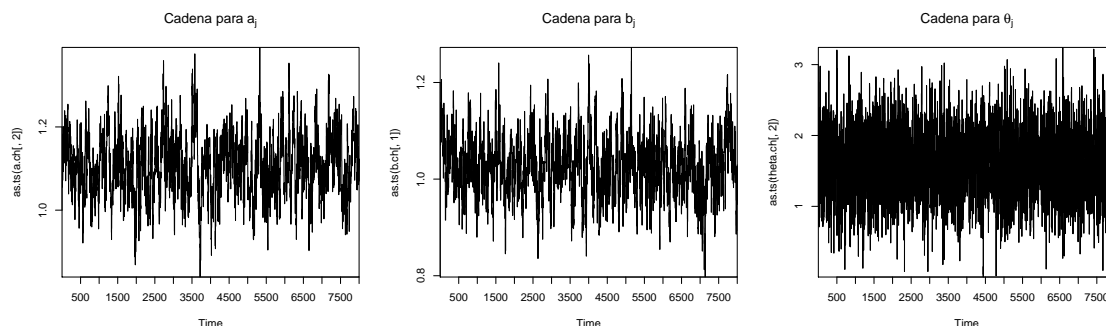


FIGURA C.1. Cadenas para los parámetros de los ítems y el trazo latente

i	1	2	3	4	5	6	7	8	9	10
$\hat{\theta}_i$	0.60	1.63	1.04	0.53	0.94	2.21	0.76	0.98	0.53	0.87
Score	2.14	3.97	3.05	2.14	2.75	4.88	2.44	3.05	1.84	3.05

TABLA C.2. Valores del trazo estimado y el score estándar para los primeros 10 registros del test

En la Figura C.2 se presentan además, las gráficas de las cadenas para algunos valores de θ , las gráficas de las densidades correspondientes, así como los autocorrelogramas, en los que se muestra que para la mayoría de las cadenas de los trazos latentes, la autocorrelación está por debajo de 0.1 entre el Lag 8 y 15.

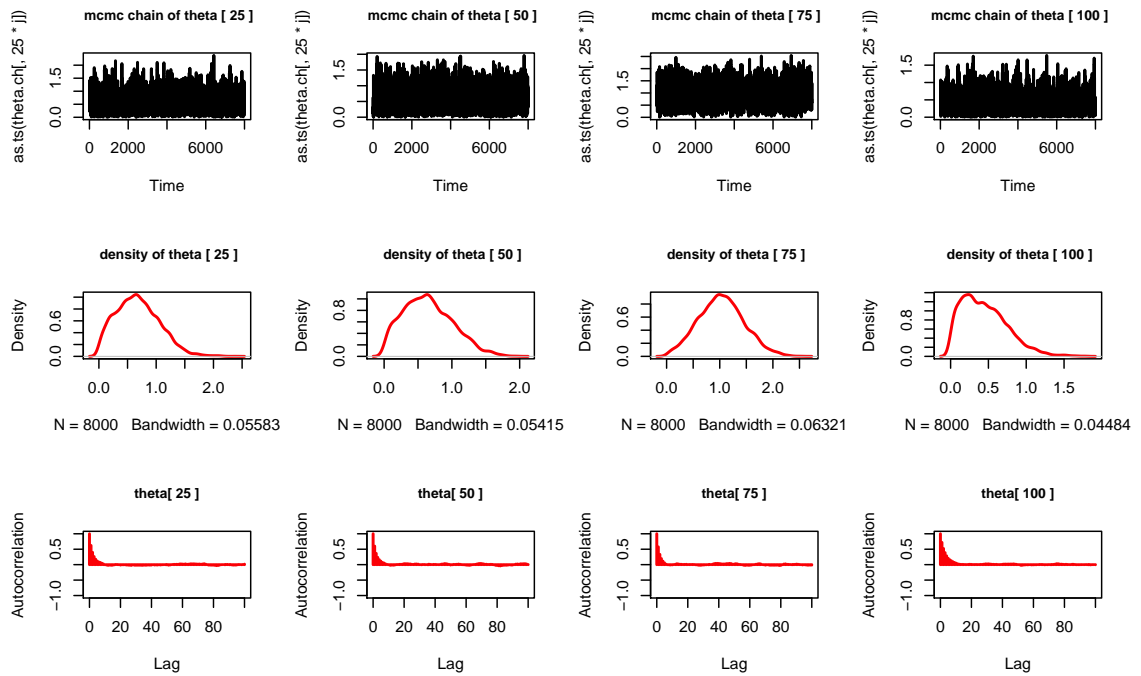


FIGURA C.2. Gráficos Cadenas, densidad y autocorrelograma para el trazo latente θ

