M. Pedro Huerta, Patricia I. Edo, Rubén Amorós, Joaquín Arnau

UN ESQUEMA DE CODIFICACIÓN PARA EL ANÁLISIS DE LAS RESOLUCIONES DE LOS PROBLEMAS DE PROBABILIDAD CONDICIONAL

Encoding scheme for the analysis of resolutions of conditional probability problems

RESUMEN

En este trabajo proponemos un esquema de variables para el análisis de las resoluciones de los problemas de probabilidad condicional. Se consideran tres variables principales: las variables del enfoque, del proceso y del producto o resultado, las cuales describen aquello que el resolutor hace desde que lee el enunciado del problema hasta que produce un resultado. Estas tres variables principales pueden ser explicadas por medio de no menos de 15 variables secundarias, la mayoría de ellas dicotómicas.

PALABRAS CLAVE:

- *Resolución de problemas*
- *Probabilidad condicional*
- *Variables del enfoque*
- *Variables del proceso*
- *Variables del producto*

ABSTRACT

In this article, a scheme of variables is proposed for analysing resolutions of conditional probability problems. Three main variables are considered, which we call: the approach variable, the process variable and the product variable. These variables try to describe solvers' behaviour since the moment they start reading the text of a problem up to the moment they give an answer. At least 15 secondary variables, being most of them dichotomous, can be defined to explain those main variables.

KEY WORDS:

- *Problem solving*
- *Conditional probability*
- *Approach variables*
- *Process variables*
- *Product variable*

RESUMO

Neste trabalho, propomos um esquema de variáveis para a análise das resoluções de problemas de probabilidade condicional. Consideram-se três variáveis principais: a variável do enfoque, a do processo e a do produto ou resultado, as quais descrevem aquilo que o resolvedor faz desde que lê o enunciado do problema até que produz um resultado. Estas três variáveis principais podem ser explicadas por, pelo menos, qinze variáveis secundárias, a maioria delas dicotómicas.

PALAVRAS CHAVE:

- *Resolução de problemas*
- *Probabilidade condicional*
- *Variável do enfoque*
- *Variável do processo*
- *Variável do produto ou resultado*

Revista Latinoamericana de Investigación en Matemática Educativa (2016) *19* (3): 335-362.
Recepción: Septiembre 18, 2013 / Aceptación: Octubre 27, 2014. DOI: 10.12802/relime.13.1934

RÉSUMÉ

Dans ce travail, nous vous proposons un système de variables pour l'analyse des résolutions des problèmes de probabilité conditionnelle. Trois principales variables sont considérées: les variables de l'approche, les variables de processus et les variables de produit, lesquelles décrivent ce qui fait le sujet depuis qu'il lit l'énoncé du problème jusqu'à ce qu'il produise un résultat. Ces trois variables principales peuvent s'expliquer par pas moins de 15 variables secondaires, la plupart d'entre elles dichotomiques.

MOTS CLÉS:

- *Résolution de problèmes*
- *Les probabilités conditionnelles*
- *Les variables d'approche*
- *Les variables de processus*
- *Les variables du produit*

1. Introducción

Desde que la probabilidad y la estadística forman parte de los programas escolares y su presencia en los currículos se ha generalizado en todo el mundo, el número de investigaciones y el interés por abordar aspectos sobre su enseñanza y aprendizaje han ido en aumento (Jones & Thornton, 2005). En particular, la probabilidad ha pasado de ser objeto de investigación casi exclusivo de la psicología a ser objeto de interés para los educadores matemáticos. Uno de los objetos más estudiados, tanto por la psicología como por los educadores matemáticos, es la probabilidad condicional. Para la psicología el centro de interés ha sido el estudio de los sujetos y los mecanismos que estos ponen en marcha para emitir juicios subjetivos sujetos a incertidumbre sobre determinados sucesos. Digamos que, en el pasado, se han dedicado muchos esfuerzos con el fin de averiguar los mecanismos por los cuales los sujetos asignan una cantidad como la medida objetiva o subjetiva de la realización de un suceso. Pero el estudio de estos procesos cognitivos no puede quedarse ahí y ser ajeno al estudio del comportamiento que tienen los sujetos (ahora resolutores de problemas) cuando la probabilidad condicional, siendo una cantidad, se ve implicada en una red compleja de relaciones con otras cantidades, esto es, cuando esas cantidades y sus relaciones son utilizadas en un contexto de resolución de problemas. Desafortunadamente, hay muy pocos trabajos de investigación que aborden esta problemática, tal vez por falta de tradición o por no disponer de una metodología adecuada que permita su estudio sistemático. En cambio gran cantidad de estudios utilizan a menudo problemas ampliamente conocidos y usados por los investigadores, por ejemplo The Taxi Cab Problem o The Disease Problem, considerados ya en Tversky y Kahneman (1982), centrándose en medir el éxito en la obtención del resultado correcto, sin informar sobre aspectos que tienen que ver con el proceso de resolución (estrategias de resolución, errores y dificultades), y su dependencia de aspectos relacionados con la tarea propuesta.

