

## La probabilidad condicional y la probabilidad conjunta en la resolución de problemas de probabilidad.

M. Pedro Huerta<sup>1</sup>, Universitat de València (España)

Joaquín Arnau<sup>2</sup>, Colegio Pío XII (Valencia, España)

Recibido el 26 de Octubre de 2017; aceptado el 10 de febrero de 2017

---

**La probabilidad condicional y la probabilidad conjunta en la resolución de problemas de probabilidad condicional.**

### **Resumen**

*En este artículo se estudian las relaciones entre las probabilidades condicionales y conjuntas en el proceso de resolución de problemas escolares, desde una perspectiva educativa antes que cognitiva. Primero mostramos cómo los problemas básicos de probabilidad condicional presentan muchas dificultades a una amplia muestra de resolutores, desde la educación secundaria a maestros y profesores de matemáticas en formación. Apuntamos hacia las complejas relaciones entre las distintas probabilidades que participan en el proceso de resolución como una de las principales causas de dichas dificultades y cómo éstas dependen de las variables de la tarea asociadas a los problemas, tales como su estructura matemática, el contexto en el que los problemas se formulan y el vocabulario usado para expresar dichas probabilidades en cada contexto. Mostramos también cómo algunos estudiantes, cuando resuelven problemas de probabilidad condicional, no tratan con sentido las probabilidades condicionales sino que éste depende de los recursos o sistemas de representación que usan para la resolución del problema.*

**Palabras clave.** Probabilidad condicional, probabilidad conjunta, problemas de probabilidad condicional, resolución de problemas.

**A probabilidade condicional e conjunta na resolução de problemas de probabilidade condicional.**

### **Resumo**

*Este artigo explora a relação entre probabilidades condicionais e conjuntas no processo de resolução de problemas de escola, de uma perspectiva educacional. Primeiro mostramos como base problemas em probabilidade condicional apresenta muitas dificuldades para uma grande amostra de estudantes, desde os estudantes do ensino secundário a professores de matemática em formação. Nosso objectivo é em direcção as complexas relações entre as diferentes probabilidades envolvidas no processo de resolução como uma das principais causas de tais dificuldades e como estes dependem de variáveis da tarefa associada os problemas, tais como sua estrutura matemática, o contexto em que os problemas são formulados e o vocabulário usado para expressar essas probabilidades em cada contexto. Também mostramos como alguns alunos não conseguem usar com sentido as probabilidades condicionais, mas isso depende de recursos ou sistemas de representação utilizados para a resolução do problema.*

**Palavras chave.** Probabilidade condicional, probabilidade conjunta, problemas de probabilidade condicional, resolução de problemas.

Para citar: Huerta, M.P. & Arnau, J. (2017). La probabilidad condicional y la probabilidad conjunta en la resolución de problemas de probabilidad. *Avances de Investigación en Educación Matemática*, 11, 87 – 106.

### **Conditional and joint probabilities in conditional probability problem solving.**

#### **Abstract**

*In this piece of work we report a research from an educational point of view, whose main objective is to study the relationships between conditional and joint probabilities in the process of solving school problems. We first show how basic conditional probability problems present many difficulties to a broader sample of solvers, from compulsory and non compulsory secondary school (15- 18 years old) to future primary and secondary teachers of mathematics. We point out to the complex relationships among absolute, joint and conditional probabilities as one of the main causes of problems difficulties, and as a consequence, they depend on task variables of problems such as their mathematical structure, context in which problems are put forward and the wording used to express these probabilities. We also show how some students, while solving problems deal with conditional probabilities, not in a proper sense, but in such a way that depends on the resources or representation systems they use for solving the problem.*

**Key words.** Conditional probability, joint probability, conditional probability problems, problem solving.

### **La probabilité conditionnelle et la probabilité conjointe dans la résolution des problèmes de probabilité conditionnelle.**

#### **Résumé**

*Cet article explore la relation entre les probabilités conjointes et conditionnelles dans le processus de résolution de problèmes scolaires, du point de vue éducatif. Nous montrons tout d'abord comment des problèmes basiques en probabilité conditionnelle présentent de nombreux difficultés à un large échantillon d'étudiants, depuis les enseignants de l'enseignement secondaire jusqu'aux professeurs de mathématiques en formation. Nous signalons les rapports complexes entre les différentes probabilités impliquées dans le processus de résolution, comme l'une des principales causes de ces difficultés et comment ceux-ci dépendent des variables de la tâche associée aux problèmes, tels que sa structure mathématique, le contexte dans lequel les problèmes sont formulés et le vocabulaire utilisé pour exprimer ces probabilités dans chaque contexte. Nous montrons aussi comment certains étudiants ne parviennent pas gérer avec du bon sens les probabilités conditionnelles, mais que cela dépend des ressources ou des systèmes de représentation utilisés pour la résolution du problème.*

**Paroles clés.** Probabilité conditionnelle, joint probabilité, problèmes du probabilité conditionnelle, résolution des problèmes.

## **1. Introducción.**

La investigación en probabilidad condicional se mueve entre corrientes que se basan en la cognición, tendencia mayoritaria en los últimos 30 años, y una más contemporánea basada en aspectos curriculares y educativos. Respecto de su enseñanza, oscila entre modelos mecanicistas y otros realistas que, desde la década de los 90 del siglo pasado, se basan en el tratamiento y análisis de datos procedentes de contextos realistas (Edo, 2014; Watson & Moritz, 2002).

Por otra parte, nadie duda de que una comprensión adecuada del concepto de probabilidad no puede considerarse completa sin una comprensión adecuada del concepto de probabilidad condicionada, pues la complementa como medida del juicio subjetivo de las personas (Borovcnik, 2012). Dos elementos, su dificultad y la necesidad de una comprensión adecuada, en el contexto de las tendencias actuales de los modelos de enseñanza, han conducido a desarrollar investigaciones (ej., Díaz & Batanero, 2009; Díaz, Batanero, & Contreras, 2010), cuyo objetivo es identificar y

explicar las razones de dichas dificultades, además de obtener conocimiento sobre el comportamiento de los estudiantes en la resolución de problemas en los que la probabilidad condicional esté implicada y poder ofrecer sugerencias que puedan mejorar su enseñanza (tales como la propuesta de Pfannkuch, Seber & Wild, 2002).

Nos movemos también en los dominios de un concepto, que es, como establece Borovcnik (2012), crucial, pero que entra en conflicto con otros, independientemente de la perspectiva desde la que uno lo contemple: filosófica, matemática o educativa. Es más, según sugieren Pollatsek, Well, Konold, Hardiman y Cobb (1987), en la mente de algunas personas no parece que este concepto sea distinto de otros conceptos de probabilidad, como la probabilidad simple, la conjunta o de la intersección e incluso de otras probabilidades condicionales, sino que, por el contrario, “parece ser una amalgama” (Ibid, p. 269) de todos ellos.

Tradicionalmente la investigación sobre la comprensión de la probabilidad condicional se ha llevado a cabo al margen del paradigma de investigación en resolución de problemas, en un sentido estricto. En consecuencia, hay un considerable conjunto de resultados que hacen referencia a heurísticas, sesgos de razonamiento y confusiones o interpretaciones equivocadas cuando los sujetos observados tratan con la probabilidad condicional (una síntesis se presenta en Díaz, Batanero, & Contreras, 2010). Las explicaciones que los investigadores ofrecen sobre las dificultades al razonar e interpretar probabilidades condicionales se han relacionado con la causalidad, o con el papel que tiene el suceso condicionante o con su confusa relación con la probabilidad conjunta (por ejemplo, Watson & Callingham, 2014; Watson & Moritz, 2002).

