



## UNA SOLUCION A LA ESTIMACION INICIAL EN LOS TESTS ADAPTATIVOS INFORMATIZADOS.

Javier Revuelta  
 Vicente Ponsoda  
 Facultad de Psicología  
 Universidad Autonoma  
 Canto Blanco  
 28049 Madrid - España  
 e-mail:Javier.Revuelta@uam.es /Vicente.Ponsoda@uam.es

### ABSTRACT.

A new method for computing ability estimates when the response vector contains only errors or correct responses is proposed. It is based on the Stepsize method (Dodd, 1990). Instead of taking as the new ability the midpoint between the last computed ability and the corresponding extreme  $b$ -value, the new method proposes as the new estimate the mean of a truncated normal (0, 1) distribution. Last ability estimate and the corresponding extreme  $b$ -value are the points in which the normal distribution truncates.

**Key words:** computerized adaptive tests, maximum-likelihood estimation, initial estimation.

\*\* El trabajo ha recibido financiación parcial de dos proyectos DGICYT (PS94-0040 y PS95-0046).

### 1.- Introducción.

Los tests adaptativos informatizados (TAIs) requieren la estimación del nivel de habilidad tras la aplicación de cada ítem. El ítem siguiente a administrar suele ser el más informativo para el nivel acabado de estimar. Un procedimiento común de estimar la habilidad consiste en el procedimiento de máxima verosimilitud (descrito en Muñiz, 1991, pág. 44 y ss.). El procedimiento no puede aplicarse cuando el vector de respuestas es un vector de unos o de ceros. Es decir, para poder obtener la estimación máximo-verosímil es preciso que al menos uno de los ítems respondidos se haya acertado y los demás fallado (o uno se haya fallado y los demás acertado). El procedimiento, por tanto, no es aplicable tras el primer ítem y tampoco tras el segundo, tercero,.. hasta que el examinando no tiene aciertos y fallos.

Una solución imaginativa a este problema ha sido propuesta por Herrando (1989). Consiste en considerar, a efectos de la estimación máximo-verosímil, que los examinandos han acertado un ítem extremadamente fácil y han fallado otro extremadamente difícil, que realmente no se han aplicado. El procedimiento añade un error y un acierto al vector de respuestas y permite la estimación máximo-verosímil incluso tras el primer ítem. En trabajos previos, con bancos de ítems reales y simulados, hemos empleado el procedimiento anterior. Los parámetros de los dos ítems ficticios han sido ( $a=3$ ,  $b=-4$ ,  $c=0$ , ítem fácil; y  $a=3$ ,  $b=4$ ,  $c=0$ , ítem difícil).

El procedimiento anterior (y/o los particulares parámetros que hemos utilizado) tiene un inconveniente: la  $\theta$  estimada tras el primer ítem alcanza valores muy extremos (por encima de 2, en caso de acierto; o por debajo de -2, en caso de fallo). Debido al procedimiento de selección de ítems, el segundo ítem del test es el más informativo del banco para ese nivel de habilidad tan extremo, y en consecuencia resulta poco informativo para el nivel de habilidad final, que suele tomar un valor intermedio (en torno a 0). En definitiva, el segundo ítem apenas contribuye a la precisión del test.

Otro procedimiento propuesto para resolver el problema de la estimación máximo-verosímil cuando sólo hay ceros o unos en el vector de respuestas ha sido denominado Stepsize, y consiste en no estimar el nivel de  $\theta$  después del primer ítem, sino sumar o restar al nivel previo una constante (Dodd, 1990). Este procedimiento se emplea hasta que en el vector de respuestas del sujeto aparecen ceros y unos.

