

# Estimación por máxima verosimilitud en la teoría de respuesta al ítem: II. Parámetros de los ítems

Iwin Leenen<sup>a,†,\*</sup>, José J. Naveja<sup>b,§</sup>, Ramsés Vázquez-Lira<sup>a,¶</sup>

Facultad de Medicina



## Resumen

La teoría de respuesta al ítem (TRI) proporciona un marco conceptual para explicar las respuestas a instrumentos psicológicos mediante modelos matemáticos que incluyen parámetros tanto para las personas que responden como para los ítems que conforman el instrumento. Este artículo, que es el segundo en una serie de tres dedicada a revisar los principios de la estimación de parámetros en modelos TRI, se centra en los procedimientos para estimar los parámetros de los ítems. En particular, revisamos tres variantes del método de estimación por máxima verosimilitud: máxima verosimilitud conjunta, máxima verosimilitud condicional y máxima verosimilitud marginal. Para cada variante resaltamos sus características principales, así como sus ventajas y limitaciones. Finalmente, ilustramos los procedimientos mediante un ejemplo de un análisis de respuestas en 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: II. Item parameters

### Abstract

Item response theory (IRT) offers a conceptual framework for explaining the responses on psychological tests through mathematical models that include parameters for the respondents as well as for the items that make up the test.

<sup>a</sup> Facultad de Psicología, Universidad Nacional Autónoma de México, Cd. Mx., México

<sup>b</sup> 3er Departamento de Medicina Interna y Centro Oncológico, Universidad Johannes Gutenberg de Mainz, Alemania.  
ORCID ID:

<sup>†</sup> <https://orcid.org/0000-0003-4807-540X>

<sup>§</sup> <https://orcid.org/0000-0001-8640-6690>

<sup>¶</sup> <https://orcid.org/0000-0003-0936-6327>

Recibido: 10-febrero-2026. Aceptado: 25-marzo-2026.

\*Autor para correspondencia: Iwin Leenen.

Correo electrónico: [iwin.leenen@gmail.com](mailto: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/>).

This article, which is the second in a series of three that review the principles underlying parameter estimation in IRT models, focuses on the procedures used to estimate the item parameters. In particular, we review three variants of maximum likelihood estimation: joint maximum likelihood, conditional maximum likelihood, and marginal maximum likelihood. For each variant, we highlight its main characteristics, together with its advantages and limitations.

Finally, we illustrate these procedures with an example of an analysis of the responses to 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

Para el análisis psicométrico de pruebas que miden rasgos psicológicos latentes (como habilidades cognitivas, competencias, actitudes, rasgos de personalidad), el enfoque de la Teoría de Respuesta al Ítem (TRI) se ha consolidado como una herramienta central. Específicamente, en el contexto de la evaluación educativa y, sobre todo, para pruebas a gran escala y de alto impacto, las ventajas conceptuales y metodológicas de la TRI han convertido este enfoque en el marco predominante para el análisis de los datos de evaluación<sup>1-6</sup>.

La aplicación de un modelo TRI requiere programas informáticos especializados que permiten, entre otros aspectos, obtener estimaciones de los parámetros del modelo, evaluar la precisión de dichas estimaciones y examinar el ajuste del modelo a los datos observados. Con respecto al método de estimación, el más utilizado en este contexto es la estimación por máxima verosimilitud (MLE, *maximum likelihood estimation*). En el número anterior de esta revista<sup>7</sup> presentamos una revisión de los principios que subyacen a la MLE y mostramos cómo se aplican para la estimación de los parámetros de las personas en el caso particular del modelo de Rasch<sup>8</sup>. En este desarrollo, se asumió que los parámetros de los ítems eran conocidos; sin embargo, en la práctica, también es necesario estimarlos. El presente trabajo se centra en los procedimientos de MLE para estimar los parámetros de los ítems.

En la siguiente sección retomamos brevemente la notación y las ecuaciones básicas del modelo de Rasch, así como el concepto de verosimilitud, que resulta necesario para las secciones subsiguientes. Posteriormente, explicamos las tres principales variantes de MLE para los parámetros de los ítems:

máxima verosimilitud conjunta, máxima verosimilitud condicional y máxima verosimilitud marginal. Tras presentar estas tres aproximaciones, ofrecemos una síntesis comparativa, destacando las principales ventajas y limitaciones de cada una. Asimismo, incluimos una breve revisión de software libre y un ejemplo de su aplicación en datos empíricos. El artículo concluye con algunos comentarios finales.

## EL MODELO DE RASCH Y EL CONCEPTO DE VEROSIMILITUD

Cuando en la práctica se quiere ajustar el modelo de Rasch a datos empíricos, generalmente se dispone de las respuestas de  $n$  personas a  $m$  ítems. Estos datos pueden organizarse en una matriz  $\mathbf{y}$  de dimensiones  $n \times m$  con valores dicotómicos (unos y ceros para respuestas correctas e incorrectas, respectivamente). El modelo de Rasch incluye para cada persona  $p$  un parámetro  $\theta_p$  (su nivel en el rasgo latente que se quiere medir) y para cada ítem  $i$  un parámetro  $\beta_i$  (su dificultad). La relación entre estos parámetros y la probabilidad de las posibles respuestas de la persona en el ítem se expresa mediante la siguiente ecuación:

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

En esta ecuación,  $Y_{pi}$  es la variable que representa la respuesta de la persona  $p$  en el ítem  $i$  y  $y_{pi}$  corresponde al valor concreto de esta variable.

La **ecuación 1** describe la probabilidad de la respuesta en un ítem individual. El modelo de Rasch también permite calcular la probabilidad de un *patrón completo de respuestas*,  $\mathbf{y}_p = (y_{p1}, y_{p2}, \dots, y_{pm})$ ; es decir, la probabilidad *conjunta* de las respuestas de la persona  $p$  a todos los  $m$  ítems de la

prueba. Para ello, el modelo asume *independencia local*, lo cual implica que la probabilidad conjunta se obtiene multiplicando las probabilidades de las respuestas en los ítems individuales:

$$\begin{aligned} \Pr(\mathbf{Y}_p = \mathbf{y}_p; \theta_p, \boldsymbol{\beta}) &= \prod_{i=1}^m \Pr(Y_{pi} = y_{pi}; \theta_p, \beta_i) \\ &= \prod_{i=1}^m \frac{\exp[y_{pi}(\theta_p - \beta_i)]}{1 + \exp(\theta_p - \beta_i)}. \end{aligned} \quad (2)$$

El problema de estimación consiste esencialmente en asignar valores a los  $n$  parámetros de las personas,  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_p, \dots, \theta_n)$ , y a los  $m$  parámetros de los ítems,  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_i, \dots, \beta_m)$ . En el artículo previo, revisamos cómo obtener, por el método de máxima verosimilitud, una estimación  $\hat{\theta}_p$  del parámetro de cada persona, asumiendo que los valores de los parámetros de los ítems en el vector  $\boldsymbol{\beta}$  eran conocidos. Para esto, se introdujo el concepto de la *verosimilitud del parámetro a la luz de los datos observados* que, en el caso de  $\theta_p$  en el modelo de Rasch, se define como:

$$\mathcal{L}(\theta_p; \mathbf{y}_p, \boldsymbol{\beta}) \stackrel{\text{def}}{=} \Pr(\mathbf{Y}_p = \mathbf{y}_p; \theta_p, \boldsymbol{\beta}). \quad (3)$$

Es decir, la verosimilitud del valor específico del parámetro  $\theta_p$  es, por definición, la probabilidad de observar los datos  $\mathbf{y}_p$  para dicho valor de  $\theta_p$  (según la **ecuación 2**). Para obtener el estimador  $\hat{\theta}_p$  por máxima verosimilitud se considera la verosimilitud como una función de  $\theta_p$  y se busca aquel valor del parámetro que maximice  $\mathcal{L}(\theta_p; \mathbf{y}_p, \boldsymbol{\beta})$ . De manera equivalente, se puede buscar también el valor  $\theta_p$  que maximice la logverosimilitud,  $\ell(\theta_p; \mathbf{y}_p) \stackrel{\text{def}}{=} \log \mathcal{L}(\theta_p; \mathbf{y}_p, \boldsymbol{\beta})$ . Dado que el logaritmo es una función estrictamente creciente, ambas funciones,  $\ell(\theta_p; \mathbf{y}_p, \boldsymbol{\beta})$  y  $\mathcal{L}(\theta_p; \mathbf{y}_p, \boldsymbol{\beta})$ , alcanzan su máximo en el mismo valor de  $\theta_p$ . En la práctica, analizar la logverosimilitud suele ser más conveniente y ofrece ventajas computacionales.

Para una explicación más extensa de los conceptos introducidos en esta sección, véase nuestro artículo publicado en el número anterior de esta revista. En las siguientes secciones presentamos los procedimientos para estimar los parámetros de los ítems en el vector  $\boldsymbol{\beta}$  utilizando los principios de MLE.

## MÁXIMA VEROSIMILITUD CONJUNTA (JML)

La verosimilitud definida en la **ecuación 3** es una función de un solo parámetro ( $\theta_p$ ). Encontrar el valor del parámetro donde esta función alcanza su máximo es relativamente directo. Sin embargo, los mismos principios de máxima verosimilitud pueden extenderse para estimar *simultáneamente* los  $n + m$  parámetros del modelo de Rasch: esto implica buscar entre todos los posibles valores para los parámetros del vector  $(\boldsymbol{\theta}, \boldsymbol{\beta}) = (\theta_1, \theta_2, \dots, \theta_n, \beta_1, \beta_2, \dots, \beta_m)$ , aquella combinación que resulta más verosímil a la luz de los datos observados  $\mathbf{y}$ .