Huerta (2014) sugiere que la resolución de problemas básicos de probabilidad condicional es una tarea difícil para una gran variedad de estudiantes, como muestran diversos trabajos (Carles, Cerdán, Huerta, Corter & Zahner, 2007; Díaz & Batanero, 2009; Fernandes, Correia, & Contreras, 2013; Huerta, 2009; Huerta, Cerdán, Lonjedo, & Edo, 2011; Lonjedo & Edo, 2009; Zahner & Corter, 2010). Estas dificultades están relacionadas, en mayor o menor medida, con las distintas variables de la tarea de los problemas (en el sentido de Kulm, 1979). Entre ellas, el contexto en el que se formulan los problemas, el formato con el que se expresan las cantidades, frecuencias, porcentajes o números entre 0 y 1, y la estructura de cantidades y relaciones entre cantidades, ya sean entre conocidas como desconocidas. La particularidad de dicha estructura es que se relacionan tres probabilidades: simples o marginales,  $P(A)$ , conjuntas  $P(A \cap B)$ , y condicionales  $P(A|B)$ , donde algunas de ellas son conocidas y otras desconocidas. La resolución de esta clase de problemas ofrece a investigadores y profesores el marco apropiado en el que explorar los roles que juegan dichas probabilidades para los resolutores y los factores que pueden influir en cómo se usan esas cantidades, ya sea en una forma apropiada, equivocada, vaga o confusa.

## 2. Objetivos.

El propósito de este artículo es informar de las dificultades que tienen los estudiantes cuando tratan conjuntamente con probabilidades condicionales y conjuntas en el seno de la resolución de problemas de probabilidad condicional. Tomando en consideración investigaciones previas, muchas de esas dificultades pueden estar causadas por una variedad de interpretaciones de las probabilidades condicionales cuando se expresan en forma verbal, y por su confusión con otras. Este trabajo es pues

complementario con otros ya citados, en el sentido de que el enfoque en el que se desarrolla nuestra investigación es la resolución de problemas verbales de probabilidad condicional. Pero, además, nuestro estudio implica a un amplio rango de estudiantes que han recibido previamente enseñanza en probabilidad condicional, incluidos los futuros maestros y profesores de secundaria, quienes no solo adquirieron enseñanza sobre el contenido sino que además sobre su enseñanza. En concreto, en este trabajo abordamos las siguientes cuestiones:

- ¿Las dificultades de los problemas son dependientes o independientes del nivel de formación de los estudiantes?
- ¿Hasta qué punto la confusión entre la probabilidad condicional y la probabilidad conjunta afectan a las dificultades de los problemas?
- ¿La confusión entre ambas probabilidades es estable y reproducible en un estudiante cualquiera o depende de factores externos al resolutor como las variables independientes del problema?

### **3. Marco teórico y antecedentes.**

#### **3.1. Antecedentes**

Es razonable pensar que la malinterpretación o confusión de la probabilidad condicional con otras probabilidades se produzca cuando la tarea o problema se presente a los estudiantes verbalmente. Esto puede afectar a los resultados de investigación; como ya señaló Shaughnessy (1992), las confusiones: “seems to occur primarily with students’ translation of conditional probability tasks, which then affects their understanding of the problem” (p. 473). Es decir, las probabilidades se interpretan a partir de un texto escrito en una lengua vernácula. Sin embargo, es menos razonable pensar que esta confusión surja cuando las probabilidades se expresen simbólicamente, pues aparentemente  $P(A|B)$  no se parece a  $P(A \cap B)$  y si se confunden la razón no está en su parecido semántico. Efectivamente, las dos expresiones: “probabilidad de que una persona que sufre una determinada enfermedad infecciosa dé positivo en un test” y “probabilidad de que una persona sufra una determinada enfermedad infecciosa y dé positivo en un test” tienen tal parecido semántico que puede ser razonable pensar que las dos expresiones se refieren al mismo fenómeno, especialmente en aquellos estudiantes que no han diferenciado claramente entre ambos conceptos de probabilidad. Es por esto que nos parece importante prestar atención a qué información verbal se ofrece a los estudiantes en los textos de estos problemas, con el fin de analizar su comportamiento cuando se les enfrenta con tareas que implican a sucesos compuestos, y sus probabilidades, y a probabilidades condicionales.

En este sentido, Watson y Moritz (2002) reducen la complejidad de las tareas propuestas en su investigación lo más posible. Así, se pide a los estudiantes que estimen probabilidades subjetivas de sucesos conjuntos, bien sea en el espacio muestral completo o en uno restringido (probabilidad condicional). Las tareas se formulan en contextos no habituales, variando el formato de expresión de la pregunta, frecuencias y probabilidades. Sin embargo, los estudiantes no disponen de datos conocidos en el texto de los problemas que permita un cálculo de la probabilidad condicional preguntada. Consecuentemente, las respuestas dadas por los estudiantes tienen un fuerte carácter subjetivo, basadas en un razonamiento intuitivo dentro del contexto en el que se formulan los problemas. Así que, las respuestas dadas no pueden

analizarse desde el punto de vista de la resolución de problemas y las conclusiones están condicionadas por el tipo de razonamiento que usan los estudiantes.

Pollatsek et al. (1987) usaron dos tipos de instrumentos para explorar la comprensión de los estudiantes sobre la probabilidad condicional, uno de elección múltiple y el otro de respuesta abierta. En ninguno de ellos los estudiantes tenían información previa con la cual encontrar una respuesta a las preguntas formuladas, por lo que tampoco en esta investigación podemos considerar las tareas que los investigadores proponen como tareas en un contexto de resolución de problemas.

En el lado opuesto de la balanza, encontramos investigaciones, y también libros de texto, en las que las tareas están formuladas usando tablas 2x2 (Contreras, Estrada, Díaz, & Batanero, 2010, Watson & Callingham, 2014, por ejemplo). En estas tareas, se proporciona información sobre las probabilidades conjuntas. Los estudiantes han de calcular una probabilidad condicional dada la información proporcionada por la tabla, generalmente por comparación entre las probabilidades conjunta y simple correspondientes, ambas presentes en la tabla como datos conocidos. En nuestra opinión, desde el punto de vista de la resolución de problemas, estas tareas están sobredimensionadas, pues a los estudiantes se les requiere leer la información contenida en la tabla de una manera conveniente. En estas tareas, solo se podría concluir sobre cómo interpretan los estudiantes la probabilidad condicional preguntada, pero no sobre cómo interpretan y usan la información sobre probabilidades condicionales conocidas en el texto del problema.

Son pocas investigaciones sobre la comprensión de la probabilidad condicional que se han llevado a cabo desde un punto de vista de la enseñanza y aprendizaje en un contexto de la resolución de problemas escolares (en el sentido de Lonjedo, Huerta, & Carles, 2012). En Carter y Zhaner (2007) y en Zahner y Corter (2010), por ejemplo, podemos reconocer perspectivas parecidas a las nuestras. Estos autores observan el uso que hacen los maestros en formación de las representaciones visuales en la resolución de problemas, tales como tablas 2x2, diagramas en árbol, diagramas de Venn, fórmulas y otras, tratando de relacionar dichos usos con la complejidad de los problemas. Intentan, también, establecer una conexión entre complejidad y tipo de representación visual elegida para la resolución del problema.

La primera evidencia sobre la amalgama de ideas relacionadas con la noción de probabilidad condicional, sugerida por Pollatsek et al. (1987), cuando los estudiantes están resolviendo problemas, es informada en Arnau (2012) y Edo (2014). Otros estudios informan también de que la resolución de problemas en los que están implicadas a la vez las probabilidades simple, conjunta y condicional, cualquiera que sea el nivel académico de los resolutores, es una tarea difícil (Carles et. al, 2009; Díaz & Batanero, 2009; Edo, 2014), incluyendo entre ellos a futuros profesores de matemáticas con una alta formación (Contreras, 2011; Huerta et al. 2011). Además, estudiantes pertenecientes a dichas muestras malinterpretan por igual cantidades conocidos como desconocidas, así como las relaciones entre ellas (Huerta, 2014). Edo (2014) ya establece que estudiantes de 15 años no parecen ser conscientes de que la probabilidad condicional y la probabilidad conjunta son dos conceptos diferentes puesto que, por ejemplo son capaces de dar respuestas numéricas correctas a probabilidades condicionales pero describiéndola como si fueran probabilidades conjuntas.