En el trabajo de Dodd se utilizan tres procedimientos: dos de ellos incorporan las constantes: 0'4, 0'7, y en el tercero se emplea un valor variable a lo largo del test. Este valor es igual a la mitad del rango entre el nivel de habilidad actual y el parámetro de dificultad del ítem más difícil ( $b_M$ ) si el ítem ha sido acertado, o más fácil ( $b_m$ ) si el ítem ha sido fallado. Es decir, si el sujeto acierta el ítem 'i', seleccionado para maximizar la información respecto a  $\theta_i$ , el nuevo valor de  $\theta$  será:

$$\hat{\theta}_{i+1} = \hat{\theta}_i + \frac{b_M - \hat{\theta}_i}{2}$$

Si falla el ítem "i", el nuevo el nuevo valor de  $\theta$  será:

$$\hat{\theta}_{i+1} = \hat{\theta}_i + \frac{b_m - \hat{\theta}_i}{2}$$

Este procedimiento presenta la ventaja, en comparación con los de incremento fijo, de que el nivel estimado de  $\theta$  nunca excederá el rango de niveles de dificultad en el banco de ítems, en el caso de que el sujeto continuase acertando (o fallando) indefinidamente.

Debido a que el objetivo principal de Dodd (1990) no es poner a prueba este método de actualización de  $\theta$ , en los resultados de su trabajo únicamente menciona que el número de casos de no convergencia en la estimación de  $\theta$  fue mayor con la constante 0'7 que con 0'4, y que no hubo ninguno con el de incremento variable. La razón está en que, cuando los incrementos son fijos, el valor estimado puede exceder del rango de  $b_M$  (o  $b_m$ ) si se producen varios aciertos (o fallos) consecutivos en los primeros ítems del test.

## 2.- Un nuevo método de actualización de $\theta$

El método de incremento variable no tiene en cuenta la forma de la distribución de  $\theta$ . En todos los casos el valor actualizado está a medio camino entre el valor anterior y el extremo correspondiente. Esto puede provocar que dicho valor sea demasiado extremo si, por ejemplo, la distribución de habilidad en el grupo de sujetos evaluados sigue la curva normal.

El nuevo método propuesto se basa en el método de incremento variable, pero cambia la forma de actualizar  $\theta$ . El valor actualizado será la media de  $\theta$  en una función de probabilidad que se define a continuación

1) Supongamos que el nivel de habilidad anterior es  $\theta_i$ , el nivel máximo de dificultad en el banco de items se corresponde con  $b_M$  y el sujeto ha acertado el item. La nueva función de probabilidad de  $\theta$  sería:

$$f(\theta) = \begin{cases} 0, & \text{si } \theta < \theta_i \\ p(\theta), & \text{si } \theta_i \leq \theta \leq b_M \\ 0, & \text{si } \theta > b_M \end{cases}$$

2) En caso de que el sujeto falle, la nueva función sería:

$$f(\theta) = \begin{cases} 0, & \text{si } \theta < b_m \\ p(\theta), & \text{si } b_m \leq \theta \leq \theta_i \\ 0, & \text{si } \theta > \theta_i \end{cases}$$

En la definición anterior falta por especificar el valor de  $p(\theta)$ . La idea fundamental es que dicha función de probabilidad sea la distribución normal (0, 1) truncada en  $\theta_i$  y  $b_M$  cuando hay acierto, y en  $b_m$  y  $\theta_i$  cuando hay error (Ríos, 1974, pág.133). Por lo tanto, siendo  $\theta$  una v.a. distribuida normal (0, 1), siendo  $g(\theta)$  la función de densidad  $N(0,1)$ , y  $G(\theta)$  la función de densidad normal acumulada:

$$p(\theta) = g(\theta)/(G(b_M)-G(\theta_i)) \quad \text{cuando hay acierto,}$$

$$\text{y } p(\theta) = g(\theta)/(G(\theta_i)-G(b_m)) \quad \text{cuando hay error.}$$