La *función de verosimilitud conjunta* de los  $n + m$  parámetros se obtiene multiplicando las probabilidades de los patrones de respuesta de las distintas personas en la **Ecuación 2**:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\beta}; \mathbf{y}) &= \prod_{p=1}^n \Pr(\mathbf{Y}_p = \mathbf{y}_p; \theta_p, \boldsymbol{\beta}) \\ &= \prod_{p=1}^n \prod_{i=1}^m \frac{\exp[y_{pi}(\theta_p - \beta_i)]}{1 + \exp(\theta_p - \beta_i)}. \end{aligned} \quad (4)$$

Esta multiplicación se base en el supuesto de *independencia experimental*, lo cual quiere decir que, condicional a los parámetros  $\boldsymbol{\beta}$ , las respuestas de una persona no aportan información sobre las respuestas de otras personas. Este supuesto es distinto del de independencia local, pero es igualmente fundamental para la verosimilitud conjunta.

Al tomar el logaritmo de ambos lados de la ecuación anterior se obtiene la función de logverosimilitud conjunta:

$$\begin{aligned} \ell(\boldsymbol{\theta}, \boldsymbol{\beta}; \mathbf{y}) &= \sum_{p=1}^n x_p \theta_p + \sum_{i=1}^m s_i (-\beta_i) \\ &\quad - \sum_{p=1}^n \sum_{i=1}^m \log(1 + \exp(\theta_p - \beta_i)), \end{aligned} \quad (5)$$

donde  $x_p = \sum_{i=1}^m y_{pi}$ , el número de respuestas correctas de la persona  $p$  en los  $m$  ítems, y  $s_i = \sum_{p=1}^n y_{pi}$ , el número de respuestas correctas en el ítem  $i$  por las  $n$  personas. La ecuación anterior implica que  $(x_1, x_2, \dots, x_n, s_1, s_2, \dots, s_m)$  son estadísticos suficientes para

estimar los  $n + m$  parámetros en  $(\theta, \beta)$ . Sin embargo, solo se requieren  $n + m - 1$  de estos estadísticos suficientes y solo hay  $n + m - 1$  parámetros libres, lo cual es consecuencia de que el modelo de Rasch no está identificado. En efecto, un análisis de la función de verosimilitud en la **ecuación 4** muestra que para cualquier constante  $c$  se cumple:

$$\begin{aligned} \mathcal{L}(\theta_1, \theta_2, \dots, \theta_n, \beta_1, \beta_2, \dots, \beta_m; \mathbf{y}) \\ = \mathcal{L}(\theta_1 + c, \theta_2 + c, \dots, \theta_n + c, \\ \beta_1 + c, \beta_2 + c, \dots, \beta_m + c; \mathbf{y}). \end{aligned}$$

Es decir, sumar la misma constante a todos los parámetros  $\theta_p$  y  $\beta_i$ , no cambia las probabilidades de acertar o fallar (en la **ecuación 1**) ya que las diferencias entre los parámetros no cambian:  $(\theta_p + c) - (\beta_i + c) = (\theta_p - \beta_i)$ . Por lo tanto, existen *múltiples soluciones para  $(\theta, \beta)$*  que maximizan la función de verosimilitud. A esto se refiere cuando se dice que el modelo no está identificado. Una forma estándar de resolver esta falta de identificación (y seleccionar una solución particular entre todas las posibles) consiste en fijar el valor de uno de los parámetros; por ejemplo, se impone  $\beta_1 = 0$ . Esto implica que no se estima  $\beta_1$  y se estiman, en total,  $n + m - 1$  parámetros del modelo.

Los estimadores por máxima verosimilitud se obtienen generalizando el procedimiento de MLE de  $\theta_p$ . Dado que ahora se trata de *múltiples parámetros*, se calculan las derivadas *parciales* de la función de logverosimilitud en la **ecuación 5** con respecto a cada uno de ellos ( $\theta_1, \theta_2, \dots, \theta_n, \beta_2, \beta_3, \dots, \beta_m$ ), se igualan las derivadas parciales a cero y se buscan los valores de los parámetros que satisfagan simultáneamente el sistema de ecuaciones que resulta. De esta manera, se obtiene el siguiente sistema de  $n + m - 1$  ecuaciones (con  $n + m - 1$  incógnitas; recuérdese que se fijó  $\beta_1 = 0$ ):

$$\left\{ \begin{array}{l} x_1 = \sum_{i=1}^m \frac{\exp(\theta_1 - \beta_i)}{1 + \exp(\theta_1 - \beta_i)} \\ \vdots \\ x_n = \sum_{i=1}^m \frac{\exp(\theta_n - \beta_i)}{1 + \exp(\theta_n - \beta_i)} \\ s_2 = \sum_{p=1}^n \frac{\exp(\theta_p - \beta_2)}{1 + \exp(\theta_p - \beta_2)} \\ \vdots \\ s_m = \sum_{p=1}^n \frac{\exp(\theta_p - \beta_m)}{1 + \exp(\theta_p - \beta_m)} \end{array} \right.$$

Una solución analítica para este sistema de ecuaciones no existe; sin embargo, se puede encontrar aplicando algoritmos iterativos.

Es importante señalar dos problemas con la estimación por JML. El primero es idéntico a un problema mencionado en nuestro artículo previo sobre la estimación de  $\theta_p$ : si, para una persona  $p$ ,  $x_p = 0$  o  $x_p = m$ , entonces no existe la MLE  $\hat{\theta}_p$ ; de manera similar, si, para un ítem  $i$ ,  $s_i = 0$  o  $s_i = n$  (es decir, nadie lo acierta o todos lo aciertan), entonces no existe la MLE  $\hat{\beta}_i$ . El segundo problema hace referencia a una propiedad teórica de los estimadores por JML: que son *inconsistentes*. En estadística se dice que un estimador  $\hat{\vartheta}$  es consistente si, conforme la muestra crece, el error de estimación (la diferencia entre  $\hat{\vartheta}$  y  $\vartheta$ ) tiende a disminuir. La idea subyacente es que muestras más grandes llevan a estimaciones más precisas. Sin embargo, los estimadores por JML en el modelo de Rasch son inconsistentes<sup>9</sup> porque, al incrementar la muestra, también se incrementan el número de parámetros por estimar (por ejemplo, cada nueva persona introduce su propio parámetro). Por esta inconsistencia, rara vez se utiliza JML como método de estimación para el modelo de Rasch (y los modelos TRI en general). En su lugar, se han desarrollado métodos que, de algún modo, “eliminan” los parámetros de las personas de la función de verosimilitud. Estos métodos se explican en las siguientes secciones.

### MÁXIMA VEROSIMILITUD CONDICIONAL (CML)

Como hemos comentado en varias ocasiones,  $x_p$ , el número de ítems que acertó la persona  $p$ , es un estadístico suficiente para estimar su parámetro  $\theta_p$  en el modelo de Rasch. Esto implica que todas las personas con el mismo número de aciertos tendrán la misma estimación  $\hat{\theta}_p$  de su nivel en el rasgo latente; si la prueba consta de  $m$  ítems, entonces es posible particionar la muestra total en  $m + 1$  grupos de personas con base en su puntaje total: aquellas personas con 0 aciertos forman el primer grupo; aquellas con 1 acierto, el segundo grupo; y así sucesivamente hasta el grupo de personas que aciertan los  $m$  ítems. Para cada grupo, solo es necesario estimar un valor:  $\hat{\theta}_x$  (el valor estimado para la  $\theta$  de una persona con puntaje  $x$ ). Así, aun cuando la muestra de personas



crece, la MLE  $\hat{\theta}_p$  para cualquier persona  $p$  solo puede adoptar uno de los  $m + 1$  valores posibles. Esta propiedad abre una vía para resolver el problema de la inconsistencia de los estimadores por JML que comentamos en la sección anterior.

CML explota esta propiedad del modelo de Rasch definiendo una función de verosimilitud condicional a los puntajes observados de las  $n$  personas en el vector  $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$ . Esta función ya no incluye los parámetros de las personas, sino únicamente los parámetros de los ítems. El siguiente ejemplo ilustra cómo se logra eliminar los parámetros de las personas de la función de verosimilitud. Nuestra

exposición aquí sigue, en lo esencial, la presentación de Verhelst<sup>10</sup>.