### **3.2. Marco teórico**

Nuestra investigación se sitúa en el marco de la resolución de problemas de matemáticas, en el contexto particular de los problemas ternarios de probabilidad condicional (Cerdán & Huerta, 2007; Huerta 2009). Estos problemas se formulan usando dos sucesos básicos,  $A$  y  $B$  y sus complementarios y considerando cuatro cantidades, tres conocidas y una desconocida, la pregunta del problema. Todas estas se eligen de un modo conveniente entre probabilidades simples, conjuntas y condicionales, estando siempre presente en el enunciado al menos una probabilidad condicional. Las relaciones entre probabilidades que se requieren durante el proceso de resolución de los problemas son fundamentalmente ternarias, aditivas y/o multiplicativas, pues conectan tres cantidades (ej.  $P(A \cap B) + P(A \cap \bar{B}) = P(A)$ ).

Hemos usado la idea de variable de la tarea introducida por Kulm (1979) para la formulación de los problemas en nuestros cuestionarios. Las variables independientes básicas consideradas son: su estructura de datos, también llamada aquí estructura conceptual, el contexto y el formato de datos. La consideración de estas variables no es casual, pues, como sabemos por investigaciones previas, cualquiera de ellas puede tener influencia en las dificultades del problema y en la actuación de los estudiantes. La variable estructura da cuenta de las cantidades que son conocidas y desconocidas en el problema, lo que implica a su vez a las relaciones que han de ser invocadas para resolver el problema. El contexto en el que los problemas están planteados pretende que las probabilidades adquieran significados particulares ligados a ellos. Los contextos que hemos considerado son: urnas, estadístico-social, estadístico-salud y test de diagnóstico, usado en investigaciones precedentes, siendo conscientes de que el uso de otros contextos podría conducir a otros resultados no alcanzables en estos. En cuanto a la variable formato de datos, el uso de frecuencias, porcentajes y probabilidades en  $[0, 1]$  dan valores a la dicha variable independiente. En el presente estudio, las tres se han tenido en cuenta, pero en todos los problemas la pregunta se ha formulado en términos de probabilidad. Sin embargo, teniendo en cuenta las sugerencias de la investigación cognitiva, con el fin de facilitar la comprensión de la información contenida en el texto de los problemas, todas las cantidades conocidas se expresan en términos de frecuencias o porcentajes (Gigerenzer & Hoffrage, 1995).

Algunos modelos de enseñanza usan sistemas de representación intermedios (mediadores) entre el lenguaje con el que se expresan las cantidades en el contexto y el lenguaje simbólico al que se traduce el problema (Zahner & Corter, 2010). Esos metalenguajes pueden ser listas organizadas de datos, diagramas en árbol, tablas de contingencia o tablas 2x2 y diagramas de conjuntos o diagramas de Venn. Con ellos la enseñanza pretende que los estudiantes puedan hacer lecturas analíticas de los problemas e introducir reglas de cálculo. Estos metalenguajes son, al mismo tiempo, fuente inagotable de información para investigadores interesados en la forma en la que los estudiantes “leen” las probabilidades conjuntas y condicionales presentes en el texto de los problemas. Ejemplo de esos usos pueden verse en Pfannkuch et al. (2002) y Watson (2011).

#### **4. Método.**

##### **4.1. Muestra**

La muestra de estudiantes implicada en esta investigación estuvo formada por 242 estudiantes cuya distribución puede verse en la Tabla 1. Los estudiantes procedentes de niveles educativos no universitarios recibieron enseñanza tradicional en probabilidad condicional según las directrices curriculares vigentes. Los recursos no

simbólicos, como tablas 2x2 y diagramas en árbol, tuvieron mayor incidencia en los niveles más bajos (15-16 años), siendo una combinación de éstos con el recurso simbólico los habituales en los niveles más altos (16-18 años). En cambio, los diagramas de Venn o cualquier otra representación conjuntista, no fueron recursos utilizados en ningún caso en la enseñanza secundaria o bachillerato.

Tabla 1. *Distribución de la muestra según su nivel educativo y formación matemática previa.*

Nivel educativo y Formación previa		Número de estudiantes	Sumas parciales según edad
ESO (15-16 años)	Matemáticas aplicadas (A)	65	107
	Matemáticas formales (B)	42	
Bachillerato (16-17 años)	Matemáticas I (formal)	57	72
	Matemáticas aplicadas a las Ciencias Sociales.	15	
Bachillerato (17-18 años)	Matemáticas aplicadas a las Ciencias Sociales II	6	6
Maestros de Educación Primaria en formación		27	27
Profesores de matemáticas de la educación secundaria en formación.		30	30
Total			242

Por otro lado, los estudiantes para maestros y profesores recibieron enseñanza especializada en didáctica de la probabilidad, que incluye no sólo enseñanza sobre el contenido, sino que, además, sobre el contenido para la enseñanza, en el sentido de Ball, Thames y Phelps (2008) y sobre resolución de problemas de probabilidad.

#### 4.2. Tareas y procedimientos.

En la Tabla 2 puede verse la descripción de las características de los 12 problemas que se diseñaron para la investigación, acorde con las variables independientes mencionadas en el marco teórico. Algunos ejemplos de estos problemas se encuentran en el anexo. Considerando dichas variables se construyeron diferentes versiones de un mismo problema que, siendo estructuralmente isomorfos, se formulan en diferentes contextos y formatos de datos.

Tabla 2. *Variables de la tarea y sus valores en los problemas de los cuestionarios. Se incluye el número de resoluciones de los estudiantes disponible de cada uno.*

Problema nº	Estructura conceptual	Contexto	Formato de datos	Nº de resoluciones
1	PC	Estadístico-social	F, p	170
2	PC	Estadístico-salud	%, p	155
3	PC	Urnas	F, p	85
4	PC	Test diagnóstico	%, p	126
5	PT	Estadístico-social	F, %, p	171
6	PT	Estadístico-salud	%, p	120
7	PT	Urnas	F, %, p	120
8	PT	Test diagnóstico	%, p	120
9	TB	Estadístico-social	F, %, p	69
10	TB	Estadístico-salud	%, p	69

11	TB	Urnas	F, %, p	64
12	TB	Test diagnóstico	%, p	64
				1333

La Tabla 3, por su parte, nos muestra los problemas básicos que fueron considerados, descritos con ayuda del lenguaje simbólico, su estructura conceptual y la tipología de las cantidades conocidas y desconocidas en ellos.

Tabla 3. *Cantidades conocidas y desconocidas en cada una de las estructuras conceptuales.*

Estructura conceptual	Cantidades conocidas	Cantidad desconocida
Probabilidad condicional (PC)	$P(A), P(B), P(\bar{A} \cap \bar{B})$	$P(A B)$
Probabilidad total (PT)	$P(A), P(B \bar{A}), P(A \cap \bar{B})$	$P(B)$
Teorema de Bayes (BT)	$P(A), P(B A), P(B \bar{A})$	$P(A B)$

Se construyeron seis cuestionarios a partir de la colección de problemas descritos por las tablas 2 y 3. Dos de ellos, con cinco problemas cada uno, fueron administrados a los estudiantes más jóvenes (15-16 años), y los otros cuatro, esta vez con seis problemas cada uno, a los estudiantes de bachillerato y a los maestros y profesores en formación. Los cuestionarios fueron administrados durante sus clases, previo acuerdo entre los respectivos profesores y alumnos. En el momento de su administración estuvieron presentes solamente los estudiantes, su profesor y uno de los firmantes de este trabajo.

Las resoluciones de los problemas se analizan según el esquema de variables y su codificación propuesta por Huerta, Edo, Amorós y Arnau (2016). Este esquema permite definir un conjunto de dificultades asociados a los problemas (Huerta, 2014). Las dificultades se expresan en porcentajes que pretenden medir la falta de éxito en una fase determinada del proceso de resolución del problema. En otras palabras, sea  $V$  una variable asociada a un problema dado  $P$ . Definimos dificultad en el problema  $P$  asociada a la variable  $V$  al valor de la expresión:

$$DV = 100 - \left( \frac{n^\circ \text{ de éxitos de } V}{N} \right) \times 100, \text{ o bien } DV = \frac{n^\circ \text{ de fracasos en } V}{N} \times 100$$

Mediante estas expresiones podemos estudiar dificultad de la realización de cálculos intermedios (DCI) o del cálculo final (DCF) en el problema  $P$ . Si hay presencia de dichos cálculos pero con errores, entonces definimos la dificultad de la realización de cálculos intermedios correctos (DCICorrectos) o en el cálculo final (DCFcorrecto). Definimos dificultad de un problema dado, (DP), en función de si la pregunta del problema es contestada o no. Igualmente definimos la dificultad de la solución correcta del problema (DSP). Finalmente, consideramos la dificultad de que la solución numérica aportada sea descrita correctamente o no en el contexto en el que se ha formulado el problema  $P$ : (DDCorrecta).

