Estimación por máxima verosimilitud en la teoría de respuesta al ítem: I. Parámetros de las personas

Iwin Leenen[a,‡], José J. Naveja[b,§], Ramsés Vázquez-Lira[a,¶]

Facultad de Medicina

Resumen

La teoría de respuesta al ítem (TRI) proporciona un marco conceptual que utiliza modelos matemáticos para explicar las respuestas a pruebas psicológicas o educativas. Estos modelos incluyen parámetros para los ítems que conforman el instrumento (p. ej., indicando su dificultad) y para las personas que lo responden (su nivel en el rasgo latente en estudio). Para aplicar un modelo TRI, es crucial la estimación de sus parámetros a partir de datos empíricos. Típicamente se utiliza para este fin el método de estimación por máxima verosimilitud. Este artículo es el primero de una serie de tres dirigida a investigadores y estudiantes que se están familiarizando con los modelos de la TRI y que quieren aprender sobre los procedimientos para su estimación. En este primer artículo damos una introducción que presenta las ideas básicas de la TRI, tomando como referencia el modelo de Rasch, y explicamos los fundamentos y la lógica que subyace a la estimación por máxima verosimilitud. Ilustramos los principios explicados mediante una aplicación a datos de un examen de opción múltiple.

Palabras clave: *Estimación de parámetros; máxima verosimilitud; modelo de Rasch; psicometría; teoría de respuesta al ítem.*

Este es un artículo Open Access bajo la licencia CC BY-NC-ND (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Maximum likelihood estimation in models from item response theory: I. Person parameters

Abstract

Item response theory (IRT) provides a conceptual framework that uses mathematical models to explain observed

[a] Universidad Nacional Autónoma de México, Facultad de Psicología, Ciudad de México, México.
[b] Departamento de Medicina Interna y Centro Oncológico, Universidad Johannes Gutenberg, Mainz, Alemania.
ORCID ID:
[‡] https://orcid.org/0000-0003-4807-540X
[§] https://orcid.org/0000-0001-8640-6690
[¶] https://orcid.org/0000-0003-0936-6327
Recibido: 7-enero-2026. Aceptado: 9-marzo-2026.
*Autor para correspondencia: Iwin Leenen.
Correo electrónico: iwin.leenen@gmail.com
Este es un artículo Open Access bajo la licencia CC BY-NC-ND (http://creativecommons.org/licenses/by-nc-nd/4.0/).

responses on psychological or educational tests. These models include parameters for both the items comprised in the test (e.g., indicating their difficulty) and the persons responding the test (measuring their ability measured by the test). The application of an IRT model implies the estimation of these parameters from empirical data. Usually, estimation in IRT models is based on maximum likelihood. This article is the first in a series of three, aimed at researchers and students who are familiarizing themselves with IRT and who want to learn about the estimation procedures. This first article provides an introduction that presents the basic concepts in IRT, taking the Rasch model as a reference and explains the foundations and logic underlying maximum likelihood estimation. We illustrate the principles explained by applying them to data of a multiple-choice test.

Keywords: *Parameter estimation; maximum likelihood; Rasch model; psychometrics; item response theory.*

This is an Open Access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

INTRODUCCIÓN

Desde su incepción en los años 50 del siglo pasado, la Teoría de Respuesta al Ítem (TRI) ha ganado importancia como enfoque psicométrico para el análisis de instrumentos que miden rasgos psicológicos latentes (es decir, no observables, tales como habilidades cognitivas, competencias, actitudes, rasgos de personalidad, etc.). Efectivamente, es cada vez más común analizar los datos utilizando modelos de la TRI para las pruebas a gran escala o de altas consecuencias, tanto en la educación médica como en otros contextos evaluativos[1-5].

El enfoque psicométrico de la TRI ofrece varias ventajas, teóricas y prácticas[6,7]. En primera instancia, los modelos TRI proveen un fundamento teórico que relaciona las respuestas a los reactivos de la prueba con el rasgo psicológico que se pretende medir. Entre las ventajas prácticas de la TRI destaca la invarianza de los parámetros: cuando hay evidencia de que un modelo TRI describe adecuadamente las respuestas de una población de personas a un universo de reactivos, las propiedades de los reactivos (por ejemplo, su dificultad) no dependen de las personas y viceversa, las propiedades de las personas (como su nivel en el rasgo latente) no dependen de los ítems. La invarianza de parámetros implica que, bajo ciertas condiciones, es posible estimar y comparar el nivel de distintas personas en el rasgo latente, incluso si no respondieron el mismo conjunto de ítems. Otra ventaja de la TRI es la manera en la que trata los errores en la medición, especificando que el error estándar depende de la persona (es decir, desde el enfoque de la TRI, la precisión de la medición puede ser mayor para una persona que para otra), lo cual contrasta con el supuesto en la teoría clásica de los test que asume el mismo error estándar para todas las personas.

Cuando se desea analizar datos recopilados de una prueba con un modelo TRI se requieren, como en cualquier análisis estadístico, algoritmos para la estimación de sus parámetros. Para los modelos TRI, el método más utilizado es la estimación por máxima verosimilitud (MLE, *maximum likelihood estimation*). A pesar de que estos procedimientos se describen rigurosamente en los artículos y reportes técnicos, son escasos los textos que los describen para un público no experto en estadística. Este artículo es el primero de una serie de tres, en la cual ofrecemos una explicación guiada y formal de los principios de MLE en modelos TRI para una audiencia hispanohablante. (El objetivo es que sea útil, p. ej., para estudiantes e investigadores que se están familiarizando con este enfoque psicométrico y desean entender mejor la estimación de parámetros en estos modelos.)

