

Modelado de árboles de procesamiento multinomiales con agregación de tiempos de reacción: elección de procedimientos de modelado según la distribución de los datos

Anahí Gutkin de Freitas

Máster en Metodología de las Ciencias de la Salud y del Comportamiento



MÁSTERES
DE LA UAM
2019 – 2020

Facultad de Psicología

**MÁSTER EN METODOLOGÍA DE LAS CIENCIAS DE LA SALUD Y DEL
COMPORTAMIENTO**



TRABAJO FIN DE MÁSTER

Junio-2020

Modelado de árboles de procesamiento multinomiales con agregación de tiempos de reacción: elección de procedimientos de modelado según la distribución de los datos.

Multinomial process tree with reaction time aggregation: choosing modeling procedure according to data distribution.

Anahí Gutkin de Freitas

Tutor: Manuel Suero Suñe

Facultad de Psicología

Universidad Autónoma de Madrid

Resumen

Los modelos de árbol de procesamiento multinomiales con tiempos de reacción (MPT-TR) son recientes y carecemos de un protocolo de modelado específico para los diferentes escenarios de distribución de los datos. Los procedimientos existentes difieren en la forma en que agregan los TR: existen los que modelan las medias de TR, sus histogramas y los que asumen distribuciones. El objetivo de este trabajo es investigar, desde la simulación de un modelo de doble umbral (2HTM), qué efectos tienen la existencia de probabilidades de procesos y de TR extremos en el sesgo y ajuste de las estimaciones, eligiendo qué procedimiento y distribución (Ex-Gauss, *Shift-Wald*, *Shift-Gamma*) es apropiada para cada caso. Al evaluar la media de los TR, *Shift-Gamma* resultó ser la mejor asunción distribucional cuando no existan probabilidades extremas ni TR típicos, mientras que al evaluar el ajuste se prefiere la *Shift-Wald* y la Ex-Gauss. Con TR extremos se encontró una ventaja en asumir Ex-Gauss, aunque dicha ventaja se pierde en ausencia de probabilidades extremas de procesos. Se concluye así que el mejor procedimiento resultó ser el que asume distribuciones, pero la distribución elegida debe depender del objetivo experimental, de la distribución de los datos categóricos y de los TR.

Palabras clave: procedimientos MPT con TR, modelo de doble umbral (2HTM); TR extremos, procesos con probabilidades extremas.

Abstract

Multinomial process tree models with reaction time data (MPT-RT) are recent and lack specific modeling protocol for different data distribution scenarios. Actual procedures differ in the way that RT data is aggregated: some aggregate mean RT, others their histograms and still others assume RT distributions. The aim of this study is to investigate the effects of extreme process probabilities and RT, using a Two High Threshold Model (2HTM) simulation, and decide which procedure and distributional assumption should be selected in each situation. Regarding mean RT evaluation, Shift-Gamma was the best distributional assumption for typical RT, whereas Ex-Gauss and Shift-Wald were preferred in terms of good fit. An Ex-Gaussian assumption proved advantageous for extreme RT, except in cases with no extreme process probabilities. In conclusion, the best procedure was the one that made distributional assumptions, nonetheless the chosen distribution should take the experimental purpose and the categorical and RT data distribution into account.

Keywords: MPT with RT procedures, 2HTM, extreme RT, process with extreme probabilities.

Índice

1. Introducción.....	4
2. Modelos de árbol de procesamiento multinomiales (MPT)	6
2.1 MPT con variables categoricas.....	6
2.2 MPT con variables categoricas y continuas	9
3. Antecedentes en el estudio del problema	13
3.1. Procedimiento de modelado 1°M (Medias)	14
3.2. Procedimiento de modelado 2°M (<i>Bins</i>).....	16
3.3 Procedimiento de modelado 3°M (Distribuciones)	18
4. Objetivos.....	19
4.1 Generales	19
4.2 Específicos.....	19
5. Metodología.....	21
5.1. Condiciones de investigación	21
5.2. Simulación de condiciones	22
5.2.1 Niveles de los parámetros de las variables categóricas	22
5.2.2 Niveles de parámetros de las variables continuas	23
5.3. Estimación de parámetros.....	26
5.3.1 Identificabilidad del modelo	26
5.3.2 Procedimiento de Estimación y de análisis	27
6. Resultados.....	28
6.1 Sesgo en las probabilidades de los procesos (D y G).....	29
6.2 Sesgo de las medias de los TR (μ_{D_o} , μ_{D_n} y de μ_G).....	30
6.3 Porcentaje de selección de los modelos 3°M.....	35
7. Discusión	37
7.1 Sesgo en las probabilidades de los procesos (D y G).....	37
7.2 Sesgo de las medias de los TR (μ_{D_o} , μ_{D_n} y de μ_G).....	38
7.3 Porcentaje de selección de los modelos 3°M.....	41
8. Conclusión (limitaciones y futuras lineas de investigación)	42
Referencias	46
Apéndice A.....	49
Apéndice B	50

1 Introducción

Con el fin de entender y explicar el comportamiento humano podemos beneficiarnos de la precisión que nos ofrecen las matemáticas. La modelización cognitiva permite desarrollar, o evaluar, teorías sobre la arquitectura de la mente mediante su formulación en forma de ecuaciones y algoritmos computacionales. De esta forma podemos establecer hipótesis sobre cómo se relacionan (*e.g.*: La variable X_1 es independiente de la variable X_2 ; la relación entre la variable X_1 e Y es exponencial) los procesos mentales y las respuestas conductuales y el grado de dicha relación (*e.g.*: En la condición A es 3 veces mayor que en condición B). En definitiva, se intenta traducir matemáticamente dicho modelo de forma que refleje la estructura y el funcionamiento de los mecanismos psicológicos involucrados en la conducta (Cavagnaro, Myung y Pitt, 2010; Marewski y Olsson, 2009; Myung, Pitt y Kim, 2005).

En el proceso de modelado se ha de tener en cuenta el tipo de variables implicadas y el tipo de respuestas u observables (variable dependiente) a modelar. Estas últimas pueden ser tanto nominales (cualitativas o categóricas) como cuantitativas (discretas o continuas).

Un ejemplo de variable categórica es la que encontramos en tareas tales como: de reconocimiento de información verbal almacenada largo-plazo (Atkinson y Shiffrin, 1968; Juola, Fischler, Wood y Atkinson, 1971), de juicios, de tomas de decisiones (Lamberts y Goldstone, 2004; Link, 1982), perceptivas, de categorización (Ashby, Prinzmetal, Ivry, y Maddox, 1996; Ashby, 2014), etc. Nos centraremos en las primeras por estar directamente relacionado con el presente trabajo.

En el paradigma de reconocimiento de información verbal existen dos fases: la fase de estudio, donde se aprenden una serie de palabras; y la fase de evaluación, donde se presentan en orden aleatorio las palabras previamente estudiadas (antiguas o targets) y no estudiadas (nuevas o distractores). En esta fase el participante debe contestar “antiguo” o “nuevo” para indicar si la palabra que le es presentada pertenece a la lista de palabras previamente estudiadas o no, respectivamente. Sabiendo el tipo de estímulo presentado en cada ensayo (antiguo/ nuevo) y la respuesta del sujeto (“antiguo”/ “nuevo”) podemos entonces distinguir cuatro categorías de respuestas. Considerándose señal al EE antiguo y ruido al EE nuevo, las 4 categorías de respuesta serán: acierto (*Hit*), si al presentarse un EE previamente presentado en la fase de estudio, el sujeto contesta “antiguo”; omisión (*Miss*), si ante un EE antiguo contesta “nuevo”; falsa alarma (*FA*), si se presenta un EE nuevo y el sujeto contesta “antiguo”; y rechazo correcto (*RC*) si contesta “nuevo” ante un EE nuevo.

Frecuentemente, para variables categóricas se han utilizado los modelos de clasificación, de juicios de categorización multidimensional y la teoría de detección de señales (TDS) (Atkinson y Juola, 1973, 1974; Ashby, 2014; Juola et al., 1971; Luce, 2012; Thurstone, 1927). Entre estos modelos se encuentran también los modelos de árbol de procesamiento multinomiales (MPT siglas en inglés), que son una herramienta de modelado de particular interés para modelizar frecuencias de categorías de respuestas observables en función de estados cognitivos o procesos discretos que no son directamente observables (Batchelder y Riefer 1990; Bröder y Schütz, 2009). Los MPT permiten abordar el estudio de la cognición humana y estudiar la viabilidad de un modelo de la arquitectura de la mente asumiendo que cada respuesta conductual observable es consecuencia de una secuencia de procesos latentes. A pesar de la utilidad que han mostrado (Erdfelder, Auer, Hilbig, Aßfalg, Moshagen, y Nadarevic, 2009), debemos admitir que, en principio, estos únicamente modelan observables categóricos. Implicando, pues, que en experimentos donde intervienen tanto variables nominales como cuantitativas, o bien no se tiene en cuenta alguna de ellas, o bien se modelan por separado.

Gracias al desarrollo de tareas computarizadas, hay una creciente facilidad de medición de variables continuas tales como: el seguimiento ocular, seguimiento del cursor del ordenador, medidas neurofisiológicas (Humphreys y Bruce, 1989; Heck, Erdfelder y Kieslich, 2018), magnitudes propias de la psicofísica sensorial, entre otras. Quizás una de las medidas continuas más estudiadas y de menor coste de medición son los tiempos de reacción (TR). Estos se suelen medir en tareas de búsqueda visual, en tareas atencionales, de memoria, entre muchas otras (Humphreys y Bruce, 1989; Heck et al., 2018; Luce, 1986).

Siguiendo la tarea ejemplificada anteriormente, pero introduciendo TR, a nivel de diseño experimental bastaría con medir en la fase de evaluación la latencia entre la presentación del estímulo y la respuesta observada de cada ensayo. De esta forma si un sujeto tiene 25 *Hits* tendrían también un vector con 25 elementos de TR asociados a la categoría *Hit*.

Podemos constatar en textos clásicos, como Townsend y Ashby (1983) o Luce (1986), diversos procedimientos de modelado de TR. Sin embargo, el modelado con agregación de datos categóricos y continuos sigue siendo un desafío. Desafío al que, recientemente, el modelado por MPT se empieza a enfrentar (Heck et al., 2018; Heck y Erdfelder, 2016; Hu, 2001; Klauer y Kellen; 2018; Schweickert y Zheng, 2019) y del que tratará el presente trabajo. Estos modelos MPT permiten agregar datos de TR (MPT-TR) o cualquier otro

dato continuo (MPT-DC) intrínsecamente asociado a los procesos que asume el modelo teórico.

En este trabajo estudiaremos uno de los modelos MPT más sencillos, el modelo de doble umbral (*Two High Threshold Model*, 2HTM siglas en inglés). En éste, se incluirán, además de las frecuencias de respuesta asociadas a cada categoría, los TR asociados. El modelo MPT-TR se aplicará en el paradigma experimental de reconocimiento de información verbal anteriormente ya descrito.

En el próximo apartado describiremos en detalle la aplicación del 2HTM para datos categóricos (ver apartado 2.1) así como su aplicación para datos continuos agregados (ver apartado 2.2). A continuación, haremos un breve repaso sobre algunas propuestas de modelado de MPT para datos categóricos y continuos (ver apartado 3) con el fin de establecer hipótesis (ver apartado 4) sobre el comportamiento de estos procedimientos en función de diferentes escenarios de distribución de los datos. Operativizaremos dichos escenarios (ver apartado 5) en condiciones de simulación concretas (ver apartado 5.1 y 5.2) y detallaremos algunos aspectos sobre el ajuste de los parámetros (ver apartado 5.3). Finalmente, en los últimos dos apartados expondremos los resultados obtenidos (ver apartado 6) y evaluaremos los procedimientos de modelado con el fin de contestar a los objetivos planteados (ver apartado 7 y 8).

2 Modelos de árbol de procesamiento multinomiales (MPT)

2.1 MPT con variables categóricas

Antes de adentrarnos a explicar el 2HTM daremos unas breves nociones generales sobre los MPT. Los MPT son aquellos modelos en los que las respuestas categóricas observadas se caracterizan postulando un conjunto finito de procesos latentes. En la Fig. 1 representamos un ejemplo de MPT. Denominaremos a cada nodo como un proceso o su complementario. Tal y como observamos, los procesos de este ejemplo tienen asociadas las probabilidades θ_1 y θ_2 , y sus complementarias, de que dichos procesos participen en la emisión de la j respuestas observadas (donde $j: 1, \dots, J$ y $J = \{C_j\}$).

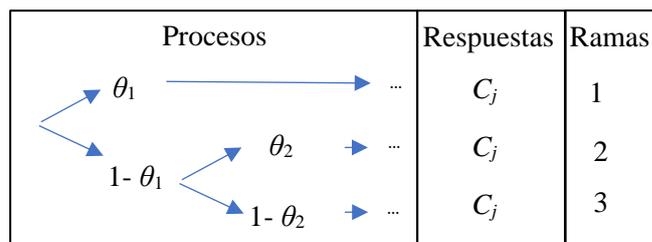


Figura 1. Representación de un modelo MPT abstracto.

A cada camino que discurre pasando por los diferentes nodos secuenciales hasta conducir a una respuesta particular lo denominamos rama. Aceptando que los procesos son independientes podemos obtener la probabilidad de cada rama como el producto de la probabilidad de los nodos contenidos en ella. A la probabilidad de cada rama z la denominaremos $\pi_z(\boldsymbol{\theta})$, y, tal y como se observa, dicha probabilidad depende del vector de parámetros $\boldsymbol{\theta}$ que incluye a las probabilidades de los procesos pertenecientes a dicha rama. Por ejemplo, la probabilidad asociada a la primera y segunda rama serán respectivamente:

$$\pi_1(\boldsymbol{\theta}) = \theta_1 \quad (1)$$

$$\pi_2(\boldsymbol{\theta}) = (1 - \theta_1)\theta_2 \quad (2)$$

Como dos ramas distintas pueden conducir a una misma categoría de respuesta observada, C_j , la probabilidad de contestar C_j , $\pi_{jz}(\boldsymbol{\theta})$, se obtiene como la combinación lineal de las probabilidades de las z ramas que conducen a una misma categoría j (Riefer y Batchelder, 1988; Schweickert y Zheng, 2019). Por consiguiente, si la categoría C_l en un modelo hipotético se alcanza únicamente desde la primera rama, la probabilidad observada para C_l sería:

$$\begin{aligned} \pi_{1z}(\boldsymbol{\theta}) &= \pi_1(\boldsymbol{\theta}) \\ &= \theta_1 \end{aligned} \quad (3)$$

Mientras que si C_l se alcanza tanto con la primera como con la segunda rama la probabilidad observada para C_l sería:

$$\begin{aligned} \pi_{1z}(\boldsymbol{\theta}) &= \pi_1(\boldsymbol{\theta}) + \pi_2(\boldsymbol{\theta}) \\ &= \theta_1 + (1 - \theta_1)\theta_2 \end{aligned} \quad (4)$$

Realizada una breve introducción general a los MPT pasaremos a explicar la lógica subyacente a su modelado utilizando directamente el modelo que nos incumbe.

Como se adelantó anteriormente, trabajaremos con el modelo de doble umbral (2HTM) aplicado a la tarea reconocimiento de información verbal explicada en el anterior apartado. El 2HTM (ver Fig. 2) es un modelo binario (donde cada nodo conduce a dos ramas). Es de extrema utilidad trabajar con el 2HTM puesto que además de ser muy sencillos tienen el beneficio de que cualquier MPT no binario pueden ser transformados a uno binario (Hu y Batchelder, 1994). El modelo en cuestión contiene dos subárboles, uno para la presentación de estímulos (EE) previamente aprendidos (antiguos) y otros para EE no aprendidos (nuevos).

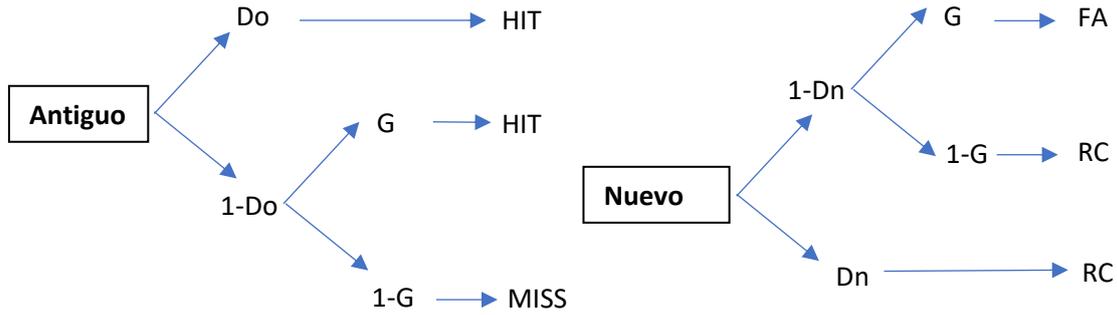


Figura 2. Representación de 2HTM para variables categóricas.

Como se puede observar en la Fig. 2, las probabilidades incluidas son: Do , probabilidad de detectar estímulos (EE) antiguos; Dn , detectar EE nuevos y G^l (*guessing old*), el sesgo hacia decir “antiguo” cuando no se ha reconocido si EE es antiguo o nuevo.