Considérese una prueba con  $m = 3$  ítems y el patrón de respuestas  $\boldsymbol{y}_p = (1,0,1)$  de la persona  $p$ ; entonces, el puntaje observado de esta persona es  $x_p = 2$ . A continuación, derivamos la probabilidad de observar este patrón de respuesta  $\boldsymbol{y}_p$  condicional al puntaje observado  $x_p$ , es decir, la probabilidad de que la persona  $p$  tenga el patrón de respuesta  $(1,0,1)$ , si se sabe que tiene dos aciertos. Nótese que el puntaje  $x_p = 2$  solo ocurre con tres patrones de respuesta:  $(1,1,0)$ ,  $(1,0,1)$  y  $(0,1,1)$ . Por lo tanto, dicha probabilidad condicional es la siguiente:

$$\begin{aligned} \Pr[\boldsymbol{Y}_p = (1,0,1) \mid x_p = 2; \theta_p, \boldsymbol{\beta}] &= \frac{\Pr[\boldsymbol{Y}_p = (1,0,1); \theta_p, \boldsymbol{\beta}]}{\Pr[\boldsymbol{Y}_p = (1,1,0) \vee \boldsymbol{Y}_p = (1,0,1) \vee \boldsymbol{Y}_p = (0,1,1); \theta_p, \boldsymbol{\beta}]} \quad (6) \\ &= \frac{\Pr[\boldsymbol{Y}_p = (1,0,1); \theta_p, \boldsymbol{\beta}]}{\Pr[\boldsymbol{Y}_p = (1,1,0); \theta_p, \boldsymbol{\beta}] + \Pr[\boldsymbol{Y}_p = (1,0,1); \theta_p, \boldsymbol{\beta}] + \Pr[\boldsymbol{Y}_p = (0,1,1); \theta_p, \boldsymbol{\beta}]} \end{aligned}$$

El símbolo  $\vee$  en la primera ecuación denota una disyunción (y se lee como “o”). Las probabilidades del lado dere-

cho de la **ecuación 6** se obtienen a partir de la **ecuación 3**. P. ej., la probabilidad del patrón  $(1,0,1)$  está dada por:

$$\begin{aligned} \Pr[\boldsymbol{Y}_p = (1,0,1); \theta_p, \boldsymbol{\beta}] &= \Pr(Y_{p1} = 1; \theta_p, \beta_1) \times \Pr(Y_{p2} = 0; \theta_p, \beta_2) \times \Pr(Y_{p3} = 1; \theta_p, \beta_3) \\ &= \frac{\exp(\theta_p - \beta_1)}{1 + \exp(\theta_p - \beta_1)} \times \frac{1}{1 + \exp(\theta_p - \beta_2)} \times \frac{\exp(\theta_p - \beta_3)}{1 + \exp(\theta_p - \beta_3)} \\ &= \frac{\exp(2\theta_p) \exp(-\beta_1 - \beta_3)}{K}, \end{aligned}$$

donde, por conveniencia, definimos  $K = \prod_{i=1}^3 [1 + \exp(\theta_p - \beta_i)]$ . Las probabilidades para los patrones  $(1,1,0)$  y  $(0,1,1)$  se obtienen de manera similar. Todas comparten el mismo denominador  $K$  e incluyen en el numerador

el factor  $\exp(2\theta_p)$ , lo cual es consecuencia de que cada uno de estos patrones de respuesta tiene exactamente dos aciertos. Al sustituir estas probabilidades en la **ecuación 6**, se obtiene:

$$\begin{aligned} \Pr[\boldsymbol{Y}_p = (1,0,1) \mid x_p = 2; \theta_p, \boldsymbol{\beta}] &= \frac{\exp(2\theta_p) \exp(-\beta_1 - \beta_3)}{K} \\ &= \frac{\exp(2\theta_p) \exp(-\beta_1 - \beta_2)}{K} + \frac{\exp(2\theta_p) \exp(-\beta_1 - \beta_3)}{K} + \frac{\exp(2\theta_p) \exp(-\beta_2 - \beta_3)}{K} \end{aligned}$$

lo cual, tras simplificar, resulta en:

$$\Pr[\mathbf{Y}_p = (1,0,1) \mid x_p = 2; \theta_p, \boldsymbol{\beta}] = \frac{\exp(-\beta_1 - \beta_3)}{\exp(-\beta_1 - \beta_2) + \exp(-\beta_1 - \beta_3) + \exp(-\beta_2 - \beta_3)}. \quad (7)$$

El punto clave de la **ecuación 7** es que, condicional al puntaje observado, la probabilidad del patrón de respuesta depende únicamente de los parámetros de los ítems  $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)$ ; ya no depende de  $\theta_p$ .

Para encontrar la estructura general de esta probabilidad condicional —para cualquier  $m$  y cualquier patrón de respuestas  $\mathbf{y}_p$ — se definen  $\varepsilon_i = \exp(-\beta_i)$  para todos los ítems ( $i = 1, \dots, m$ ), y el vector  $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m)$ . En el ejemplo anterior, el numerador de la **ecuación 7** es igual a  $\varepsilon_1\varepsilon_3$ , el producto de las  $\varepsilon_i$  correspondientes a los ítems acertados por la persona  $p$ . En general, el numerador tendrá la forma:  $\prod_{i=1}^m \varepsilon_i^{y_{pi}}$ . En el denominador de la **ecuación 7** se tiene  $\varepsilon_1\varepsilon_2 + \varepsilon_1\varepsilon_3 + \varepsilon_2\varepsilon_3$ ; es decir, la suma de todos los productos de dos  $\varepsilon_i$  que pueden formarse a partir de las tres ( $\varepsilon_1, \varepsilon_2, \varepsilon_3$ ). Son productos de dos  $\varepsilon_i$  porque corresponden a patrones de respuesta con dos aciertos y la suma consta de tres términos porque con tres ítems se pueden formar tres pares distintos. Para el caso general, el denominador es la suma de todos los posibles productos de  $x_p$  elementos  $\varepsilon_i$ 's (en total,  $m!/x_p! (m - x_p)!$  términos) y se conoce como la *función elemental simétrica de orden  $x_p$*  del vector  $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m)$ , representada por  $\gamma_{x_p}(\boldsymbol{\varepsilon})$ . Se define como sigue:

$$\begin{aligned} \gamma_0(\boldsymbol{\varepsilon}) &= 1, \\ \gamma_1(\boldsymbol{\varepsilon}) &= \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_m, \\ \gamma_2(\boldsymbol{\varepsilon}) &= \varepsilon_1\varepsilon_2 + \varepsilon_1\varepsilon_3 + \dots + \varepsilon_1\varepsilon_m \\ &\quad + \varepsilon_2\varepsilon_3 + \varepsilon_2\varepsilon_4 + \dots + \varepsilon_2\varepsilon_m \\ &\quad + \dots + \varepsilon_{m-1}\varepsilon_m, \\ &\vdots \\ \gamma_m(\boldsymbol{\varepsilon}) &= \varepsilon_1\varepsilon_2\varepsilon_3 \dots \varepsilon_m. \end{aligned}$$

Ahora, la probabilidad del patrón de respuesta  $\mathbf{y}_p$  condicional al puntaje  $x_p$  observado en dicho patrón de respuesta, puede escribirse de forma general como:

$$\Pr[\mathbf{Y}_p = \mathbf{y}_p \mid x_p; \boldsymbol{\varepsilon}] = \frac{\prod_{i=1}^m \varepsilon_i^{y_{pi}}}{\gamma_{x_p}(\boldsymbol{\varepsilon})}.$$

Considerando los patrones de respuesta de todas las personas, se obtiene la función de verosimilitud de los parámetros  $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m)$  a la luz de la matriz  $n \times m$  de datos observados  $\mathbf{y}$ , condicional a los puntajes observados  $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$ :

$$\begin{aligned} \mathcal{L}(\boldsymbol{\varepsilon}; \mathbf{y}|\boldsymbol{x}) &= \prod_{p=1}^n \Pr[\mathbf{Y}_p = \mathbf{y}_p \mid x_p; \boldsymbol{\varepsilon}] \\ &= \prod_{p=1}^n \left( \frac{\prod_{i=1}^m \varepsilon_i^{y_{pi}}}{\gamma_{x_p}(\boldsymbol{\varepsilon})} \right). \end{aligned}$$

Reiteramos que esta función no depende de los parámetros de las personas y que, al maximizarla con respecto a  $\boldsymbol{\varepsilon}$ , se obtendrán estimadores consistentes. Para encontrar el máximo, se considera la función de logverosimilitud,

$$\begin{aligned} \ell(\boldsymbol{\varepsilon}; \mathbf{y}|\boldsymbol{x}) &= \log \mathcal{L}(\boldsymbol{\varepsilon}; \mathbf{y}|\boldsymbol{x}) \\ &= \sum_{i=1}^m s_i \log(\varepsilon_i) - \sum_{p=1}^n \log(\gamma_{x_p}(\boldsymbol{\varepsilon})), \end{aligned} \quad (8)$$



y se calcula, para cada ítem, la derivada parcial de la logverosimilitud con respecto a  $\varepsilon_i$ . Este cálculo requiere derivar la función elemental simétrica  $\gamma_{x_p}(\boldsymbol{\varepsilon})$ . Considérese, como ejemplo, la función elemental simétrica de tercer orden para el vector  $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4)$ ,

$$\begin{aligned} \gamma_3(\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4) \\ = \varepsilon_1 \varepsilon_2 \varepsilon_3 + \varepsilon_1 \varepsilon_2 \varepsilon_4 + \varepsilon_1 \varepsilon_3 \varepsilon_4 + \varepsilon_2 \varepsilon_3 \varepsilon_4. \end{aligned}$$

Su derivada parcial con respecto a  $\varepsilon_2$ , por ejemplo, es:

$$\frac{\partial \gamma_3(\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4)}{\partial \varepsilon_2} = \varepsilon_1 \varepsilon_3 + \varepsilon_1 \varepsilon_4 + \varepsilon_3 \varepsilon_4,$$

lo cual es la función elemental simétrica de segundo orden para el vector  $(\varepsilon_1, \varepsilon_3, \varepsilon_4)$ , es decir, el vector  $\boldsymbol{\varepsilon}$  original del cual se ha quitado  $\varepsilon_2$ . En general, la derivada parcial de  $\gamma_{x_p}(\boldsymbol{\varepsilon})$  con respecto a  $\varepsilon_i$  es:

$$\frac{\partial \gamma_{x_p}(\boldsymbol{\varepsilon})}{\partial \varepsilon_i} = \gamma_{x_p-1}(\boldsymbol{\varepsilon}^{(-i)}),$$

donde  $\boldsymbol{\varepsilon}^{(-i)}$  denota el vector  $\boldsymbol{\varepsilon}$  del cual se ha quitado  $\varepsilon_i$ . De esta manera, las derivadas parciales de la logverosimilitud en la **ecuación 8** con respecto a  $\varepsilon_i$  son:

$$\frac{\partial \ell(\boldsymbol{\varepsilon}; \mathbf{y}|\mathbf{x})}{\partial \varepsilon_i} = \frac{s_i}{\varepsilon_i} - \sum_{p=1}^n \frac{\gamma_{x_p-1}(\boldsymbol{\varepsilon}^{(-i)})}{\gamma_{x_p}(\boldsymbol{\varepsilon})} \quad (9)$$

para  $i = 1, \dots, m$ .