Por el contrario, nosotros estamos interesados en lo que ocurre durante el proceso de resolución del problema y no solo en el éxito o el fracaso. Es decir, estamos interesados en la observación, en detalle, del comportamiento de un resolutor a lo largo del proceso de resolución y de la influencia que pudieran tener las variables de la tarea, en el sentido de Kulm (1979), en dicho comportamiento. Para ello, se necesita disponer de un número suficiente de variables explicativas que faciliten la identificación de estrategias de resolución con éxito, errores y dificultades. Estas variables serán, a buen seguro, dependientes de algunas o de todas las variables de la tarea consideradas. Éste es el objetivo principal de este trabajo: proporcionar un conjunto, posible y viable, de variables dependientes que permitan explicar la actuación de cualquier resolutor, así como una manera de codificación de las mismas para su posterior tratamiento, ya sea cualitativo o cuantitativo.

2. Antecedentes

Hasta donde nosotros sabemos, muy pocos artículos abordan de un modo sistemático y metódico la investigación de aspectos relacionados con la resolución de problemas de probabilidad, en los que se muestre la metodología de investigación seguida y en la que se citen las variables que se han considerado y la manera en la que se han codificado para su posterior tratamiento. Una investigación así, que comparte con la nuestra elementos teóricos, como el modelo de fases de Polya, y la consideración de variables dependientes, en el sentido de Kulm (1979), se puede encontrar en Corter y Zahner (2007) y Zahner y Corter (2010). Lo que nos separa de estos investigadores es, de una parte, los problemas sobre los que se investiga: problemas de probabilidad de cualquier tipología (incluido algún caso de probabilidad condicionada) en Zahner y Corter (2010) y los *problemas ternarios de probabilidad condicional* (Cerdán y Huerta, 2007), escolares y de enunciado verbal (Lonjedo, Huerta & Carles, 2012) en nuestras investigaciones; de otra, la codificación de las variables que describen el proceso de resolución, que en Zahner y Corter (2010) se centran en los sistemas de representación y en los métodos de cómputo de la probabilidad pedida, y que en el nuestro se considera el proceso completo de resolución de los problemas, en relación con todas sus fases, desde que el estudiante se enfrenta al problema hasta que redacta la respuesta a la pregunta formulada. Otras investigaciones que, por ejemplo, abordan la comprensión de la probabilidad condicional se limitan a considerar aspectos puntuales de las resoluciones, generalmente el éxito o el fracaso, interpretándolos en términos de alguna de las variables de la tarea propuesta. No sería razonable citar la multitud de investigaciones en las que se da este fenómeno, que desde la perspectiva de la investigación en resolución de problemas constituye una cierta debilidad metodológica (Huerta, 2009).

La investigación en resolución de problemas comienza con el análisis de aquellos problemas que son objeto de investigación y que proporcionan las variables independientes. Así, hablamos de problemas que llamamos *ternarios de probabilidad condicional* y que se caracterizan por ser problemas escolares, de probabilidad condicional, de enunciado verbal, formulados utilizando el menor número posible de cantidades conocidas, tres, y que permiten proporcionar una respuesta a una cantidad desconocida por la que se pregunta. Todas estas cantidades son convenientemente escogidas entre probabilidades marginales, conjuntas o de intersección y condicionales (Cerdán y Huerta, 2007; Huerta, 2014). En estos problemas, siempre se menciona al menos una probabilidad condicionada, ya sea como cantidad conocida o desconocida o como ambas. Más adelante, en el apartado 4 pueden verse ejemplos de estos problemas.