Para describir formalmente el modelo, llamemos z a cada rama (donde z : 1, ..., Z ; y $Z = 6$), i a cada participante (donde i : 1, ..., I) y j a las categorías de respuesta (donde j : 1, ..., J y $J = \{Hit, Miss, FA, RC\}$). La frecuencia de cada rama, x_{ijz} , es una variable aleatoria discreta (*v.a.d*) y se distribuyen tal que,

$$x_{ijz} \sim B(N, \pi_{ijz}(\theta_i)) \quad (5)$$

Donde N es el número de ensayos por tipo de EE (nótese que se estableció que N es común a todos los participantes) y $\pi_{ijz}(\theta_i)$, la probabilidad de cada una de las ramas. Aceptando que existen diferencias individuales entre los participantes asumiremos que habrá variabilidad en los parámetros del modelo que representan las probabilidades de los procesos. En consecuencia, las probabilidades θ_i serán variables aleatorias continuas (*v.a.c*) y se definirán con el subíndice i referido al participante (ver apartado 5.2.1). Obsérvese que $\pi_{ijz}(\theta_i)$ tiene el subíndice i por lo que también será una *v.a.c* (ver apartado 5.2.1). Para el 2HTM, $\pi_{ijz}(\theta_i)$ sería:

$$\pi_{i11}(\theta_i) = Do_i \quad (6)$$

$$\pi_{i12}(\theta_i) = (1 - Do_i)G_i \quad (7)$$

$$\pi_{i23}(\theta_i) = (1 - Do_i)(1 - G_i) \quad (8)$$

$$\pi_{i34}(\theta_i) = (1 - Dn_i)G_i \quad (9)$$

$$\pi_{i45}(\theta_i) = (1 - Dn_i)(1 - G_i) \quad (10)$$

$$\pi_{i46}(\theta_i) = Dn_i \quad (11)$$

¹ Autores como Hu (2001) distinguen la probabilidad Go , sesgo hacia decir “antiguo” cuando no se ha reconocido si EE es antiguo, de la probabilidad Gn , sesgo hacia decir “nuevo” cuando no se ha reconocido si EE es nuevo.

Finalmente, las frecuencias observadas, x_{ij} , se obtienen sumando las x_{ijz} latentes que conducen a la misma respuesta j . Es decir:

$$x_{ij} = \sum_{z=1}^Z x_{ijz} \quad (12)$$

A modo de ilustración, tenemos que, por ejemplo, la frecuencia observada de la categoría de respuesta *Hit* es:

$$x_{i1} = x_{i11} + x_{i12} \quad (13)$$

Si x_{ijz} es una variable aleatoria y asumimos independencia, entonces x_{ij} también será una variable aleatoria, y la probabilidad asociada a cada categoría observable se puede obtener desde la probabilidad de rama como,

$$\pi_{ij}(\boldsymbol{\theta}_i) = \sum_{z=1}^Z \pi_{ijz}(\boldsymbol{\theta}_i) \quad (14)$$

Y la frecuencia de las respuestas de categoría observada será una *v.a.d* que sigue:

$$x_{ij} \sim B(N, \pi_{ij}(\boldsymbol{\theta}_i)) \quad (15)$$

2.2 MPT con variables categóricas y continuas

Los procesos incluidos no son instantáneos, sino que discurren en el tiempo y, además, desde un modelo MPT se asume que lo hacen serialmente (*i.e* solamente cuando un proceso de acaba se inicia el siguiente). Teniendo en cuenta que el tiempo de procesamiento es variable, asumiremos que dicha variabilidad es estocástica.

Para estudiar el tiempo de procesamiento, tendremos pues que agregar los datos continuos de tiempo de procesamiento a nuestro modelo MPT. Para ello bastaría asociar a cada nodo del árbol una distribución de tiempo de procesamiento. De esta forma, para cada nodo tendremos que el tiempo de procesamiento, y_θ , es una *v.a.c* que queda caracterizada por una función de densidad de probabilidad (*f.d.p.*), representada por g , y por su respectivo vector de parámetros $\boldsymbol{\eta}$. Los tiempos de respuesta (TR), tanto observados (por categoría) como latentes (por rama), serán pues una función de los tiempos de procesamiento de cada proceso, y_θ .

Dependiendo de qué distribución sea g tendremos diferentes tipos de parámetros a definir. Algunas de las funciones de distribución más empleadas y de mejor ajuste en el modelado de TR son: la *Ex-Gaussian*, *Shift-Wald* y *Shift-Gamma* (Luce, 1986; Cousineau, Brown y Heathcote, 2004; Matzke y Wagenmakers, 2009; Palmer, Horowitz, Torralba y Wolfe, 2011).

Por otro lado, debemos también destacar que al existir una variabilidad entre los sujetos asumimos que existirá una variabilidad en los parámetros η (ver apartado 5.2.2.).

Con el fin de facilitar la explicación de cómo se enfrentan los modelos MPT al modelado de la variable continua tiempo de respuesta (MPT-TR) nos referiremos de aquí en adelante específicamente al 2HTM (ver Fig.3.).

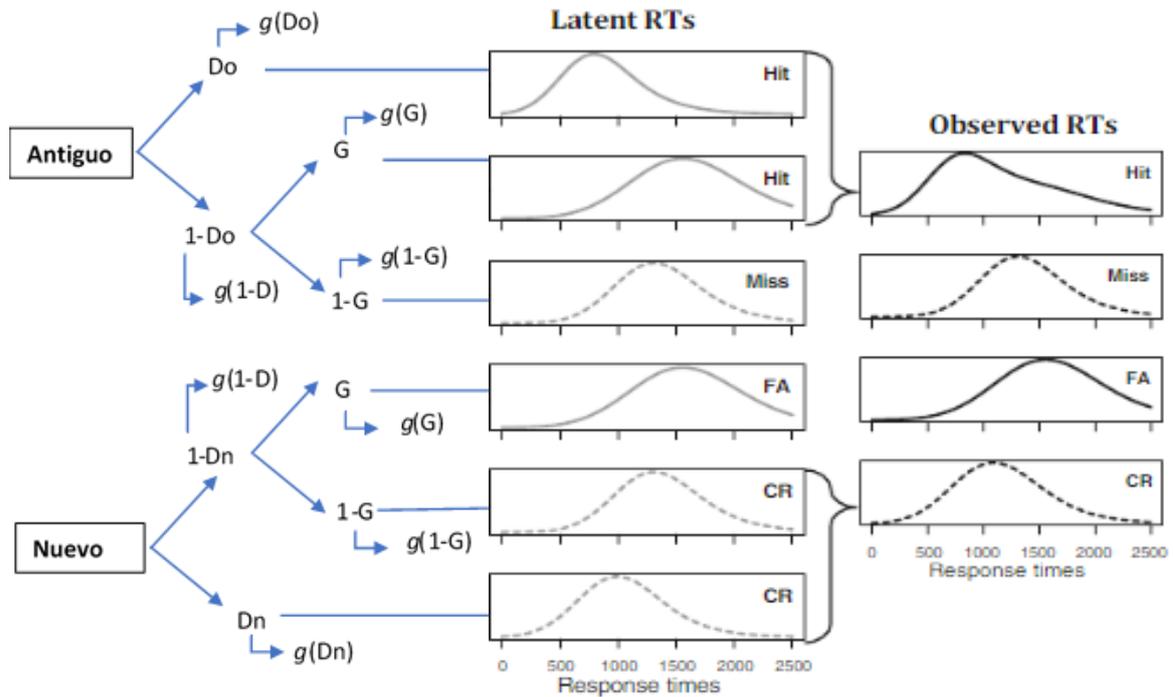


Figura 3. Representación del 2HTM para variables categóricas y continuas. Heck y Erdfelder (2016). [Figura]. Recuperado de: <https://link.springer.com/article/10.3758/s13423-016-1025-6>.

Como se observa en la Fig.3. existe una distribución de los TR asociados a cada rama y una distribución de tiempos de procesamiento de asociada a los nodos. Se asume que los TR de las ramas son la suma de los tiempos de procesamiento de los estados contenidas en ellas. Así pues, asumiendo independencia entre estados, la distribución de los TR de la rama z será la convolución de las *f.d.p* asociada a los tiempos de procesamiento de los nodos contenidos en z . Para cada individuo i hay un vector y_{ijz} con x_{ijz} elementos correspondientes al conjunto de TR de cada rama z , e y_{ijz} será la *v.a.c* resultante de la suma de los tiempos de procesamiento correspondientes a los estados, y_{θ} , contenidos en una misma rama z . En nuestro modelo, para cada rama tendremos que y_{ijz} es:

$$y_{i11} = y_{(Do)_i} \quad (16)$$

$$y_{i12} = y_{(1-D)_i} + y_{(G)_i} \quad (17)$$

$$y_{i23} = y_{(1-D)_i} + y_{(1-G)_i} \quad (18)$$

$$y_{i34} = y_{(1-D)_i} + y_{(G)_i} \quad (19)$$

$$y_{i45} = y_{(1-D)_i} + y_{(1-G)_i} \quad (20)$$

$$y_{i46} = y_{(Dn)_i} \quad (21)$$

Sabemos también que los TR observados por categoría de respuesta, y_{ij} , provendrán de una o más ramas latentes, por lo que y_{ij} es la unión del conjunto de TR de cada rama z que conducen a la misma categoría j , es decir:

$$y_{ij} = \bigcup_{z=1}^Z y_{ijz} \quad (22)$$

Siguiendo la lógica anterior, como y_{ijz} es una *v.a.c* también quedará caracterizada por la función g , y por su respectivo vector de parámetros η .

Un asunto que subrayar acerca de la parametrización de las distribuciones g es que resulta útil reagrupar las distribuciones de los TR de ramas latentes que contienen los mismos parámetros (Frühwirth-Schnatter, 2006; como citado en Heck et al., 2018). Como nuestro modelo teórico (siguiendo a Heck y Erdfelder, 2016) asume que los tiempos de procesamiento asociados a la no detección de EE antiguos y nuevos siguen la misma distribución, tendremos que,

$$\begin{aligned} g_{(1-Do)_i} &= g_{(1-Dn)_i} \\ &= g_{(1-D)_i} \end{aligned} \quad (23)$$

En consecuencia, solamente 4 de las 6 distribuciones g de TR asociados a las ramas tendrán parámetros diferenciados entre sí. A estos componentes distribucionales diferentes los denominaremos: g_{Do} , g_{Go} , g_{Gn} y g_{Dn} . En concreto, los y_{ij1} seguirán una distribución que denominamos g_{Do} , y_{ij2} e y_{ij4} seguirán la denominada g_{Go} , y_{ij3} e y_{ij5} seguirán la función g_{Gn} e y_{ij6} seguirá la función g_{Dn} . Donde los subíndices Go y Gn permitirán diferenciar las distribuciones de los TR asociados a las ramas que incluyen los procesos *guessing old* (cuya probabilidad es G) de los que incluyen procesos de *guessing new* (cuya probabilidad es $1 - G$).

Realizada esta aproximación a los fundamentos básicos de este proceso de modelado, destacamos una de las particularidades de elevado interés que tienen los MPT que modelizan conjuntamente datos categóricos y continuos. El interés reside en que los *finite mixture models* están naturalmente relacionados los modelos MPT con medidas cuantitativas y categóricas, por lo que potencialmente podríamos aplicar a los MPT

herramientas ya implementadas en los *finite mixture models*. La naturaleza de esta relación se debe a que si los TR observados y_{ji} son una combinación lineal de los TR latentes y_{jiz} , entonces la *f.d.p.* de TR observada $g(y_{ji})$ es una mezcla de Z distribuciones de TR latentes, $g(y_{ijz})$, con sus probabilidades, π_{ijz} . Formalmente esto se traduce en que,

$$g(y_{ij}|\boldsymbol{\theta}_i, \boldsymbol{\eta}_i) = \sum_{z:Z \text{ conduce a } J} \pi_{ijz}(\boldsymbol{\theta}_i) \cdot g(y_{ijz}|\boldsymbol{\eta}_i) \quad (24)$$

Tal y como se define un *finite mixture model* (ver, por ejemplo, McLachlan y Peel , 2004) las probabilidades $\pi_{iz}(\boldsymbol{\theta})$ deben cumplir tanto que,

$$0 < \pi_{iz}(\boldsymbol{\theta}_i) < 1 \quad (z = 1, \dots, Z) \quad (25)$$

como que,

$$\sum_{z=1}^Z \pi_{iz}(\boldsymbol{\theta}_i) = 1 \quad (26)$$

Debido a lo anterior, para que nuestro modelo MPT-TR sea un *finite mixture model*, las probabilidades de las ramas tendrán que ser ponderadas respecto a las probabilidades de cada una de las categorías de respuestas observadas. Por ejemplo, como la categoría de respuesta $j = 1$ (*i.e.* $C_1 = Hit$) se puede alcanzar mediante la rama $z = 1$ y $z = 2$, tendremos para la primera rama que $\pi_{i1z}(\boldsymbol{\theta}_i)$ será:

$$\pi_{i11}(\boldsymbol{\theta}_i) = \frac{Do_i}{Do_i + (1-Do_i) \cdot G_i} \quad (27)$$

y que para la segunda que $\pi_{i12}(\boldsymbol{\theta}_i)$ será:

$$\pi_{i12}(\boldsymbol{\theta}_i) = \frac{(1-Do_i) \cdot G_i}{Do_i + (1-Do_i) \cdot G_i} \quad (28)$$

A modo de síntesis de la nomenclatura empleada hasta el momento, véase la Tabla 1. Recordemos que todas las variables que incluyan el subíndice i son variables aleatorias.

Tabla 1

Variables implicadas en la generación de datos categóricos y continuos

	<i>Frecuencia</i>	<i>Probabilidad</i>	<i>Tiempo</i>
<i>Categorías de Respuesta</i> <i>Ramas</i>	x_{ij}	π_{ij}	y_{ij}
	x_{ijz}	π_{ijz}	y_{ijz}
<i>Nodos</i> <i>(los procesos y sus complementarios)</i>		Do_i	$y_{(Do)i}$
		$1 - Do_i$	$y_{(1-D)i}$
		Dn_i	$y_{(Dn)i}$
		$1 - Dn_i$	$y_{(1-D)i}$
		G_i	$y_{(G)i}$
		$1 - G_i$	$y_{(1-G)i}$

Recuérdese que los y_{ijz} siguen cuatro componentes distribucionales diferentes, que hemos denominado como $g_{Do}(y_{ij1})$, $g_{Gn}(y_{ij2} \text{ e } y_{ij4})$, $g_{Go}(y_{ij3} \text{ e } y_{ij5})$ y $g_{Dn}(y_{ij6})$.

3 Antecedentes en el estudio del problema

El primer modelo MPT-DC, propuesto en 2001 por Hu, es una generalización del modelo de Link (1982). Este último autor corregía los TR medios estimados para *Hits*, en un paradigma MPT de decisión léxica, mediante la exclusión de la proporción de estos TR que se debía a un estado de *guessing*. Desarrollando esta idea, Hu (2001) descompone los TR medios de las categorías de respuestas observadas en medias de tiempos de respuesta latentes, proponiendo un *finite mixture model* para describir los componentes aditivos latentes que dan lugar las medias de TR observadas.

Otros autores como Heck y Ederfelder (2016), también tratan de modelizar un MPT-TR, pero en vez de analizar las medias de TR observadas, discretizan la variable continua en intervalos, ordenados de más a menos rápidos, y estiman la probabilidad de dichos intervalos, denominados *bins*. Obsérvese que de esta forma se transforma la variable continua en una con categorías ordenadas. Además, en contraposición al primer procedimiento de modelado, este tiene la particularidad de poder usarse para representar la forma aproximada (histograma) de la distribución de los TR sin necesidad de asumir ninguna *f.d.p.* (Heck y Ederfelder; 2016).

Con los procedimientos hasta ahora mencionados obtenemos una información limitada acerca de los tiempos de reacción; no obstante, tienen el beneficio de apenas realizar asunciones acerca de las distribuciones de los tiempos. Hay procedimientos de modelado (bayesianos, multinivel, etc.) que sí parten de realizar asunciones paramétricas sobre la forma de la distribución los TR (Heck et al., 2018; Klauer y Kellen, 2018). Estos procedimientos aprovechan y aportan más información de los datos, pero, en contrapartida, podrían no ser métodos adecuados para situaciones en las que hay importantes discrepancias entre lo que se asume y cómo se distribuyen los datos (Klauer y Kellen, 2018).

En el presente trabajo nos referiremos al MPT Extendido con medias de TR de Hu (2001) como el primer procedimiento (1ºM), al MPT-*extended free-histogram, discrete state* de Heck y Ederfelder (2016) como el segundo método (2ºM), y a la propuesta un MPT-TR general que asume distribuciones como el tercer procedimiento (3ºM), para el cual nos hemos basado en el estudio de (Heck et al., 2018). Tras esta breve introducción a los procedimientos de modelado, pasaremos a describirlos formalmente centrándonos en el 2HTM.

3.1 Procedimiento de modelado 1°M (Medias)

El primer procedimiento de modelado estima el modelo en dos pasos secuenciales. Primero se estima los parámetros θ del MPT, teniendo en cuenta solamente las variables de tipo categóricas. Después, en un segundo paso, se estiman las medias de los TR de las ramas latentes, que denominamos μ_{jz} (Hu, 2001). En la Fig. 4 representamos el 2HTM bajo el procedimiento 1°M:

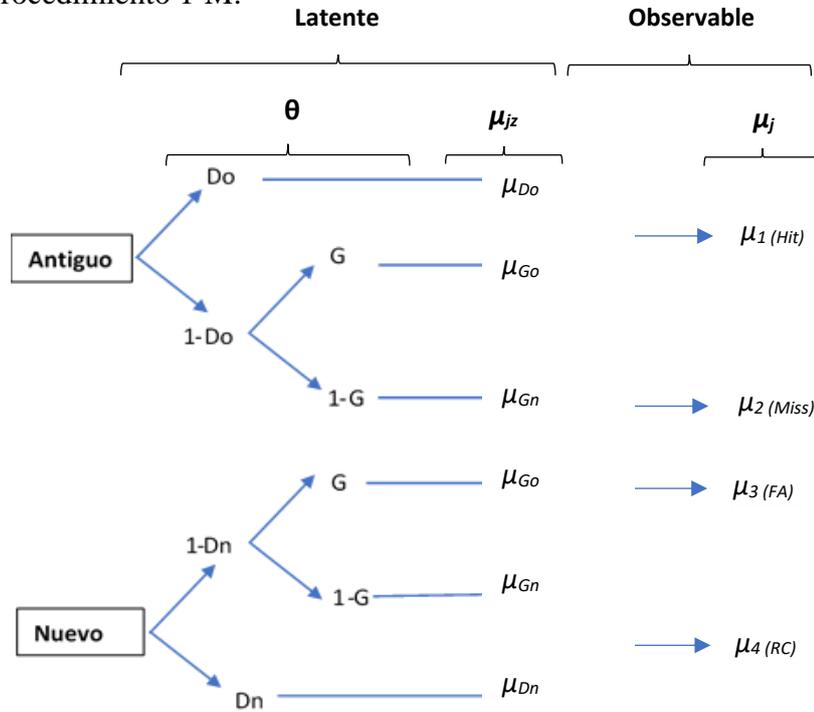


Figura 4. Representación del 2HTM con agregación de medias de TR (1°M) Nótese que partimos un modelo en el que de las 6 ramas solamente hay 4 componentes de tiempos diferenciados; μ_{Do} , μ_{Dn} , μ_{Go} y μ_{Gn} (ver Ecuación 23).