## 5. Resultados. Análisis e interpretación.

### 5.1 Resultados sobre las dificultades de los problemas y análisis.

La Tabla 4 muestra los diferentes índices de dificultad descritos, asociados a cada uno de los 12 problemas del cuestionario. En negrita se resaltan aquellos valores más altos en las dificultades. Los que en ella se consideran son aquellas dificultades que hemos considerado que se pueden relacionar directamente con una confusión entre las probabilidades condicionales y conjuntas, derivada a su vez, de una inapropiada



interpretación de alguna de ellas o de ambas. Son, en primera instancia, las dificultades de llevar a cabo cálculos intermedios para producir nuevas cantidades con éxito y las dificultades de considerar un cálculo final que responda a la pregunta del problema. En segunda instancia, la dificultad de dar una solución al problema mediante un número que no solamente se desea que sea correcto sino que esté correctamente descrito en el contexto del problema.

Tabla 4. *Dificultades de los problemas medidas en porcentajes*

Dificultad	Problema											
	1	2	3	4	5	6	7	8	9	10	11	12
Cálculos intermedios	1,8	22,7	7,1	<b>35,8</b>	13,3	5,3	2,6	19,4	7,8	14,3	7,8	20,3
Cálculos intermedios correctos	7,8	24,8	12,7	31,4	25,9	<b>91,7</b>	6,3	66,3	8,5	16,7	16,9	8,5
Cálculos finales	5,9	27,0	10,6	<b>39,4</b>	24,2	20,2	8,7	26,2	10,9	11,1	12,5	35,6
Cálculos finales correctos	40,3	53,4	<b>61,8</b>	50,0	16,8	12,1	12,4	31,6	21,1	35,7	55,4	57,9
Problema	4,1	14,9	4,7	19,3	18,2	11,4	7,8	19,4	10,9	7,9	9,4	<b>25,4</b>
Solución	43,2	65,8	65,4	68,2	38,5	<b>94,1</b>	17,9	85,5	29,8	50,0	62,1	63,6
Descripción correcta	60,5	53,3	<b>66,7</b>	59,1	18,5	14,9	23,6	24,1	47,4	37,9	65,5	63,6

Entre la amplia variedad de datos proporcionados por la Tabla 4 se intuye que no hay gran dificultad en proporcionar una solución a los problemas (la dificultad del problema se mantiene siempre por debajo del 25.4%). No obstante, la alta dispersión mostrada en la Tabla 5 nos puede hacer pensar que ésta es debida a los diferentes niveles de formación de los estudiantes. Pero esto es cierto sólo parcialmente, como veremos más tarde.

Tabla 5. *Dificultades medias de los problemas en porcentajes, según la estructura conceptual del problema.*

Dificultad	Media			Desviación típica		
	PC	PT	TB	PC	PT	TB
Cálculos intermedios	16,9	10,2	12,6	13,4	6,6	5,2
Cálculos intermedios correctos	19,2	<b>47,6</b>	12,7	9,4	<b>33,4</b>	4,2
Cálculos finales correctos	<b>51,4</b>	18,2	<b>42,5</b>	7,7	7,9	<b>15,1</b>
Problema	10,8	14,2	13,4	6,5	4,8	7,0
Solución	<b>60,7</b>	<b>59,0</b>	<b>51,4</b>	10,1	<b>31,8</b>	13,5
Descripción correcta	<b>59,9</b>	20,3	<b>53,6</b>	4,8	3,8	<b>11,5</b>

En general, los estudiantes son capaces de incluir, sin muchas dificultad, cálculos intermedios en la resolución de los problemas, lo que implica básicamente usar las cantidades conocidas en el enunciado de una forma más o menos competente. El uso, competente o no, incluye la posibilidad de que las cantidades sean interpretadas correctamente y no sólo operativamente, aunque depende de la estructura de cantidades que estas dificultades aparezcan en mayor o menor grado, como puede verse al comparar la dificultad de los cálculos intermedios correctos en PC, PT y TB (Tabla 5). Además, en este caso, la alta dispersión indica que, dada una familia de

problemas, a la hora de interpretar esas cantidades, un contexto es más influyente que otro (P6 y P8 frente a P5 y P7, Tabla 4). En cambio, esta influencia no se observa tanto en las otras dos familias de problemas.

Vemos, en la Tabla 5, cómo las dificultades en la solución de los problemas se mantienen en promedio superiores al 50% en todos los casos, excepto en la descripción correcta de la solución de un problema (Descripción correcta del 20,3% para la clase de problemas PT). No obstante, si bien la pregunta se interpreta y se describe correctamente, se hace sobre cantidades interpretadas incorrectamente, como puede deducirse al comparar las dificultades en los cálculos intermedios. La Tabla 5 muestra también que la probabilidad de que un estudiante cometa errores en los cálculos intermedios es mucho más alta en los problemas PT, en donde una probabilidad conjunta y una probabilidad condicional son datos del problema, comparado con aquellos en los que esto no ocurre (ver problemas PC o problemas TB en Tabla 5). Este hecho sugiere que una causa probable de los errores que se cometen en los cálculos intermedios descansa en la mala interpretación de las probabilidades conjunta y condicional que aparecen en los textos de los problemas, pero no así en la tercera cantidad, una probabilidad simple. Los estudiantes tienden a interpretar ambos datos como probabilidades conjuntas o, a veces, ambos como probabilidades condicionales, lo que les conduce a establecer relaciones entre cantidades malinterpretadas. Más adelante veremos algunos ejemplos de estas actuaciones.

Por otro lado, es mucho más probable que los estudiantes cometan errores en el cálculo final cuando se pregunta por una probabilidad condicional (Figura 1), como también observan Contreras et al. (2010). A pesar de los niveles de formación tan dispares entre los estudiantes, sus comportamientos se relacionan fuertemente respecto de esta variable. De nuevo, la presencia de una probabilidad condicional en los problemas de la clase PC y TB, aunque esta vez como cantidad preguntada, es una causa probable de los valores altos de dificultad, en general, debido a que se malinterpretan como probabilidades conjuntas. Obsérvese como las dificultades decrecen cuando se pregunta por una probabilidad simple (problemas PT).

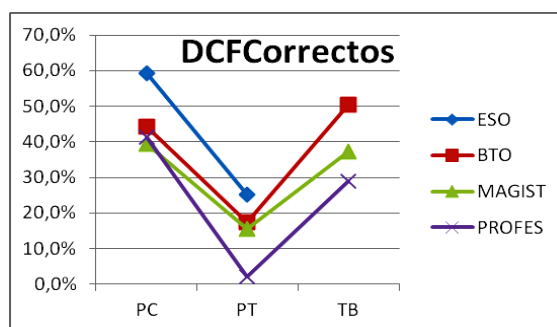


Figura 1. Dificultad de proporcionar un cálculo final correcto por estructura y por nivel educativo.

Obtener la respuesta correcta a un problema dado es relativamente difícil para nuestros estudiantes, independientemente de su nivel educativo (ver Figura 2.a), pues no hay grandes diferencias entre las dificultades a la hora de dar una solución correcta al problema (DSP), según el nivel educativo que se trate. Esto nos sugiere que, aunque la dificultad media varíe, un problema que sea difícil de responder correctamente por un grupo de estudiantes es probablemente difícil también para cualquiera de los otros.

Resultados parecidos obtiene Contreras (2011) con estudiantes de psicología y futuros profesores de matemáticas.