En este primer artículo, aunque asumimos que el lector ya está familiarizado conceptualmente con la TRI y específicamente con el modelo de Rasch[8], empezamos con una breve revisión de los conceptos básicos de estos modelos. A continuación, explicamos el concepto de verosimilitud y el principio de MLE y damos un ejemplo sencillo de MLE fuera

del contexto de los modelos psicométricos. Posteriormente, aplicamos estos principios al caso de la estimación del parámetro (θ_p) de una persona p en el modelo de Rasch a partir de sus respuestas en m ítems y suponiendo que se conocen los parámetros de los ítems. Ilustramos los principios explicados con una aplicación a datos de un examen de opción múltiple. Concluimos este artículo con una breve discusión.

En el segundo artículo de la serie (en el siguiente número de la revista), nos enfocamos en la MLE de los parámetros de los ítems. El tercer artículo aborda el tema de la precisión de las estimaciones obtenidas por MLE.

CONCEPTOS BÁSICOS

Principios y conceptos generales de la teoría de respuesta el ítem

En la TRI se modela el comportamiento de una persona (dentro de una población de personas) respondiendo a un ítem (de un universo de ítems); más precisamente, los modelos TRI definen dos o más categorías de respuesta para el ítem i y especifican para cualquier persona p de la población la probabilidad de que responda en la categoría j de este ítem. Para esto, en el modelo se definen uno o más parámetros —características numéricas que cuantifican ciertas propiedades— para la persona p y uno o más parámetros para el ítem i y se combinan estos parámetros de la persona y del ítem (a través de una ecuación) para asignar una probabilidad a cada una de las posibles categorías de respuesta. De manera más formal, se escribe:

$$\Pr(Y_{pi} = j) = f(\text{parámetro(s) de la persona } p, \text{ parámetro(s) del ítem } i) \quad (1)$$

donde Y_{pi} es la variable que representa la (categoría de) respuesta de la persona p en el ítem i.

Dado que los detalles de las pruebas y de los fenómenos psicológicos subyacentes pueden ser muy variados, la TRI posee una gran variedad de modelos, que difieren entre sí con respecto a cómo definen las categorías de respuesta para los ítems (p. ej., existen modelos para ítems dicotómicos, politómicos, etc.), el número y el significado de los parámetros para personas e ítems y la función (f, en la Ecuación 1) exacta para obtener las probabilidades a partir de los parámetros.

Utilizaremos el modelo de Rasch[8] en el resto del texto, que es uno de los modelos más utilizados y elegantes de la TRI. Para una discusión más amplia acerca de los conceptos básicos de la TRI y del modelo de Rasch en particular, referimos al lector interesado a Abad et al.[9], Naveja y Leenen[10], y el libro de van der Linden[11]. Este último incluye una descripción de los modelos TRI más importantes.

El modelo de Rasch

El modelo de Rasch considera ítems dicotómicos (es decir, ítems con dos categorías de respuesta), de tal manera que la variable Y_{pi} en la Ecuación 1 asume el valor de 0 o 1. Es común etiquetar a estas categorías como "respuesta incorrecta" y "correcta", respectivamente —y lo haremos también en este artículo—, aunque se puede aplicar el modelo también a ítems no cognitivos (p. ej., al medir actitudes, las categorías pueden corresponder a "en desacuerdo" y "de acuerdo" con la afirmación planteada en el ítem). En el modelo de Rasch, se asigna un parámetro a cada persona (comúnmente representado por θ_p, con $\theta_p \in \mathbb{R}$, e interpretado como el nivel de la persona p en el rasgo latente que el test mide) y un parámetro a cada ítem ($\beta_i \in \mathbb{R}$, el grado de dificultad del ítem i). La ecuación que permite calcular la probabilidad de responder en la categoría $Y_{pi} = 1$ es:

$$\Pr(Y_{pi} = 1; \theta_p, \beta_i) = \frac{\exp(\theta_p - \beta_i)}{1 + \exp(\theta_p - \beta_i)}, \quad (2a)$$

donde $\exp(x)$ significa e^x ($e \approx 2.71828$ siendo la base de los logaritmos naturales). Puesto que Y_{pi} es dicotómica, a partir de la Ecuación 2a se obtiene directamente la probabilidad de responder en la categoría con valor 0:

$$\Pr(Y_{pi} = 0; \theta_p, \beta_i) = 1 - \Pr(Y_{pi} = 1; \theta_p, \beta_i) = \frac{1}{1 + \exp(\theta_p - \beta_i)}. \quad (2b)$$

Las dos ecuaciones anteriores se pueden compactar en una, como se muestra a continuación:

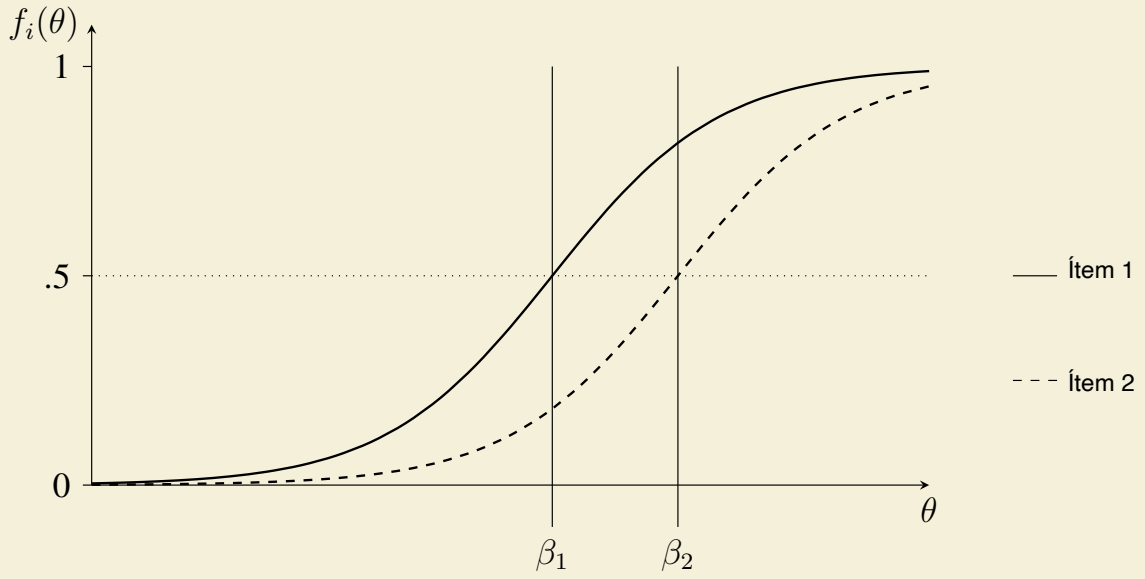


Figura 1. Ejemplos de la función característica de dos ítems en el modelo de Rasch. La abscisa representa la dimensión latente de una persona (θ), mientras que la ordenada indica la probabilidad de acertar los ítems. En este ejemplo, el ítem 1 es más fácil que el ítem 2 ($\beta_1 < \beta_2$).

$$\Pr(Y_{pi} = y_{pi}; \theta_p, \beta_i) = \frac{\exp[y_{pi}(\theta_p - \beta_i)]}{1 + \exp(\theta_p - \beta_i)}, \quad (2c)$$

donde y_{pi} es el valor (0 o 1) de la variable Y_{pi}.

Con base en la Ecuación 2a se define la *función característica del ítem*. Esta función, $f_i(\theta)$, describe cómo la probabilidad de acertar el ítem *i* cambia en función del nivel θ de la persona en el rasgo latente. En el caso del modelo de Rasch, esta función es continua y estrictamente creciente (véase la **figura 1**, que muestra la función característica para unos ítems). Es una propiedad del modelo de Rasch que las funciones características de diferentes ítems (con diferentes parámetros β) no se cruzan.

Las Ecuaciones 2a-2c permiten derivar la probabilidad de cierta respuesta (correcta o incorrecta) en un ítem en particular. Sin embargo, a menudo el interés está en conocer la probabilidad de un *patrón de respuestas*, $y_p = (y_{p1}, y_{p2}, ..., y_{pm})$; es decir, la probabilidad *conjunta* de las respuestas de la persona a ítems (en función del nivel de la persona θ_p y las dificultades de los ítems $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_m)$):

$$\begin{aligned}\Pr(\boldsymbol{Y}_p = \boldsymbol{y}_p; \theta_p, \boldsymbol{\beta}) &= \Pr[(Y_{p1} = y_{p1}) \wedge (Y_{p2} = y_{p2}) \wedge ... \wedge (Y_{pm} = y_{pm}); \\ &\quad \theta_p, \beta_1, \beta_2, ..., \beta_m],\end{aligned}$$

donde ∧ denota una conjunción (y se lee como "y"). Para esta probabilidad conjunta el modelo de Rasch, como casi todos los modelos de la TRI, incluye un supuesto de independencia local, por el cual:

$$\begin{aligned}\Pr(\boldsymbol{Y}_p = \boldsymbol{y}_p; \theta_p, \boldsymbol{\beta}) &= \Pr(Y_{p1} = y_{p1}; \theta_p, \beta_1) \times \Pr(Y_{p2} = y_{p2}; \theta_p, \beta_2) \times ... \\ &\quad \times \Pr(Y_{pm} = y_{pm}; \theta_p, \beta_m) \\ &= \prod_{i=1}^{m} \Pr(Y_{pi} = y_{pi}; \theta_p, \beta_i) \\ &= \prod_{i=1}^{m} \frac{\exp[y_{pi}(\theta_p - \beta_i)]}{1 + \exp(\theta_p - \beta_i)}\end{aligned} \quad (3)$$

Independencia local quiere decir que *si se conociera el nivel* θ_p *de la persona en el rasgo latente*—por esta condición, la independencia es "local" o "con-

dicional"—, las respuestas que da en los otros ítems no añadirían información sobre su respuesta en el ítem *i*.

ESTIMACIÓN POR MÁXIMA VEROSIMILITUD

El siguiente paso, después de haber especificado el modelo, es *ajustarlo* a los datos observados. Esto significa que se encuentran los valores para los parámetros del modelo que, de cierta manera, "explican mejor" las observaciones en la muestra[12]. Si bien existen diferentes métodos para estimar los parámetros de un modelo, MLE[13] es el procedimiento estándar en la TRI. MLE tiene viertes propiedades útiles; por ejemplo, como abordaremos en el tercer artículo de esta serie, sobre precisión de los estimadores, MLE permite cuantificar la precisión de los estimadores utilizando la función de información. Aplicado al modelo de Rasch, MLE encuentra los valores en los parámetros (θ_p's y β_i's de las distintas personas e ítems) que son *máximo-verosímiles* a la luz de las respuestas observadas de una muestra de *n* personas que contestaron los *m* ítems dicotómicos[14].