La recuperación de los parámetros de los ítems también se obtuvo a partir de la media. A continuación se presentan los gráficos de las cadenas, autocorrelación y densidades de las mismas para los parámetros de los ítems. Las Figuras C.3 y C.4 corresponden a los parámetros de dificultad y discriminación respectivamente.

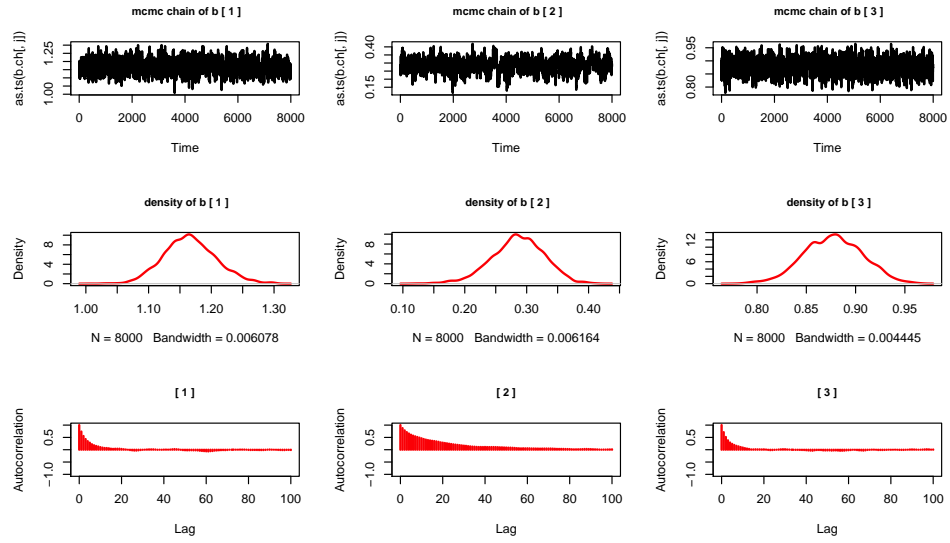


FIGURA C.3. Gráficos Cadenas, densidad y autocorrelograma para el parámetro de dificultad

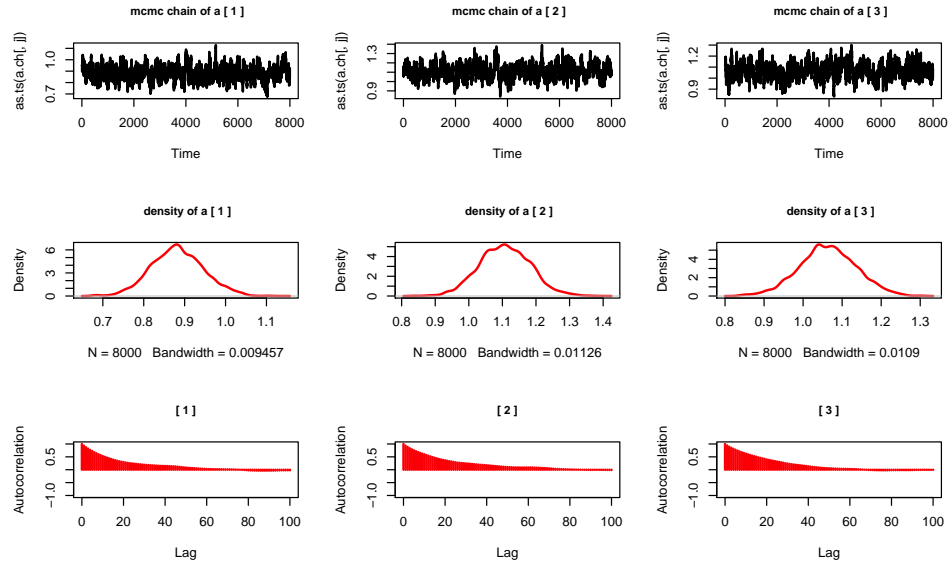


FIGURA C.4. Gráficos Cadenas, densidad y autocorrelograma para el parámetro de discriminación