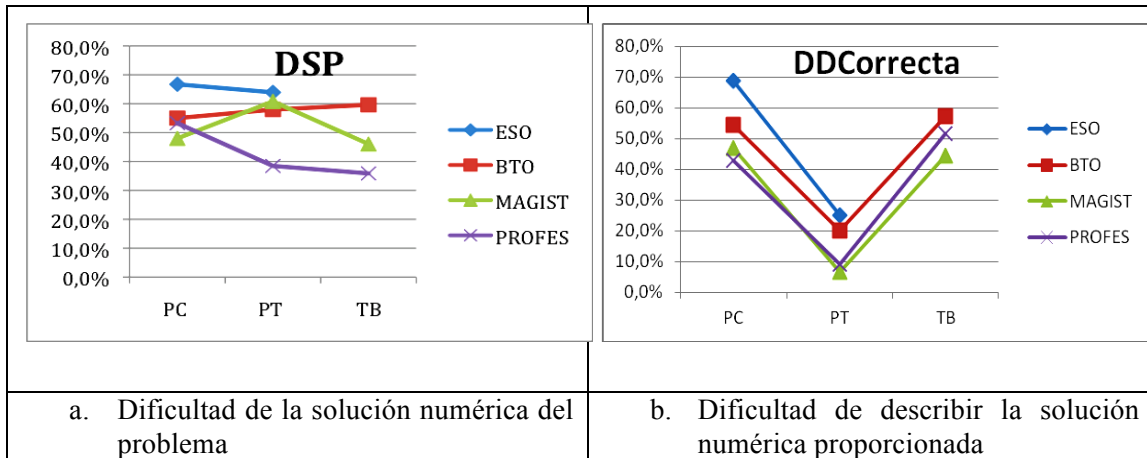


Figura 2. Dificultad por clases de problemas y por niveles de los resolutores

La Figura 2.b) nos proporciona evidencias de cómo la estructura del problema es un factor influyente en la capacidad del estudiante de describir correctamente la respuesta numérica dada. Mientras que es poco probable describir incorrectamente una probabilidad simple (por ejemplo en los problemas PT), es más probable que esto ocurra con las probabilidades condicionales. Además, esta dificultad crece a medida que el nivel educativo de los estudiantes decrece, sugiriéndonos que se requiera cierto nivel de madurez para describir correctamente probabilidades condicionales. En consecuencia, se sugiere que los modelos de enseñanza deberían prestar atención a este elemento en la enseñanza desde los niveles formativos más bajos.

**5.2. Resultados derivados de las actuaciones de los estudiantes en las distintas clases de problemas y análisis.**

**5.2.1. Problemas de probabilidad condicional**

En las clases de problemas PC, los estudiantes deben interpretar correctamente tres cantidades conocidas: dos probabilidades simples y una probabilidad conjunta, y una desconocida que se corresponde con una probabilidad condicional. Las dificultades medias para la obtención de una respuesta correcta al conjunto de problemas 1 a 4 se presentan en la Tabla 6 y en las Figura 2.

Tabla 6. Medias y desviaciones típicas de la dificultad al dar una respuesta numérica correcta a los problemas PC, según el nivel educativo.

	ESO	Bachillerato	Magisterio	Profesorado
Media	66,7	55,0	48,5	53,3
Desviación típica	10,6	10,7	13,8	15,0

Estos resultados nos indican que era poco probable que los estudiantes proporcionasen una respuesta correcta a estos problemas, aunque con algunas diferencias. Además, la dificultad aparece en todos los niveles educativos, sin que la preparación previa sea un factor determinante, pues resulta ser un poco menor en los futuros maestros que en los estudiantes de bachillerato y futuros profesores de

secundaria, quienes muestran medidas parecidas, pero razonablemente menores que las medidas de los estudiantes más jóvenes. Nuestra hipótesis es que esa dificultad surge debido a que la pregunta del problema se malinterpreta al considerarla como una probabilidad conjunta en lugar de una probabilidad condicional.

Por otro lado, la dispersión de esta dificultad sugiere que hay algunas características de los problemas que permiten que los estudiantes interpreten la pregunta de una manera más competente. Esas características solo pueden ser el contexto o el formato de datos, o a ambos, pues los problemas son isomorfos estructuralmente (Tabla 3).

El contexto puede provocar que los estudiantes consideren que las dos expresiones siguientes son semánticamente equivalentes: “Si un ciudadano no padece tuberculosis, ¿qué probabilidad tiene de que un test le dé negativo?” y “¿qué probabilidad hay de que un ciudadano padezca tuberculosis y dé negativo en un test?” (Problema 4, anexo). En ese caso, los estudiantes proporcionan una respuesta a la segunda como si fuera la primera. Esta confusión se produce en el 75% de los estudiantes de secundaria, en el 56% de los estudiantes de bachillerato, el 63% de los maestros en formación y en el 80% de los futuros profesores de matemáticas de educación secundaria. Por contra, los porcentajes disminuyen, respectivamente, hasta el 46.5%, 39.4%, 23.1% y 46.7% en problemas isomorfos en los que el contexto es menos influyente (Carles et al., 2009) y en los que no existe una relación o bien causal o bien temporal entre los sucesos que fuerce a los estudiantes a considerar semánticamente equivalentes expresiones del tipo: “Si un estudiante de 4º usa gafas, ¿qué probabilidad hay de que sea una chica?” y “¿cuál es la probabilidad de que un estudiante de 4º use gafas y sea chica?”

Entre los recursos que los modelos de enseñanza proporcionan a los estudiantes, y cuyo uso está ampliamente generalizado, predominan las tablas 2x2 (también llamadas de contingencia) y los diagramas en árbol, y en menor medida los diagramas de Venn o el modelo del cuadrado unidad (Böcherer-Linder, Eichler, & Vogel, 2016). Esos recursos facilitan la lectura analítica del problema y permiten organizar la información disponible por medio de estructuras en las que dicha información puede visualizarse. Por tanto, posibilitan la correcta interpretación de las cantidades conocidas y el cálculo de las cantidades desconocidas. La respuesta a una probabilidad condicional, una vez el resto de probabilidades implicadas se han interpretado correctamente, puede obtenerse de dos maneras dependiendo del recurso usado: a) en el caso de tablas 2x2, por asignación de una probabilidad que se calcula dividiendo la probabilidad conjunta con la correspondiente probabilidad simple o marginal; b) o por medio del producto de dos probabilidades, la regla del producto, derivada de la definición de la probabilidad condicional, en el caso de los diagramas en árbol.

Una mala interpretación, o una confusión entre las cantidades implicadas, puede hacer que el estudiante dé como respuesta a la probabilidad condicional alguna de las cantidades disponibles en la tabla o en el árbol. En el caso de las tablas 2x2 suele ser una de las cantidades “interiores”, es decir, una probabilidad conjunta (Figura 3). En el ejemplo, el estudiante no confunde la probabilidad conjunta y la probabilidad condicional preguntada cuando las describe simbólicamente pero sí, en cambio, cuando le asigna un valor a ésta última. El mismo comportamiento se ha observado en profesores de matemáticas.

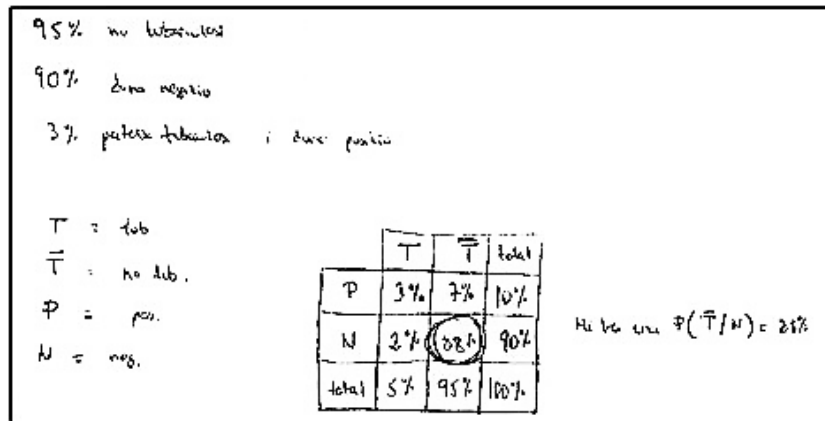


Figura 3. Resolución del Problema 4 por un estudiante de secundaria (15-16 años) usando una tabla 2x2.