Para introducir el concepto de verosimilitud, consideremos el siguiente ejemplo: un mago nos presta una de sus monedas de utilería y queremos conocer la probabilidad de que, al lanzarla, caiga con la cara A (águila) hacia arriba. Representemos esta probabilidad por π (un número real desconocido entre 0 y 1). Para estimar este número π, se decide realizar un pequeño estudio y recopilar datos: se lanza la moneda $m = 5$ veces y se observa en cada lanzamiento el resultado, formalizándolo como la realización de una variable , que asume el valor de 1 si en el *i*-ésimo lanzamiento salió la cara A y 0 en el caso contrario. Supongamos que como resultado de esta recolección de datos se obtiene: $\boldsymbol{y} = (1,0,1,1,1,)$, es decir, salió la cara A en los lanzamientos 1, 3, 4 y 5 y la cara B en el segundo lanzamiento. La estimación de π requiere, además de estos datos, la especificación de un modelo, el cual formaliza los supuestos sobre el fenómeno. En este ejemplo, suponemos que π es constante (es decir: $\Pr(Y_i = 1) = \pi$, para cualquier lanzamiento *i*) y que las variables Y_i son mutuamente independientes. A partir de estos supuestos, se puede calcular la probabilidad de cualquier conjunto de resultados (observados en los *m* lanzamientos); por ejemplo, podemos calcular cuál hubiera sido la probabilidad de los datos observados $\boldsymbol{y}$ (previo a su recolección):

$$\begin{aligned}
&\Pr[\boldsymbol{Y} = (1,0,1,1,1); \pi] \\
&\quad = \Pr(Y_1 = 1 \wedge Y_2 = 0 \wedge Y_3 = 1 \ \wedge \ Y_4 = 1 \ \wedge \ Y_5 = 1; \pi) \\
&\quad = \Pr(Y_1 = 1; \pi) \times \Pr(Y_2 = 0; \pi) \times \Pr(Y_3 = 1; \pi) \\
&\qquad \times \Pr(Y_4 = 1; \pi) \times \Pr(Y_5 = 1; \pi) \\
&\quad = \pi\,(1-\pi)\,\pi\,\pi\,\pi \ = \ \pi^4(1-\pi).
\end{aligned} \tag{4}$$

Es decir, si se conociera el valor numérico que representa π, la ecuación anterior daría la probabilidad de observar el resultado $\boldsymbol{y} = (1,0,1,1,1,)$, antes de realizar los cinco lanzamientos. Sin embargo, como ya se llevó a cabo la recopilación de datos, ya no es de tanto interés la probabilidad de $\boldsymbol{y}$ (u otros posibles resultados del proceso de recopilación de datos) a partir de cierto valor para π; más bien, el interés estará en π, y particularmente, en la *verosimilitud* de los distintos valores que π puede asumir a la luz de los datos observados $\boldsymbol{y}$. Nótese que esto implica un cambio de perspectiva: para la verosimilitud exploramos cómo la expresión en la Ecuación 4 cambia *en función de* π. Efectivamente, la función de verosimilitud $\mathcal{L}$ es una función del parámetro que queremos estimar. En este ejemplo:

$$\mathcal{L}[\pi; \boldsymbol{y} = (1,0,1,1,1)] \stackrel{\text{def}}{=} \Pr[\boldsymbol{Y} = (1,0,1,1,1); \pi] = \ \pi^4(1-\pi). \tag{5}$$

A partir de esta función $\mathcal{L}$ se puede, por ejemplo, concluir que, *a la luz de los datos que se observaron* (cuatro veces el lado A de la moneda y una vez el lado B), el valor de .90 para π es más verosímil que el valor de .50, ya que $\mathcal{L}[.90; \boldsymbol{y} = (1,0,1,1,1)] = .06561 > \mathcal{L}[.50; \boldsymbol{y} = (1,0,1,1,1)] = .03125$. Además, se puede buscar entre *todos los posibles valores* para π (es decir, los números reales entre 0 y 1) cuál es el valor con *máxima verosimilitud*. La **figura 2** grafica la función $\mathcal{L}$ y muestra que este valor es .80; se dice que, considerando los datos observados, .80 es la MLE del parámetro π y se escribe: $\hat{\pi} = .80$. (Nótese que el valor de π sigue siendo desconocido; escribimos $\hat{\pi}$ para dejar claro que se trata de una estimación.)

La expresión $\pi^4\,(1 - \pi)$ en las Ecuaciones 4 y 5 tiene esta forma porque observamos en cuatro lan-

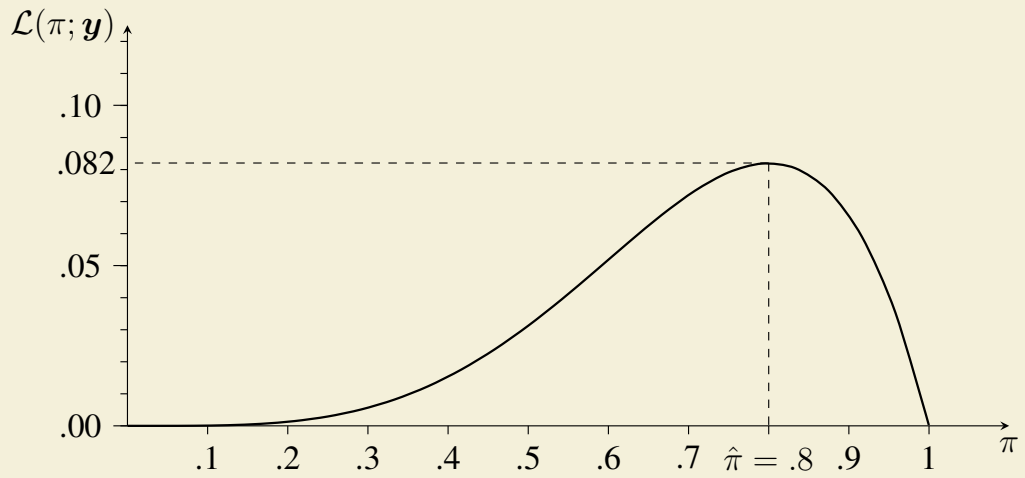


Figura 2. Función de verosimilitud del parámetro π a la luz de los datos observados $\boldsymbol{y}$ = (1,0,1,1,1). Se muestra $\hat{\pi}$ = .80, la estimación por máxima verosimilitud de $\hat{\pi}$.

zamientos la cara A y en un lanzamiento la cara B. De manera más general, para cualquier vector y de caras observadas en m lanzamientos, la función de verosimilitud se da por:

$$\mathcal{L}(\pi; \boldsymbol{y}) = \pi^{x}(1-\pi)^{m-x}, \tag{6}$$

donde $x = \sum_{i=1}^{m} y_i$ es el número de lanzamientos que muestran la cara A (y $m - x$ el número de lanzamientos que muestran la cara B). A continuación, derivamos una expresión general para la MLE de π; es decir, el valor de π que maximice la verosimilitud en la Ecuación 6.