En la Figura C.3 se muestra que para los parámetros de dificultad, la mayoría de autocorrelaciones está por debajo de 0.1 entre los Lags 60 y 80. Para los parámetros de discriminación la Figura C.4 muestra que la mayoría de autocorrelaciones están por debajo de 0.1 entre los Lags 60 y 80.

C.3. Comparación entre los valores simulados y recuperados

Se procede ahora a comparar los valores de los parámetros estimados con los de la simulación inicial para determinar si la metodología los recupera de manera apropiada.

En primer lugar, se compararon los parámetros de dificultad, la correlación en este caso resultó ser igual a 0.9938, en tanto que el error absoluto estuvo entre 0.0169 y 0.4551 como se indica en la Tabla C.3, adicionalmente, en la Figura C.5 se contrastan los valores simulados y recuperados para el parámetro b_j .

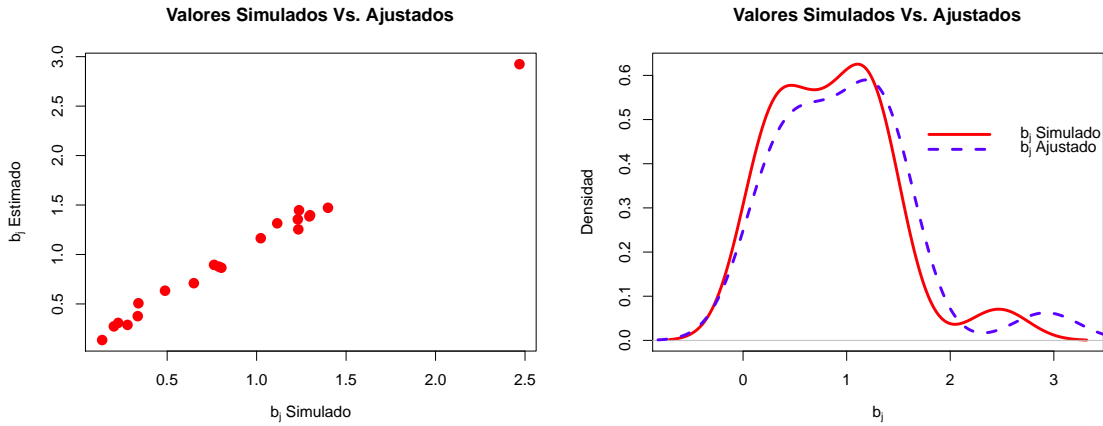


FIGURA C.5. Comparación entre los valores simulados y recuperados para el parámetro de dificultad

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	sd
0.001695	0.062510	0.089710	0.114000	0.142500	0.455100	0.0992

TABLA C.3. Estadísticas descriptivas para el Error Absoluto del parámetro de dificultad

Para el parámetro de discriminación se obtuvo una correlación de 0.8825 entre los valores simulados y los recuperados, además el error absoluto estuvo entre 0.015 y 0.25 con una desviación de 0.071 como se indica en la Tabla C.4. Al igual que se hizo con el parámetro de dificultad, en la Figura C.6 se realiza una comparación entre los valores recuperados y simulados para a_j .