Finalmente, el nuevo valor de  $\theta$  estimado para el sujeto es el valor esperado de  $f(\theta)$ :

$$\hat{\theta}_{i+1} = \int_{-\infty}^{\infty} \theta f(\theta) d\theta$$

El cálculo de las funciones  $G(\theta)$  supone calcular una integral de forma numérica para cada estimación. Este procedimiento puede simplificarse calculando las funciones  $G(\theta)$  y  $g(\theta)$  mediante la aproximación logística a la normal (Santesteban, 1990, pág. 269), de la siguiente forma:

$$G(\theta) \approx \frac{e^{(1.7\theta)}}{1 + e^{(1.7\theta)}}$$

$$g(\theta) \approx \frac{1.7 e^{(1.7\theta)}}{(1 + e^{(1.7\theta)})^2}$$

### 3.- Método

A partir de un banco de 221 items de vocabulario inglés (descrito en Olea, Ponsoda, Revuelta y Belchi, 1996) se simuló la aplicación de un test adaptativo en dos condiciones diferentes:

- 1) En la primera condición se estima el nivel de habilidad desde el primer item utilizando el procedimiento de máxima verosimilitud y la estrategia propuesta por Herrando (1989).
- 2) En la segunda condición se estima por el método descrito hasta que en el vector de respuestas aparezca al menos un acierto y un error.

En cada condición se simularon las respuestas de los mismos 500 sujetos, cuya habilidad se distribuía  $N(0, 1)$ . Se analizó la precisión obtenida a lo largo del test y al final del mismo.

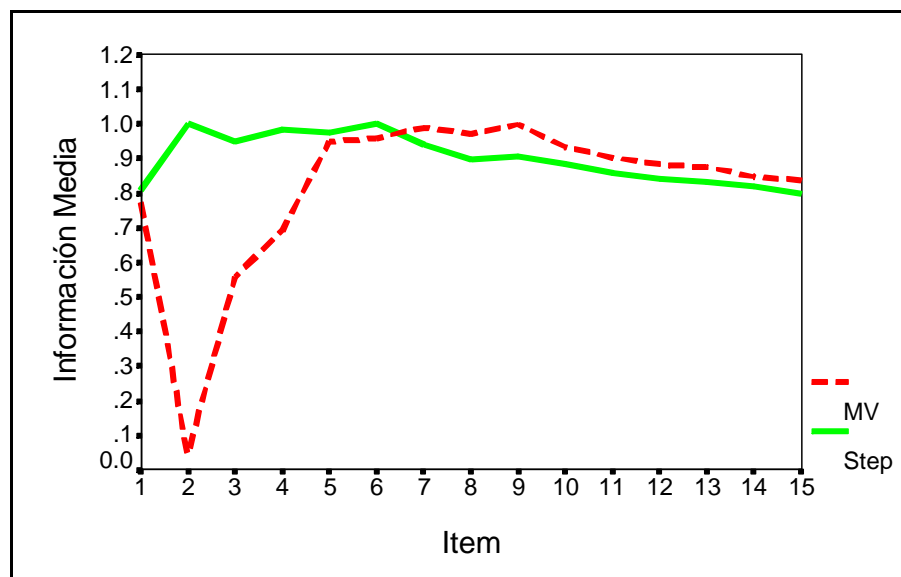
### 4.- Resultados

La precisión final del test medida por la raíz cuadrada de la media de las desviaciones entre cada habilidad real y estimada (RMSD) fue igual con ambos métodos. En cambio, el error típico de estimación de la habilidad fue ligeramente inferior al utilizar el procedimiento propuesto anteriormente. La Tabla 1 muestra las medias de las habilidades reales (primera fila), de las estimadas (segunda), de los errores típicos de estimación (tercera) y de los índices RMSD, en las dos condiciones MV (estimación máximo-verosímil desde el comienzo) y “Stepsize” (método propuesto).

Variable	MV	Stepsize
$\theta$	-0.01	-0.01
$\hat{\theta}$	0.04	0.01
$S_{\theta}$	0.30	0.28
RMSD	0.33	0.33
Items	15.00	15.00

**Tabla 1**

La Gráfica 1 muestra la información media del primer ítem, segundo,..último para la habilidad final (es decir, la estimada tras el último ítem). Respecto a la precisión a lo largo del test, puede apreciarse que al estimar la habilidad por el método MV el segundo ítem apenas aporta información respecto a la  $\theta$  final, y los ítems tercero y cuarto también son menos informativos con MV que con Stepsize. Este efecto se debe a que el segundo ítem se selecciona a partir de un nivel de habilidad provisional muy extremo, y por lo tanto alejado del valor final, hasta el quinto ítem del test los valores estimados de habilidad no se aproximan a los reales. Por el contrario, con el procedimiento propuesto (“Step”, en la gráfica), en el segundo ítem no se produce una disminución de la información.