En general, para encontrar el máximo de una función derivable, se calcula la primera derivada de esta función, se iguala a 0 y se resuelve la expresión obtenida por el argumento de la función. Cuando se aplica este método en el contexto de MLE, dicha función es la función de verosimilitud; sin embargo, en vez de analizar la función de verosimilitud, es más común analizar el *logaritmo* (natural) de esta función (es decir, la log-verosimilitud): $\ell(\pi;\boldsymbol{y}) \stackrel{\text{def}}{=} \mathcal{L}(\pi;\boldsymbol{y})$. (Nótese que, puesto que el logaritmo es una transformación estrictamente creciente, $\ell(\pi;\boldsymbol{y})$ y $\mathcal{L}(\pi;\boldsymbol{y})$ tienen su máximo para el mismo valor de π. En este sentido, no importa si se analiza la función de logverosimilitud en vez de la función de verosimilitud.) Para nuestro ejemplo:

$$\ell(\pi; \boldsymbol{y}) = x \log \pi + (m - x) \log(1 - \pi). \tag{7}$$

Calculamos la derivada de esta logverosimilitud con respecto a π:

$$\frac{\mathrm{d}\ell(\pi; \boldsymbol{y})}{\mathrm{d}\pi} = \frac{x}{\pi} - \frac{m-x}{1-\pi}.$$

Igualamos este resultado a 0 y resolvemos la ecuación resultante para obtener π:

$$\frac{x}{\pi} - \frac{m-x}{1-\pi} = 0$$

$$\iff \qquad \pi = \frac{x}{m}. \tag{8}$$

Se llama $\hat{\pi} = \frac{x}{m}$ la MLE para el parámetro π en este ejemplo.[a] Aplicado a nuestros datos, $\boldsymbol{y}$ = (1,0,1,1,1) y $x = 4$ con $m = 5$, la estimación es $\hat{\pi} = \frac{4}{5} = .80$.

[a] Cabe señalar que, para asegurar que el valor de π para el cual la primera derivada es 0 sea un máximo, es necesario evaluar también derivadas de orden superior. En este caso basta con mostrar que la segunda derivada de la función de log-verosimilitud es negativa para todos los valores entre 0 y 1.

Concluimos esta sección con dos comentarios. Primero, retomamos la diferencia conceptual entre verosimilitud y probabilidad. El concepto de probabilidad aplica a variables aleatorias (discretas; para variables aleatorias continuas se obtienen probabilidades para eventos a partir de la función de *densidad*), por ejemplo, las variables Y_i en el ejemplo anterior: podríamos decir que (quizás, si $\pi = .60$) es más probable el valor de 1 en Y_i que el valor de 0 (o que es más probable observar (1,0,1,1,1) que (0,1,0,0,0) en el vector de variables $\boldsymbol{Y}$). No obstante, dado que un parámetro (en la estadística frecuentista) *no* es una variable aleatoria, sino una constante, sería incorrecto decir que, a la luz de los datos observados $\boldsymbol{y} = (1,0,1,1,1)$, es más probable el valor de .90 para el parámetro π que el valor de .50; pero sí, se puede decir que .90 es el valor más verosímil. Dicho en otras palabras, en el marco frecuentista, π es una constante, aunque desconocida. No obstante, al ser desconocida, podemos "probar" diferentes valores y evaluar cuál maximizaría la probabilidad de los datos observados y sería, en este sentido, el más plausible o verosímil.

El segundo comentario tiene que ver con que, en las expresiones al lado derecho de la igualdad en las Ecuaciones 6 y 7, se utiliza el estadístico x y no las observaciones originales $\boldsymbol{y} = (y_1, y_2,..., y_m)$. Esto implica que, en la función de (log-)verosimilitud (que se maximiza para tener una estimación de π), no se requiere otra información presente en el vector $\boldsymbol{y}$ de observaciones más allá de x, el número de veces que se observó la cara A en los m lanzamientos. Es decir, x contiene toda la información relevante en los datos observados para estimar π; otros aspectos del vector $\boldsymbol{y}$ (p. ej., en cuáles lanzamientos exactamente se observó la cara A) no aportan información para estimar π. Por eso, se dice que el número de águilas es un *estadístico suficiente* para estimar π. Veremos que también en el modelo de Rasch existen estadísticos suficientes para estimar los parámetros del modelo. No todos los modelos de la TRI permiten obtener estadísticos suficientes, pero, cuando existen, el proceso de estimación se simplifica considerablemente.

ESTIMACIÓN DEL PARÁMETRO DE LA PERSONA EN EL MODELO DE RASCH

En esta sección se aplican los principios de MLE para estimar el parámetro θ_p de una persona p en el modelo de Rasch a partir de sus respuestas observadas en m ítems dicotómicos. Además de asumir que el modelo se cumple para estos datos, suponemos por ahora que conocemos los valores de los m parámetros de los ítems $\beta_1, \beta_2, ..., \beta_m$.