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	sd
0.01560	0.03944	0.04881	0.07990	0.08852	0.24910	0.07155093

TABLA C.4. Estadísticas descriptivas para el Error Absoluto del parámetro de discriminación

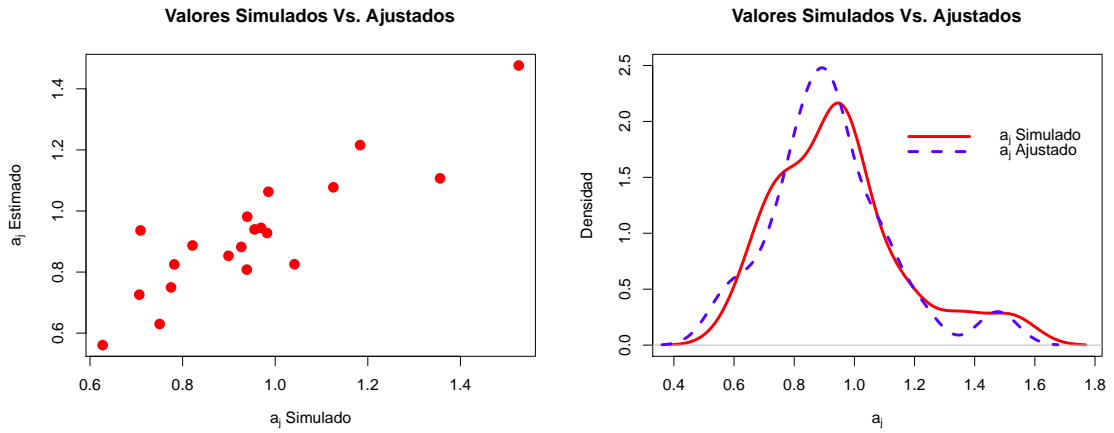


FIGURA C.6. Comparación entre los valores simulados y estimados para el parámetro de discriminación

Realizando el mismo procedimiento para el trazo, se encontró que el error absoluto fue en promedio de 0.2520 con una desviación estándar igual a 0.237, mientras que la correlación entre el simulado y el recuperado fue de 0.79164. Como se presenta Tabla C.5 y la Figura C.7.

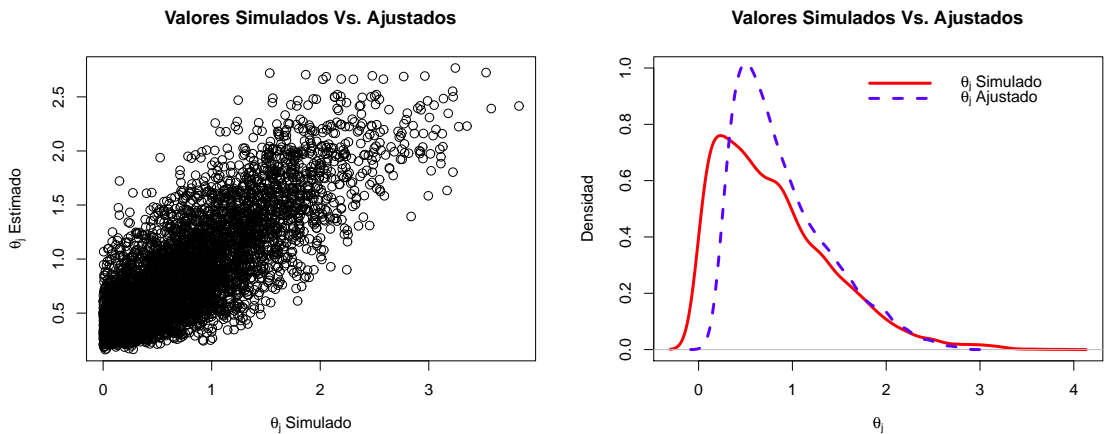


FIGURA C.7. Comparación entre los valores simulados y recuperados del Trazo

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	sd
0.000024	0.120500	0.252500	0.305000	0.433000	1.572000	0.23709

TABLA C.5. Estadísticas descriptivas para el Error Absoluto del Trazo

En la Tabla C.6 se presentan los valores simulados, recuperados y el error absoluto para cada uno de los parámetros de los ítems.