En el primer paso, se aplica la Ecuación 16 para descomponer la probabilidad de cada categoría observada, $\pi_{ij}(\theta_i)$, como la suma de las probabilidades de las z ramas, $\pi_{ijz}(\theta_i)$ que conduce a j . A su vez, aplicando ahora las Ecuaciones 6 a 11, descomponemos cada $\pi_{ijz}(\theta_i)$ como el producto de las probabilidades de los nodos que conducen a dichas z . En el primer árbol esto se traduce en:

$$\pi_{i1}(\theta_i) = Do_i + (1 - Do_i)G_i \quad (29)$$

$$\pi_{i2}(\theta_i) = (1 - Do_i)(1 - G_i) \quad (30)$$

Y en el segundo árbol en:

$$\pi_{i3}(\theta_i) = (1 - Dn_i)G_i \quad (31)$$

$$\pi_{i4}(\theta_i) = Dn_i + (1 - Dn_i)(1 - G_i) \quad (32)$$

Si el modelo es identificable (ver apartado 5.3.1) se estiman los parámetros del modelo MPT sin tener en cuenta los datos de TR, por métodos de estimación tales como mínimos

cuadrados o máxima verosimilitud. Habiendo librerías implementadas en R (R Core Team 2018) que permiten estimar los parámetros θ , tales como “*mpt*” (Wickelmaier, 2011), “*MPTinR*” (Singmann y Kellen, 2013) o “*TreeBUGS*” (Heck, Arnold y Arnold, 2018).

En base a los θ estimados, $\hat{\theta}$, en un segundo paso se descomponen las esperanzas de los TR de las categorías observadas como una suma ponderada de las esperanzas de los TR de las z ramas² que conducen a cada categoría j , tal que,

$$\mu_{ij} = \sum_{z=1}^Z \pi_{ijz} \cdot \mu_{ijz} \quad (33)$$

Para el 2HTM (ver Fig. 4) tendríamos pues que μ_{ij} sería:

$$\mu_{i1} = (1 - Do_i)G_i \cdot \mu_{Go} + Do_i \cdot \mu_{Do} \quad (34)$$

$$\mu_{i2} = (1 - Do_i)(1 - G_i) \cdot \mu_{Gn} \quad (35)$$

$$\mu_{i3} = (1 - Dn_i)G_i \cdot \mu_{Go} \quad (36)$$

$$\mu_{i4} = (1 - Dn_i)(1 - G_i) \cdot \mu_{Gn} + Dn_i \cdot \mu_{Dn} \quad (37)$$

Finalmente, si el modelo final es identificable, se sustituye las $\hat{\theta}$ y se estiman las medias de TR latentes por mínimos cuadrados, tal y cómo propone originalmente el autor del modelo (Hu, 2001).

Al respecto del procedimiento de modelado que acabamos de describir, cabe mencionar que estimar únicamente las medias de los TR puede resultar problemático por varias razones. En primer lugar, al describir el conjunto de TR mediante su media se pierde información acerca de los momentos de mayor orden y, por lo tanto, desconocemos la forma en que distribuyen los ensayos (*e.g.* su variabilidad, asimetría...). En segundo lugar, los TR suelen seguir distribuciones asimétricas positivas (Luce, 1986), por lo que la media no tendría por qué ser representativa de la distribución subyacente. Además, es habitual que en contextos experimentales existan valores de TR extremos positivos, que se han interpretado como problemas de ejecución de motora (McGill and Gibbon, 1965) o como una incapacidad de inhibición de EE distractores (Horowitz & Wolfe, 2005). Mediante el 1ºM estos *outliers* que podrían ser influyentes y alterar sustancialmente la media estimada. Razón por la cual, en las condiciones que acabamos de mencionar, la media de los TR podría constituir un estadístico sesgado para el modelado de los TR.

² Nótese que para que sea un mixture model las π_{ijz} tendrían que ser ponderadas por π_{ij} (ver Ecuación 23).

3.2 Procedimiento de modelado 2^oM (Bins)

Este procedimiento parte del hecho de que cualquier distribución, incluidas las asociadas a los TR, puede ser representada mediante un histograma (Van Zandt, 2000). Para ello, se categorizan los vectores de TR observados en b bins. Después se calcula la frecuencia relativa de cada bin contando el número de TR dentro de este y dividiéndolo por el número de TR asociados a dicha categoría de respuesta. Los bins se etiquetan según una escala de rapidez establecida según el número de bins y sus fronteras. Si para $b=2$ estableciéramos la frontera en la media de los TR observados, entenderíamos como rápidos aquellos TR mayores a la media y como lentos a todos los demás.

Con los TR discretizados en bins, se reparametriza el modelo, formando lo que se denomina un MPT-extendido. En el MPT-extendido se subdividen las ramas del MPT original en tantas sub-ramas como bins. De esta forma cada categoría de respuesta observada se subdivide en b bins, de más rápido a más lento. Finalmente se introduce un nuevo nodo con probabilidad L_{ijzb} ; donde L_{ijzb} es la probabilidad de que una rama z que conduzca a la categoría j caiga en el bin b . La representación de nuestro modelo 2HTM-2BIN se muestra en la Fig. 5:

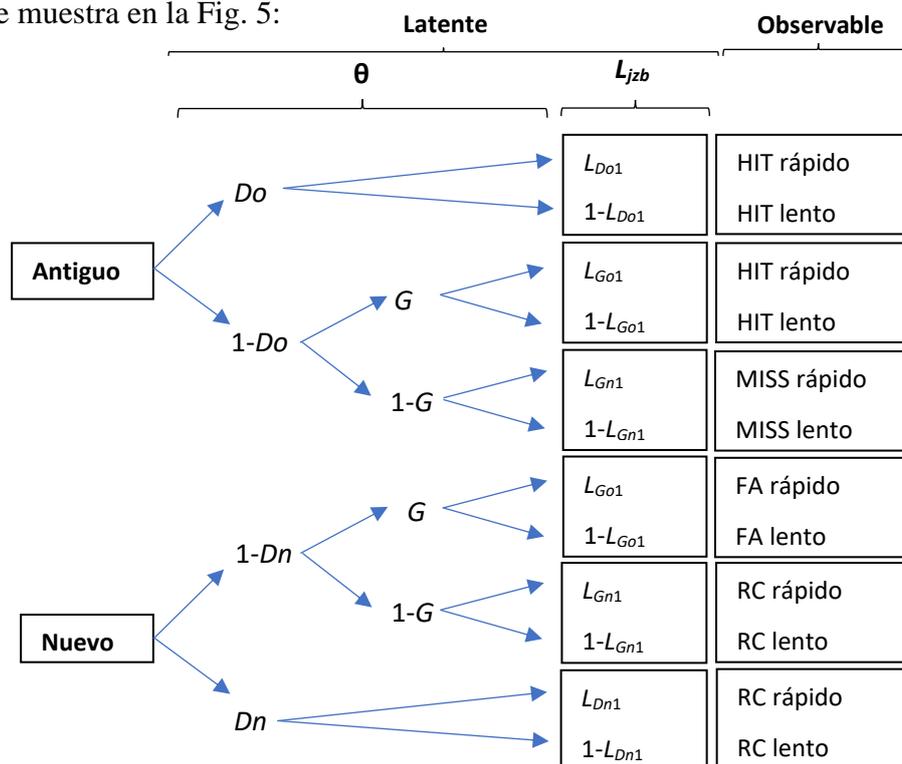


Figura 5. Representación de 2HTM-2BIN (2^oM). Nótese que partimos de un modelo en el que de las 6 ramas solamente hay 4 componentes de tiempos diferenciales; $L(Do)$, $L(Dn)$, $L(Go)$ y $L(Gn)$ (ver Ecuación 23.)

A mayor número de *bins* mayor precisión, puesto que más se aproximará el histograma de los tiempos a la distribución de los datos. No obstante, al aumentar el número de *bins* aumenta el número de parámetros a estimar y ello suele conllevar problemas de identificabilidad del modelo. Así pues, para evitar este problema trabajaremos $b=2$ *bins*.

Podemos obtener ahora la probabilidad de que una respuesta j caiga en un *bin* b , π_{ijb} , como la suma de las probabilidades de que la rama z que conduzca a la categoría j caiga en el *bin* b ponderada por la probabilidad de que la categoría j caiga en la rama z . Por lo que π_{ijb} se obtiene tal que,

$$\pi_{ijb} = \sum_{z:Z \text{ conduce a } j} \pi_{ijz} \cdot L_{ijzb} \quad (38)$$

Que en nuestro modelo se traduciría a:

$$\pi_{i11} = D o_i \cdot L_{iD o1} + (1 - D o_i) \cdot G_i \cdot L_{iG o1} \quad (39)$$

$$\pi_{i12} = D o_i \cdot (1 - L_{iD o1}) + (1 - D o_i) \cdot G_i \cdot (1 - L_{iG o1}) \quad (40)$$

$$\pi_{i21} = (1 - D o_i) \cdot (1 - G_i) \cdot L_{iG n1} \quad (41)$$

$$\pi_{i22} = (1 - D o_i) \cdot (1 - G_i) \cdot (1 - L_{iG n1}) \quad (42)$$

$$\pi_{i31} = (1 - D n_i) \cdot G_i \cdot L_{iG o1} \quad (43)$$

$$\pi_{i32} = (1 - D n_i) \cdot G_i \cdot (1 - L_{iG o1}) \quad (44)$$

$$\pi_{i41} = D n_i \cdot L_{iD n1} + (1 - D n_i) \cdot (1 - G_i) \cdot L_{iG n1} \quad (45)$$

$$\pi_{i41} = D n_i \cdot (1 - L_{iD n1}) + (1 - D n_i) \cdot (1 - G_i) \cdot (1 - L_{iG n1}) \quad (46)$$

Finalmente, si el modelo es identificable (ver apartado 5.3.1), se estiman los parámetros θ y L_b conjuntamente mediante el algoritmo de esperanza-maximización (*EM*), por ejemplo, utilizando el *software MPTinR*'' (Singmann y Kellen, 2013).

El procedimiento 2^oM consigue un aumento de precisión de las estimaciones sin la necesidad de asumir ninguna distribución y dando más información sobre la distribución de los datos que 1^oM. No obstante, esta precisión está relacionada directamente con el número de *bins*, y tal y como hemos apuntado anteriormente, aumentar el número de *bins* suele conllevar problemas de identificabilidad del modelo. Como solución a este problema se suelen establecer restricciones, pero estas podrían poner en riesgo la validez del modelo teórico (Heck y Erdelfelder, 2006; Kellen, Singmann, Vogt y Klauer, 2015).

Por otro lado, si las probabilidades de los procesos son demasiado extremas podríamos encontrarnos que, en algunos *bins* donde no hay observaciones, el modelo no sería estimable o, si hay muy pocas observaciones, dicha estimación podría ser imprecisa.

3.3. Procedimiento de modelado 3^oM (Distribuciones)

En el tercer proceso se empieza por asumir una *f.d.p.* para la distribución de los TR con pesos determinado por el modelo MPT. La validez del modelo dependerá de las asunciones distribucionales que establezcamos respecto a g . Se suelen asumir distribuciones positivas con asimetría derecha (Luce, 1986).

Teniendo en cuenta el modelo representado en Fig.6, bastaría con aplicar la Ecuación 24 para formular la función de distribución conjunta de los datos categóricos y continuos.

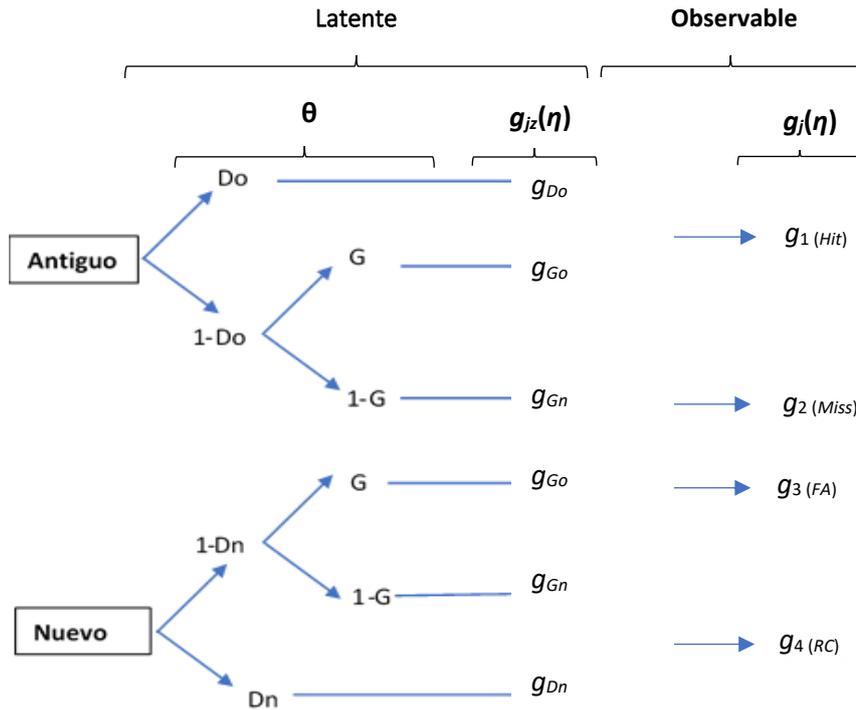


Figura 6. Representación de 2HTM en 3^oM. Nótese que partimos de un modelo en el que de las 6 ramas solamente hay 4 componentes de tiempos diferenciales; g_{Do} , g_{Go} , g_{Gn} y g_{Dn} (ver Ecuación 23)

Que para el 2HTM sería:

$$g(y_{ij}|j = 1) = Do_i \cdot g(y_{i11}|\eta_i) + (1 - Do_i)(G_i) \cdot g(y_{i12}|\eta_i) \quad (47)$$

$$g(y_{ij}|j = 2) = (1 - Do_i) \cdot (1 - G_i) \cdot g(y_{i23}|\eta_i) \quad (48)$$

$$g(y_{ij}|j = 3) = (1 - Dn_i)(G_i) \cdot g(y_{i34}|\eta_i) \quad (49)$$

$$g(y_{ij}|j = 4) = (1 - Dn_i)(1 - G_i) \cdot g(y_{i45}|\eta_i) + (Dn_i) \cdot g(y_{i46}|\eta_i) \quad (50)$$

Finalmente, si el modelo es identificable, se estiman los parámetros que definen las variables categóricas y continuas conjuntamente, por ejemplo, por *EM*. Utilizando en este caso la librería “*gpt*” (Heck et al., 2018).

Frente a los anteriores procedimientos, el 3^oM aprovecha más información sobre la distribución teórica de los TR, pudiendo dar cuenta de las asimetrías de las distribuciones

e incluso, dependiendo de la *f.d.p.* asumida, consideramos podría modelizar correctamente TR extremos (*i.e.* no habría una tendencia de *outliers* positivos). Pero incluso cuando tenemos un conocimiento previo de cómo se distribuirá la variable continua (*e.g.* los TR son positivos y se caracterizan por tener asimetrías positivas) sigue sin estar claro qué *f.d.p.* se debe asumir.

La validez del modelo depende, entre otras cosas, de la *f.d.p.* asumida. Esto sugiere que en las situaciones en las que no se tiene información clara que permitan asumir una *f.d.p.* podría ser preferible utilizar el 2ºM. Así pues, cabría preguntarse en qué situaciones de distribuciones de los datos no asumir ninguna distribución (2ºM) puede ser dar lugar a un mejor ajuste que asumirla (3ºM).

Por último, debemos comentar que con el 2ºM podemos aumentar la precisión de sus estimaciones aumentando el número de *bins*. En contrapartida, frente al 2ºM, el procedimiento 3ºM no exige una discretización en *bins* y la creación de los asociados parámetros de probabilidad. Por lo que un 2ºM con un elevado número de *bins* seguramente tendrá problemas de identificabilidad. En definitiva, para evitar el riesgo de establecer un elevado número de restricciones que violen el modelo teórico, puede que el número de *bins* tenga que ser bajo, perdiendo por lo tanto precisión. En consecuencia, para que exista una equivalencia en las restricciones paramétricas aplicadas en 3ºM y 2ºM, se establecerán pocos *bins* en 2ºM y sus predicciones podrían ser más imprecisas que las obtenidas con el 3ºM.

4 Objetivos

4.1 Generales

Estudiar cuál de los tres procedimientos de modelización de MPT-RT (*i.e.* 1ºM, 2ºM y 3ºM) y qué asunciones distribucionales (Ex-Gaussian, *Shift-Wald* y *Shift-Gamma*) son más adecuadas, bajo el criterio de estimaciones insesgadas y de buen ajuste, para cada escenario de distribución de los TR. Los escenarios están definidos por variables a las que pueden ser sensibles los tres métodos. Este trabajo tiene como fin último proponer un protocolo de modelización específico para los diferentes escenarios de distribución de los datos observados.

4.2 Específicos

- a) Analizar los cambios en el sesgo de los parámetros estimados ante condiciones de simulación con niveles paramétricos que generan probabilidades extremas de procesos; y estudiar qué procedimiento da lugar a estimaciones más insesgadas en cada situación:

Hipótesis 1: Todos los modelos estimarán peor sus parámetros θ y TR cuando su valor verdadero es 0 o 1 o un valor cercano, pudiendo haber casos donde el modelo no converge.