Ilustraremos el procedimiento con un ejemplo concreto, que involucra únicamente $m = 4$ ítems, con los siguientes valores para sus parámetros:

$$\beta_1 = 2.0, \quad \beta_2 = 0.0, \quad \beta_3 = 1.0, \quad \beta_4 = 0.5 \tag{9}$$

y el siguiente patrón de respuestas por parte de la persona p: $\boldsymbol{y}_p = (0,1,1,0)$; es decir, falló el primero y cuarto ítem y acertó el segundo y tercero. El procedimiento para encontrar el MLE para θ_p es muy similar al ejemplo de la moneda trucada en la sección anterior.

El primer paso es explicitar la función de verosimilitud de θ_p: $\mathcal{L}(\theta_p; \boldsymbol{y}_p)$. Por definición, se obtiene a través de la probabilidad (en la Ecuación 3) del patrón de respuestas observado, dados los valores para los parámetros. Elaborando esta ecuación en función del patrón de respuestas $\boldsymbol{y}_p = (0,1,1,0)$ y sustituyendo las cuatro β_i por los valores especificados en la Ecuación 9, se obtiene:

$$\begin{aligned}\mathcal{L}[\theta_p; \boldsymbol{y}_p = (0,1,1,0), \boldsymbol{\beta}] \\ = \frac{1}{1 + \exp(\theta_p - 2.0)} \times \frac{\exp(\theta_p - 0.0)}{1 + \exp(\theta_p - 0.0)} \\ \times \frac{\exp(\theta_p - 1.0)}{1 + \exp(\theta_p - 1.0)} \times \frac{1}{1 + \exp(\theta_p - 0.5)}.\end{aligned}$$

Aunque esta expresión es bastante más compleja que la función de verosimilitud en el ejemplo de la moneda (Ecuación 5), solo incluye un elemento desconocido: el parámetro θ_p. La **figura 3** representa la función de verosimilitud gráficamente y se puede apreciar que llega a su máximo cuando el parámetro θ_p es igual 0.861. En otras palabras, entre todos los posibles valores para θ_p (en principio, todos los

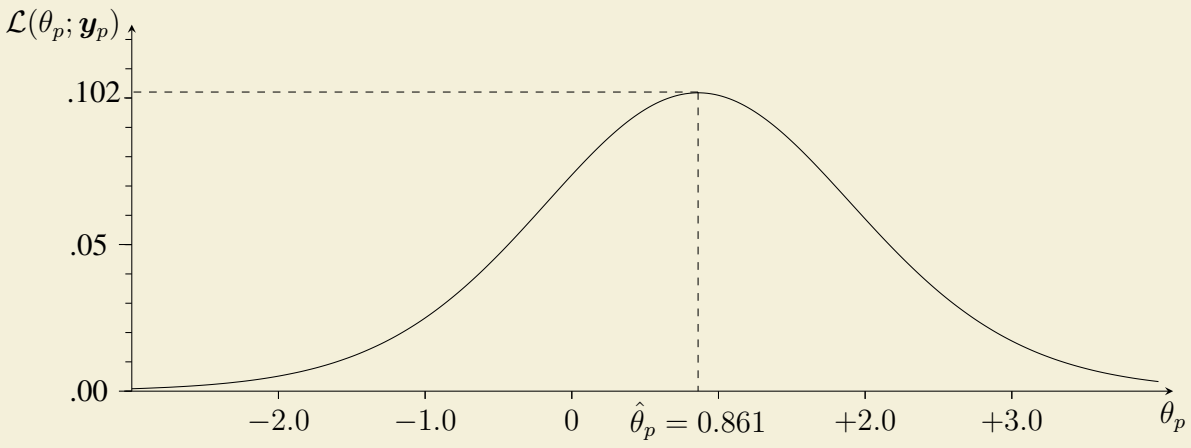


Figura 3. Función de verosimilitud del parámetro θ_p a la luz de los datos observados $\boldsymbol{y}_p$ = (0,1,1,0) de la persona p y para los siguientes valores para los parámetros de los ítems: $\beta_1 = 2.0$, $\beta_2 = 0.0$, $\beta_3 = 1.0$ y $\beta_4 = 0.5$. Se muestra $\theta_p = 0.861$, la estimación por máxima verosimilitud de θ_p.

números reales), 0.861 es el más verosímil a la luz de las respuestas observadas de la persona p en los cuatro ítems. Por lo tanto, la MLE es $\theta_p = 0.861$.

Para obtener una expresión general para el estimador por máxima verosimilitud de θ_p, definimos la función de logverosimilitud, tomando el logaritmo de la expresión en la Ecuación 3, lo cual resulta en:

$$\ell(\theta_p; \boldsymbol{y}_p, \boldsymbol{\beta}) = x_p\theta_p + \sum_{i=1}^{m} y_{pi}(-\beta_i) - \sum_{i=1}^{m} \log[1 + \exp(\theta_p - \beta_i)] \tag{10}$$

donde $x_p = \sum_{i=1}^{m} y_{pi}$ indica el número de ítems acertados por la persona p. Para hallar el valor de θ_p que maximice la función en la Ecuación 10, calculamos la primera derivada con respecto a θ_p:

$$\frac{d\ell(\theta_p; \boldsymbol{y}_p, \boldsymbol{\beta})}{d\theta_p} = x_p - \sum_{i=1}^{m} \frac{\exp(\theta_p - \beta_i)}{1 + \exp(\theta_p - \beta_i)}$$

y buscamos el valor de θ_p para el cual esta expresión sea igual a cero, que es el mismo para el cual se cumple la siguiente igualdad:

$$x_p = \sum_{i=1}^{m} \frac{\exp(\theta_p - \beta_i)}{1 + \exp(\theta_p - \beta_i)} \tag{11}$$

(Recuérdese que en esta expresión tanto x_p como las β_i son valores numéricos conocidos y θ_p es la única incógnita. Nótese también que la fracción en la Ecuación 11 coincide con $\Pr(Y_{pi} = 1; \theta_p, \beta_i)$, definida en la Ecuación 2.) Resulta imposible despejar θ_p y obtener una fórmula cerrada que permita encontrar dicho valor con una operación algebraica sencilla (como la Ecuación 8 para π en el ejemplo de la moneda). No obstante, se puede encontrar una solución por métodos numéricos utilizando algoritmos iterativos[15].