j	a_j	\hat{a}_j	$ a_j - \hat{a}_j $	b_j	\hat{b}_j	$ b_j - \hat{b}_j $
1	0.927	0.882	0.044	1.023	1.164	0.141
2	1.356	1.107	0.249	0.278	0.288	0.010
3	0.985	1.063	0.078	0.788	0.877	0.089
4	0.940	0.981	0.042	1.295	1.385	0.090
5	0.709	0.936	0.227	0.339	0.507	0.168
6	0.939	0.808	0.131	2.469	2.924	0.455
7	1.183	1.216	0.032	0.226	0.308	0.082
8	0.782	0.825	0.043	1.233	1.254	0.021
9	0.982	0.928	0.054	0.803	0.866	0.063
10	0.821	0.887	0.066	0.649	0.710	0.061
11	1.042	0.826	0.216	1.237	1.447	0.210
12	1.126	1.078	0.048	0.762	0.895	0.133
13	0.969	0.945	0.025	1.230	1.355	0.125
14	0.899	0.853	0.046	1.399	1.472	0.073
15	0.955	0.940	0.016	1.299	1.397	0.098
16	0.750	0.630	0.120	0.137	0.135	0.002
17	0.706	0.726	0.020	0.488	0.634	0.146
18	0.775	0.750	0.025	1.115	1.315	0.200
19	0.627	0.561	0.066	0.202	0.273	0.071
20	1.526	1.476	0.050	0.336	0.377	0.041

TABLA C.6. Comparación entre los valores simulados y recuperados de los parámetros de los ítems

De acuerdo a lo anterior, el algoritmo parece recuperar de manera apropiada los parámetros simulados.

C.4. Pruebas de Bondad y Ajuste

Se realizan las pruebas de bondad y ajuste para los parámetros recuperados y se presenta consigo la calidad de las cadenas muestreadas, basadas en la prueba estacionaria de Heidelberg & Welch (1983).

La calidad de las cadenas muestreadas son analizadas con un nivel de significancia de $\alpha = 5\%$ y haciendo la comparación con el p -valor, se observa que cantidad de cadenas son rechazadas. Es decir, si el p -valor es menor que α la cadena no pasa el prueba Heidelberg-Welch. A continuación se presenta los resultados de esta prueba para las cadenas de cada uno de los parámetros.

En la Tabla C.7 y en la Tabla C.8 se presentan las medidas descriptivas del p -valor para las cadenas de los parámetros de dificultad y discriminación respectivamente. De acuerdo a estos resultados, se puede observar que ninguna cadena es rechazada, es decir, las cadenas presentan una buena calidad de convergencia.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.05891	0.19660	0.34420	0.42840	0.61070	0.96750

TABLA C.7. Estadísticas descriptivas para el p -valor de Heidelberg-Welch para el parámetro de dificultad

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.1439	0.2964	0.4466	0.4995	0.6818	0.9768

TABLA C.8. Estadísticas descriptivas para el p -valor de Heidelberg–Welch para el parámetro de discriminación

Para el trazo latente, se encuentra que 42 de las 4995 cadenas o el 0.84% no pasan la prueba con un nivel de 0.05. Dado que el porcentaje es muy bajo, se puede considerar que las autocorrelaciones para el trazo en su mayoría convergen. Por lo que se puede continuar con el método de recuperación.

Ahora, el método de Bondad y Ajuste, está basado de acuerdo al algoritmo descrito en la Sección 2.3. Así, para los parámetros recuperados de los ítems, tenemos que sus valores están entre 0.45 y 0.52 (ver Tabla C.9) lo que indica que tenemos un buen ajuste. Además, se calculó la diferencia entre las probabilidades observadas y las probabilidades recuperadas de respuesta correcta para cada ítem, y se tiene que sus valores están entre -9.5×10^{-4} y 5.15×10^{-4} , es decir, una diferencia mínima.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.4580	0.4836	0.5004	0.4985	0.5102	0.5259

TABLA C.9. Estadísticas descriptivas Bondad y Ajuste de los parámetros de los ítems recuperados

Por otro lado, de acuerdo a las pruebas de Bondad y Ajuste, se encontró que 202 de los 4995 trazos latentes recuperados tienen valores por fuera del intervalo (0.05, 0.95), es decir, un 4% del 100% de los trazos no presentarían un buen ajuste. Por lo tanto, se puede concluir que el modelo propuesto ajusta de manera adecuada los parámetros de los datos simulados. Es decir, que los parámetros recuperados tienen una buena aproximación a los parámetros poblacionales.