Los valores para las  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m$  que maximizan la función de logverosimilitud condicional en la **ecuación 8** son aquellos para los cuales las derivadas parciales en la **ecuación 9** son iguales a cero. La solución de este sistema de ecuaciones se encuentra utilizando algoritmos iterativos. Finalmente, como explicamos en la sección de JML, para identificar el modelo y obtener una solución única es necesario imponer una restricción adicional, por ejemplo, fijar el parámetro  $\beta_1 = 0$ , lo cual corresponde con fijar  $\varepsilon_1 = 1$ . Esto significa que se elimina la derivada parcial con respecto a  $\varepsilon_1$  y que el sistema en la **ecuación 9** contiene solo  $m - 1$  ecuaciones.

### MÁXIMA VEROSIMILITUD MARGINAL (MML)

El segundo método también elimina los parámetros  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_n)$  de la función de verosimilitud; para ello, agrega dos supuestos: (1) las  $n$  personas que respondieron los  $m$  ítems forman una muestra aleatoria de una población y (2) sus parámetros se extrajeron de cierta distribución teórica para  $\boldsymbol{\theta}$  en esta población.

Retómese el caso de una persona  $p$  con su patrón de respuestas observadas  $\mathbf{y}_p$ . El parámetro  $\theta_p$  de esta persona sigue representando un valor numérico desconocido, pero supongamos, por motivos ilustrativos, que únicamente puede asumir tres valores<sup>10</sup>, por ejemplo:  $\theta_{p1} = -0.5$ ,  $\theta_{p2} = 1.2$  y  $\theta_{p3} = 2.0$ ; además, las probabilidades con las que el parámetro asume estos valores son .20, .50 y .30, respectivamente. Esto quiere decir que se define una variable  $\Theta_p$  discreta, con su distribución de probabilidad:

$$\begin{cases} \Pr(\Theta_p = -0.5) = .20 \\ \Pr(\Theta_p = 1.2) = .50 \\ \Pr(\Theta_p = 2.0) = .30. \end{cases} \quad (10)$$

Este supuesto permite calcular, mediante el teorema de la probabilidad total<sup>11</sup>, la probabilidad marginal del patrón de respuesta  $\mathbf{y}_p$ :

$$\begin{aligned} \Pr(\mathbf{Y}_p = \mathbf{y}_p; \boldsymbol{\beta}) \\ = \sum_{j=1}^3 [\Pr(\mathbf{Y}_p = \mathbf{y}_p | \Theta_p = \theta_{pj}; \boldsymbol{\beta}) \Pr(\Theta_p = \theta_{pj})] \\ = \Pr(\mathbf{Y}_p = \mathbf{y}_p | \Theta_p = -0.5; \boldsymbol{\beta}) \times \Pr(\Theta_p = -0.5) \\ + \Pr(\mathbf{Y}_p = \mathbf{y}_p | \Theta_p = 1.2; \boldsymbol{\beta}) \times \Pr(\Theta_p = 1.2) \\ + \Pr(\mathbf{Y}_p = \mathbf{y}_p | \Theta_p = 2.0; \boldsymbol{\beta}) \times \Pr(\Theta_p = 2.0). \end{aligned} \quad (11)$$

Una probabilidad marginal puede interpretarse como una probabilidad que promedia (en este caso, tres) probabilidades condicionales. Puesto que se conocen los valores de  $\Theta_p$  y las probabilidades asociadas (por la **ecuación 10**), la probabilidad marginal en la

**ecuación 11** depende únicamente de los parámetros de los ítems  $\beta$ . Los posibles valores para  $\Theta_p$  se han integrado analíticamente de tal manera que dicha probabilidad ya no depende de  $\Theta_p$ .

En el ejemplo anterior, es poco plausible que  $\Theta_p$  solo pueda asumir tres valores y que las probabilidades asociadas sean las de la ecuación 10. Es más común y apropiado considerar  $\Theta_p$  como una variable continua, que puede asumir cualquier valor real (ya que en el modelo de Rasch  $\theta_p \in \mathbb{R}$ ), y añadir el supuesto de que la distribución teórica de  $\Theta_p$  es normal. En general, una distribución normal específica la función de densidad para una variable continua y se caracteriza por dos parámetros: su media  $\mu$  y su varianza  $\sigma^2$ . La densidad normal de  $\Theta_p$  se define por:

$$\varphi_{\mu,\sigma^2}(\theta_p) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2} \frac{(\theta_p - \mu)^2}{\sigma^2}\right] \quad (12)$$

donde  $\pi \approx 3.14159$ . Esta densidad normal tiene el mismo papel que la función de probabilidad en la **ecuación 10**: permite calcular la probabilidad marginal del patrón de respuesta  $\mathbf{y}_p$  integrando los posibles valores para  $\Theta_p$ . En particular:

$$\begin{aligned} \Pr(\mathbf{Y}_p = \mathbf{y}_p; \beta, \mu, \sigma^2) \\ = \int_{-\infty}^{+\infty} \Pr(\mathbf{Y}_p = \mathbf{y}_p | \Theta_p = \theta_p; \beta) \varphi_{\mu,\sigma^2}(\theta_p) d\theta_p. \end{aligned} \quad (13)$$

Puesto que  $\Theta_p$  ahora es una variable continua, la probabilidad marginal se calcula por una integral definida (entre todos los posibles valores reales, de a  $-\infty$  a  $+\infty$ ) en vez de una suma (entre los tres posibles valores discretos de  $\Theta_p$  en el ejemplo anterior). Sin embargo, nótese la similitud en la estructura de las Ecuaciones 11 y 13: Se suma o integra el producto de la probabilidad del patrón de respuesta  $\mathbf{y}_p$  condicional a un valor de  $\Theta_p$  por la probabilidad o densidad de este valor. Por otro lado, una diferencia con el ejemplo de la distribución discreta para  $\Theta_p$  en la **ecuación 10** es que esta se conoce por completo, mientras que la densidad normal en la **ecuación 12** incluye dos parámetros desconocidos ( $\mu$  y  $\sigma^2$ ). Sin embargo, como veremos en el siguiente párrafo, es-

tos parámetros se incluyen en el proceso de estimación, suponiendo que tienen el mismo valor para todas las personas. De esta manera, la probabilidad marginal para el patrón de respuesta de cualquier persona  $p$  solo depende de los parámetros  $\beta$  de los ítems y de los parámetros  $\mu$  y  $\sigma^2$  de la distribución normal; ya no depende de  $\theta_p$ .

Hasta ahora, solo consideramos el patrón de respuesta de una persona. Generalizando la idea en los párrafos anteriores, en MML se suele suponer que los parámetros  $\theta_p$  de *todas* las personas se han extraído aleatoriamente de la *misma* distribución normal; se escribe:

$$\Theta_p \stackrel{iid}{\sim} \mathcal{N}(\mu, \sigma^2) \quad \text{para } p = 1, \dots, n,$$

lo cual se puede leer como “las  $\Theta_p$  de las distintas personas se distribuyen normalmente, con media  $\mu$  y varianza  $\sigma^2$ , de forma idéntica e independiente entre sí”. A partir de este supuesto, la función de verosimilitud de los parámetros a la luz de las respuestas observadas  $\mathbf{y}$  de *todas* las personas es:

$$\begin{aligned} \mathcal{L}(\beta, \mu, \sigma^2; \mathbf{y}) &= \prod_{p=1}^n \Pr(\mathbf{Y}_p = \mathbf{y}_p; \beta, \mu, \sigma^2) \\ &= \prod_{p=1}^n \int_{-\infty}^{+\infty} \Pr(\mathbf{Y}_p = \mathbf{y}_p | \Theta_p = \theta_p; \beta) \varphi_{\mu,\sigma^2}(\theta_p) d\theta_p \end{aligned}$$

y la logverosimilitud:

$$\begin{aligned} \ell(\beta, \mu, \sigma^2; \mathbf{y}) \\ = \sum_{p=1}^n \log \left[ \int_{-\infty}^{+\infty} \Pr(\mathbf{Y}_p = \mathbf{y}_p | \Theta_p = \theta_p; \beta) \varphi_{\mu,\sigma^2}(\theta_p) d\theta_p \right]. \end{aligned} \quad (14)$$