La Ecuación 11 deja claro que el número de aciertos de la persona p contiene toda la información relevante en los datos $\boldsymbol{y}_p$ para estimar θ_p. Esto quiere decir que x_p es un estadístico suficiente para estimar θ_p: solo importa cuántos ítems responde una persona correctamente; no tiene mayor relevancia para el problema de estimación conocer exactamente cuáles acertó. Cabe enfatizar que esta es una propiedad del modelo de Rasch, no de todos los modelos TRI.

Cerramos esta sección señalando dos inconvenientes de la MLE de θ_p. Primero, no siempre existe dicha estimación: cuando $x_p = 0$ o $x_p = m$ (es decir, cuando la persona no responde ningún ítem correc-

tamente, o bien, cuando todas sus respuestas son correctas), no existe la MLE de θ_p; no está definida. En estos casos, las estimaciones tenderían a $-\infty$ y $+\infty$, respectivamente (por ejemplo, para $x_p = 0$, valores para θ_p más bajos siempre serán más verosímiles). Por ello, los *softwares* que utilizan máxima verosimilitud para estimar θ_p dan un valor faltante si $x_p = 0$ o $x_p = m$. Segundo, el estimador $\hat{\theta}_p$ por máxima verosimilitud suele estar muy sesgado. En la Estadística, un estimador es insesgado si su valor esperado es igual al valor del parámetro; en este caso, $\hat{\theta}_p$ sería un estimador insesgado de θ_p, si se cumpliera que $\mathbb{E}(\hat{\theta}_p) = \theta_p$ (el símbolo $\mathbb{E}$ indica "valor esperado"; $\mathbb{E}(\hat{\theta}_p) = \theta_p$ representa la media teórica de las $\hat{\theta}_p$ obtenidas en un número infinito de réplicas de la medición). Para la MLE de θ_p, generalmente no se cumple $\mathbb{E}(\hat{\theta}_p) = \theta_p$: las θ_p relativamente bajas se subestiman ($\mathbb{E}(\hat{\theta}_p) > \theta_p$) y las θ_p relativamente altas se sobreestiman ($\mathbb{E}(\hat{\theta}_p) > \theta_p$). Warm[16] propuso una variante, máxima verosimilitud ponderada (WLE, *weighted o Warm likelihood estimation*), que resuelve ambos inconvenientes. Otra solución consiste en utilizar el estimador *máximo a posteriori* (MAP) o *esperado a posteriori* (EAP), que se sitúan en el marco Bayesiano[17].

SOFTWARE Y EJEMPLO ILUSTRATIVO

Existen varios programas que implementan MLE para modelos TRI, tanto comerciales[18,19] como *software* libre; actualmente son muy populares los paquetes del *software* R para realizar análisis TRI[20-23]. Aquí ilustramos el procedimiento explicado en la sección anterior aplicando el paquete mirt[20] del *software* R a datos de un examen de opción múltiple. Este examen, que evalúa el concepto de homeostasis y consiste en 20 preguntas de cuatro opciones cada una, fue desarrollado y validado por McFarland y colaboradores[24] en una muestra de 669 estudiantes de 12 instituciones que ofertaban cursos relacionados con ciencias de la vida y de la salud. Los datos se encuentran disponibles como parte del paquete ShinyItemAnalysis en R, mientras que el código en R para descargar y analizar estos datos está disponible como *Material Suplementario* en un proyecto de *Open Science Framework* (https://doi.org/10.17605/OSF.IO/23J7U). Para los tres artículos de esta serie se utilizan estos mismos datos y código.

Para los conceptos tratados en este artículo nos enfocamos en la segunda parte del código (bajo el encabezado "PARTE 2. Estimación de los parámetros de las personas"; en la primera parte se realiza la estimación de parámetros de ítems, los cuales —como se explicó en la sección anterior— se requieren para estimar los parámetros de las personas). El código, que incluye comentarios explicativos extensos, muestra cómo obtener las estimaciones $\hat{\theta}_p$ para las distintas personas de dos maneras: (1) utilizando el paquete mirt y (2) implementando la función de verosimilitud y encontrando su máximo (gráficamente o a través de un procedimiento iterativo disponible en R).

COMENTARIOS CONCLUYENTES

MLE es uno de los métodos más utilizados para estimar los valores de los parámetros en modelos estadísticos. En este artículo, explicamos e ilustramos cómo el método se aplica cuando el objetivo es estimar los parámetros de las personas en los modelos TRI. Aunque nos enfocamos en el modelo de Rasch, los principios explicados aplican de manera similar a otros modelos de la TRI. Lo que cambia es la expresión exacta de la función de (log-)verosimilitud (en la Ecuación 10), pero la idea de maximizar esta expresión en función de θ_p es la misma en todos los modelos (unidimensionales) de la TRI. (Incluso en los modelos TRI multidimensionales, los principios siguen siendo los mismos; en este caso, cada persona se caracteriza con un vector de d parámetros $\theta_p = (\theta_{p1}, \theta_{p1}, ..., \theta_{pd})$ donde d es el número de dimensiones, y la función de verosimilitud es una función multivariable cuyo máximo se busca en función de los d parámetros de la persona.)