Discusión y Conclusiones

En la TRI el supuesto de simetría en el trazo latente es muy común y con base a éste se realiza el proceso de estimación de la variable no observada, así como de los parámetros de los ítems. En las estimaciones clásicas el trabajo computacional es demasiado arduo, por esta razón el uso de la estadística Bayesiana es una opción que ha arrojado resultados aceptables.

Son diversos los trabajos e investigaciones sobre el uso de la estadística Bayesiana en la TRI en los que se proponen métodos de estimación que computacionalmente son más eficientes, por ejemplo, en los trabajos realizados por Mislevy (1985), Fox (2010), Andrade et al. (2000), entre otros, se presentan algunos métodos y algoritmos Bayesianos para la TRI, pero todos siguen enmarcados en el supuesto de la simetría para el trazo latente θ .

Por experiencia se ha notado que algunos de los trazos y scores presentan algunas variaciones en su simetría según Woods (2008). Por esta razón, a lo largo de este documento se propone un modelo logístico de dos parámetros en el que el trazo latente sigue una distribución Normal Truncada, debido a que existen variables latentes que presentan comportamientos distintos al explicado por una distribución normal estándar y, por tanto, realizar el proceso de estimación sin tener en cuenta el verdadero comportamiento del trazo puede llevar a su incorrecta interpretación.

Bajo este supuesto, el proceso de estimación es implementado en R, siguiendo las metodologías propuestas por Fox (2010), Kim & Bolt (2007), Robert & Casella (2009), entre otros, en las que se implementan métodos MCMC que permiten obtener las estimaciones de parámetros de manera más eficiente. El algoritmo bayesiano aquí presentado realiza la estimación usando un muestreo M-H donde la distribución proposal del trazo latente y de los parámetros de los ítems siguen una distribución Normal Truncada.

Al realizar el proceso de simulación con los supuestos ya nombrados, se encontró que la recuperación tanto de los parámetros de los ítems como del trazo latente es óptima. Por ejemplo, se tiene que la correlación entre los simulados y los recuperados es de 0.94 y 0.99 para los parámetros de discriminación y dificultad respectivamente, y con unas pruebas de bondad y ajuste se encontraron valores entre 0.47 y 0.52. Con relación al trazo latente, tenemos que el simulado y el recuperado tiene una correlación de 0.78 y apenas 3.5 % de los trazos presentarían un mal ajuste con nivel de confianza del 95 %, un porcentaje bajo para la cantidad de trazos estimados.

Observando estos resultados en la simulación, se decide realizar la implementación del modelo propuesto a los datos reales descritos en la Sección 4.1, y se encontró que la estimación de los parámetros también es buena, ya que, para los parámetros de los ítems las

prueba de bondad y ajuste arrojaron valores entre 0.27 y 0.5 y el 4.5 % de los 5210 trazos estimados presentarían un mal ajuste con un nivel de confianza del 95 %. Nuevamente un porcentaje muy bajo con relación a la cantidad de trazos estimados.

Trabajo futuro

- Extender la metodología propuesta a modelos TRI jerárquicos y longitudinales que sean aplicables en la evaluación de la eficacia escolar en Colombia.
- Implementar el modelo propuesto en datos politómicos y multidimensionales.
- Ampliar el número de distribuciones que pueda asumir el trazo latente bajo el procedimiento descrito en este documento.