Para hallar los valores de los parámetros ( $\beta, \mu, \sigma^2$ ) que maximizan esta función, se aplica esencialmente el mismo procedimiento que para el caso de JML y CML: calcular las derivadas parciales con respecto a cada uno de los parámetros, igualar estas derivadas parciales a 0 y resolver el sistema de ecuaciones resultante. Al igual que para JML y CML, se debe



incluir también para MML una restricción a los parámetros para identificar el modelo. En este caso, casi siempre se realiza esta restricción a través de  $\mu$ , fijando  $\mu = 0$ . Esto quiere decir que se estiman  $m + 1$  parámetros:  $(\beta_1, \beta_2, \dots, \beta_m, \sigma^2)$ . El cálculo de estas derivadas parciales es matemáticamente demandante; el lector interesado lo encuentra en el *material suplementario (anexo 1)*. Aquí solo presentamos el resultado:

$$\left\{ \begin{aligned} s_i &= \sum_{x=0}^m n_x \mathbb{E}_{\Theta_p|x,\beta,\sigma^2} \left[ \frac{\exp(\theta_p - \beta_i)}{1 + \exp(\theta_p - \beta_i)} \right] \\ &\text{para } i = 1, \dots, m \\ \sigma^2 &= \frac{1}{n} \sum_{x=0}^m n_x \mathbb{E}_{\Theta_p|x,\beta,\sigma^2} (\theta_p^2), \end{aligned} \right. \quad (15)$$

donde  $n_x$  es el número de personas en la muestra con puntaje total igual a  $x$  y los valores esperados (denotados con el símbolo  $\mathbb{E}$ ) son con respecto a la distribución posterior de  $\Theta_p$  a la luz de los datos observados. Dichos valores esperados requieren

resolver integrales definidas que no tienen una solución cerrada, por lo cual encontrar su máximo es un problema complicado que se suele aproximar con algoritmos iterativos.

Resumiendo, en MML se añade el supuesto de que los parámetros de las personas son variables (o efectos aleatorios) con una distribución específica, lo cual lleva a una función de verosimilitud que no incluye dichos parámetros y permite obtener estimadores consistentes de los parámetros de los ítems. Aunque se puede escoger cualquier distribución teórica para los parámetros de las personas, casi siempre se supone una distribución normal.

### COMPARACIÓN ENTRE MÁXIMA VEROSIMILITUD CONJUNTA, CONDICIONAL Y MARGINAL

La **tabla 1** contrasta, de manera sintética, las principales diferencias entre las tres variantes de MLE para los parámetros de los ítems en el modelo de Rasch descritas en este artículo, resaltando las funciones de verosimilitud y las ecuaciones de estimación. Debido a que los estimadores obtenidos mediante JML no son consistentes, en la gran mayoría de las aplicaciones se utilizan CML o MML.

**Tabla 1.** Comparación de las variantes de estimación por máxima verosimilitud para los parámetros de los ítems en el modelo de Rasch

	Método de estimación		
	Máxima verosimilitud conjunta (JML)	Máxima verosimilitud condicional (CML)	Máxima verosimilitud marginal (MML)
Función de verosimilitud	$\mathcal{L}(\theta, \beta; \mathbf{y}) = \prod_{p=1}^n \Pr(\mathbf{Y}_p = \mathbf{y}_p; \theta_p, \beta)$	$\mathcal{L}(\epsilon; \mathbf{y} \mathbf{x}) = \prod_{p=1}^n \Pr[\mathbf{Y}_p = \mathbf{y}_p   x_p; \epsilon]$	$\mathcal{L}(\beta, \mu, \sigma^2; \mathbf{y})$ $= \prod_{p=1}^n \int_{-\infty}^{+\infty} \Pr(\mathbf{Y}_p = \mathbf{y}_p   \theta_p; \beta) \varphi_{\mu, \sigma^2}(\theta_p) d\theta_p$
Restricción para la identificación <sup>‡</sup>	$\beta_1 = 0$	$\beta_1 = 0$	$\mu = 0$
Ecuaciones de estimación	$\left\{ \begin{aligned} x_p &= \sum_{i=1}^m \frac{\exp(\theta_p - \beta_i)}{1 + \exp(\theta_p - \beta_i)} \quad * \\ s_i &= \sum_{p=1}^n \frac{\exp(\theta_p - \beta_i)}{1 + \exp(\theta_p - \beta_i)} \quad ** \end{aligned} \right.$ * para $p = 1, \dots, n$ ** para $i = 2, \dots, m$	$s_i = \epsilon_i \sum_{p=1}^n \frac{y_{x_p-1}(\epsilon^{(-i)})}{y_{x_p}(\epsilon)}$ para $i = 2, \dots, m$	$\left\{ \begin{aligned} s_i &= \sum_{x=0}^m n_x \mathbb{E}_{\Theta_p x,\beta,\sigma^2} \left[ \frac{\exp(\theta_p - \beta_i)}{1 + \exp(\theta_p - \beta_i)} \right] \\ &\text{para } i = 1, \dots, m \\ \sigma^2 &= \frac{1}{n} \sum_{x=0}^m n_x \mathbb{E}_{\Theta_p x,\beta,\sigma^2} (\theta_p^2) \end{aligned} \right.$

<sup>‡</sup>Se presenta la restricción utilizada en el resto del texto. (Existen otras maneras para identificar el modelo.)

Al elegir entre CML y MML son relevantes dos consideraciones. En primera instancia, CML requiere un estadístico suficiente para  $\theta_p$  —como el número de aciertos en el modelo de Rasch— y solo una subfamilia de los modelos de la TRI posee esta propiedad. Otros ejemplos en los que es posible aplicar CML incluyen el modelo logístico lineal de rasgo latente (LLTM)<sup>12</sup> y el modelo de crédito parcial<sup>13</sup>; sin embargo, para muchos otros modelos en la TRI, como los modelos logísticos de dos y tres parámetros<sup>14</sup> o el modelo de respuesta graduada<sup>15</sup>, no existe un estadístico suficiente, por lo cual no es posible realizar la estimación por CML, y se suele utilizar MML. Además, en la presencia de datos faltantes, la aplicación de CML puede resultar muy compleja (o incluso inviable) mientras que MML ofrece una mayor flexibilidad en este sentido.

La segunda consideración hace referencia a los supuestos adicionales en MML, que no se requieren en CML. Estrictamente hablando, la validez de las estimaciones de los parámetros, así como de las inferencias y decisiones que se deriven de ellas, se sustenta teóricamente en el cumplimiento de los supuestos del modelo. En este sentido, el uso de MML puede hacer al analista (ligeramente) más vulnerable en comparación con el uso de CML.

### SOFTWARE Y EJEMPLO ILUSTRATIVO

Debido a que casi nunca se utiliza JML, la estimación por máxima verosimilitud en la TRI generalmente implica un procedimiento de dos pasos: Primero, se estiman los parámetros de los ítems con CML o MML y, posteriormente, tomando estas estimaciones como valores “conocidos”, se obtienen las estimaciones de los parámetros de las personas mediante el método descrito en nuestro artículo en el número anterior de esta revista. Existen varios programas que implementan MLE para modelos TRI, incluyendo paquetes del *software* R como eRm<sup>16</sup>, mirt<sup>17</sup>, ltm<sup>18</sup> y TAM<sup>19</sup>.

Para ilustrar la estimación de los parámetros de los ítems utilizamos los datos introducidos en nuestro artículo previo. El código en R (específicamente bajo el encabezado “PARTE 1: Estimación de los parámetros de los ítems”), que se encuentra disponible en el *material suplementario* alojado en un proyecto de Open Science Framework ([https://](https://doi.org/10.17605/OSF.IO/23J7U)

[doi.org/10.17605/OSF.IO/23J7U](https://doi.org/10.17605/OSF.IO/23J7U)), ajusta el modelo de Rasch a un examen de 20 reactivos de opción múltiple sobre el concepto de homeostasis respondido por 669 estudiantes<sup>20</sup>. Se emplean tanto el paquete eRm, que implementa estimación por CML, como el paquete mirt, que utiliza MML. Este código, que incluye comentarios explicativos detallados, ilustra las diferencias entre ambos enfoques; específicamente, aquellas relacionadas con las distintas maneras de identificar el modelo y con la estimación adicional, en mirt, de la varianza de la distribución normal asumida para los parámetros de las personas.

### COMENTARIOS CONCLUYENTES

MLE es uno de los métodos más utilizados para obtener los valores de los parámetros en modelos estadísticos. En este artículo revisamos tres variantes de MLE para estimar los parámetros de los ítems en los modelos TRI. Aunque el desarrollo se centró en el modelo de Rasch, los principios expuestos aplican de manera análoga a otros modelos de este marco.