### 5.2.2. Problemas de probabilidad total

En los problemas 5 al 8, la clase de probabilidad total (PT), la cantidad preguntada es una probabilidad simple que puede obtenerse por medio de la suma de dos probabilidades conjuntas. Los enunciados de los problemas proporcionan información sobre tres cantidades que se han de interpretar, respectivamente, como una probabilidad simple, conjunta y condicional (Tabla 3). Así que, en cada problema de esta clase, los estudiantes se enfrentan al dilema de interpretar correctamente dichas probabilidades (especialmente la conjunta y la condicional).

Tabla 7. Medias y desviación típica de la dificultad de la solución correcta de los problemas de la clase PT, por niveles educativos.

	ESO	Bachillerato	Magisterio	Profesorado
Media	63,9	57,9	55,1	38,4
Desviación típica	27,8	33,3	28,3	25,3

La diferencia en los valores medios nos permite afirmar que este conjunto de problemas presenta una alta dificultad a los estudiantes de niveles más bajo (Tabla 7). Son notables las altas dispersiones que aparecen. Con los problemas 5 y 7 las dificultades medias son 35% en educación secundaria, 20% en bachillerato, 26% en los futuros maestros y 12% en futuros profesores de matemáticas y valores de dispersión más bajos. Pero, en los problemas 6 y 8, los valores medios son, respectivamente, 94%, 95.1%, 83.3% y 65.3% e incluso con valores de dispersión menores. Para explicar estas diferencias, es de suma importancia prestar atención a los datos proporcionados por el problema, relacionados tanto con el contexto como con su formato. En efecto, en los problemas 5 y 7 la interpretación correcta del porcentaje condicional está favorecida por el hecho de operar multiplicativamente sobre una frecuencia absoluta. Sin embargo, en los problemas 6 y 8 el porcentaje condicional debe operar multiplicativamente sobre otro porcentaje, lo que incrementa la dificultad de la interpretación correcta. Esta dificultad se presenta incluso en un mismo resolutor, como puede verse en las resoluciones que presenta un estudiantes de 15-16 años de los problemas 5 y 6 (Figura 6).

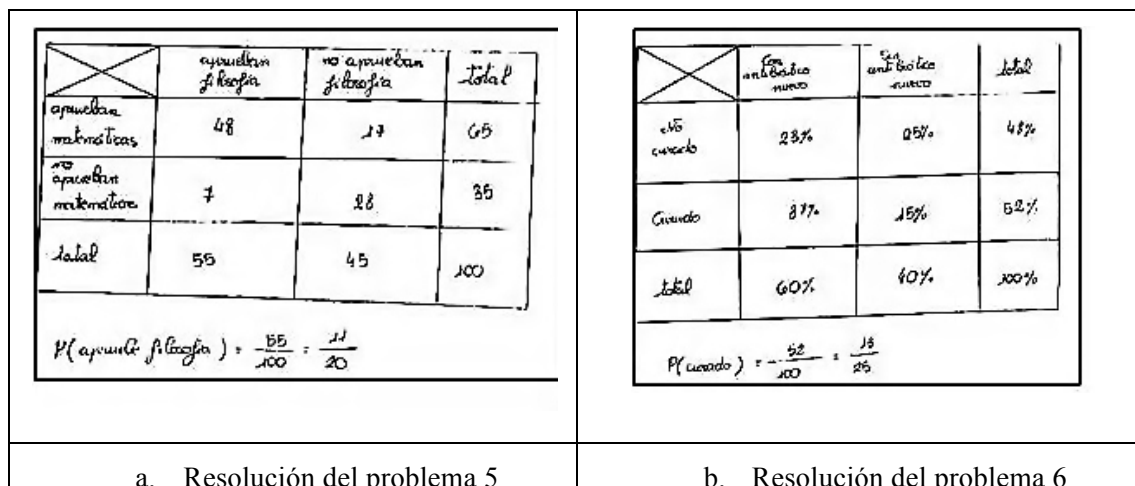


Figura 6. Resolución de los problemas 5 y 6 por el mismo estudiante de ESO de 15-16 años.

Esta interpretación de la probabilidad condicional como una probabilidad conjunta no se presenta solamente en los estudiantes más jóvenes, lo que podría ser comprensible, dada su limitada competencia en el uso de porcentajes en problemas similares, sino que también en los estudiantes de nivel superior. Puede pensarse entonces que utilizar las tablas 2x2, frecuentes en los maestros y profesores de secundaria, puede influir sobre los estudiantes, empujándoles a interpretar incorrectamente una probabilidad condicional conocida. Pero la evidencia nos enseña que usando las tablas 2x2 la interpretación es a veces correcta, como en los problemas 5 (Figura 4.a) y 7, o incorrecta si las cantidades están expresadas en porcentajes (problemas 6 (Figura 4.b) y 8). Esta confusión ocurre, por tanto, al margen del recurso usado pudiéndose deber a la interpretación competente de las cantidades en favor de las frecuencias ante los porcentajes (Gigerenzer & Hoffrage, 1995).

### 5.2.3. Problemas de Bayes

Para finalizar nuestros análisis, disponemos de cuatro tareas tradicionales en la enseñanza del teorema de Bayes, isomorfos en su estructura (Tabla 3), que solamente difieren en el contexto y el formato de datos. Dado que los estudiantes más jóvenes, de 15-16 años, no habían recibido enseñanza relacionada con el Teorema de Bayes, la muestra de estudiantes se restringió a estudiantes de Bachillerato y futuros maestros y profesores. Las dificultades medias para este conjunto de problemas pueden verse en la Tabla 8.

Tabla 8. Medias y desviaciones típicas de la dificultad en la obtención de una solución numérica correcta para los problemas TB, por niveles educativos.

	Bachillerato	Magisterio	Profesorado
Media	59,6	46	36,4
Desviación típica	11,7	29,6	8,4

En estos problemas, el enunciado proporciona dos probabilidades condicionales y una probabilidad simple. Las primeras están expresadas en porcentajes y la absoluta

tanto en frecuencias (problemas 9 y 11) como en porcentajes (problemas 10 y 12). El problema pide una probabilidad condicional, inversa de una de las condicionales conocidas. Por tanto, los estudiantes tienen tres ocasiones para interpretar correctamente las tres probabilidades condicionales que participan en el problema, dos como conocidas y una desconocida. A partir de la Tabla 8 podemos aventurarnos a decir que, en tanto que las tres se interpretan correctamente, en el 40,4% de los estudiantes de bachillerato, el 54% de los futuros maestros y el 63,6% de los futuros profesores de secundaria, se obtiene una respuesta correcta. En cualquier otro caso, el error aparece. Sea como sea, es la condicional preguntada la que mayoritariamente se malinterpreta, respectivamente, con estos porcentajes: 46,3%, 38,7% y 29,5% (ver Figura 2).

Tradicionalmente, algunos modelos de enseñanza abogan por ofrecer los diagramas en árbol como el recurso más eficiente en la resolución de problemas de Bayes, aunque Böcherer-Linder, Eichler, y Vogel (2016) discuten dicha eficacia frente al modelo del cuadrado unidad cuando los datos se expresan en frecuencias. En aquel caso, se necesitarían dos diagramas en árbol, o bien uno con el que leer las cantidades del problema y usar la fórmula donde resolver la probabilidad condicional inversa preguntada. Si un estudiante usa únicamente un diagrama en árbol, podría interpretar correctamente las probabilidades condicionales conocidas pero interpretar incorrectamente la desconocida como una probabilidad conjunta, resolviendo el problema simplemente aplicando la regla de la multiplicación en la rama correspondiente del árbol, como vemos en la Figura 5.a. Estas confusiones no solo ocurren con el recurso del diagrama en árbol. En la Figura 5.b podemos observar la misma interpretación equivocada usando tablas 2x2 para el mismo problema.