Son problemas que se clasifican en cuatro familias y veinte subfamilias así que proporcionan una primera variable independiente asociada a la tarea: la variable estructura (Huerta, 2009). Esta variable da cuenta del tipo de datos conocidos y desconocidos con el que se formula el problema.

Como problemas verbales, siempre están formulados en algún contexto particular (Carles & Huerta, 2007) del que se insta a una situación más general como las descritas en Henry (2005). El contexto, como variable independiente de la tarea, siempre habrá que tenerlo en cuenta para el análisis de las actuaciones de los resolutores, como en Carles, Cerdán, Huerta, Lonjedo y Edo (2009).

Suele decirse que los resolutores abordan mejor las tareas en las que se implican probabilidades condicionales cuando los datos conocidos se expresan en términos de frecuencias condicionales y porcentajes (Jones, Langrall & Mooney, 2007; Watson & Kelly, 2007) o frecuencias naturales (Hoffrage, Gigerenzer, Krauss & Martignon, 2002). Siendo probablemente el formato de expresión de las cantidades conocidas y desconocidas un factor influyente en el éxito en la resolución de problemas de probabilidad condicionada, como informan Huerta y Lonjedo (2007), está por saber si también tiene cierta influencia en otras fases del proceso. La variable formato de datos, pues, es otra de las posibles variables independientes a tener en cuenta.

Por otra parte, resolver un problema es un proceso. Polya ya lo describió, desde el resolutor ideal, mediante lo que es conocido como el modelo de las cuatro fases. Puig y Cerdán (1988) parten de él y lo reinterpretan para la familia de problemas aritméticos de enunciado verbal, considerando el modelo ampliado a seis: lectura, comprensión, traducción, cálculo, solución y revisión-comprobación.

Dado que los problemas de probabilidad pueden verse de alguna manera, en algún instante de la resolución, como una clase particular de problemas aritméticos-algebraicos, las fases consideradas por Puig y Cerdán son reinterpretadas para los problemas de probabilidad condicional, como veremos en el apartado siguiente.

3. El esquema de variables dependientes y las fases

En Kulm (1979) se presenta un esquema teórico mostrando cómo las variables de la tarea pueden influir sobre el proceso de resolución de un problema dividido en fases. Entonces, aceptando esta posibilidad teórica, es posible considerar un conjunto de variables dependientes con las que describir todo el proceso de resolución del problema. Este conjunto de variables dependientes forma parte de las que allí se citan como variables del proceso y del producto. Por medio de los valores que estas variables puedan tomar será posible describir el comportamiento de un resolutor dado, desde que se le proporciona la tarea de resolver un problema hasta que da por terminada ésta, proporcionando, o no, una respuesta a la pregunta formulada. Lo que es particular aquí, para los problemas de los que hablamos, es que aquello que se observa es el uso que hace el resolutor de las probabilidades conocidas en el problema y de sus relaciones con el fin de dar una respuesta a lo que se pregunta, otra probabilidad, siendo siempre una de ellas una probabilidad condicionada.

Como, por lo general, la información de la que se va a disponer consiste en una resolución escrita, conviene descomponerla comparándola con aquello de lo que dispondríamos por escrito de una resolución ideal que recorriera las fases (Huerta y Arnau, 2013). Lo que significa que no toda resolución podrá explicarse por la totalidad del conjunto de las variables que citaremos, pues podría haber resoluciones reales que no las recorrieran todas. Así, por ejemplo, podríamos encontrar resoluciones en las que la respuesta a la pregunta del problema fuera una simple asignación subjetiva a la probabilidad preguntada, sin que para ello hubiera habido en el proceso ninguna fase previa de cálculos intermedios y ni siquiera un cálculo final.