Un enfoque alternativo para la estimación de los parámetros en modelos estadísticos y psicométricos es la estimación bayesiana. La diferencia más importante con el enfoque frecuentista (del cual MLE es un ejemplo paradigmático) radica en el tratamiento de la incertidumbre. Un bayesiano considera todos los parámetros como variables aleatorias y define distribuciones para representar la incertidumbre sobre estos parámetros. (Existe un paralelismo conceptual con cómo MML trata los parámetros de las personas, aunque el enfoque bayesiano lleva esta idea hasta sus últimas consecuencias). El objetivo principal de un análisis bayesiano es caracterizar la distribución posterior, la cual refleja la incertidumbre sobre los parámetros después de conocer los datos y que se obtiene combinando, a través del teorema de Bayes, la distribución previa (que representa el conocimiento y la incertidumbre antes de conocer los datos) con la función de verosimilitud. El lector que quiere aprender sobre este enfoque puede consultar, entre otros, los libros de Fox<sup>21</sup> y Levy y Mislavy<sup>22</sup>.

Un análisis psicométrico en el marco de la TRI no termina con la estimación de los parámetros del modelo. Una pregunta natural que surge es qué tan confiables o precisas son estas estimaciones, ya que el psicómetra sabe que, en general, el valor estimado  $\hat{\theta}$

de un parámetro no coincide exactamente con su valor verdadero  $\vartheta$ ; es decir, casi siempre hay un error de estimación. En el tercer y último artículo de esta serie abordamos este tema con mayor detalle con un énfasis particular en la función de información de Fisher.

### 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 Muñoz-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. doi: 10.1016/S2007-5057(14)72724-3
7. Leenen I, Naveja JJ, Vázquez-Lira R. Estimación por máxima verosimilitud en la teoría de respuesta al ítem: Parámetros de las personas. *Inv Ed Med.* 2026;(15):118-127. doi: 10.22201/fm.20075057e.2026.58.26791
8. Rasch G. *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests.* Copenhagen: Nielsen & Lydiche; 1960.
9. Haberman SJ. Models with nuisance and incidental parameters. En: van der Linden WJ (Ed.) *Handbook of item response theory (Vol. 2): Statistical tools.* Boca Raton, FL: CRC Press; 2016. p. 151-170.
10. Verhelst ND. Itemresponstheorie [Teoría de respuesta al ítem]. En: Eggen TJHM, Sanders PF (Ed.) *Psychometrie in de praktijk.* Arnhem, NL: Cito; 1993. p. 83-178.
11. Ross S. *A first course in probability (10ª ed.).* Harlow, UK: Pearson; 2020.
12. Fischer GH. Logistic latent trait models with linear constraints. *Psychometrika* 1983;48:3-26. doi: 10.1007/BF02314674
13. Masters GN. A Rasch model for partial credit scoring. *Psychometrika* 1982;47:149-174. doi: 10.1007/BF02296272
14. Birnbaum A. Some latent trait models and their use in inferring an examinee's ability. En: Lord FM, Novick MR (Ed.) *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley; 1968. p. 397-479.
15. Samejima F. Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph No. 17.* Psychometric Society; 1969.
16. 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
17. 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
18. 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
19. Robitzsch A, Kiefer T, Wu M. Package 'TAM': Test analysis modules (R package version 4.14). <https://bit.ly/42T6GpI>
20. 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
21. Fox JP. *Bayesian item response modeling: Theory and applications.* Nueva York: Springer; 2010.
22. Levy R, Mislevy RJ. *Bayesian psychometric modeling.* Boca Raton, FL: Chapman & Hall/CRC; 2016.

## ANEXO 1

Se derivan las **ecuaciones 15** de estimación de parámetros por MML a partir de la función de logverosimilitud en la **ecuación 14**. Para razones de identificación, se fija  $\mu = 0$  y se estiman entonces los parámetros  $(\beta_1, \beta_2, \dots, \beta_m, \sigma^2)$ .

Calculamos la derivada parcial de la logverosimilitud con respecto a cada uno de los  $m + 1$  parámetros. Primero, obtenemos la derivada parcial con respecto a  $\beta_i$  para cualquier ítem  $i$  ( $i = 1, \dots, m$ ).

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\beta}, \sigma^2; \mathbf{y})}{\partial \beta_i} &= \sum_{p=1}^n \frac{\partial}{\partial \beta_i} \log \left[ \int_{-\infty}^{+\infty} \Pr(\mathbf{Y}_p = \mathbf{y}_p \mid \Theta_p = \theta_p; \boldsymbol{\beta}) \varphi_{0, \sigma^2}(\theta_p) d\theta_p \right] \\ &= \sum_{p=1}^n \frac{\frac{\partial}{\partial \beta_i} \int_{-\infty}^{+\infty} \Pr(\mathbf{Y}_p = \mathbf{y}_p \mid \Theta_p = \theta_p; \boldsymbol{\beta}) \varphi_{0, \sigma^2}(\theta_p) d\theta_p}{\int_{-\infty}^{+\infty} \Pr(\mathbf{Y}_p = \mathbf{y}_p \mid \Theta_p = \theta_p; \boldsymbol{\beta}) \varphi_{0, \sigma^2}(\theta_p) d\theta_p} \\ &= \sum_{p=1}^n \frac{\int_{-\infty}^{+\infty} \frac{\partial}{\partial \beta_i} [\Pr(\mathbf{Y}_p = \mathbf{y}_p \mid \Theta_p = \theta_p; \boldsymbol{\beta}) \varphi_{0, \sigma^2}(\theta_p)] d\theta_p}{\Pr(\mathbf{Y}_p = \mathbf{y}_p; \boldsymbol{\beta}, \sigma^2)}, \end{aligned}$$

donde el denominador es la probabilidad marginal (para  $\mu = 0$ ) en la **ecuación 13**. Recordando que, en general para cualquier función  $f$ ,  $d \log f(x)/dx = 1/f(x) \times df(x)/dx$  y, por lo tanto,  $df(x)/dx = f(x) \times d \log f(x)/dx$ , y aplicándolo al integrando en el numerador, se obtiene:

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\beta}, \sigma^2; \mathbf{y})}{\partial \beta_i} &= \sum_{p=1}^n \frac{\int_{-\infty}^{+\infty} \Pr(\mathbf{Y}_p = \mathbf{y}_p \mid \Theta_p = \theta_p; \boldsymbol{\beta}) \varphi_{0, \sigma^2}(\theta_p) \frac{\partial}{\partial \beta_i} \log[\Pr(\mathbf{Y}_p = \mathbf{y}_p \mid \Theta_p = \theta_p; \boldsymbol{\beta}) \varphi_{0, \sigma^2}(\theta_p)] d\theta_p}{\Pr(\mathbf{Y}_p = \mathbf{y}_p; \boldsymbol{\beta}, \sigma^2)} \\ &= \sum_{p=1}^n \int_{-\infty}^{+\infty} \frac{\Pr(\mathbf{Y}_p = \mathbf{y}_p \mid \Theta_p = \theta_p; \boldsymbol{\beta}) \varphi_{0, \sigma^2}(\theta_p)}{\Pr(\mathbf{Y}_p = \mathbf{y}_p; \boldsymbol{\beta}, \sigma^2)} \frac{\partial \log[\Pr(\mathbf{Y}_p = \mathbf{y}_p \mid \Theta_p = \theta_p; \boldsymbol{\beta}) \varphi_{0, \sigma^2}(\theta_p)]}{\partial \beta_i} d\theta_p. \end{aligned}$$

Aplicando el teorema de Bayes a la primera fracción, resulta en:

$$\frac{\partial \ell(\boldsymbol{\beta}, \sigma^2; \mathbf{y})}{\partial \beta_i} = \sum_{p=1}^n \int_{-\infty}^{+\infty} \phi(\theta_p \mid \mathbf{Y}_p = \mathbf{y}_p; \boldsymbol{\beta}, \sigma^2) \frac{\partial \log[\Pr(\mathbf{Y}_p = \mathbf{y}_p \mid \theta_p; \boldsymbol{\beta}) \varphi_{0, \sigma^2}(\theta_p)]}{\partial \beta_i} d\theta_p, \quad (\text{A.1})$$

donde  $\phi(\theta_p \mid \mathbf{Y}_p = \mathbf{y}_p; \boldsymbol{\beta}, \sigma^2)$  representa la densidad de  $\theta_p$  condicional a las respuestas observadas  $\mathbf{y}_p$  de la persona  $p$  y en función de los parámetros  $(\boldsymbol{\beta}, \sigma^2)$ ; es decir, considerando la distribución posterior de  $\Theta_p$ . Al final de este anexo elaboraremos esta distribución posterior.