Hipótesis 2: Se espera encontrar que ante estas condiciones los procedimientos 1°M y el 2°M harán estimaciones de TR más sesgadas que 3°M, ya que no parten del conjunto de TR observados sino de sus medias, y estas últimas serán sesgadas debido a que habrá menos observaciones de TR y, además, estas no se distribuyen simétricamente. Siguiendo la misma lógica, dado que en 2°M se categorizan las respuestas observadas en rápidas o lentas en base a la media de TR observados, y si dichas medias están sesgadas también lo estarán las fronteras que separan los *bins* entre sí. Además, habrá *bins* con pocas o ninguna observación, estimando los parámetros de forma sesgada.

- b) Deseamos estudiar las variaciones del sesgo de los parámetros estimados ante condiciones que generan TR típicos del paradigma 2HTM (Heck et al., 2018) o TR extremos, así como elegir cuál de los diferentes procedimientos de modelización MPT-RT estima de forma menos sesgada.

Hipótesis 3: Hipotetizamos que tanto 1°M como 2°M se verán especialmente perjudicados en las condiciones con TR extremos, especialmente en condiciones donde también hay probabilidades extremas de los procesos.

Hipótesis 4: Ante la existencia TR y θ extremos positivos posiblemente asumir una distribución *Ex-Gaussian* permita la obtención de un menor sesgo de estimación de los TR, debido a que el componente exponencial, que refleja la cola de distribución, permite modelizarlos con mayor facilidad (Matzke y Wagenmakers, 2009) que la distribución *Shift-Gamma* o *Shift-Wald*. Por otro lado, ante TR extremos y θ no extremos no está claro si el 3°M_ *Ex-Gaussian* será más adecuado que el 2°M, ya que, bajo nuestras hipótesis, 2°M funcionaría mejor si no hay θ extremos y, además, no requiere establecer asunciones distribucionales.

- c) Estudiar para todas las condiciones mencionadas qué distribución de 3°M debemos asumir, ahora bajo un criterio de parsimonia y mínimo desajuste, teniendo o no en cuenta la distribución generadora de TR:

Hipótesis 5: Se espera que el desajuste sea menor si la *f.d.p.* subyacente a los datos coincide con la asumida por el 3°M. Sin embargo, ante el desconocimiento de la distribución generadora de los TR se espera encontrar el mismo patrón de selección de *f.d.p.* asumidas que se hipotetizó en lo referido al sesgo.

5 Metodología

Para la simulación del experimento se generaron por cada individuo la frecuencia de cada categoría de respuesta y el vector de TR asociado a esta. Para ello hay que definir tanto las distribuciones de los TR como de las probabilidades de las categorías de respuestas. Asimismo, como se asume que hay variabilidad entre los participantes, los parámetros implicados serán *v.a.c* que a su vez vendrán caracterizados por una *f.d.p* y con unos hiperparámetros. A continuación, especificaremos detalladamente lo anterior para cada escenario de investigación.

5.1 Condiciones de investigación

Mediante la manipulación de los parámetros que caracterizan las *f.d.p.* generadoras de datos propondremos un total de 24 escenarios de simulación diferentes (ver Tabla 2) que serán fruto de la combinación de las condiciones de dos variables independientes (VIs) y de una variable control (VC).

Tabla 2

Resumen de condiciones de Simulación.

VI_{θ}		VI_{TR}					
Do y Dn	Go	A			B		
		<i>Ex-Gaussian</i>	<i>Shifed-Wald</i>	<i>Shifed-Gamma</i>	<i>Ex-Gaussian</i>	<i>Shifed-Wald</i>	<i>Shifed-Gamma</i>
A	A	AA_A	AA_A	AA_A	AA_B	AA_B	AA_B
A	B	AB_A	AB_A	AB_A	AB_B	AB_B	AB_B
B	A	BA_A	BA_A	BA_A	BA_B	BA_B	BA_B
B	B	BB_A	BB_A	BB_A	BB_B	BB_B	BB_B

Las siglas A referidas a los parámetros θ indican probabilidades extremas y B a no extremas. Las siglas A referidas a los parámetros η indican TR con distribuciones típicas y B a distribuciones con TR extremos.

Con respecto a las probabilidades de los procesos, θ , establecemos para la primera VI, denominada “ VI_{θ} ”, cuatro condiciones posibles:

- Elevada probabilidad de generar parámetros Do_i , Dn_i y G_i extremos (AA)
- Elevada probabilidad de generar parámetros Do_i , Dn_i extremos (AB)
- Elevada probabilidad de generar parámetros G_i extremos (AB)
- Baja probabilidad de generar parámetros Do_i , Dn_i y G_i extremos (BB)

Con respecto a la forma de las distribuciones de los TR, en la segunda VI, denominada “ VI_{TR} ”, habrá dos condiciones posibles:

- Distribuciones de TR típicas (A).
- Distribuciones que generan TR extremos positivos (B).

A su vez, para cada condición los TR serán generados por tres de las funciones de distribución más empleadas en modelado de TR, la *Ex-Gaussian*, *Shift-Wald* y *Shift-Gamma* (Luce, 1986; Van Zandt, 2000; Palmer et al., 2011).

Considerando lo expuesto, cada una de las 4 condiciones posibles de distribuciones de frecuencias de respuesta categórica se combinará con dos condiciones de distribución de TR ($4 \times 2 = 8$ condiciones a investigar) que seguirán las tres funciones de distribución antes mencionadas ($4 \times 2 \times 3 = 24$ escenarios a simular).

5.2 Simulación de condiciones

En cada uno de los escenarios de simulación se trabaja con una muestra de 200 sujetos. Para cada sujeto se generan 100 respuestas: 50 a EE antiguos y 50 a EE nuevos ($N=50$). Cada conjunto de datos individual contiene por cada ensayo el tipo de EE presentado la respuesta categórica observada (“antiguo”/“nuevo”) y el TR asociado a esta. Se justifica la ausencia de réplicas por sujetos por varios motivos. En primer lugar, se está simulando un paradigma experimentalmente verosímil. En segundo lugar, la estimación de parámetros se realizará individualmente, ya que se comprobó que el ajuste individual de un modelo MPT es más adecuado que el obtenido por la agregación grupal de los datos, siempre y cuando se establezca un elevado número de ensayos individuales por condición ($N > 40$) (Chechile, 2009). Y, en tercer lugar, debido al elevado tiempo de estimación de todos los conjuntos de datos individuales, resultó inviable realizar suficientes réplicas por sujeto y ajustarnos al tiempo y condiciones materiales disponibles.

5.2.1 Niveles de los parámetros de las variables categóricas

Como se adelantó en anteriores apartados, se asume variabilidad en la población respecto a los parámetros θ . Por lo tanto, cada parámetro es una variable aleatoria, y se asume que sigue una distribución Beta con hiperparámetros (α, β) tal que,

$$Do_i \sim Beta(\alpha_{Do}, \beta_{Do}) \quad (51)$$

$$Dn_i \sim Beta(\alpha_{Dn}, \beta_{Dn}) \quad (52)$$

$$G_i \sim Beta(\alpha_G, \beta_G) \quad (53)$$

Para elegir los niveles de los hiperparámetros α, β estableceremos que tanto para la condición de simulación con probabilidades extremas (A) de procesos como la de no extremos (B) la media de cada parámetro Do, Dn, G será constante. En concreto, las medias de las probabilidades θ , que se establecieron en base a estudios previos (Heck et al., 2018), fueron $Do = Dn = .6$ y $G = .5$. Sin embargo, la variabilidad de los procesos Do, Dn y G será mayor en la condición A que en la condición B.

Teniendo en cuenta lo anterior y que $\sigma^2_A = 0.1$; $\sigma^2_B = 0.001$, entonces, aplicando las relaciones entre los parámetros α, β y la media y la varianza de la distribución Beta,

$$\beta = (1 - \mu) \left(\frac{\mu(1-\mu)}{\sigma^2} - 1 \right) \quad (54)$$

y

$$\alpha = \mu \left(\frac{\mu(1-\mu)}{\sigma^2} - 1 \right) \quad (55)$$

obtenemos que los hiperparámetros α, β que cumplen las condiciones anteriores para la condición A son: $\alpha_{Do} = \alpha_{Dn} = .84$; $\beta_{Do} = \beta_{Dn} = .56$; $\alpha_G = .75, \beta_G = .75$; y para la condición B son: $\alpha_{Do} = \alpha_{Dn} = 143.4$; $\beta_{Do} = \beta_{Dn} = 95.5$; $\alpha_G = 124.5$; $\beta_G = 124.5$.

En la Fig. 7 se representan las funciones de densidad asociadas a los θ para cada uno de los tres procesos (Do , Dn y G) en los dos escenarios de su distribución (A y B):

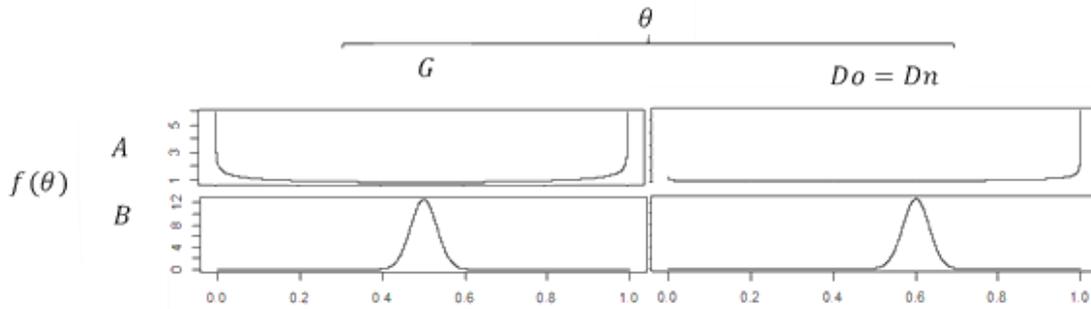


Figura 7. Funciones de densidad de la Beta para los procesos Do (primera fila), Dn (segunda fila) y G (tercera fila) en condición A (columna izquierda) y B (columna derecha).

5.2.2 Niveles de los parámetros de las variables continuas.

Como se mencionó en el apartado 2, dependiendo de la forma de la distribución g de los tiempos de reacción habrá diferentes conjuntos de parámetros que forman el vector η . En el presente apartado describiremos la lógica subyacente a la generación de los tiempos de reacción.

En concreto generaremos los tiempos de reacción, y_{ijz} , de las 6 ramas y, tal y como se expresó formalmente en la Ecuación 23, cada y_{ijz} seguirá uno de los 4 componentes distribucionales diferentes. No obstante, adelantemos que pesar de que en el modelo teórico podamos distinguir 4 componentes distribucionales, en este trabajo, basándonos en estudios anteriores (Kellen, Singmann, Vogt y Klauer, 2014; Heck, Erdelfelder y Kieslich, 2018; Heck y Erdelfelder, 2016), hemos establecido como restricción que $\eta_{Gn} = \eta_{Go}$. Así pues, tendremos en cuenta solamente 3 componentes distribucionales diferenciados (ver apartado 3.2.1).

Hay recordar que tenemos dos condiciones (A y B de VI_{TR}) de distribución de los TR. Por ello, para generar los datos de TR en los escenarios de distribuciones de TR típicas (A) y de elevada probabilidad de generar TR extremos positivos (B), vamos a establecer los siguientes criterios: a) los componentes de distribución diferentes de TR pertenecientes a una misma condición tienen la misma forma (o muy aproximada), solamente difieren entre sí en su media.; y b) las distribuciones de los TR de cada componente diferencial de TR en las dos condiciones tendrán la misma media, solamente difieren en su forma. Donde las esperanzas poblacionales³ de las distribuciones de los tres componentes son: $E(Y_{Do}) = 800$, $E(Y_{Dn}) = 900$, $E(Y_G) = 1000$.

Seguidamente pasaremos a definir los parámetros η de las tres distribuciones generadoras de datos.

- *Shift-Wald*

La distribución *Shift-Wald*, o gaussiana inversa desplazada, viene definida por los parámetros μ , λ (forma) y s (*shift* o desplazamiento). Así pues, los TR de los tres componentes distribucionales diferentes antes mencionados se distribuirán tal que,

$$y_{ijz} \sim \text{Shift} - \text{Wald}(\mu, \lambda, s_{ijz}). \quad (56)$$

Con el fin de manipular la forma de la distribución en las condiciones A y B se estableció para la condición A $\lambda_A = 2000$ y para la condición B $\lambda_B = 400$. En ambas condiciones mantendremos el mismo parámetro $\mu = 400$.

Para establecer los valores de s de cada componente (s_{jz}), manteniendo las esperanzas establecida a priori, tendremos en cuenta que en esta distribución se cumple la relación:

$$E(Y) = \mu + s \quad (57)$$

Despejando s en la Ecuación 57, tendremos que $s_{Do} = 400$; $s_{Dn} = 500$ y $s_G = 600$.

En el Apéndice A, presentamos en la Fig.A.1 las funciones de densidad para cada proceso y escenario.

Como además hay variabilidad entre los sujetos en los valores de los parámetros, estableceremos que dicha variabilidad afectará únicamente a la media de las distribuciones, a través del parámetro s . Así pues, el parámetro desplazamiento de cada componente será una *v.a.c* que se distribuirá tal que,

$$s_{ijz} \sim N(s_{jz}, \sigma) \quad (58)$$

Donde la variabilidad intersujeto es $\sigma = 10$.

³ Denominamos esperanza poblacional, $E(Y)$, a la tendencia central de la distribución de esperanzas individuales, $E(Y_i)$. Donde $E(Y_i) \sim N(E(Y), 10)$

En resumen, el vector $\boldsymbol{\eta}$ en el modelo que asumen una distribución *Shift-Wald* es $\boldsymbol{\eta}_{A_i} = (\mu, \lambda_A, s_{Do_{A_i}}, s_{Dn_{A_i}}, s_{G_{A_i}})$ y $\boldsymbol{\eta}_{B_i} = (\mu, \lambda_B, s_{Do_{B_i}}, s_{Dn_{B_i}}, s_{G_{B_i}})$, para las condiciones A y B respectivamente.

Los parámetros que generaron los TR para una *Shift-Wald* en la condición A (TR típicos) se han establecido basándonos en los parámetros usados por Heck et al. (2018) ya que, según este autor, reproducen TR típicos del 2HTM⁴.

- Ex-Gaussiana

Bajo una *Ex-Gaussian*, los parámetros que pertenecen al vector $\boldsymbol{\eta}$ son la media μ de la distribución Gaussian, su desviación típica σ y el parámetro τ que es la media y la desviación típica de la distribución exponencial. Así pues, los TR de los tres componentes distribucionales diferentes, y_{ijz} , seguirán:

$$y_{ijz} \sim \text{Ex-Gaussian}(\mu_{ijz}, \sigma, \tau) \quad (60)$$

Con el fin de modelar las condiciones A y B alteraremos la forma de las distribuciones con el parámetro τ manteniendo constante el parámetro $\sigma = 50$. En la condición A, caracterizada por TR típicos (Heck et al., 2018), establecimos que $\tau_A = 300$, mientras que en la condición B, que $\tau_B = 500$. Como sabemos además que,

$$E(Y) = \mu + \tau \quad (61)$$

Despejando μ , podemos obtener los valores que cumplen con las esperanzas poblacionales establecidas a priori para cada uno de los componentes distribucionales. Entonces, para la condición A los valores de μ_{jz} de cada componente distribucional serán: $\mu_{Do_A} = 500$; $\mu_{Dn_A} = 600$; $\mu_{G_A} = 700$. Mientras que en la condición B los valores μ_{jz} asociados a cada componente distribucional serán: $\mu_{Do_B} = 300$; $\mu_{Dn_B} = 400$; $\mu_{G_B} = 500$. Lo dicho se representa gráficamente a en el Apéndice A en la Fig. A.2.

Nuevamente, para generar la variabilidad intersujeto ($\sigma = 10$) estableceremos que, para escenario y componente μ es una *v.a.c.*, que se distribuye tal que,

$$\mu_{ijz_A} = N(\mu_{jz_A}, \sigma) \quad (62)$$

$$\mu_{ijz_B} = N(\mu_{jz_B}, \sigma) \quad (63)$$

En resumen, en el modelo que asume que los TR siguen una distribución Ex-Gaussian el vector $\boldsymbol{\eta}$ será $\boldsymbol{\eta}_{A_i} = (\mu_{Do_{A_i}}, \mu_{Dn_{A_i}}, \mu_{G_{A_i}}, \sigma, \tau_A)$ y $\boldsymbol{\eta}_{B_i} = (\mu_{Do_{B_i}}, \mu_{Dn_{B_i}}, \mu_{G_{B_i}}, \sigma, \tau_B)$, en las condiciones A y B respectivamente.

⁴ Los parámetros no reportados en Heck et al., 2018 se establecieron en base a la información contenida en el repositorio que fue facilitado por los autores tras una comunicación personal con los mismos.

- *Shift-Gamma*

Siguiendo la lógica anterior,

$$y_{ijz} \sim \text{Gamma}(\alpha, \beta, s_{ijz}) \quad (64)$$

Donde $\alpha = 1,5$ y donde en las condiciones A y B se manipulará β . En concreto, en la condición A, $\beta_A = 300$ y en la condición B $\beta_B = 500$. Sabiendo que la relación entre los parámetros α y β y el valor esperado de una distribución gamma es:

$$E(Y) = \frac{\beta}{\alpha} \quad (65)$$

Así pues, para mantener las esperanzas poblacionales previamente estipuladas tendremos para la condición A que el parámetro *shift* de cada componente (s_{jz}) será: $s_{Do_A} = 600$; $s_{Dn_A} = 700$; $s_{G_A} = 800$ y para la B será $s_{Do_B} = 466.67$; $s_{Dn_B} = 566.67$; $s_{G_B} = 666.67$.

En la Fig.A.3 del Apéndice A se representa gráficamente los tres componentes g en las condiciones A y B.

Finalmente, para generar variabilidad intersujeto estableceremos que el parámetro desplazamiento de cada componente será una *v.a.c* que se distribuirá tal que

$$s_{ijz_B} \sim N(s_{jz_B}, \sigma) \quad (66)$$

$$s_{ijz_A} \sim N(s_{jz_A}, \sigma) \quad (67)$$

Donde nuevamente la variabilidad intersujeto es $\sigma = 10$.