Bibliografía

- Andrade, D. F., Tavares, H. R. & Valle, R. C. (2000). *Teoria da resposta ao item: conceitos e aplicações*, São Paulo: Associação Brasileira de Estatística.
- Attorresi, H., Lozzia, G., Abal, F., Galibert, M. & Aguerri, M. (2009). Teoría de respuesta al ítem. conceptos básicos y aplicaciones para la medición de constructos psicológicos, *Revista Argentina de Clínica Psicológica* **XVIII**: 179–188.
- Bock, R. D. & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an em algorithm, *Psychometrika* **46**(4): 443–459.
URL: <http://dx.doi.org/10.1007/BF02293801>
- Casella, G. & Berger, R. (2002). *Statistical Inference*, Duxbury advanced series in statistics and decision sciences, Thomson Learning.
URL: https://books.google.com.co/books?id=0x_vAAAAMAAJ
- Chib, S. & Greenberg, E. (1995). Understanding the metropolis-hastings algorithm.
- Fox, J. (2010). *Bayesian Item Response Modeling: Theory and Applications*, Statistics for Social and Behavioral Sciences, Springer New York.
URL: <https://books.google.com.co/books?id=BZcPc4ffSTEC>
- Gamerman, D. & Lopes, H. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, Second Edition*, Chapman & Hall/CRC Texts in Statistical Science, Taylor & Francis.
URL: https://books.google.com.co/books?id=yPvECi_L3bwC
- Gelfand, A., Dey, D. & Chang, H. (1992). *Model Determination Using Predictive Distributions with Implementation Via Sampling-based Methods*, Stanford University Department of Statistics and Defense Technical Information Center.
URL: <https://books.google.com.co/books?id=HApBHwAACAAJ>
- Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. (2003). *Bayesian Data Analysis, Second Edition (Chapman & Hall/CRC Texts in Statistical Science)*, 2 edn, Chapman and Hall/CRC.
URL: <http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/158488388X>
- Gelman, A., li Meng, X. & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies, *Statistica Sinica* pp. 733–807.

- Hambleton, R. K. & Jones, R. W. (1993).]comparison of classical test theory and item response theory and their applications to test development, *Educational Measurement: Issues and Practice* **12**(3): 38–47.
URL: <http://dx.doi.org/10.1111/j.1745-3992.1993.tb00543.x>
- Hambleton, R. K., Swaminathan, H. & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*, Measurement Methods for the Social Science, SAGE Publications.
URL: <http://books.google.com.co/books?id=cmJU9SHCzecC>
- Hambleton, R. & Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*, Evaluation in education and human services, Springer.
URL: <http://books.google.com.co/books?id=jKFMUFI-e1UC>
- Harris, D. (1989). Comparison of 1-, 2-, and 3-parameter irt models, *Educational Measurement: Issues and Practice* **8**(1): 35–41.
URL: <http://dx.doi.org/10.1111/j.1745-3992.1989.tb00313.x>
- Heidelberger, P. & Welch, P. D. (1983). Simulation run length control in the presence of an initial transient, *Operations Research (INFORMS)* **31**: 1109–1144.
- Hoff, P. D. (2009). *A First Course in Bayesian Statistical Methods*, 1st edn, Springer Publishing Company, Incorporated.
- Horn, J. (1965). A rationale and test for the number of factors in factor analysis, *Psychometrika* **30**(2): 179–185.
URL: <http://dx.doi.org/10.1007/BF02289447>
- Kim, J.-S. & Bolt, D. M. (2007). Estimating item response theory models using markov chain monte carlo methods, *Educational Measurement: Issues and Practice* **26**(4): 38–51.
- Mislevy, R. J. (1985). Bayes modal estimation in item response models*, *ETS Research Report Series* **1985**(2): i–54.
URL: <http://dx.doi.org/10.1002/j.2330-8516.1985.tb00118.x>
- Páyer, M. (2013). Características de personalidad asociadas a estudiantes de permanencia prolongada, *Revista de Psicología* **29**(1).
- Ponce, M. & Granell, P. (2006). Manual del inventario pihema r-2, *Universidad Central de Venezuela. Facultad de Humanidades y Educación. Comisión Técnica de Admisión. Segunda Edición*.
- Robert, C. P. (1995). Simulation of truncated normal variables, *Statistics and Computing* **5**(2): 121–125.
URL: <http://dx.doi.org/10.1007/BF00143942>
- Robert, C. P. & Casella, G. (2009). *Introducing Monte Carlo Methods with R (Use R)*, 1st edn, Springer-Verlag, Berlin, Heidelberg.
- Woods, C. M. (2006b). Ramsay-curve item response theory (rc-irt) to detect and correct for nonnormal latent variables, *Psychometrika* **11**(3): 253–270.
- Woods, C. M. (2008). Ramsay-curve item response theory for the three-parameter logistic item response model, *Applied Psychological Measurement* **32**(6): 447–465.
URL: <http://dx.doi.org/10.1177/0146621607308014>