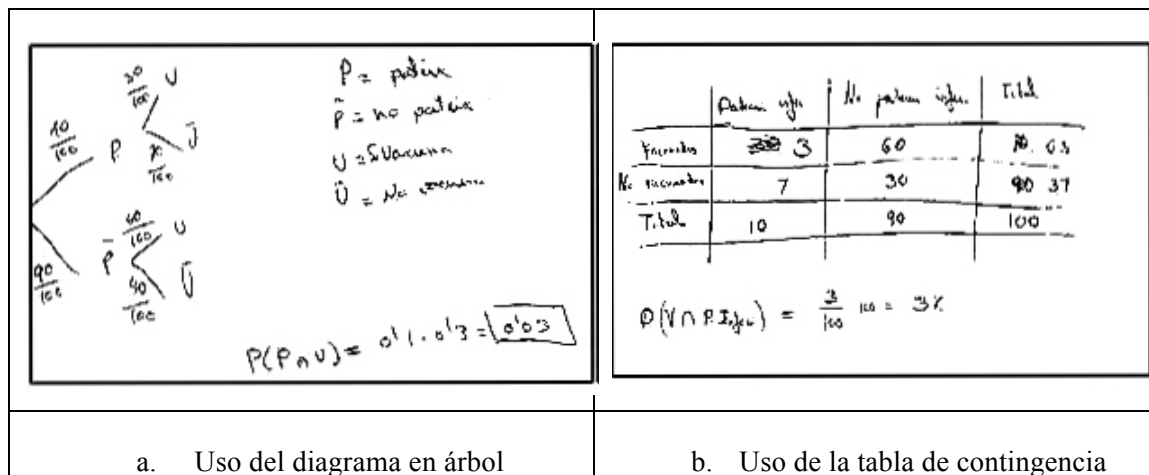


Figura 5. Dos resoluciones del problema 10 que evidencian la confusión entre probabilidad condicional y conjunta provocada por el recurso empleado y las reglas de cálculo asociadas al mismo.

## 6. Conclusiones

En el razonamiento probabilístico, y por tanto en la enseñanza, la probabilidad condicional y la probabilidad conjunta están condenadas a entenderse. En efecto, tanto en la formulación de problemas como en sus resoluciones, las probabilidades conjuntas se constituyen como mediadoras entre probabilidades condicionales y

simples, incluso sin ser calculadas. Para un alumno puede no tener sentido preguntar por la “probabilidad de que una persona padezca una enfermedad y dé positivo para dicha enfermedad”, pero en cambio sí da sentido a la pregunta de “qué probabilidad hay de que se padezca una enfermedad, siendo que el test para dicha enfermedad dio positivo”. Aun no teniendo en sí misma interés, dicha probabilidad conjunta resulta necesaria para la determinación de la probabilidad condicional relacionada con ella. Esta relación se deriva de la propia definición de la probabilidad condicional y de los modelos de enseñanza usuales. En efecto, en dichos modelos la proporcionalidad suele estar ausente en la enseñanza de la probabilidad condicional, lo que imposibilita el estudio de otras relaciones posibles, como la presencia de la proporcionalidad inversa entre probabilidades condicionales y simples, en las que la probabilidad conjunta es su razón de proporcionalidad.

Hemos visto cómo las interpretaciones erróneas de la probabilidad condicional en los estudiantes pueden aparecer en cualquiera de las fases del proceso de resolución de problemas: en su lectura como datos conocidos y desconocidos y en su uso para el cálculo de otras cantidades intermedias. Estas malas interpretaciones están influenciadas mayoritariamente por el contexto y el formato de expresión de los datos en el enunciado del problema y su mayor o menor incidencia no depende del nivel formativo de los estudiantes. Así, los resultados nos han mostrado como los estudiantes tratan el concepto de probabilidad condicional como una amalgama de otros conceptos de probabilidad, fundamentalmente la probabilidad conjunta, tal y como ya sugirieron Pollatsek et al. (1987).

De esta manera, no es inusual observar a estudiantes que interpreten dos condicionales dadas en el problema una correctamente y la otra como conjunta, o ambas como conjuntas, y al mismo tiempo interpretando o malinterpretando la probabilidad condicional preguntada. Además, hemos observado cómo un mismo estudiante puede no considerar idénticos problemas estructuralmente isomorfos pero formulados en diferentes contextos. Esto nos sugiere que la confusión no se debe exclusivamente a las estructuras gramaticales usadas sino también al lenguaje particular usado en cada contexto para referirse a la probabilidad condicional. Si el contexto es bien conocido, tales como urnas o aspectos sociales, los estudiantes encuentran menos dificultades en las interpretaciones, usándolas apropiadamente. Sin embargo, en ciertos contextos, como en el de salud, las dificultades crecen. Si los estudiantes confían más en los recursos (tablas 2x2 y diagramas en árbol), que en la comprensión del concepto, entonces la estructura interna de los recursos en lugar de ayudar a leer apropiadamente las cantidades puede forzarles a las interpretaciones erróneas.

Con todo, la pobre familiaridad con el lenguaje relacionado con el contexto y una fuerte influencia de los recursos proporcionados por la enseñanza parecen ser posibles causas de la confusión entre probabilidad condicional y conjunta, provocando que los problemas que hemos usado en esta investigación hayan proporcionado tan altos grados de dificultad que la enseñanza actual no parece capaz de reducir. Muy llamativos resultan los datos sobre los futuros profesores de enseñanza secundaria, que confirman la dificultad ya detectada en estos problemas por otras investigaciones, como en Contreras (2011).

Como sugieren Lonjedo, Huerta y Carles (2012), y ha mostrado Edo (2014), un enfoque de la enseñanza basado en la exploración fenomenológica de los conceptos de probabilidad mediante la resolución de problemas formulados en contexto puede



ayudar a los estudiantes en la interpretación y usos de esas probabilidades. Las probabilidades condicionales se refieren a fenómenos diferentes de aquellos que se describen mediante probabilidades conjuntas y están expresados de maneras distintas, dependiendo del contexto, pues lo que se aprende en un contexto no siempre se extiende con facilidad a otro contexto diferente. Esto, pues, nos permite meditar sobre la actual enseñanza y preguntarnos si podemos mejorarla. En este sentido, hemos intentado aportar algunas ideas con este trabajo.

## Referencias

- Arnau, J. (2012). *Un estudio exploratorio de la resolución de problemas de probabilidad condicional centrado en la fase de cálculo*. Memoria de Máster. Universitat de València.
- Böcherer-Linder K., Eichler, A., & Vogel, M. (2016). The impact of visualization on understanding conditional probabilities. Comunicación presentada en el *13th International Congress on Mathematics Education, Hamburgo, Alemania*. Disponible en: [http://iase-web.org/documents/papers/icme13/ICME13\\_S1\\_Boechererlinder.pdf](http://iase-web.org/documents/papers/icme13/ICME13_S1_Boechererlinder.pdf)
- Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, 59 (5), 389-407.
- Borovcnik, M. (2012). Multiple perspectives on the concept of conditional probability. *Avances de Investigación en Didáctica de la Matemática*, 2, 5-27.
- Carles, M., Cerdán, F., Huerta, M. P., Lonjedo, M<sup>a</sup> A., & Edo, P. (2009). Influencia de la estructura y el contexto en las dificultades de los problemas de probabilidad condicional de nivel  $N_0$ . Un estudio exploratorio con estudiantes sin enseñanza previa. En M. J. González, M. T. González, & J. Murillo (Eds.), *Investigación en Educación Matemática XIII*, 173-185. Santander: SEIEM.
- Cerdán, F., & Huerta, M. P. (2007). Problemas ternarios de probabilidad condicional y grafos trinomiales. *Educación Matemática* 19(1), 27-62.
- Contreras, J. M. (2011). *Evaluación de conocimientos y recursos didácticos en la formación de profesores sobre probabilidad condicional*, Tesis doctoral. Universidad de Granada.
- Contreras, J.M., Estrada, A., Díaz, C., & Batanero, C. (2010). Dificultades de futuros profesores en la lectura y cálculo de probabilidades en tablas de doble entrada. En M.M. Moreno, A. Estrada, J. Carrillo, & T.A. Sierra (Eds.), *Investigación en Educación Matemática XIV* (pp. 271-280). Lleida: SEIEM.
- Corter, J. E. & Zahner, D. (2007). Use of external visual representation in probability problem solving. *Statistics Education Research Journal* 6, 22-50.
- Díaz, C., & Batanero, C. (2009). Students' formal knowledge and biases in conditional probability reasoning. Do they improve with instruction? *International Electronic Journal of Mathematics Education*, 4(3), 131-162.
- Díaz, C., Batanero, C., & Contreras, J. M. (2010). Teaching independence and conditional probability. *Boletín de Estadística e Investigación Operativa*, 26 (2), 149-162.
- Edo, P. (2014). *Estudios sobre los problemas ternarios de probabilidad condicional de nivel  $N_0$* . Tesis doctoral. Universitat de València.
- Fernandes, J.A., Correia, P.F., & Contreras, J.M. (2013). Ideas intuitivas de alumnos do 9º ano em probabilidade condicionada e probabilidade conjunta. *Avances de Investigación en Educación Matemática*, 4, 5-26.
- Gigerenzer, G. & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review* 102(4), 684-704.