En resumen, en el modelo que asume que los TR siguen una distribución *Shift-Gamma* el vector η será $\eta_{A_i} = (\mu, \lambda_A, s_{Do_{A_i}}, s_{Dn_{A_i}}, s_{G_{A_i}})$ y $\eta_{B_i} = (\mu, \lambda_B, s_{Do_{B_i}}, s_{Dn_{B_i}}, s_{G_{B_i}})$, en las condiciones A y B respectivamente.

5.3 Estimación de parámetros

5.3.1 Identificabilidad del Modelo

Como regla general, para que un modelo sea identificable y dé soluciones únicas el número de parámetros libres debe ser menor o igual que el de categorías observables ($k < J$) (Bamber y van Santen, 1985). Para solucionar el problema de identificabilidad se suele recurrir al establecimiento de restricciones.

En el presente trabajo se ha impuesto que las probabilidades de detección (D) asociadas a EE antiguos y EE nuevos es la misma (*i.e* $Do = Dn$). Una vez establecidas las restricciones referidas a θ , el modelo MPT, que sólo tiene en cuenta variables categóricas, es identificable. Ahora bien, tras la introducción de los parámetros relativos a los TR y una nueva evaluación de la identificabilidad, tendremos que establecer nuevas

restricciones. En nuestro caso, asumiremos que los tiempos de *guessing* “antiguo” y de *guessing* “nuevo” siguen una misma distribución g , por lo que sus esperanzas serían también iguales (*i.e.* $E(y_{i_G}) = E(y_{i_{1-G}})$). Por consiguiente, también se cumpliría que $g(y_{i_{23}}) = g(y_{i_{34}})$ y que $E(y_{i_{23}}) = E(y_{i_{34}})$. Es decir, el primer procedimiento cumpliría que $\mu_{Go} = \mu_{Gn}$, el segundo que $L_{Go1} = L_{Gn1}$ y el tercer que $\eta_{Go} = \eta_{Gn}$.

Todas las restricciones se establecieron en base a los estudios revisados (Kellen et al., 2014; Heck et al., 2018; Heck y Erdelfelder, 2016).

5.3.2 Procedimiento de Estimación y de análisis

Para el procedimiento 1ºM hay que recordar que las estimaciones de los parámetros se hacen en dos pasos. En el primer paso utilizamos la librería “*MPTinR*” (Singmann y Kellen, 2013) para obtener las estimaciones de las probabilidades de los procesos, $\hat{\theta}$, desde las frecuencias de categorías observadas. A partir de estas estimaciones se calculan las $\hat{\pi}_{ijz}$ aplicando la Ecuación 5.

En el segundo paso estimamos las medias de los TR asociados a las ramas, $\hat{\mu}_{ijz}$. Debido a que no se ha implementado el primer procedimiento de estimación en ninguna librería en R (R Core Team, 2018), hemos desarrollado una función dentro de este *software* donde, mediante métodos de aproximación numérica, se buscan los parámetros que hacen mínima la siguiente matriz de residuos,

$$|M - \Lambda m^T|, \quad (68)$$

donde M es el vector de medias de TR observadas, m en el vector de medias de TR de rama desconocidos y Λ es una matriz $J \times Z$ que incluye en cada casilla (j, z) o bien $\hat{\pi}_{ijz}$ si la z rama conduce a la j respuesta o bien un 0 si no lo hace.

Para el procedimiento 2ºM se siguieron los siguientes pasos:

- 1) Para cada una de las 4 categorías originales, se crearon dos nuevas atendiendo a los TR, según estos fuesen rápidos o lentos (*i.e.* TR mayor o menor que la media). De esta forma las frecuencias de cada categoría original se distribuyeron entre las dos nuevas categorías. De esta forma se crean dos nuevos árboles con 4 categorías
- 2) Los parámetros de cada uno de los árboles se estiman mediante la librería de R “*MPTinR*” (Singmann y Kellen, 2013)

Para el procedimiento 3ºM se utiliza la librería “*gpt*” (Heck et al., 2018).

Como ya se comentó, para los tres procedimientos se estiman los parámetros de cada participante. Una vez estimados los parámetros para cada procedimiento e individuo, se

calculó el sesgo absoluto (δ) de los parámetros θ (δ_D y δ_G) individualmente. En lo referente a los TR, debido a que los parámetros asociados a los TR son diferentes en cada procedimiento, se calculó para cada procedimiento la media de los TR estimados. A partir de esta medida común, se calculó el sesgo individual de las medias de los TR por rama ($\delta_{\mu_{Do}}$, $\delta_{\mu_{Dn}}$ y δ_{μ_G}).

Posteriormente, se analizó el sesgo de cada parámetro mediante un modelo lineal mixto con tres factores de medidas repetidas con una matriz de varianza-covarianza de simetría compuesta, donde los efectos fijos del modelo saturado fueron: el procedimiento de estimación, “*Procedimiento*”; las condiciones de probabilidad de los procesos, “*VI_θ*” (AA/AB/BA/BB); las condiciones de distribución de los TR, “*VI_TR*” (A/B); y las interacciones entre estas variables. Para los efectos significativos se realizaron comparaciones múltiples *post-hoc* mediante la prueba de Tukey con corrección de Bonferroni. También se evaluó la relación entre sesgos mediante correlaciones de *Pearson* (r). En todos los contrastes realizados se estableció un nivel de confianza del 95%. Como medida de tamaño del efecto (TE) de los efectos fijos del modelo mixto se empleó el índice eta cuadrado parcial (η_p^2) mientras que para las correlaciones se empleó el coeficiente de determinación corregido (R_c^2).

Por otro lado, se obtuvieron medidas de bondad de ajuste. Después, para cada conjunto de datos individual se compararon los modelos 3°M (*i.e.* Ex -Gaussian, *Shift*-Gamma y *Shift*-Wald) y se seleccionó, bajo los índices AIC y BIC, el que mejor ajustaba. Se contabilizó la frecuencia de selección de cada modelo para cada condición. Y, de forma adicional, para comprobar si con el 3°M el mejor ajuste se da cuando el modelo asume el mismo escenario distribucional que el que generan los TR, se contabilizó la frecuencia de selección de cada modelo 3°M para para cada escenario distribucional y condición.

La comparación y selección de modelos no incluyó el ajuste de 2°M porque no existe una equivalencia con 3°M. Esto se debe a que 2°M no evalúa el desajuste de los TR sino el desajuste de la probabilidad L_b . Tampoco se incluyó 1°M ya que, siguiendo las propuestas del autor del procedimiento, Hu (2001), la estimación se realizó por mínimos cuadrados.

6 Resultados

Pasaremos ahora a exponer los resultados de los análisis del sesgo de cada parámetro.

6.1 Sesgo en las probabilidades de los procesos (\hat{D} y \hat{G}).

En primer lugar, se expondrán aquellos relativos a los sesgos de los parámetros de probabilidades de procesos (ver Fig. 11).

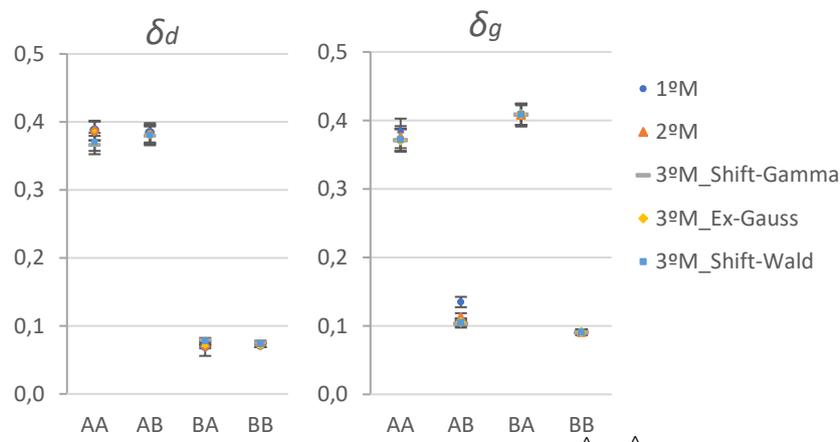


Figura 11. Medias del sesgo de las probabilidades \hat{D} y \hat{G} . Donde AA, AB, BA y BB representan las condiciones de probabilidad de los procesos son (VI_θ) . Las barras representan el IC95%.

Para el sesgo del parámetro \hat{D} (δ_D) el modelo restringido incluye únicamente los factores de efectos fijos “ VI_θ ”, “*Procedimiento*” y “ $VI_\theta * Procedimiento$ ”. Bajo un criterio de parsimonia, y manteniéndose la hipótesis nula de que el ajuste del modelo saturado y el restringido es igual ($X^2(20) = 151.866$; $p = 1$) nos quedamos con el modelo más simple. Aplicando la misma lógica respecto al sesgo del parámetro \hat{G} (δ_G), no hay evidencia de pérdida de ajuste ($X^2(32) = 198.158$; $p = 1$) entre el modelo saturado y el modelo que incluye únicamente los efectos fijos “ VI_θ ” y “*Procedimiento*”, quedándonos pues con el modelo restringido.

Se rechaza la hipótesis nula de ausencia de diferencias entre las 4 condiciones VI_θ para δ_D y para el δ_G (ver Tabla 3).

Tabla 3

Efecto de VI_θ para δ_D y para el δ_G .

δ_θ	<i>Procedimiento</i>	<i>F</i>	<i>gl</i>	<i>p</i>	η_p^2
δ_D	1ºM	1200.68	3, 1823	< .001	.664
	2ºM	1245.8	3, 1819	< .001	.673
	3ºM_Ex-Gaussian	1245.8	3, 1862	< .001	.651
	3ºM_Shift -Gamma	1198.7	3, 1847	< .001	.661
	3ºM_Shift -Wald	1183.62	3, 1860	< .001	.656
δ_G		4142.41	3, 8983	< .001	.316

Respecto al patrón de diferencias entre las condiciones “ VI_{θ} ” del δ_D , se mantienen las hipótesis nulas de que el δ_D en AA es igual que en AB ($p = 1$) y que el sesgo en BA es igual que BB ($p = 1$). Encontramos también que el δ_D es significativamente mayor en AA que en BA ($p < .001$) y que en BB ($p < .001$). Así como sucede con AB, donde encontramos evidencias de que el δ_d en dicha condición es significativamente mayor que el encontrado en BA ($p < .001$) y que en BB ($p < .001$). Además, este patrón de diferencias que se repite con todos los procedimientos de estimación.

Siguiendo con los efectos principales del factor “ VI_{θ} ”, encontramos que, con independencia del procedimiento de estimación empleado, el δ_G es significativamente mayor en la condición BA que en las demás condiciones (AA, AB y BB) ($p < .001$) y que en AA el δ_G es significativamente mayor que en AB ($p < .001$) y BB ($p < .001$).

Por otro lado, encontramos que únicamente en la condición BA se rechaza la hipótesis nula de que las diferencias de δ_D entre los diferentes procedimientos es la misma ($F(4, 2428) = 6.611, p < .001, \eta_p^2 = .004$). En concreto se encuentra que tanto el δ_D estimado mediante el procedimiento $3M_{Ex-Gaussian}$ es menor que en $2^{\circ}M$ ($p = .002$) y $1^{\circ}M$ ($p = .001$) como que el procedimiento $3M_{Shift-Wald}$ es menor que en $2^{\circ}M$ ($p = .007$) y que en $1^{\circ}M$ ($p = .05$).

Analizando ahora los efectos principales de la variable *Procedimiento* para el δ_G , rechazamos la hipótesis nula de que el δ_D es igual para todos los procedimientos ($F(4, 4181) = 3.662, p = .006, \eta_p^2 = .003$). Las evidencias indican que el δ_D con el procedimiento $1^{\circ}M$ es significativamente menor que en $3^{\circ}M_{Ex-Gaussian}$ ($p = .008$) que $3^{\circ}M_{Shift-Gamma}$ ($p = .04$) y que con $3^{\circ}M_{Shift-Wald}$ ($p = .02$).

Cabe también mencionar que con independencia del procedimiento existe una correlación negativa entre δ_D y δ_G , ($r = -.08; p < .001; N=24000; R^2_c = .006$), que, en ausencia de un contexto de comparación, serán interpretados con los criterios de Cohen (1992) correspondiendo a una relación de pequeña intensidad.

6.2 Sesgo de las medias de los TR ($\hat{\mu}_{Do}$, $\hat{\mu}_{Dn}$ y de $\hat{\mu}_G$)

A continuación, pasaremos a reportar los análisis del sesgo de las medias de los TR. En la Fig. 12. encontramos la representación de los tres sesgos de $\hat{\mu}_{Do}$ ($\delta_{\mu_{Do}}$), $\hat{\mu}_{Dn}$ ($\delta_{\mu_{Dn}}$) y de $\hat{\mu}_G$ (δ_{μ_G}).

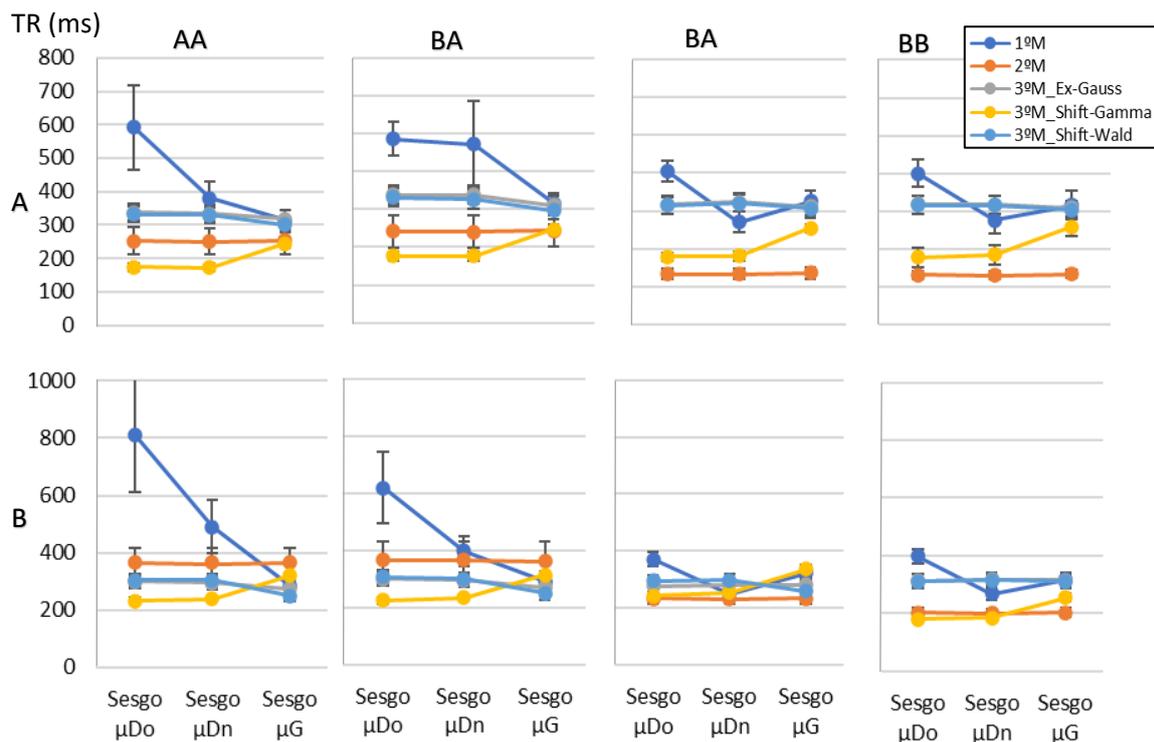


Figura 12. Sesgo medio de las de las medias de los TR estimados por rama. Donde se representan en dos filas las condiciones A y B de distribución de los TR (VI_{TR}) y en las columnas las condiciones AA, AB, BA y BB de probabilidad de los procesos son (VI_{θ}). Las barras representan el IC95%

Teniendo en cuenta el espacio disponible en el presente trabajo y con la finalidad de no dar información redundante sobre el comportamiento del sesgo de cada rama bajo las diferentes condiciones, solamente analizaremos el $\delta_{\mu_{Do}}$. Consideramos que no analizar el sesgo de los TR de las demás ramas está justificado por tres motivos. El primero es que, con la excepción del procedimiento 1ºM, dentro de una misma condición no parecen existir diferencias de sesgos medios entre ramas. En segundo lugar, el patrón de diferencias entre condiciones parece ser el mismo en los diferentes sesgos medios de TR de rama. Y, en tercer lugar, y nuevamente excluyendo 1ºM, las correlaciones entre los sesgos de los TR de las diferentes ramas son o próximos a 1, o por lo menos, bajo los criterios de Cohen (1992), de una elevada magnitud (ver Apéndice B). Por último, atendiendo a que el objetivo principal del presente trabajo es encontrar qué procedimiento debemos emplear en cada escenario, no resulta muy problemático que lo dicho hasta ahora no se cumpla en 1ºM porque, adelantamos desde ya, resulta ser el procedimiento menos indicado en casi todas las condiciones.

Para $\delta_{\mu_{Do}}$ (ver Fig. 13) nos quedaremos con el modelo saturado, donde los efectos fijos incluidos son “ VI_{θ} ”, “ VI_{TR} ”, “*Procedimiento*”, “ $VI_{\theta} * Procedimiento$ ”, “ $VI_{\theta} * VI_{TR}$ ”, “ $VI_{TR} * Procedimiento$ ” y $VI_{\theta} * VI_{TR} * Procedimiento$.