Los algoritmos de MLE para la estimación de los parámetros de las personas en los modelos TRI requieren valores "conocidos" para los parámetros de los ítems. En los ejemplos anteriores (tanto en la Ecuación 9 como en la ilustración) especificamos de antemano los valores de los parámetros $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_m)$. En principio, estos parámetros no se conocen y deben estimarse a partir de los datos, previo a la aplicación de los procedimientos presentados en este artículo para estimar las 's de las personas. En el siguiente artículo de esta serie se explicará cómo se estiman los parámetros de los ítems.

Una tarea importante después de la estimación de los parámetros de un modelo consiste en evaluar la confiabilidad (o precisión) de las estimaciones y la validez de las inferencias que se realizan a partir de las mismas. En el último artículo de esta serie presentaremos las herramientas que se suelen utilizar para este fin.

CONTRIBUCIÓN INDIVIDUAL

- IL: Idea original para el proyecto, redacción y revisión del manuscrito.
- JN: Redacción y revisión del manuscrito.
- RVL: Idea original para el proyecto, revisión del manuscrito.

AGRADECIMIENTOS

Los autores agradecen a Alicia Muñiz-Jiménez y Christian Francisco Badillo-Hernández por sus valiosos comentarios sobre una versión previa del manuscrito.

PRESENTACIONES PREVIAS

Sin ninguna presentación previa.

FINANCIAMIENTO

Ninguno.

CONFLICTO DE INTERESES

Los autores declaran no tener ningún conflicto de interés.

REFERENCIAS

1. Egberink IJL, Meijer R. Het nut van de item respons theorie bij de constructie en evaluatie van niet-cognitieve instrumenten voor selectie en assessment binnen organisaties [La utilidad de la teoría de respuesta al ítem en la construcción y evaluación de instrumentos no cognitivos para la selección y evaluación en organizaciones]. Gedrag Organ. 2012;(25):87-107. doi: 10.5117/2012.025.001.087
2. Reise SP, Waller NG. Item response theory and clinical measurement. Annu Rev Clin Psychol. 2009;(5):27-48. doi: 10.1146/annurev.clinpsy.032408.153553
3. Bulut O. Applying item response theory models to entrance examination for graduate studies: Practical issues and insights. J Meas Eval Educ Psychol. 2015;(6):313-330.
4. Instituto Nacional para la Evaluación de la Educación. Plan Nacional para la Evaluación de los Aprendizajes (PLANEA). INEE; 2018.
5. Hernández-Torrano D, Courtney MGR. Modern international large-scale assessment in education: An integrative review and mapping of the literature. Large Scale Assess Educ. 2021;(9):17. doi: 10.1186/s40536-021-00109-1
6. Leenen I. Virtudes y limitaciones de la teoría de respuesta al ítem para la evaluación educativa en ciencias médicas. Inv Ed Med. 2014;(3):40-55.
7. doi: 10.1016/S2007-5057(14)72724-3
8. Muñiz J. Las teorías de los tests: Teoría clásica y teoría de respuesta a los ítems. Papel Psicol. 2010;(31):57-66.
9. Rasch G. Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests. Copenhague: Nielsen & Lydiche; 1960.
10. Abad FJ, Olea J, Ponsoda V, García C. Medición en ciencias sociales y de la salud. Madrid: Síntesis; 2011.
11. Naveja JJ, Leenen I. Una introducción a la teoría de respuesta el ítem para el análisis psicométrico de exámenes. En: Sánchez-Mendiola M, Martínez-González A (Ed.) Evaluación y aprendizaje en educación universitaria: estrategias e instrumentos. México: UNAM; 2022. p. 265-284.
12. van der Linden WJ (Ed.) Handbook of item response theory (Vol. 1): Models. Boca Raton: CRC Press; 2016.
13. Myung IJ. Tutorial on maximum likelihood estimation. J Math Psychol. 2003;(47):90-100. doi: 10.1016/S0022-2496(02)00028-7
14. Fisher R. On the mathematical foundations of theoretical statistics. Philos Trans R Soc Lond A. 1922;(222):309-368. doi: 10.1098/rsta.1922.0009
15. Baker FB, Kim SH (Ed.) Item response theory: Parameter estimation techniques. Nueva York: Marcel Dekker; 2004.
16. Burden RL, Faires DJ, Burden AM. Numerical analysis. Boston: Cengage Learning; 2015.
17. Warm TA. Weighted likelihood estimation of ability in item response theory. Psychometrika. 1989;(54):427-50. doi: 10.1007/BF02294627
18. Glas CAW. Maximum-likelihood estimation. En: van der Linden WJ (Ed.) Handbook of item response theory (Vol. 2): Statistical tools. Boca Raton: CRC Press; 2016. p. 197-216.
19. Muthén LK, Muthén BO. Mplus user's guide (eighth edition). Los Ángeles: Muthén & Muthén; 19982017.
20. Wu EJC, Bentler, PM. EQSIRT: A comprehensive item response theory program. Encino: Multivariate Software; 2013.
21. Chalmers RP. mirt: A multidimensional item response theory package for the R environment. J Stat Softw. 2012;(48-6):1-29. doi: 10.18637/jss.v048.i06
22. Mair P, Hatzinger R. Extended Rasch modeling: The eRm package for the application of IRT models in R. J Stat Softw. 2007;(20-9):1-20. doi: 10.18637/jss.v020.i09
23. Rizopoulos D. ltm: An R package for latent variable modelling and item response theory analyses. J Stat Softw. 2006;(17-5);1-25. doi: 10.18637/jss.v017.i05
24. Robitzsch A, Kiefer T, Wu M. Package 'TAM': Test analysis modules (R package version 4.14). https://CRAN.R-project.org/package=TAM
25. McFarland JL, Price RM, Wenderoth MP, Martinková P, Cliff W, Michael J, Modell H, Wright A. Development and validation of the homeostasis concept inventory. CBE Life Sci Educ. 2017;(16-2):ar35. doi: 10.1187/cbe.16-10-0305