- Huerta, M. P. (2009). On conditional probability problem solving research —structures and context. *International Electronic Journal of Mathematics Education* 4(3), 163-194.
- Huerta, M. P. (2014). Researching conditional probability problem solving. En E. J. Chernoff & B. Sriraman (Eds.), *Probabilistic thinking: Presenting plural perspectives* (pp. 613-639). Dordrecht: Springer.
- Huerta, M. P., Cerdán, F., Lonjedo, M<sup>a</sup>. A., & Edo, P. (2011). Assessing difficulties of conditional probability problems. En M. Pytlak; T. Rowland, & E. Swoboda (Eds.) *Proceedings of the Seventh Congress of the European Society for Research in Mathematics Education* (pp. 807-817). Rzeszów, Poland: ERME
- Huerta, M. P., Edo, P., Amorós, R., & Arnau, J. (2016). Un esquema de codificación para el análisis de las resoluciones de los problemas de probabilidad condicional. *Revista Latinoamericana de Investigación en Matemática Educativa* 19(3). DOI: 10.12802/relime.13.1934.
- Kulm, G. (1979). The classification of problem-solving research variables. En G. A. Goldin & C. E. McClintock (Ed.), *Task variables in mathematical problem solving* (pp. 1-22). Columbus, OH: ERIC.
- Lonjedo, M<sup>a</sup> A., Huerta, M. P., & Carles, M. (2012). Conditional probability problems in textbooks: An example from Spain. *Revista Latinoamericana de Investigación en Matemática Educativa* 15(3), 319-338.
- Pfannkuch, M., Seber, G. A. F., & Wild, Ch. (2002). Probability with less pain. *Teaching Statistics* (24)1, 24-30.
- Pollatsek, A., Well, A. D., Konold, C., Hardiman, P., & Cobb, G. (1987). Understanding conditional probability. *Organizational behaviour and human decision processes* 40, 255-269.
- Shaughnessy, M. (1992). Research in probability and statistics: Reflections and directions. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 465-494). New York: Macmillan.
- Watson J. (2011). Cheating partners, conditional probability and contingency tables. *Teaching Statistics* 33, 66-70.
- Watson, J. M. & Moritz, J. B. (2002). School students' reasoning about conjunction and conditional events. *International Journal of Mathematical in Science and Technology* 33, 59-84. DOI: 10.1080/00207390110087615
- Watson, J., & Callingham, R. (2014). Two-way tables: Issues at the heart of statistics and probability for students and teachers. *Mathematical Thinking and Learning*, 16, 254–284.
- Zahner, D., & Corter, J. E. (2010). The process of probability problem solving: use of external visual representations. *Mathematical Thinking and Learning* 12, 177-204.

### **Anexo. Enunciados de algunos problemas.**

**Problema 1.** En una clase hay 100 estudiantes, de los cuales 52 son chicas. Sabemos que 35 estudiantes usan gafas y que hay 33 estudiantes que son chicos y no usan gafas. Si elegimos a un estudiante de esta clase y usa gafas, ¿qué probabilidad hay de que sea chica?

**Problema 4.** 95% of la población no padece de tuberculosis. Se ha desarrollado un A test para diagnosticar esta enfermedad. Las evidencias mostraron que el 90% of los resultados fueron negativos, y que el 3% de la población padece tuberculosis y dieron

positivo en el test. Si un ciudadano no padece de tuberculosis, ¿qué probabilidad hay de que dé negativo en el test?

**Problema 5.** En una clase hay 100 estudiantes, de los cuales 65 aprobaron Matemáticas. Entre los que no aprobaron Matemáticas, el 20% aprobó Filosofía. Al mismo tiempo sabemos que 17 estudiantes aprobaron Matemáticas pero no aprobaron Filosofía. Si un estudiante de este curso es elegido al azar, ¿qué probabilidad hay de que haya aprobado Filosofía?

**Problema 6.** En una población que padece una enfermedad infecciosa, el 60% ha sido tratado con un antibiótico nuevo. Entre los que no han sido tratados con el antibiótico nuevo, el 15% se ha curado. Además, hay un 23% de la población que no ha sido tratada con el antibiótico nuevo y no se ha curado. Si se escoge una persona de esta población al azar, ¿qué probabilidad hay de que se haya curado?

### Referencias a los autores

M. Pedro Huerta<sup>1</sup>, Universitat de València (España). [manuel.p.huerta@uv.es](mailto:manuel.p.huerta@uv.es)

Joaquín Arnau<sup>2</sup>, Colegio Pío XII (Valencia, España). [ximo\\_arrededors@hotmail.com](mailto:ximo_arrededors@hotmail.com)

## **Conditional and joint probabilities in conditional probability problem solving.**

M. Pedro Huerta, Universitat de València (España).

Joaquín Arnau, Colegio Pío XII (Valencia, España).

In this research we are interested in studying how students use and apply the relationships between probabilities along the process of solving a particular family of school probability problems. In particular, we pay attention to the relationship between conditional probabilities and joint probabilities in the context of solving basic and common conditional probability problems, which often appear in textbooks.

Our starting point is the suggestion made many years ago by Pollasek et. al (1987), affirming that many people (and also students) have an amalgam of ideas related to conditional probability, mainly with joint probabilities, but even with other conditional probabilities or simple probabilities. Some years later, this paper tries to show that this idea is still strongly present at schools, and how the students' performance when solving conditional probability problems depends on the task variables considered.

A set of 12 conditional probability problems, structurally isomorphic in pairs, were designed in order to investigate whether task variables, such as data structure, context and data format, could be influential factors in using and interpreting conditional probabilities in a proper, vague or confuse sense. A sample of 242 students tried to solve the aforementioned problems organized in six questionnaires with five or six problems each, depending on the participants' school level. Secondary School Students, aged 15-18, and primary and secondary school teacher students form our sample. This heterogeneity of the sample also allows us to investigate whether the hypothesized amalgam of ideas depends or not on the students' level.

We have found that students can misinterpret and misuse conditional probabilities in every stage in the process of solving a problem: while they are reading analytically the problem or when they are using them for calculating new probabilities. This fact may be one of the most relevant causes explaining why these problems are generally so difficult for most students. In general, in those stages, conditional probabilities are mostly replaced by joint probabilities and we have not found big differences in students' performances and levels. Therefore, in the reading stage, these confusions have a strong conceptual character and misinterpretations are mainly influenced by the context and data format for the quantities in verbal problems. For example, we have also found that in contexts such as urns or related to social aspects, difficulties in interpreting accurately a conditional probability are lower than in health contexts. We have also found that, given a problem, in which more than one conditional probability has to be meaningfully interpreted and used for solving the question, a student can accurately give a proper sense to some of these probabilities and not to others. Sometimes, not only the contexts and data format influence these misinterpretations: the calculation rules in resources such as 2x2 tables and tree diagrams can also force students to misuse conditional probabilities as joint probabilities or, on the contrary, joint probabilities as conditional probabilities.