El esquema es fruto, de una parte, de la reflexión teórica anterior y, de otra, de la observación de 960 resoluciones de estudiantes de 15-16 años y 651 resoluciones de estudiantes graduados en Matemáticas e Ingenierías resolviendo problemas de probabilidad condicional en los que las variables de la tarea (independientes) que se han considerado tienen que ver con el contexto, la estructura de datos y el formato de expresión de éstos. Así, podemos considerar tres conjuntos de variables dependientes cuyos valores permiten describir lo que ocurre, ya sea a lo largo de todo el proceso de resolución o bien en alguna de las fases en las que éste puede descomponerse y que pueda interesar observar con más detenimiento (Figura 1).

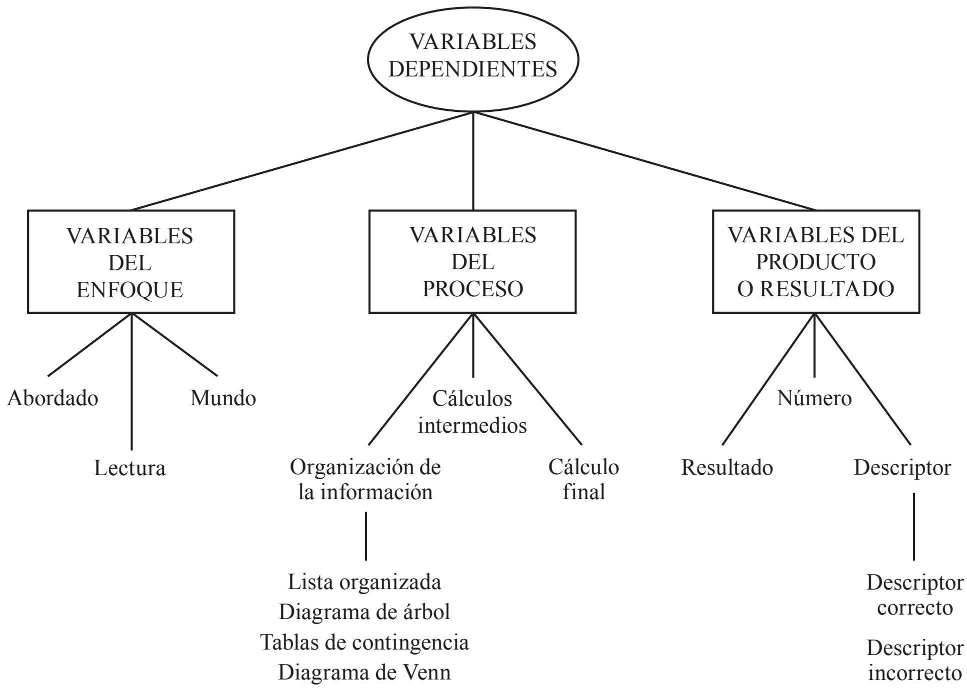


Figura 1. Esquema de variables dependientes para el análisis del proceso de resolución de un problema de probabilidad condicional

Si bien entraremos en el detalle en la próxima sección, en la Figura 1 se ponen en relación las variables que se pueden considerar con las distintas fases del proceso de resolución. El proceso de resolución, llevado a cabo por un resolutor ideal, se corresponde con una lectura de izquierda a derecha de las variables descritas en la Figura 1. Explicaremos, a continuación, estas fases por las variables consideradas, señalando en cursiva en qué fase podrían observarse.

a) Variables del enfoque. Conjunto de variables que pretende identificar las decisiones previas que el resolutor toma cuando aborda un problema particular. Estas variables son:

1. Abordado: variable que indicará si el resolutor ha abordado la resolución de un problema particular o no.

2. Lectura: variable que informa de la presencia o no del álgebra en la resolución.

3. Mundo: variable que determina el protagonismo que tienen las probabilidades en la resolución del problema, si es que lo tienen.

b) Variables del proceso. Este conjunto de variables está compuesto por:

1. Organización de la información. Es usual que la enseñanza proporcione medios para la representación de los datos conocidos en el enunciado de los problemas que facilitan la lectura y comprensión del problema, proporcionando, en algunos casos, además, un plan para su resolución. Pueden ser medios de organización disponibles:

1.1 Lista organizada
1.2 Diagramas de árbol
1.3 Tablas de contingencia
1.4 Diagramas de Venn

Por diagramas de árbol y tablas de contingencia, o tablas 2x2, nos referimos a lo que habitualmente es proporcionado por la enseñanza como recursos para la resolución de problemas. Por lista organizada nos referimos a la presencia en la resolución, generalmente al principio de la misma, de un listado con las cantidades conocidas y desconocidas presentes en el enunciado del problema. Este listado puede ser el resultado de una lectura analítica del problema, es decir, el resultado de una lectura en la que el resolutor se fija solamente en qué cantidades son conocidas y desconocidas. En cualquier caso es la respuesta a una lectura intencionada del texto del problema. Por otra parte, por diagramas de Venn nos referimos a la representación de los datos mediante conjuntos que representan a los sucesos y a sus operaciones, como unión e intersección. Generalmente, el resolutor asocia a estos conjuntos un número que representa la medida del suceso (en frecuencias o porcentajes). Es preciso decir, finalmente, que muchas resoluciones contienen más de una manera de organizar la información, al combinar más de una de las formas que hemos mencionado.

Del análisis de esta parte inicial del proceso de resolución, que se correspondería con las fases de *lectura, comprensión y traducción*, va a depender en gran medida el análisis posterior que se haga. Es aquí donde puede verse si el resolutor interpreta correctamente los datos conocidos, a partir de la forma en que describe los números que aparecen en el enunciado. Es decir, si el resolutor interpreta o malinterpreta las probabilidades conocidas y desconocidas proporcionadas por el problema y si, además, la lectura organizada le proporciona un plan para resolver el problema o no, como se correspondería, por ejemplo, con el uso de los diagramas de árbol, las tablas de contingencia o el uso de expresiones formales.

> 2. Cálculos intermedios. Con esta variable se pretende identificar si el resolutor produce nuevas cantidades, llamadas cantidades intermedias, a partir de las cantidades conocidas inicialmente por el problema. Estas cantidades intermedias no resuelven la pregunta sino que contribuyen a su resolución. Si la lectura que hace el resolutor del problema es algebraica, entonces es posible ver si produce nuevas expresiones algebraicas como cantidades intermedias.

Para esto, el resolutor ha de producir las cantidades con sentido, partiendo de una correcta interpretación de las que ya son conocidas. Es aquí, en esta etapa de la *fase de cálculos*, donde es posible confirmar si las cantidades han sido interpretadas correctamente, pues son usadas para la producción de las cantidades intermedias mediante el establecimiento de las relaciones pertinentes entre ellas.

> 3. Cálculo final. De especial interés cuando en el problema se pregunta por una probabilidad condicional. Con esta variable se pretende observar si el resolutor realiza un cálculo específico para proporcionar una respuesta a la pregunta del problema.

En ocasiones, este cálculo es substituido por una de las cantidades intermedias o por un juicio subjetivo o por cualquier otra cantidad distinta de aquella por la que se pregunta. Es esperable, no obstante, que la respuesta sea producida por un cálculo *ad hoc*, lo que permite a su vez determinar hasta qué punto el resolutor interpreta o malinterpreta la pregunta del problema en función de si la obtiene o no por medio de la relación pertinente entre las cantidades apropiadas. En probabilidad condicional esta observación es totalmente pertinente por la variedad de interpretaciones que los estudiantes proporcionan a este concepto que, de una parte, convierte los problemas en tareas difíciles de resolver con éxito para amplias muestras de estudiantes y, de otra, permite observar algunas de las causas de estas interpretaciones equivocadas que se pueden situar también en el lenguaje particular con el que se expresa.

c) Variables del resultado. Variables que nos permiten observar y analizar lo que el resolutor declara como respuesta a la pregunta formulada por el problema. Esta variable la dividimos en otras secundarias, como sigue:

1. Resultado. Variable que nos indica si el resolutor ha proporcionado una respuesta a la pregunta del problema, sea ésta del tipo que sea, numérica o no.