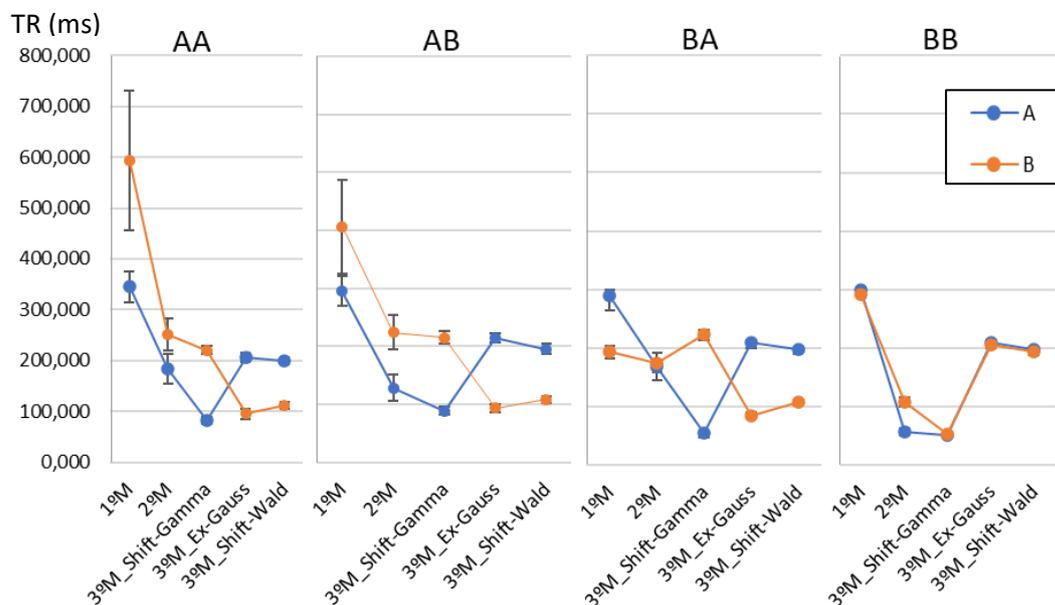


Figura 13. Sesgo medio de las medias de los TR de la primera rama. Donde se representan en dos colores las condiciones A y B de distribución de los TR (VI_{TR}) y en diferentes figuras las condiciones AA, AB, BA y BB de probabilidad de los procesos son (VI_{θ}). Las barras representan el IC95%.

En primer lugar, se rechaza la hipótesis nula de ausencia de diferencias de $\delta_{\mu_{Do}}$ entre las condiciones de “ VI_{θ} ” para todos los procedimientos (ver Tabla 4).

Tabla 4

Efecto de VI_{θ} en $\delta_{\mu_{Do}}$

<i>Procedimiento</i>	<i>F</i>	<i>gl</i>	<i>p</i>	η_p^2
1°M	70.52	3, 410	< .001	.34
2°M	134.58	3, 905	< .001	.309
3°M_Ex-Gaussian	242.28	3, 837	< .001	.465
3°M_Shift -Gamma	528.72	3, 1211	< .001	.567
3°M_Shift -Wald	124.89	3, 907	< .001	.292

En concreto, encontramos que el $\delta_{\mu_{Do}}$ en AA es mayor que en BB, ocurriendo esto para 1°M ($p < .001$), en 2°M ($p < .001$) y 3°M_Shift-Gamma ($p < .001$). No obstante, encontramos que el $\delta_{\mu_{Do}}$ en AA es menor que en BB en 3°M_Ex-Gauss ($p < .001$) y en 3°M_Shift-Wald ($p < .001$). Por otra parte, encontramos que el $\delta_{\mu_{Do}}$ en AA es significativamente mayor que en BA tanto para 1°M ($p < .001$) como para 2°M ($p = .001$) y para 3°M_Shift-Gamma ($p = .015$), pero mantenemos la hipótesis nula de igualdad

entre ambas condiciones cuando se estima con 3°M-Ex-Gauss ($p = 1$) y 3°M_Shift-Wald ($p = 1$). También observamos que el $\delta_{\mu_{Do}}$ en BA es mayor que en BB, tanto utilizando el 1°M ($p < .001$) como con 2°M ($p < .001$) y con 3°M_Shift-Gamma ($p < .001$), mientras que en 3°M_Shift-Wald y 3°M-Ex-Gauss, sucede lo contrario, siendo en BB mayor que en BA tanto en 3°M-Ex-Gauss ($p < .001$) como en 3°M_Shift-Wald ($p < .001$). Cabe también destacar que con 3°M_Shift-Gamma, AB es mayor que BB ($p < .001$), ocurriendo esta misma diferencia con 2°M ($p < .001$). Sin embargo, el sesgo en AB es menor que BB, si estimamos con 3°M-Ex-Gauss ($p < .001$) o 3°M_Shift-Wald ($p < .001$). No se rechaza la hipótesis nula de diferencia entre AB y BA para 2°M ($p = 1$), 3°M_Shift-Wald ($p = 1$) o 3°M-Ex-Gauss ($p = .178$). Frente a que AB es más sesgado que BA en 1°M ($p < .001$) y 3°M_Shift-Gamma ($p = .001$).

Pasaremos ahora a analizar el efecto de “VI_TR” en el $\delta_{\mu_{Do}}$. Para analizar el comportamiento de $\delta_{\mu_{Do}}$ de los diferentes procedimientos ante las condiciones A y B de la variable “VI_TR” tendremos que especificar el patrón de diferencias (A-B) para cada nivel de “VI_θ” (ver Tabla 5). Respecto a las diferencias A-B más relevantes (*i.e.* con un nivel de TE moderado o elevado (Cohen, 1992), destacamos que se rechaza la hipótesis de igualdad de diferencias de $\delta_{\mu_{Do}}$ entre las condiciones de “VI_TR” cuando se estima por 3°M_Shift-Gamma en las condiciones AA, AB y BA. También con 3°M-Ex-Gauss se rechaza esta hipótesis nula en AA, AB y BA. Así como con 3°M_Shift-Wald en AA, y BA. Los efectos que encontramos en 2°M y 1°M son de pequeño tamaño.

Tabla 5

Efecto de VI_TR en $\delta_{\mu_{Do}}$

VI_θ	Procedimiento	gl	F	p	η_p^2	VI_θ	Procedimiento	gl	F	p	η_p^2
AA	1°M	1 258	11,96	<.001	.04	BA	1°M	1 501	188,73	<.001	.27
	2°M	1 604	9,52	<.001	.02		2°M	1 742	0,42	.52	0
	3°M-Ex-Gaussian	1 517	385,35	<.001	.43		3°M-Ex-Gaussian	1 556	1276,12	<.001	.7
	3°M_Shift-Gamma	1 709	455,01	<.001	.39		3°M_Shift-Gamma	1 613	959,81	<.001	.61
	3°M_Shift-Wald	1 546	309,60	<.001	.36	3°M_Shift-Wald	1 572	623,68	<.001	.52	
AB	1°M	1 358	6,66	.001	.02	BB	1°M	1 421	2,46	.12	.01
	2°M	1 505	24,77	<.001	.05		2°M	1 479	105,89	<.001	.18
	3°M-Ex-Gaussian	1 497	481,15	<.001	.49		3°M-Ex-Gaussian	1 419	1,47	.23	0
	3°M_Shift-Gamma	1 676	354,71	<.001	.34		3°M_Shift-Gamma	1 385	0,08	.78	0
	3°M_Shift-Wald	1 480	241,45	<.001	.33	3°M_Shift-Wald	1 423	1,36	.24	0	

En particular (ver Fig. 13), observamos que con 3°M_Shift-Gamma los $\delta_{\mu_{Do}}$ son mayores en B que en A ($p < .001$ para AA, AB y BA), mientras que con 3°M-Ex-Gauss 3 y con 3°M_Shift-Wald ocurre lo contrario. Es decir, el $\delta_{\mu_{Do}}$ es mayor en A que en B,

tanto en la condición AA ($p < .001$ para $3^{\circ}\text{M_Shift-Wald}$ y $3^{\circ}\text{M_Ex-Gauss}$, respectivamente) como en AB ($p < .001$ para $3^{\circ}\text{M_Shift-Wald}$ y $3^{\circ}\text{M_Ex-Gauss}$) como en BA ($p > .001$ para $3^{\circ}\text{M_Shift-Wald}$ y $3^{\circ}\text{M_Ex-Gauss}$). En BB mantenemos la hipótesis de igualdad de $\delta_{\mu_{Do}}$ entre A y B cuando se estima con el procedimiento $3^{\circ}\text{M_Shift-Gamma}$ ($p = .780$), $3^{\circ}\text{M_Ex-Gauss}$ ($p = .266$), $3^{\circ}\text{M_Shift-Wald}$ ($p = .244$).

Hasta el momento se han descrito los resultados de las comparaciones de condiciones de “VI $_{\theta}$ ”, “VI $_{TR}$ ”, para los $\delta_{\mu_{Do}}$ de cada procedimiento. A continuación, pasaremos a describir las comparaciones de los procedimientos entre sí (ver Tabla 6).

Tabla 6

Efecto de Procedimientos en $\delta_{\mu_{Do}}$

VI $_{TR}$	A				VI $_{\theta}$				B			
	gl	F	p	η_p^2	gl	F	p	η_p^2	gl	F	p	η_p^2
AA	4 310	223,80	<.001	.74	4 381	121,57	<.001	.56	4 381	121,57	<.001	.56
AB	4 397	173,36	<.001	.64	4 417	114,42	<.001	.52	4 417	114,42	<.001	.52
BA	4 416	893,74	<.001	.9	4 694	214,12	<.001	.55	4 694	214,12	<.001	.55
BB	4 373	1726,12	<.001	.95	4 398	1212,50	<.001	.92	4 398	1212,50	<.001	.92

En la condición AA de “VI $_{\theta}$ ” se rechaza la hipótesis nula de que todos los procedimientos tengan el mismo $\delta_{\mu_{Do}}$, tanto para las condiciones A de “VI $_{TR}$ ” como en B (ver Tabla 6). Para la condición A de “VI $_{TR}$ ”, encontramos por un lado que $3^{\circ}\text{M_Shift-Gamma}$ tiene menor $\delta_{\mu_{Do}}$ que todos los demás procedimientos ($p < .001$ para todas las comparaciones) y, por otro lado, mantenemos la hipótesis de ausencia de diferencias entre $3^{\circ}\text{M-Ex-Gauss}$ y $3^{\circ}\text{M_Shift-Wald}$ ($p = 1$), entre $3^{\circ}\text{M-Ex-Gauss}$ y 2°M ($p = 1$) y entre $3^{\circ}\text{M_Shift-Wald}$ y 2°M ($p = 1$). En B, tanto $3^{\circ}\text{M-Ex-Gauss}$ como $3^{\circ}\text{M_Shift-Wald}$ tiene menos sesgos que $3^{\circ}\text{M_Shift-Gamma}$ y 2°M ($p < .001$ para todos). Se mantiene la hipótesis nula de igualdad entre 2°M y $3^{\circ}\text{M_Shift-Gamma}$ ($p = .6$). También en B encontramos que $3^{\circ}\text{M_Shift-Wald}$ da lugar a estimaciones más sesgadas que $3^{\circ}\text{M-Ex-Gauss}$ ($p = .015$). Por último, en AA, encontramos que el 1°M procedimiento tiene mayor $\delta_{\mu_{Do}}$ que los demás procedimientos, patrón que ocurre en A ($p < .001$ para todas las comparaciones) y B ($p < .001$ para todas las comparaciones).

En la condición AB de “VI $_{\theta}$ ” también se rechaza la hipótesis nula de que todos los procedimientos tengan el mismo $\delta_{\mu_{Do}}$, tanto en las condiciones A de “VI $_{TR}$ ” como en las B (ver Tabla 6). El procedimiento 1°M vuelve a tener estimaciones más sesgadas que los demás procedimientos, tanto en A ($p < .001$ para todas las comparaciones) como en

B ($p < .001$ para todas las comparaciones). En la condición A, 3°M_ *Shift*-Gamma vuelve a ser el procedimiento con menor $\delta_{\mu_{Do}}$, ($p < .015$ para la comparación con el procedimiento 2°M y $p < .001$ para las demás comparaciones), mientras que en B, 3°M-Ex-Gauss vuelve a tener un menor $\delta_{\mu_{Do}}$ que los demás procedimientos ($p = .02$ en comparación con Wald; $p < .001$ para las demás comparaciones). Asimismo, volvemos a encontrar que en B 3°M_ *Shift*-Wald tiene menor $\delta_{\mu_{Do}}$ que 3°M_ *Shift*-Gamma, 2°M y 1°M ($p < .001$) en B y, además, mantenemos la hipótesis nula de igualdad de sesgo entre 3°M_ *Shift*-Gamma y 2°M ($p = 1$).

En la condición BA se rechaza nuevamente la hipótesis nula de que todos los procedimientos tengan el mismo $\delta_{\mu_{Do}}$ en A y en B (ver Tabla 6). Nuevamente en A, 3°M_ *Shift*-Gamma es el procedimiento con menor $\delta_{\mu_{Do}}$ ($p < .001$ para todas las comparaciones) y 2°M el segundo procedimiento con menor $\delta_{\mu_{Do}}$ ($p < .001$ para todas las comparaciones), mientras que en B, 3°M-Ex-Gauss es el procedimiento con menor $\delta_{\mu_{Do}}$ ($p < .001$ para todas las comparaciones) y 3°M_ *Shift*-Wald es el segundo procedimiento con menor $\delta_{\mu_{Do}}$ ($p < .001$ para todas las comparaciones).

En BB también se rechaza la hipótesis nula de que todos los procedimientos tengan el mismo $\delta_{\mu_{Do}}$ en A y B (ver Tabla 6). En A se mantiene la hipótesis nula de que el $\delta_{\mu_{Do}}$ estimado con el procedimiento con 3°M_ *Shift*-Gamma y 2°M es igual ($p = 1$) y, además, ambos procedimientos tienen menor sesgo que con los demás ($p < .001$ para todas las comparaciones). En B encontramos que el $\delta_{\mu_{Do}}$ estimado mediante 3°M_ *Shift*-Gamma es menor que el estimado con los demás procedimientos ($p < .001$) y que, 2°M tiene un menor $\delta_{\mu_{Do}}$ que 3°M_ *Shift*-Wald, 3°M-Ex-Gauss y 1°M ($p < .001$ para todas las comparaciones). Siendo ahora además que $\delta_{\mu_{Do}}$ es mayor con el procedimiento 3°M-Ex-Gauss que con 3°M_ *Shift*-Wald ($p = .011$).

6.3 Porcentaje de selección de los modelos 3°M.

A continuación, pasaremos a describir los resultados de la comparación y selección de los modelos 3°M mediante los índice AIC y BIC (donde las frecuencias de selección fueron idénticas para ambos índices). Los porcentajes de selección de cada modelo se presentan: a) no teniendo en cuenta la distribución generadora de datos (selección de modelo sin condicionar por la distribución generadora de datos); y b) teniéndola en cuenta (selección del modelo condicionado a la distribución generadora de datos).

Si no se tiene en cuenta la distribución generadora de datos (ver Tabla 7) encontramos que en la condición A (TR extremos), el modelo seleccionado en mayor porcentaje es Ex-Gaussian. En las condiciones A los porcentajes de elección de *Shift-Wald* y Ex-Gaussian son parecidos. En concreto encontramos un ligero mayor porcentaje de elección de Ex-Gaussian en AA BA y BB y de *Shift-Wald* en AB. Para *Shift-Wald* y *Shift-Gamma* hay una tendencia a elegirse menos estos modelos en las condiciones B que en las A.

Tabla 7

Selección de los modelos 3^oM, mediante AIC y BIC.

<i>VI_θ</i>	<i>VI_TR</i>	<i>Shift-Wald</i>	Ex-Gaussian	<i>Shift-Gamma</i>
AA	A	36,0	42,5	21,5
AB	A	40,7	39,7	19,7
BA	A	38,0	44,5	17,5
BB	A	37,5	43,3	19,2
AA	B	26,2	55,0	18,8
AB	B	24,3	54,8	20,8
BA	B	22,8	65,0	12,2
BB	B	40,7	42,5	16,8

En negrita se indican los mayores porcentajes de selección (%) por condición.

Las casillas sombreadas representan las condiciones con TR típicos (A).

Teniendo ahora en cuenta la distribución generadora de datos (ver Tabla 8), encontramos que en las condiciones A hay un mayor porcentaje de selección de aquellos modelos que asumen la distribución que genera los datos. En las condiciones B, cuando los datos generados siguen una Ex-Gaussian encontramos que para todas las condiciones de “*VI_θ*” el modelo que se selecciona en mayor porcentaje es el que asume una distribución Ex-Gaussian. En las condiciones B, cuando los datos siguen una *Shift-Gamma* el modelo seleccionado en mayor porcentaje es el que asumen *Shift-Gamma*, tanto en AA AB y BB. En BA hay poca diferencia de porcentaje de selección entre los diferentes modelos. En condiciones B de “*VI_TR*”, cuando TR son generados por una *Shift-Wald* encontramos que el modelo que asume dicha distribución solamente es elegido en mayor porcentaje en BB, siendo el modelo que asume una Ex-Gaussian el más elegido en las demás condiciones de “*VI_θ*”.

Tabla 8

Selección de los modelos 3^oM, mediante AIC y BIC, teniendo en cuenta g

VI_{θ}	VI_{TR}	g	Shift-Wald	Ex-Gaussian	Shift-Gamma
AA	A	Shift-Wald	66,0	33,0	1,0
AB	A	Shift-Wald	72,0	27,5	0,5
BA	A	Shift-Wald	65,0	35,0	0,0
BB	A	Shift-Wald	69,5	30,5	0,0
AA	B	Shift-Wald	31,5	67,5	1,0
AB	B	Shift-Wald	28,0	71,5	0,5
BA	B	Shift-Wald	16,0	83,5	0,5
BB	B	Shift-Wald	69,0	31,0	0,0
AA	A	Ex-Gaussian	17,0	78,5	4,5
AB	A	Ex-Gaussian	23,5	74,0	2,5
BA	A	Ex-Gaussian	14,0	85,0	1,0
BB	A	Ex-Gaussian	16,5	83,5	0,0
AA	B	Ex-Gaussian	19,0	75,5	5,5
AB	B	Ex-Gaussian	23,0	70,0	7,0
BA	B	Ex-Gaussian	18,0	79,5	2,5
BB	B	Ex-Gaussian	15,0	84,5	0,5
AA	A	Shift-Gamma	25,0	16,0	59,0
AB	A	Shift-Gamma	26,5	17,5	56,0
BA	A	Shift-Gamma	34,0	13,5	52,5
BB	A	Shift-Gamma	26,5	16,5	57,0
AA	B	Shift-Gamma	28,0	22,0	50,0
AB	B	Shift-Gamma	22,0	23,0	55,0
BA	B	Shift-Gamma	34,5	32,0	33,5
BB	B	Shift-Gamma	38,0	12,0	50,0

En negrita se indican los mayores porcentajes de selección (%) por condición. Las casillas sombreadas representan las condiciones con TR típicos (A).