Ahora, seguimos elaborando la **ecuación A.1**:

$$\begin{aligned}
\frac{\partial \ell(\boldsymbol{\beta}, \sigma^2; \mathbf{y})}{\partial \beta_i} &= \sum_{p=1}^n \int_{-\infty}^{+\infty} \phi(\theta_p | \mathbf{Y}_p = \mathbf{y}_p; \boldsymbol{\beta}, \sigma^2) \frac{\partial \log[\Pr(\mathbf{Y}_p = \mathbf{y}_p | \theta_p = \theta_p; \boldsymbol{\beta}) \varphi_{0, \sigma^2}(\theta_p)]}{\partial \beta_i} d\theta_p \\
&= \sum_{p=1}^n \int_{-\infty}^{+\infty} \phi(\theta_p | \mathbf{Y}_p = \mathbf{y}_p; \boldsymbol{\beta}, \sigma^2) \frac{\partial \log[\Pr(\mathbf{Y}_p = \mathbf{y}_p | \theta_p = \theta_p; \boldsymbol{\beta})] + \log \varphi_{0, \sigma^2}(\theta_p)}{\partial \beta_i} d\theta_p \\
&= \sum_{p=1}^n \int_{-\infty}^{+\infty} \phi(\theta_p | \mathbf{Y}_p = \mathbf{y}_p; \boldsymbol{\beta}, \sigma^2) \frac{\partial \log[\Pr(\mathbf{Y}_p = \mathbf{y}_p | \theta_p = \theta_p; \boldsymbol{\beta})]}{\partial \beta_i} d\theta_p.
\end{aligned} \tag{A.2}$$

Con respecto a la derivada en la última expresión, se obtiene:

$$\begin{aligned}
\frac{\partial \log[\Pr(\mathbf{Y}_p = \mathbf{y}_p | \theta_p = \theta_p; \boldsymbol{\beta})]}{\partial \beta_i} &= \frac{\partial}{\partial \beta_i} \log \left[ \prod_{i=1}^m \frac{\exp [y_{pi}(\theta_p - \beta_i)]}{1 + \exp (\theta_p - \beta_i)} \right] \\
&= \frac{\partial}{\partial \beta_i} \sum_{i=1}^m \log \frac{\exp [y_{pi}(\theta_p - \beta_i)]}{1 + \exp (\theta_p - \beta_i)} \\
&= \frac{\partial y_{pi}(\theta_p - \beta_i)}{\partial \beta_i} - \frac{\partial \log[1 + \exp (\theta_p - \beta_i)]}{\partial \beta_i} \\
&= -y_{pi} + \frac{\exp (\theta_p - \beta_i)}{1 + \exp (\theta_p - \beta_i)}.
\end{aligned}$$

Sustituir este resultado en la **ecuación A.2**, resulta en:

$$\begin{aligned}
\frac{\partial \ell(\boldsymbol{\beta}, \sigma^2; \mathbf{y})}{\partial \beta_i} &= \sum_{p=1}^n \int_{-\infty}^{+\infty} \left[ -y_{pi} + \frac{\exp (\theta_p - \beta_i)}{1 + \exp (\theta_p - \beta_i)} \right] \phi(\theta_p | \mathbf{Y}_p = \mathbf{y}_p; \boldsymbol{\beta}, \sigma^2) d\theta_p \\
&= \sum_{p=1}^n \int_{-\infty}^{+\infty} (-y_{pi}) \phi(\theta_p | \mathbf{Y}_p = \mathbf{y}_p; \boldsymbol{\beta}, \sigma^2) d\theta_p \\
&\quad + \sum_{p=1}^n \int_{-\infty}^{+\infty} \frac{\exp (\theta_p - \beta_i)}{1 + \exp (\theta_p - \beta_i)} \phi(\theta_p | \mathbf{Y}_p = \mathbf{y}_p; \boldsymbol{\beta}, \sigma^2) d\theta_p
\end{aligned}$$

$$\begin{aligned}
 \frac{\partial \ell(\boldsymbol{\beta}, \sigma^2; \mathbf{y})}{\partial \beta_i} &= \sum_{p=1}^n \left[ (-y_{pi}) \int_{-\infty}^{+\infty} \phi(\theta_p | \mathbf{Y}_p = \mathbf{y}_p; \boldsymbol{\beta}, \sigma^2) d\theta_p \right] \\
 &\quad + \sum_{p=1}^n \int_{-\infty}^{+\infty} \frac{\exp(\theta_p - \beta_i)}{1 + \exp(\theta_p - \beta_i)} \phi(\theta_p | \mathbf{Y}_p = \mathbf{y}_p; \boldsymbol{\beta}, \sigma^2) d\theta_p \\
 &= \sum_{p=1}^n (-y_{pi}) + \sum_{p=1}^n \int_{-\infty}^{+\infty} \frac{\exp(\theta_p - \beta_i)}{1 + \exp(\theta_p - \beta_i)} \phi(\theta_p | \mathbf{Y}_p = \mathbf{y}_p; \boldsymbol{\beta}, \sigma^2) d\theta_p \\
 &= -s_i + \sum_{p=1}^n \int_{-\infty}^{+\infty} \frac{\exp(\theta_p - \beta_i)}{1 + \exp(\theta_p - \beta_i)} \phi(\theta_p | \mathbf{Y}_p = \mathbf{y}_p; \boldsymbol{\beta}, \sigma^2) d\theta_p,
 \end{aligned}
 \tag{A.3}$$

donde  $s_i$  se define como en la **ecuación 5**. La integral en la última expresión es el valor esperado de  $\frac{\exp(\theta_p - \beta_i)}{1 + \exp(\theta_p - \beta_i)}$  (la probabilidad de que la persona  $p$  acierte el ítem  $i$ ) con respecto a la distribución posterior del parámetro  $\theta_p$  a la luz de las respuestas observadas  $\mathbf{y}_p$  de la persona  $p$ :

$$\mathbb{E}_{\theta_p | \mathbf{y}_p, \boldsymbol{\beta}, \sigma^2} \left[ \frac{\exp(\theta_p - \beta_i)}{1 + \exp(\theta_p - \beta_i)} \right] = \int_{-\infty}^{+\infty} \frac{\exp(\theta_p - \beta_i)}{1 + \exp(\theta_p - \beta_i)} \phi(\theta_p | \mathbf{Y}_p = \mathbf{y}_p; \boldsymbol{\beta}, \sigma^2) d\theta_p.
 \tag{A.4}$$

Igualar la Ec. (A.3) a 0 y reacomodar términos resulta en:

$$s_i = \sum_{p=1}^n \mathbb{E}_{\theta_p | \mathbf{y}_p, \boldsymbol{\beta}, \sigma^2} \left[ \frac{\exp(\theta_p - \beta_i)}{1 + \exp(\theta_p - \beta_i)} \right].
 \tag{A.5}$$

A continuación, aplicamos una estrategia similar para la derivada parcial de la logverosimilitud con respecto a  $\sigma$ :

$$\begin{aligned}
 \frac{\partial \ell(\boldsymbol{\beta}, \sigma^2; \mathbf{y})}{\partial \sigma} &= \sum_{p=1}^n \frac{\partial}{\partial \sigma} \log \left[ \int_{-\infty}^{+\infty} \Pr(\mathbf{Y}_p = \mathbf{y}_p | \Theta_p = \theta_p; \boldsymbol{\beta}) \varphi_{0, \sigma^2}(\theta_p) d\theta_p \right] \\
 &= \sum_{p=1}^n \frac{\frac{\partial}{\partial \sigma} \int_{-\infty}^{+\infty} \Pr(\mathbf{Y}_p = \mathbf{y}_p | \Theta_p = \theta_p; \boldsymbol{\beta}) \varphi_{0, \sigma^2}(\theta_p) d\theta_p}{\int_{-\infty}^{+\infty} \Pr(\mathbf{Y}_p = \mathbf{y}_p | \Theta_p = \theta_p; \boldsymbol{\beta}) \varphi_{0, \sigma^2}(\theta_p) d\theta_p}
 \end{aligned}$$

$$\begin{aligned}
\frac{\partial \ell(\boldsymbol{\beta}, \sigma^2; \mathbf{y})}{\partial \sigma} &= \sum_{p=1}^n \frac{\int_{-\infty}^{+\infty} \frac{\partial}{\partial \sigma} [\Pr(\mathbf{Y}_p = \mathbf{y}_p | \Theta_p = \theta_p; \boldsymbol{\beta}) \varphi_{0, \sigma^2}(\theta_p)] d\theta_p}{\Pr(\mathbf{Y}_p = \mathbf{y}_p; \boldsymbol{\beta}, \sigma^2)} \\
&= \sum_{p=1}^n \frac{\int_{-\infty}^{+\infty} \Pr(\mathbf{Y}_p = \mathbf{y}_p | \Theta_p = \theta_p; \boldsymbol{\beta}) \varphi_{0, \sigma^2}(\theta_p) \frac{\partial}{\partial \sigma} \log[\Pr(\mathbf{Y}_p = \mathbf{y}_p | \Theta_p = \theta_p; \boldsymbol{\beta}) \varphi_{0, \sigma^2}(\theta_p)] d\theta_p}{\Pr(\mathbf{Y}_p = \mathbf{y}_p; \boldsymbol{\beta}, \sigma^2)} \\
&= \sum_{p=1}^n \int_{-\infty}^{+\infty} \frac{\Pr(\mathbf{Y}_p = \mathbf{y}_p | \Theta_p = \theta_p; \boldsymbol{\beta}) \varphi_{0, \sigma^2}(\theta_p)}{\Pr(\mathbf{Y}_p = \mathbf{y}_p; \boldsymbol{\beta}, \sigma^2)} \frac{\partial \log[\Pr(\mathbf{Y}_p = \mathbf{y}_p | \Theta_p = \theta_p; \boldsymbol{\beta}) \varphi_{0, \sigma^2}(\theta_p)]}{\partial \sigma} d\theta_p \\
&= \sum_{p=1}^n \int_{-\infty}^{+\infty} \phi(\theta_p | \mathbf{Y}_p = \mathbf{y}_p; \boldsymbol{\beta}, \sigma^2) \frac{\partial \log[\Pr(\mathbf{Y}_p = \mathbf{y}_p | \Theta_p = \theta_p; \boldsymbol{\beta}) \varphi_{0, \sigma^2}(\theta_p)]}{\partial \sigma} d\theta_p \tag{A.6} \\
&= \sum_{p=1}^n \int_{-\infty}^{+\infty} \phi(\theta_p | \mathbf{Y}_p = \mathbf{y}_p; \boldsymbol{\beta}, \sigma^2) \frac{\partial \log[\Pr(\mathbf{Y}_p = \mathbf{y}_p | \Theta_p = \theta_p; \boldsymbol{\beta})]}{\partial \sigma} \frac{\partial \log \varphi_{0, \sigma^2}(\theta_p)}{\partial \sigma} d\theta_p \\
&= \sum_{p=1}^n \int_{-\infty}^{+\infty} \phi(\theta_p | \mathbf{Y}_p = \mathbf{y}_p; \boldsymbol{\beta}, \sigma^2) \frac{\partial \log \varphi_{0, \sigma^2}(\theta_p)}{\partial \sigma} d\theta_p.
\end{aligned}$$