**Gráfica 1**

A partir del séptimo ítem, los valores de la información son menores en el método Step que en el MV, lo que puede deberse a que con el método Step los ítems más informativos se emplean en las primeras selecciones, mientras que con MV es menos probable que salgan estos ítems en las primeras selecciones y en consecuencia están disponibles para las posteriores. Sin embargo, como hemos visto, la información media suministrada por los quince ítems del test es mayor en el método Step, de ahí que el error típico de estimación sea inferior en este caso.

## 5.- Conclusiones

El método propuesto parece ser una alternativa viable al problema de la estimación al comienzo del test. Resulta necesario realizar un estudio más detallado en el que el método propuesto se compare con el de Herrando (1989), el bayesiano y los tres propuestos por Dodd (1990). Convendría hacer las comparaciones en bancos reales y simulados de distintas características, para poder determinar la posible generalización de las conclusiones.

El método propuesto consiste en calcular el valor esperado de la distribución normal truncada en ciertos valores. En casos de distribuciones de probabilidad tan asimétricas como las utilizadas aquí, la media puede ser menos representativa que la mediana como indicador de tendencia central, por lo que puede proponerse un nuevo procedimiento.

La idea básica es tomar como nuevo valor del nivel de habilidad la mediana (en vez de la media) de la distribución doblemente truncada definida antes. Resulta fácil comprobar que, siendo  $G$  la función de distribución normal  $(0, 1)$  y  $G^{-1}$  su inversa, cuando hay acierto,

$$\hat{\theta}_{i+1} = G^{-1} \left( \frac{G(b_M) + G(\theta_i)}{2} \right)$$

y cuando se produce error,

$$\hat{\theta}_{i+1} = G^{-1} \left( \frac{G(b_m) + G(\theta_i)}{2} \right)$$

Al igual que en el caso anterior, puede utilizarse la aproximación logística a la normal para simplificar los cálculos de  $G(\theta)$ . En la distribución logística,

$$G^{-1}(p) = \frac{1}{1.7} \ln \left( \frac{p}{1-p} \right)$$

Por lo tanto, en el caso de acierto, la nueva  $\theta$  sería:

$$\hat{\theta}_{i+1} = (1/1.7) \ln \left( \frac{G(b_M) + G(\theta_i)}{2 - G(b_M) - G(\theta_i)} \right)$$

y cuando se produce error,

$$\hat{\theta}_{i+1} = (1/1.7) \ln \left( \frac{G(b_m) + G(\theta_i)}{2 - G(b_m) - G(\theta_i)} \right)$$

## 6.- Referencias

- Dodd, B. G. 1990. The Effect of Item Selection Procedure and Stepsize on Computerized Adaptive Attitude Measurement Using the Rating Scale Model. *Applied Psychological Measurement* 14, 4, 355-366.
- Herrando, S. 1989. Tests adaptativos computerizados: una sencilla solución al problema de la estimación con puntuaciones perfecta y cero. *II Conferencia Española de Biometría*. Biometric Society. Segovia.
- Muñiz, J. 1990. *Teoría de Respuesta a los Items*. Madrid: Pirámide.
- Olea, J., Ponsoda, V., Revuelta J. y Belchi, J. 1996. Propiedades psicométricas de un test adaptativo informatizado de vocabulario inglés. *Estudios de Psicología*. No 55. Pp 61-73.
- Rios, S. 1974. *Métodos estadísticos*. Sexta edición. Madrid: Ediciones del Castillo.
- Santisteban, C. 1990. *Psicometría. Teoría y práctica en la construcción de tests*. Ed: Norma. Madrid.