2. Número. Con este nombre nos referimos a la variable que nos indicará si el resolutor ha dado una respuesta numérica correcta a la pregunta del problema.

3. Descriptor correcto. Con esta variable observamos si el resolutor interpreta correctamente el número dado como respuesta. Esta interpretación la observamos por medio de la descripción verbal que hace de él, es decir, si se dota al número del significado de una probabilidad o una frecuencia condicional o un porcentaje, o no.

4. Descriptor incorrecto. Con la misma intención que en la variable anterior, pero en este caso el resolutor proporciona una descripción verbal que no parece proporcionar al número el significado deseado.

El hecho de observar la variable resultado por medio de ésas cuatro se fundamenta en la noción de cantidad que hemos introducido para los problemas de probabilidad condicional (Edo, Huerta y Cerdán, 2011). Teniendo su origen en el trabajo de Cerdán (2007), definimos cantidad como una terna (x, S, f) en la que x es un número, S es una proposición que describe el número (dicho de otra forma, que x "mide" algo expresado por S) y f el formato con el que se expresa el número. Dependiendo del formato de expresión estamos hablando de frecuencias, porcentajes o probabilidades entre 0 y 1. Es por eso que observamos si la respuesta dada por el resolutor se realiza en términos de cantidades o no, lo mismo que la interpretación y uso de la información proporcionada por el enunciado del problema. Es decir, se observa si, durante la resolución del problema, el problema y el resolutor comparten o no los significados de las probabilidades implicadas y de sus relaciones. Este último conjunto de variables aparecen en la fase de *solución* del problema y *comprobación-revisión*, al exigir del resolutor la respuesta en términos de cantidades.

4. Codificación de las variables. ejemplos

Descritas las variables y su relación con las fases procede sugerir un esquema para su codificación, mostrando ejemplos de cómo se ha hecho.

4.1. *El conjunto de las variables del enfoque*

Recordemos que lo que hemos llamado variables del enfoque de resolución es un conjunto de tres variables: variable abordado, variable mundo y variable lectura. La codificación de las variables que describirán el proceso de resolución del problema puede depender de los valores que tomen éstas.

4.1.1. *Variable abordado*

Esta variable informa sobre si un problema ha sido abordado por un resolutor o no. Este conocimiento permite referir los resultados de la investigación bien a toda la muestra de resolutores a quienes se les proporciona un problema o bien restringirlos a aquellos que lo abordaron. Por otra parte, puede suponerse que si un estudiante no aborda un problema dado es debido a que, después de la lectura del mismo, percibe la dificultad de resolverlo y decide no abordarlo. Esta variable pues nos dará una aproximación de la *dificultad percibida o apreciada* del problema antes de que puedan considerarse otras dificultades. La dificultad apreciada se sustenta en la hipótesis anterior, aunque se es consciente de que otras causas de tipo emocional pueden provocar el abandono de un problema antes de comenzar: cansancio, falta de motivación, frustración, etc.

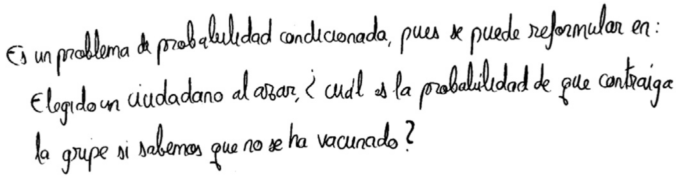
Dependiendo de la muestra de estudiantes, se considerará, en general, que un problema no ha sido abordado si el espacio reservado para la respuesta está en blanco. El investigador puede ser más o menos exigente en cuanto al contenido mínimo que debe tener una resolución para considerar que sí se ha abordado. Siendo una variable binaria, se codificará con 1 si existe algún tipo de registro de escritura numérica, literal o gráfica no tachado. Se codificará con 0 en caso contrario.