La última derivada es igual a:

$$\begin{aligned}
\frac{\partial \log \varphi_{0, \sigma^2}(\theta_p)}{\partial \sigma} &= \frac{\partial}{\partial \sigma} \log \left[ \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\theta_p^2}{2\sigma^2}\right) \right] \\
&= \frac{\partial}{\partial \sigma} \left[ -\log(\sqrt{2\pi}) - \log \sigma - \frac{\theta_p^2}{2\sigma^2} \right] \\
&= -\frac{1}{\sigma} + \frac{\theta_p^2}{\sigma^3}.
\end{aligned}$$

Sustituir este resultado en la **ecuación A.6** resulta en:

$$\begin{aligned}
\frac{\partial \ell(\boldsymbol{\beta}, \sigma^2; \mathbf{y})}{\partial \sigma} &= \sum_{p=1}^n \int_{-\infty}^{+\infty} \left( -\frac{1}{\sigma} + \frac{\theta_p^2}{\sigma^3} \right) \phi(\theta_p | \mathbf{Y}_p = \mathbf{y}_p; \boldsymbol{\beta}, \sigma^2) d\theta_p \\
&= \sum_{p=1}^n \int_{-\infty}^{+\infty} \left( -\frac{1}{\sigma} \right) \phi(\theta_p | \mathbf{Y}_p = \mathbf{y}_p; \boldsymbol{\beta}, \sigma^2) d\theta_p + \sum_{p=1}^n \int_{-\infty}^{+\infty} \frac{\theta_p^2}{\sigma^3} \phi(\theta_p | \mathbf{Y}_p = \mathbf{y}_p; \boldsymbol{\beta}, \sigma^2) d\theta_p
\end{aligned}$$

$$\begin{aligned}
 \frac{\partial \ell(\boldsymbol{\beta}, \sigma^2; \mathbf{y})}{\partial \sigma} &= \sum_{p=1}^n \left[ \left(-\frac{1}{\sigma}\right) \int_{-\infty}^{+\infty} \phi(\theta_p | \mathbf{Y}_p = \mathbf{y}_p; \boldsymbol{\beta}, \sigma^2) d\theta_p \right] + \sum_{p=1}^n \int_{-\infty}^{+\infty} \frac{\theta_p^2}{\sigma^3} \phi(\theta_p | \mathbf{Y}_p = \mathbf{y}_p; \boldsymbol{\beta}, \sigma^2) d\theta_p \\
 &= \sum_{p=1}^n \left(-\frac{1}{\sigma}\right) + \frac{1}{\sigma^3} \sum_{p=1}^n \int_{-\infty}^{+\infty} \theta_p^2 \phi(\theta_p | \mathbf{Y}_p = \mathbf{y}_p; \boldsymbol{\beta}, \sigma^2) d\theta_p \tag{A.7} \\
 &= -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{p=1}^n \int_{-\infty}^{+\infty} \theta_p^2 \phi(\theta_p | \mathbf{Y}_p = \mathbf{y}_p; \boldsymbol{\beta}, \sigma^2) d\theta_p .
 \end{aligned}$$

Reconociendo que la última integral representa el valor esperado de  $\theta_p^2$  con respecto a la distribución posterior de  $\Theta_p$  a la luz de las respuestas observadas  $\mathbf{y}_p$  de la persona  $p$ , escribimos:

$$\mathbb{E}_{\Theta_p | \mathbf{y}_p, \boldsymbol{\beta}, \sigma^2}(\theta_p^2) = \int_{-\infty}^{+\infty} \theta_p^2 \phi(\theta_p | \mathbf{Y}_p = \mathbf{y}_p; \boldsymbol{\beta}, \sigma^2) d\theta_p \tag{A.8}$$

e igualando la **ecuación A.7** a 0 y reacomodando términos, obtenemos:

$$\sigma^2 = \frac{1}{n} \sum_{p=1}^n \mathbb{E}_{\Theta_p | \mathbf{y}_p, \boldsymbol{\beta}, \sigma^2}(\theta_p^2) . \tag{A.9}$$

Por último, derivamos la distribución posterior  $\phi(\theta_p | \mathbf{Y}_p = \mathbf{y}_p; \boldsymbol{\beta}, \sigma^2)$  con respecto a la cual se definen los valores esperados en las **ecuaciones A.4** y **A.8**:

$$\phi(\theta_p | \mathbf{Y}_p = \mathbf{y}_p; \boldsymbol{\beta}, \sigma^2) = \frac{\Pr(\mathbf{Y}_p = \mathbf{y}_p | \Theta_p = \theta_p; \boldsymbol{\beta}) \varphi_{0, \sigma^2}(\theta_p)}{\Pr(\mathbf{Y}_p = \mathbf{y}_p; \boldsymbol{\beta}, \sigma^2)} . \tag{A.10}$$

En esta ecuación  $\Pr(\mathbf{Y}_p = \mathbf{y}_p | \theta_p; \boldsymbol{\beta})$  es igual a:

$$\begin{aligned}
 \Pr(\mathbf{Y}_p = \mathbf{y}_p | \theta_p; \boldsymbol{\beta}) &= \prod_{i=1}^m \frac{\exp [y_{pi}(\theta_p - \beta_i)]}{1 + \exp (\theta_p - \beta_i)} \\
 &= \exp(x_p \theta_p) \left[ \prod_{i=1}^m \exp(-y_{pi} \beta_i) \right] \mathcal{H}(\theta_p, \boldsymbol{\beta})
 \end{aligned}$$

donde  $\mathcal{H}(\theta_p, \boldsymbol{\beta}) = 1/\prod_{i=1}^m [1 + \exp(\theta_p - \beta_i)]$ . Sustituyendo este resultado en (A.10), se obtiene:

$$\begin{aligned}
 \phi(\theta_p | \mathbf{Y}_p = \mathbf{y}_p; \boldsymbol{\beta}, \sigma^2) &= \frac{\Pr(\mathbf{Y}_p = \mathbf{y}_p | \Theta_p = \theta_p; \boldsymbol{\beta}) \varphi_{0, \sigma^2}(\theta_p)}{\int_{-\infty}^{+\infty} \Pr(\mathbf{Y}_p = \mathbf{y}_p | \Theta_p = \vartheta_p; \boldsymbol{\beta}) \varphi_{0, \sigma^2}(\vartheta_p) \varphi_{0, \sigma^2}(\vartheta_p) d\vartheta_p} \\
 &= \frac{\exp(x_p \theta_p) [\prod_{i=1}^m \exp(-y_{pi} \beta_i)] \mathcal{H}(\theta_p, \boldsymbol{\beta}) \varphi_{0, \sigma^2}(\theta_p)}{\int_{-\infty}^{+\infty} \exp(x_p \vartheta_p) [\prod_{i=1}^m \exp(-y_{pi} \beta_i)] \mathcal{H}(\vartheta_p, \boldsymbol{\beta}) \varphi_{0, \sigma^2}(\vartheta_p) d\vartheta_p} \\
 &= \frac{[\prod_{i=1}^m \exp(-y_{pi} \beta_i)] \exp(x_p \theta_p) \mathcal{H}(\theta_p, \boldsymbol{\beta}) \varphi_{0, \sigma^2}(\theta_p)}{[\prod_{i=1}^m \exp(-y_{pi} \beta_i)] \int_{-\infty}^{+\infty} \exp(x_p \vartheta_p) \mathcal{H}(\vartheta_p, \boldsymbol{\beta}) \varphi_{0, \sigma^2}(\vartheta_p) d\vartheta_p} \\
 &= \frac{\exp(x_p \theta_p) \mathcal{H}(\theta_p, \boldsymbol{\beta}) \varphi_{0, \sigma^2}(\theta_p)}{\int_{-\infty}^{+\infty} \exp(x_p \vartheta_p) \mathcal{H}(\vartheta_p, \boldsymbol{\beta}) \varphi_{0, \sigma^2}(\vartheta_p) d\vartheta_p} \\
 &= \phi(\theta_p | X_p = x_p; \boldsymbol{\beta}, \sigma^2).
 \end{aligned}$$

La última línea indica que la distribución posterior depende de los datos únicamente a través del puntaje total de la persona ( $x_p$ ). Por lo tanto, los valores esperados en las **ecuaciones A.4** y **A.8** son idénticos para todas las personas con el mismo puntaje total. Esto nos permite escribir las ecuaciones de estimación para MML en las **ecuaciones A.5** y **A.9** como en la **ecuación 15**.