7 Discusión

7.1 Sesgo en las probabilidades de los procesos (\hat{D} y \hat{G}).

En primer lugar, desde el análisis de los efectos simples de “ VI_{θ} ” encontramos un patrón de diferencias antagónico para el sesgo de las estimaciones de las probabilidades de los procesos *detección* (\hat{D}) y *guessing* (\hat{G}) que se repite en todos los procedimientos de estimación. Este patrón muestra que las \hat{D} más sesgadas se dan cuando el valor verdadero de d es próximo a 1 y 0 (*i.e.* AA y AB), mientras que las dos condiciones donde los \hat{G} verdaderos son extremos (AA y BA) son precisamente aquellas donde el sesgo de \hat{G} es mayor. Los resultados concuerdan con nuestra primera hipótesis, y podría explicarse por el hecho de que las probabilidades extremas afectan al número de respuestas por rama, y, por lo tanto, a la estimación de todos los parámetros pertenecientes a esta.

Considerando lo dicho en el párrafo anterior, hay que matizar que en el caso de \hat{G} la condición más sesgada es BA, no AA. Consideramos que esto se podría explicar por un pequeño *trade-off* (relación negativa) entre los sesgos, ya que encontramos r negativas significativas, aunque con un TE reducido (nótese que esta pequeña correlación podría no representar la magnitud de relación entre las dos probabilidades por un problema de restricción de rango). Las estimaciones de \hat{D} no se verían tan afectadas por dicho *trade-off* ya que el parámetro G aparece en 4 ramas en las que siempre hay al menos un parámetro D , mientras que d aparece en 6 ramas en las que hay dos que no incluyen el parámetro G . Así pues, las \hat{G} se verían más afectadas por las D extremas que las \hat{D} por las G extremas. El hecho de que la correlación entre ambos sesgos sea negativa se debe a que los procesos de *guessing* se dan precisamente cuando no hay *detección*.

Respecto a los \hat{D} y de \hat{G} en cada condición VI_θ , queda por contestar qué procedimiento da lugar estimaciones menos sesgadas. Para δ_D únicamente hay diferencias entre procesos en la condición BA, sin embargo, teniendo en cuenta el pequeño TE ($\eta_p^2 = .004$) de estas diferencias y el elevado número de muestra, podrían ser meras variaciones espurias. Y, siguiendo la misma línea de razonamiento, las diferencias entre los procedimientos en el δ_G ($\eta_p^2 = .003$) también parecen ser irrelevantes. Por lo que en lo relativo a los sesgos de \hat{D} y de \hat{G} no parece haber ningún procedimiento preferible.

7.2 Sesgo de las medias de los TR ($\hat{\mu}_{D0}$, $\hat{\mu}_{Dn}$ y de $\hat{\mu}_G$)

Una vez discutido cómo las probabilidades extremas de los procesos afectan al sesgo de los $\hat{\theta}$ pasaremos a evaluar su afectación a las medias de los TR, teniendo en cuenta los procesos de estimación utilizados.

En primer lugar, como ya se expuso anteriormente, el patrón de diferencias entre las condiciones de VI_θ en el sesgo de $\hat{\mu}_{D0}$ es diferente según el procedimiento de estimación empleado. Tal y como sucede con las $\hat{\theta}$, para $\hat{\mu}_{D0}$ encontramos que los TR medios estimados por los procedimientos 1ºM, 2ºM y 3ºM_Shift_Gamma son compatibles con nuestra primera hipótesis ya que tendemos a encontrar un mayor sesgo con θ extremos. Sin embargo, tanto para 3ºM_Ex-Gauss y 3ºM_Shift-Wald, sucede lo opuesto, ya que en BB (θ no-extremos) hay un menor sesgo que en todas las demás condiciones de VI_θ . Así pues, contrariamente a lo que esperábamos, los procedimientos 3ºM_Ex-Gauss y 3ºM_Shift-Wald parecen beneficiarse de la existencia de θ extremos. La razón por la cual

se da este último patrón no está del todo clara; no obstante, como veremos posteriormente, para entenderlo mejor habría que considerar cómo funciona cada procedimiento y el nivel de condición de la variable “ VI_{TR} ”. Por lo que antes de comentar este inesperado patrón pasaremos a estudiar qué efecto tienen los TR extremos (B) y típicos (A) en el sesgo de \hat{D} , \hat{G} y $\hat{\mu}_{Do}$.

Sobre el efecto de “ VI_{TR} ” en el sesgo, constatamos que mediante la razón de verosimilitudes entre los modelos saturados y los restringidos que analizan δ_D y δ_G , no hallamos pérdida de ajuste al excluir el factor “ VI_{TR} ”. Así pues, la variable “ VI_{TR} ” no parece afectar al sesgo de las probabilidades de los procesos. Sin embargo, sí encontramos un efecto de “ VI_{TR} ” en el sesgo de $\hat{\mu}_{Do}$.

En particular encontramos que únicamente hay diferencias A-B (ver Fig. 13), con TE moderados o elevados, en los procedimientos 3ºM. Cuando se asume una *Shift-Gamma*, en las condiciones AA, AB y BA, el sesgo de $\hat{\mu}_{Do}$ es mayor cuando los TR son extremos (B) que cuando son típicos (A), pero si asumimos una Ex-Gaussian o una *Shift-Wald* encontramos que el sesgo de $\hat{\mu}_{Do}$ es mayor cuando los TR son típicos. Por otro lado, comparando los procedimientos entre sí encontramos resultados compatibles con nuestras hipótesis 4 y 2. Si los TR son típicos (A) y los parámetros θ extremos (AA, AB y BA), las estimaciones menos sesgadas de $\hat{\mu}_{Do}$ se dan con 3ºM_*Shift-Gamma*. Además, cuando los TR son típicos (A) y los θ no son extremos 3ºM_*Shift-Gamma* y 2ºM dan las estimaciones menos sesgadas. Lo cual indica que, tal y como esperábamos 2ºM sólo funciona bien en condiciones donde no hay θ extremos (hipótesis 2). Respecto a la hipótesis 4, observamos que, si los TR son extremos (B) y también lo son los parámetros θ (AA, AB y BA), el procedimiento con menos sesgo de $\hat{\mu}_{Do}$ es 3ºM_Ex-Gauss. No obstante, en discordancia con esta última hipótesis, la ventaja de asumir una distribución Ex-Gauss ante TR extremos positivos (B) se pierde cuando los parámetros θ no son extremos (BB), ya que encontramos que 3ºM_*Shift-Gamma* aporta los $\hat{\mu}_{Do}$ menos sesgados.

Como hemos mencionado antes, para explicar estos resultados tenemos que atenernos a las particularidades de cada procedimiento. Pasaremos ahora a describir algunas de ellas.

En primer lugar, como vimos anteriormente, las condiciones con θ extremas afectarían al sesgo de las $\hat{\theta}$, por lo que el sesgo en \hat{D} podría provocar un sesgo en $\hat{\mu}_{Do}$. Es decir,

aunque exista una independencia entre los parámetros podría existir una relación entre sus estimadores. Respaldo esta explicación, encontramos evidencias de que la correlación entre δ_d y $\delta_{\mu_{Do}}$ es especialmente pequeña en los procedimientos 3°M ($r = .017$; $p = .25$ para 3°M_Ex-Gauss; $r = .019$; $p = .198$; para 3°M_Shift-Gamma; $r = .028$; $p = .5$; para 3°M_Shift-Wald) en comparación con las obtenidas en 1°M y 2°M (ver Apéndice B). De esta forma, si en 3°M el efecto del sesgo en \hat{D} en $\hat{\mu}_{Do}$ es menor que en los demás procedimientos, cuando los θ son extremos el modelo menos sesgado es siempre uno de los 3°M. No obstante, aunque lo anterior fuese cierto, seguiríamos desconociendo por qué la relación entre δ_d y $\delta_{\mu_{Do}}$ es menor en 3°M.

En segundo lugar, en 1°M encontramos que el proceso de estimación no siempre obtiene resultados en las condiciones con parámetros θ extremos (un 32,92% de los modelos no se estiman en AA, un 17,58% en AB un 17,5% en BA y 0% en BB; no habiendo diferencias entre A y B mayores al 0,02%). Dado que dicho procedimiento estima las medias de los TR latentes desde la media de los TR observados, cuando hay θ extremos la media de los TR observados se hará en base a muy pocas o incluso ninguna respuesta de una o más ramas del modelo. En 2°M las probabilidades extremas afectan también al número de elementos de TR que encontramos por *bin* lo cual podría dar lugar a estimaciones sesgadas. En definitiva, ambos métodos trabajan con las medias, 1°M parte de las medias de los TR observados y 2°M divide en rápidas o lentas las categorías de respuesta según estas sean mayores o menores que la media de los TR observados.

El hecho de partir de la media observada, especialmente teniendo pocas observaciones en algunas ramas (como sucede cuando hay θ extremos), es preocupante en la modelización de los TR. Las distribuciones empíricas de TR suelen ser asimétricas positivas, por lo que cualquier valor extremo tendrá mucha influencia en la media muestra. Por lo contrario, los procedimientos 3°M parten del conjunto de TR observados y estiman conjuntamente la distribución de los TR de las ramas y la probabilidad asociada a esta, sin recurrir a la media muestral de los TR. Por esa razón, pese a que existan menos observaciones de alguna de las ramas cuando hay θ extremos, al no recategorizar las respuestas en *bins* basados en la media, posiblemente estos procedimientos no se ven tan afectados como el 2°M de que alguna categoría tenga pocas observaciones. Además, al no resumir el conjunto de los TR observados en su media, como con 1°M, asegura una mejor representación de la distribución de los TR subyacentes. Así pues, nuestras hipótesis 2 y 3 son congruentes con el hecho que el procedimiento 1°M funcione peor

cuando las probabilidades θ son extremas, especialmente si también son extremos los TR. E igualmente, estas dos hipótesis son compatibles con el hecho de que solamente cuando no hay θ ni TR extremos el procedimiento 2ºM es uno de los procedimientos menos sesgados (juntamente con 3ºM_*Shift*-Gamma).

Con el objeto de comparar los modelos 3ºM entre sí, debemos tener en cuenta que los resultados encontrados no evalúan realmente el sesgo de la distribución de los TR por rama, sino su media (información que no podemos obtener desde 1ºM y 2ºM). Por lo que, dependiendo de la distribución que asumamos, pueden ser más o menos probables los valores entorno a la media. De hecho, si tenemos en cuenta la selección de modelos, contrariamente a lo indicado en la hipótesis 5, no encontraremos el mismo patrón analizando el sesgo medio de los TR y el ajuste del conjunto de los TR. Por consiguiente, pasaremos a relacionar los resultados hasta ahora expuestos con los obtenidos en la comparación y selección de modelos.

7.3 Porcentaje de selección de los modelos 3ºM.

Por las razones que acabamos de exponer, atenderemos ahora a los porcentajes en los que se elige cada modelo 3ºM (ver Tabla 7). En primer lugar, observamos que, efectivamente, hay una tendencia a seleccionar menos los modelos que asumen una *Shift*-Gamma cuando los TR son extremos frente a cuando son típicos. Pero en oposición a lo que ocurre al analizar el sesgo de $\hat{\mu}_{Do}$, cuando hay TR típicos se seleccionan los modelos que asumen una Ex-Gaussian y *Shift*-Wald en un porcentaje superior a *Shift*-Gamma. En segundo lugar, al estudiar el sesgo de $\hat{\mu}_{Do}$ antes veíamos que las estimaciones de 3ºM_*Shift*-Wald eran más sesgadas en condiciones con TR típicos (A) que con TR extremos (B), pero al evaluar el ajuste global, observamos que se selecciona menos *Shift*-Wald en las condiciones B que en las condiciones A. Es decir, si evaluamos la media de los TR, *Shift*-Wald funciona mejor con TR extremos, mientras que si evaluamos el ajuste del conjunto de TR funciona mejor con TR típicos. Lo último podría deberse a que *Shift*-Wald se ajusta a TR extremos aumentando el parámetro λ (Matzke y Wagenmakers, 2009), que provoca un alargamiento de la cola de la distribución *Shift*-Wald, pero, simultáneamente, hay un aumento del apuntamiento en zonas cercanas a la media, disminuyendo así el sesgo de $\hat{\mu}_{Do}$ (ver Apéndice A para una comparación visual de las *f.d.p* en las condiciones A y B). Curiosamente la tendencia de las demás distribuciones es exactamente la contraria. Por ejemplo, la distribución *Shift*-Gamma disminuye sustancialmente el apuntamiento entorno a su media cuando alarga su cola, explicando

así el aumento del sesgo de $\hat{\mu}_{Do}$ en condiciones con TR extremos. Mientras que en la condición con TR típicos tiene un apuntamiento entorno a la media superior a las demás distribuciones, explicando porqué es el procedimiento con menor sesgo de $\hat{\mu}_{Do}$.

Así pues, para elegir qué distribución vamos a asumir deberemos tener en cuenta qué queremos evaluar y cómo se distribuyen los datos. Si estamos en un escenario con TR típicos y nos interesa evaluar $\hat{\mu}_{Do}$ encontramos que resulta preferible asumir una *Shift-Gamma*, pero si evaluamos el ajuste del conjunto de los TR elegiríamos una *Ex-Gaussian* o una *Shift-Wald*. Siguiendo con la comparativa de resultados, hay una convergencia de evidencias acerca de la ventaja en modelar asumiendo una *Ex-Gaussian* cuando los TR son extremos, con la excepción, nuevamente, de la condición BB (i.e θ no extremo), reafirmando que la ventaja de asumir una *Ex-Gaussian* en condiciones donde los TR son extremos solo existe cuando también lo es algún parámetro θ .

Por último, para analizar el posible sobreajuste de algunas de estas asunciones distribucionales, conviene aquí atender a la Tabla 8, donde se evalúa la selección de modelos teniendo en cuenta la distribución generadora de TR. Específicamente cuando la distribución generadora es una *Shift-Gamma* y *Ex-Gaussian*, encontramos que hay una tendencia a seleccionarse correctamente la distribución mediante los índices BIC y AIC. No obstante, en contra de lo esperado, cuando los TR siguen una *Shift-Wald* y nos hallamos en condiciones con TR extremos (B), hay una tendencia a seleccionar más a los modelos que asumen una *Ex-Gaussian* que los que asumen una *Shift-Wald*. Así pues, en escenarios de TR extremos generados por *Shift-Wald* podría darse un sobreajuste de *Ex-Gaussian*. Esto posiblemente se deba a que no se está contemplando la complejidad de la *Ex-Gaussian*, aunque tenga el mismo número que parámetros que las demás distribuciones (Myung, Pitt, Zhang y Balasubramanian, 2001). La razón por la cual el sobreajuste no ocurre en condiciones con θ no extremos (BB), no está clara. No obstante, se ha encontrado en un estudio (Van Zandt, 2000) que compara el nivel de precisión y sesgo del modelado de TR con distribuciones Gamma, *Ex-Gaussian* y Wald, entre otras, que ante muestras pequeñas la distribución Wald da lugar a un mayor sesgo e impresión en la estimación de sus parámetros que las demás. Algo que, en coherencia con nuestros datos, podría explicar por qué únicamente cuando hay un elevado número de observaciones de todas ramas (BB) se selecciona correctamente la distribución Wald como la generadora de los TR.

8 Conclusiones

Los MPT-DC se caracterizan por permitir estudiar la viabilidad de un modelo que asume unos procesos latentes particulares, el orden en el que surgen, cuánto duran, su recorrido y muchas otras cuestiones relativas a la estructura y al funcionamiento de los mecanismos psicológicos. Dado el potencial uso que tienen estos modelos en diversos campos de la psicología, y vista la recencia de los procedimientos MPT-DC, consideramos que existe una necesidad de establecer un consenso sobre el procedimiento de modelado más adecuado para cada escenario experimental, siendo esta nuestra ambición principal.

A propósito del objetivo que nos incumbe, sacamos conclusiones tales como la necesidad de un control experimental que permita evitar en la medida de lo posible que las probabilidades de los procesos sean extremas, ya que, con independencia del procedimiento empleado, estas afectan negativamente a la estimación de las θ . Ejemplo de tales controles, aplicables al 2HTM, serían: la evitación de instrucciones o claves que pudiesen penalizar o recompensar determinadas categorías de respuesta, puesto que podrían afectar a una tendencia de *guessing* hacia dicha categoría; el control de la dificultad de la tarea (*e.g.* controlar el número de ítems antiguos que recordar (señal) presentes en la primera fase del experimento, y el número de ítems nuevos (ruido) presentados en la segunda); etc. Sin embargo el control experimental no siempre será posible o de interés (*e.g.* la dificultad de la tarea puede ser una manipulación experimental necesaria para poner a prueba hipótesis). Así pues, consideramos que la validez del MPT dependerá de la combinación entre el modelo teórico de partida, de las condiciones experimentales, y de la población a la que se aplica (*e.g.* puede haber población experta con elevada probabilidad de *detección*). A pesar de ello, debido a que no hay diferencias entre procedimientos en las θ estimadas, pero sí en TR, concluimos que el 3^oM resultó ser el mejor procedimiento.

Sin embargo, dependiendo de si el objetivo experimental es evaluar la media de los TR o el ajuste del conjunto de TR, si las probabilidades de los procesos y/o los TR son extremos o no, posiblemente habría que optar por asumir una u otra distribución. Algo a destacar respecto a cuando el interés reside en evaluar tendencias centrales de los TR, es que incluso en estos casos el modelo 1^oM (que evalúa directamente la media) no es el que obtiene mejores predicciones. Por último, en situaciones con TR extremos hemos constatado un sobreajuste de los datos mediante la Ex-Gauss por lo que recomendaríamos que, al comparar diferentes asunciones distribucionales, habría que penalizar no sólo el

número de parámetros k , sino otras dimensiones de la complejidad del modelo (Myung et al., 2001).

Como limitación a la generalizabilidad de nuestras conclusiones debemos admitir que la variable continua estudiada en el MPT-DC fueron los TR. Partiendo del conocimiento previo que existen de cómo se distribuyen es factible asumir distribuciones asimétricas y positivas en los procedimientos 3^oM. No obstante, en caso de no tener fuertes indicios acerca de la distribución de la variable continua quizás 2^oM fuese más ventajoso que 3^oM. Asimismo, no podemos generalizar nuestros resultados a otros modelos ya que no partimos de un modelo abstracto, sino del 2HTM. Otra limitación, que podría afectar a la validez de nuestras conclusiones, se relaciona con la imposibilidad de una medida de imprecisión y sesgo común para todos los procedimientos, razón por la cuál se estudió el sesgo mediante la media de los TR estimados. Tampoco se estudió la proporción de veces que un parámetro entra dentro de su IC95% porque cuando la distribución asumida y la generadora de datos no coincide, tampoco lo hacen los parámetros η . Igualmente, queda pendiente agregar índices para la selección de modelos (*e.g.* como MDL, BMS, entre otros), que penalicen el grado de complejidad en base a la forma en que se combinan los parámetros en la ecuación del modelo (Myung et al., 2001; Grunwald, 2004; Zeugner, 2011), cuestión que, aun estando relacionado con las hipótesis del trabajo, excede el objetivo final de éste. Asimismo sería interesante indagar sobre la efectividad de una modificación de los procedimientos 1^oM y 2^oM que parta de la mediana de los TR observados en vez de sus medias. Por un lado, posiblemente la mediana fuese más representativa de la distribución de los TR de los sujetos y, por otro lado, tendría la ventaja de que no hubiesen *bins* con pocas observaciones en casos de θ extremos. Asimismo, una ventaja adicional de utilizar la mediana como punto de corte entre *bins* sería la de obtener una misma interpretación entre sujetos.

Dentro de las limitaciones generales de los MPT-DC destacamos el problema de la identificabilidad de los parámetros. Aunque la mayor parte de los autores propone realizar restricciones paramétricas, aquí defendemos una solución desde el diseño experimental. Bajo esta última perspectiva, se pueden proponer nuevas condiciones experimentales (*i.e.* un nuevo árbol con otras categorías de respuestas) que incluye un número mínimo de nuevos parámetros a la vez que mantiene el máximo de parámetros posibles de los otros árboles. Otra de las limitaciones generales de los modelos MPT-DC que debemos destacar es que se han asumido supuestos cuestionables para los cuales, desde nuestro

conocimiento, no hay propuesta de modelado ante su incumplimiento (*i.e.* serialidad e independencia).

Como futuras líneas de investigación, queda la propuesta de proponer MPT-DC que permitan tener en cuenta otros aspectos, tales como: la variación de los parámetros a lo largo de los ensayos, la utilización de covariables, etc. Como ya se mencionó, los MPT-DC, son un caso de *finite mixture model*. Si bien es cierto que autores que han desarrollado los MPT-DC lo mencionan, hasta donde se ha revisado no se han encontrado trabajos que traten de aplicar los procedimientos asociados a los finite mixture models a los MPT-DC. Por otra parte, queda todavía en vías de investigación si ante un mismo paradigma experimental todos los sujetos comparten un mismo modelo o si hay subpoblaciones con arquitecturas cognitivas representadas por modelos distintos. Así pues, cabe preguntarnos: ¿Qué consecuencias tendría que los sujetos satisficieran diferentes supuestos distribucionales y, sin embargo, asumiésemos que son los mismos?; ¿Qué indicios nos permitirían sospechar de la existencia de dos o más modelos en una misma población?; y ¿Cuál es la mejor manera para discriminar a los sujetos en función de los modelos que siguen?.

En definitiva, no está del todo claro qué procedimiento resulta más eficaz, puesto que, como hemos intentado demostrar en el presente trabajo, depende de nuestro modelo teórico, la población a la que se aplica y la distribución de los datos. Pero además queda pendiente mejorar los procedimientos ya existentes, explorar nuevas aplicaciones y hacerlos más accesibles a la comunidad científica.

Referencias

- Ashby, F. G. (Ed.). (2014). *Multidimensional models of perception and cognition*. Psychology Press.
- Ashby, F. G., Prinzmetal, W., Ivry, R., y Maddox, W. T. (1996). A formal theory of feature binding in object perception. *Psychological review*, 103(1), 165.
- Atkinson, R. C., y Juola, J. F. (1973). Factors influencing speed and accuracy of word recognition. In S. Kornblum (Ed.), *Attention and Performance IV*. New York: Academic Press.
- Atkinson, R. C., y Juola, J. F. (1974). Search and decision processes in recognition memory. In D. Krantz, R. Atkinson, R. Luce, and P. Suppes (Eds.). *Contemporary Developments in Mathematical Psychology*, Vol. 1
- Atkinson, R. C., y Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes.
- Bamber, D., y Van Santen, J. P. (1985). How many parameters can a model have and still be testable?. *Journal of Mathematical Psychology*, 29(4), 443-473.
- Batchelder, W. H., y Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review*, 97, 548–564.
- Bröder, A. y Schütz, J. (2009). Recognition ROCs are curvilinear – or are they? On premature arguments against the two-high-threshold model of recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 587-606.
- Cavagnaro, D. R., Myung, J. I., y Pitt, M. a. (2010). Mathematical Modeling. In *The Oxford Handbook of Quantitative Methods in Psychology*, Vol. 1 (pp. 437–452).
- Chechile, R. A. (2009). Pooling data versus averaging model fits for some prototypical multinomial processing tree models. *Journal of Mathematical Psychology*, 53(6), 562-576.
- Cohen, J. (1992). A power primer. *Psychological bulletin*, 112(1), 155.
- Cousineau, D., Brown, S. y Heathcote, A. Fitting distributions using maximum likelihood: Methods and packages. *Behavior Research Methods, Instruments, & Computers* 36, 742–756 (2004). <https://doi.org/10.3758/BF03206555>
- Erdfelder, E., Auer, T. S., Hilbig, B. E., Aßfalg, A., Moshagen, M., y Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Zeitschrift für Psychologie/Journal of Psychology*, 217(3), 108-124.

- Grunwald, P. (2004). A tutorial introduction to the minimum description length principle. *arXiv preprint math/0406077*.
- Heck*, D. W., Arnold*, N. R., y Arnold, D. (2018). TreeBUGS: An R package for hierarchical multinomial-processing-tree modeling. *Behavior Research Methods*, 50, 264-284. <https://doi.org/10.3758/s13428-017-0869-7>
- Heck, D. W., Erdfelder, E., y Kieslich, P. J. (2018). Generalized processing tree models: Jointly modeling discrete and continuous variables. *Psychometrika*, 83(4), 893-918.
- Heck, D. W., y Erdfelder, E. (2016). Extending multinomial processing tree models to measure the relative speed of cognitive processes. *Psychonomic Bulletin & Review*, 23(5), 1440-1465.
- Horowitz, T. S., y Wolfe, J. M. (2005). Visual Search: The role of memory for rejected distractors.
- Hu, X. (2001). Extending general processing tree models to analyze reaction time experiments. *Journal of Mathematical Psychology*, 45(4), 603-634.
- Hu, X., y Batchelder, W. H. (1994). The statistical analysis of general processing tree models with the EM algorithm. *Psychometrika*, 59(1), 21-47.
- Humphreys, G. W., y Bruce, V. (1989). *Visual cognition: Computational, experimental and neuropsychological perspectives*. Psychology Press.
- Juola, J. F., Fischler, I., Wood, C. T., y Atkinson, R. C. (1971). Recognition time for information stored in long-term memory. *Perception & Psychophysics*, 10, 8-14.
- Klauer, K. C., y Kellen, D. (2018). RT-MPTs: Process models for response-time distributions based on multinomial processing trees with applications to recognition memory. *Journal of Mathematical Psychology*, 82, 111-130.
- Lamberts, K., y Goldstone, R. (Eds.). (2004). *Handbook of cognition*. Sage.
- Link, S. W. (1982). Correcting response measures for *Guessing* and partial information. *Psychological Bulletin*, 92(2), 469.
- Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization* (No. 8). Oxford University Press on Demand.
- Luce, R. D. (2012). *Individual choice behavior: A theoretical analysis*. Courier Corporation.
- Marewski, J. N., y Olsson, H. (2009). Beyond the null ritual: Formal modeling of psychological processes. *Zeitschrift für Psychologie/Journal of Psychology*, 217(1), 49-60.

- McGill, W. J., y Gibbon, J. (1965). The general-gamma distribution and reaction times. *Journal of mathematical Psychology*, 2(1), 1-18.
- McLachlan, G. J., y Peel, D. (2004). *Finite mixture models*. John Wiley & Sons.
- Myung, I. J., Pitt, M. A., y Kim, W. (2005). Model evaluation, testing and selection. *Handbook of cognition*, 422-436.
- Myung, I. J., Pitt, M. A., Zhang, S., & Balasubramanian, V. (2001). The use of MDL to select among computational models of cognition. In *Advances in Neural Information Processing Systems* (pp. 38-44).
- Palmer, E. M., Horowitz, T. S., Torralba, A., y Wolfe, J. M. (2011). What are the shapes of response time distributions in visual search?. *Journal of experimental psychology: human perception and performance*, 37(1), 58.
- Riefer, D. M., y Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, 95(3), 318.
- Schweickert, R., y Zheng, X. (2019). Tree Inference: Response time in Multinomial Processing Trees, representation and uniqueness of parameters. *Journal of Mathematical Psychology*, 91, 51-69.
- Singmann, H., y Kellen, D. (2013). MPTinR: Analysis of multinomial processing tree models in R. *Behavior Research Methods*, 45(2), 560-575.
- Team, R. C. (2019). R: a language and environment for statistical computing, version 3.0. 2. Vienna, Austria: R Foundation for Statistical Computing; 2013.
- Townsend, J. T., y Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. CUP Archive.
- Wickelmaier F (2011). "Multinomial processing tree models in R." Presented at the R User Conference 2011, August 16-18, Coventry, UK.
- Zeugner, S. (2011). Bayesian model averaging with BMS. *Tutorial to the R-package BMS 1e30*.

Apéndice A

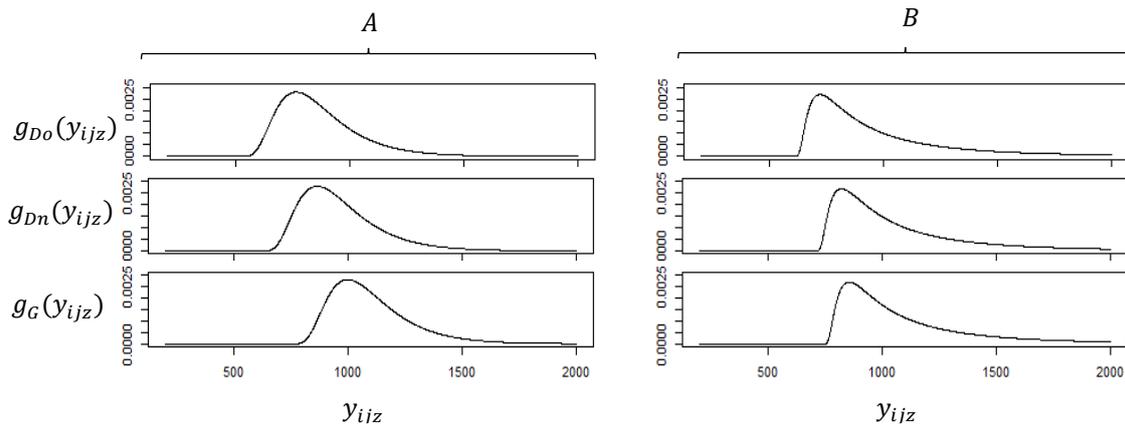


Figura A.1. Función de densidad de los componentes g_{Do} (primera fila), g_{Dn} (segunda fila) y g_{Gn}/g_{Go} (tercera fila) en condición A (columna izquierda) y en la condición B (columna derecha) bajo una de Shift-Wald.

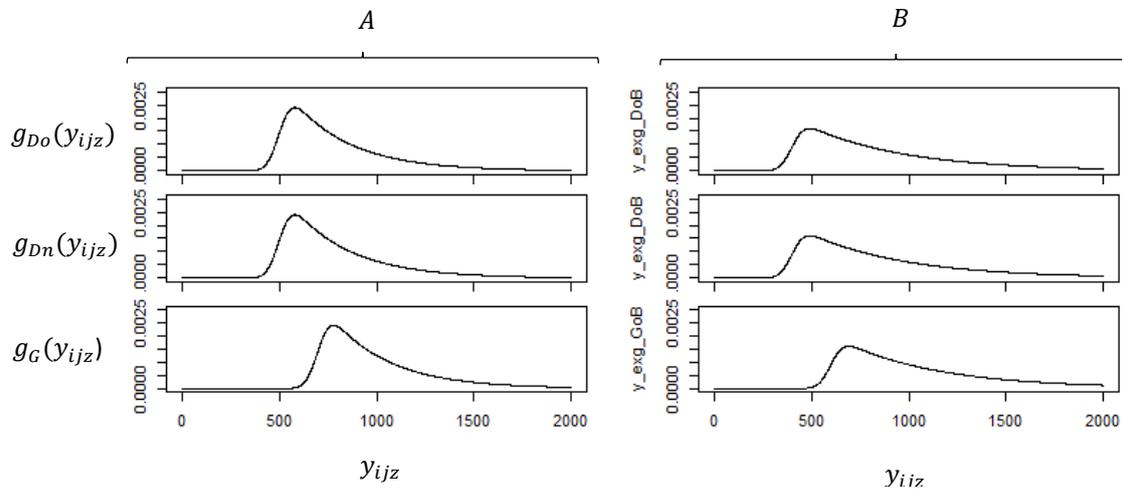


Figura A.2. Función de densidad de los componentes g_{Do} (primera fila), g_{Dn} (segunda fila) y g_{Go}/g_{Gn} (tercera fila) en condición A (columna izquierda) y en la condición B (columna derecha) bajo una de Ex-Gaussian.

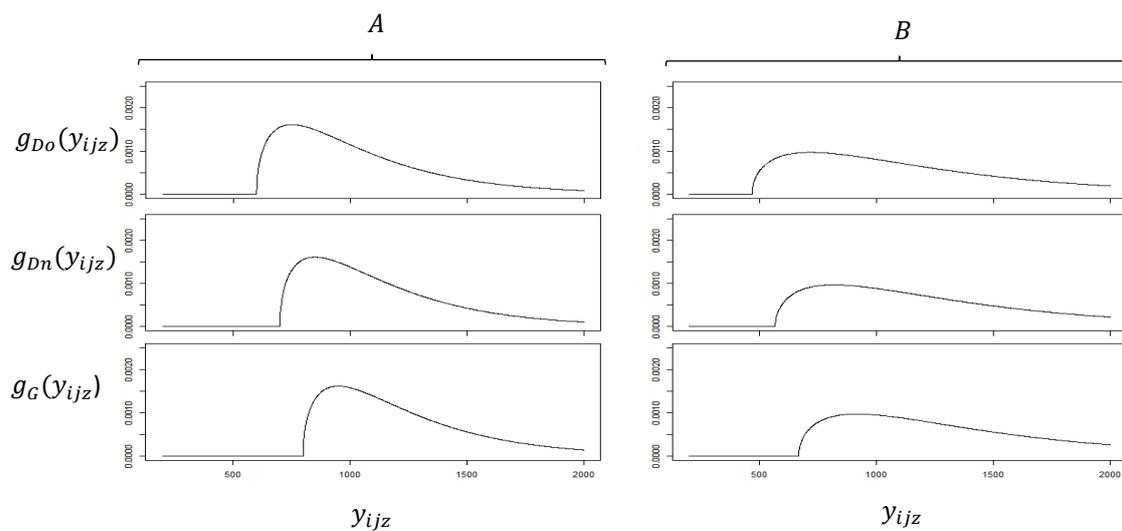


Figura A.3. Función de densidad de los componentes g_{Do} (primera fila), g_{Dn} (segunda fila) y g_{Go}/g_{Gn} (tercera fila) en condición A (columna izquierda) y en la condición B (columna derecha) bajo una de Shift-Gamma

Apéndice B

Procedimientos	VDs	1. \hat{D}	2. \hat{G}	4. $\hat{\mu}_{Dn}$	4. $\hat{\mu}_{Dn}$
1°M	2. \hat{G}	-,103**			
	3. $\hat{\mu}_{Do}$,198**	,035*		
	4. $\hat{\mu}_{Dn}$,182**	,007	,326**	
	5. $\hat{\mu}_G$	-,028	,012	,256**	,308**
2°M	2. \hat{G}	-,084**			
	3. $\hat{\mu}_{Do}$,259**	-,019		
	4. $\hat{\mu}_{Dn}$,259**	-,02	,999**	
	5. $\hat{\mu}_G$,256**	-,019	,999**	,999**
3°M_Ex-Gaussian	2. \hat{G}	-,073**			
	3. $\hat{\mu}_{Do}$,029*	-,011		
	4. $\hat{\mu}_{Dn}$,017	-,013	,935**	
	5. $\hat{\mu}_G$	-,012	,014	,893**	,895**
3°M_Shift-Gamma	2. \hat{G}	-,071**			
	3. $\hat{\mu}_{Do}$,019	,026		
	4. $\hat{\mu}_{Dn}$	-,006	,031*	,882**	
	5. $\hat{\mu}_G$	-,031*	,036*	,785**	,897**
3°M_Shift-Wald	2. \hat{G}	-,069**			
	3. $\hat{\mu}_{Do}$,028*	-,011		
	4. $\hat{\mu}_{Dn}$,013	-,01	,959**	
	5. $\hat{\mu}_G$	-,016	-,012	,945**	,947**

* indica $p < ,05$; ** indica $p < ,001$