4.1.2. *Variable mundo*

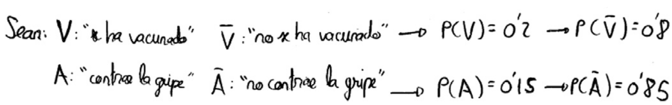
Llamamos así al campo de las matemáticas en el que el resolutor ubica la resolución del problema. Esta distinción entre campos se puede observar por el sistema de signos utilizado en la resolución. En efecto, la resolución de problemas como el que figura a continuación puede tener lugar tanto en el campo de la aritmética como en el de la probabilidad. Ésto depende solamente de cómo sean percibidos por el resolutor. Así, un problema como el siguiente:

> *Problema 1. El 20% de los ciudadanos se vacunan para la gripe común. Por otra parte, el 15% de los ciudadanos contrae la gripe común y un 70% ni se vacuna ni contrae la gripe común. Entre los ciudadanos que no se vacunan, ¿qué porcentaje contrae la gripe común?*

es percibido como un problema de aritmética por la totalidad de los 165 estudiantes de 15-16 años, sin enseñanza previa en probabilidad, a los que se les propuso su resolución (Carles et al., 2009), porcentaje que se reduce al 53,7% en una muestra de 54 estudiantes para profesores de matemáticas de la enseñanza secundaria graduados en Matemáticas, Ingenierías o Arquitectura (Amorós, 2012). Pero, al mismo tiempo un 46,3% de estos futuros profesores percibieron el problema como de un problema de probabilidades:



traduciendo los datos conocidos y desconocidos a probabilidades de sucesos:



Entonces, el sistema de signos y las matemáticas que emplean los resolutores al usar esos sistemas de signos son diferentes y perfectamente identificables, aunque compartan la intención de encontrar una respuesta al mismo problema. Lo que sigue (Figuras 2 y 3) son ejemplos paradigmáticos de resoluciones en ambos mundos.

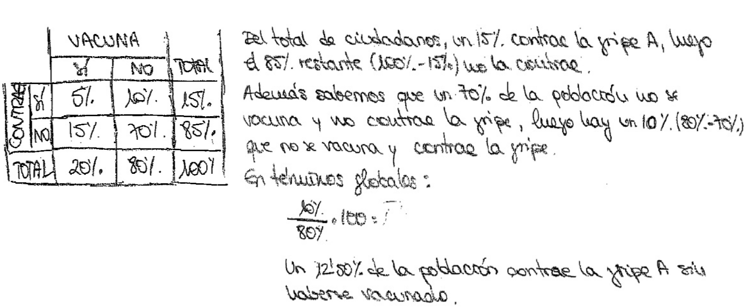


Figura 2. Una resolución aritmética del Problema 1

Mientras que la resolución en el mundo de la aritmética (Figura 2) se ha producido en el mismo contexto en el que se formula el problema, en el que las cantidades intermedias y final se obtienen mediante relaciones aritméticas, la resolución en el mundo de la probabilidad ha requerido de una traducción de las cantidades conocidas y desconocidas en el contexto del problema a cantidades en el lenguaje formal de las probabilidades (Figura 3). Esta traducción puede hacerse explícita o contener los registros suficientes con los que apreciar si el papel de las probabilidades en la resolución del problema es determinante, escaso o nulo. Así, esta variable se codificará atendiendo a este papel protagonista de las cantidades y sus relaciones en alguno de estos dos mundos o bien en una posible "transición" entre ellos (Huerta, 2009).

Hacemos una tabla de contingencia:

	V	$\bar{V}$	
A	5	10	15
$\bar{A}$	15	70	85
	20	80	$\boxed{100}$

Por tanto, $P(A|\bar{V}) = \dfrac{P(A \cap \bar{V})}{P(\bar{V})} = \dfrac{\frac{10}{100}}{\frac{80}{100}} = \boxed{\frac{1}{8}} = 0'125$

Solución: un 12'5%

Figura 3. Una resolución en probabilidad del Problema 1

Además, el enfoque dado a una resolución, en uno u otro *mundo*, e incluso en mundos intermedios, puede ocurrir tanto si en el problema aparecen referencias explícitas a las probabilidades como si no, como ocurre con la versión del problema anterior en la investigación de Arnau (2012):

> *El 20% de los ciudadanos se vacunan para la gripe común. Por otra parte, el 15% de los ciudadanos contrae la gripe común y un 70% ni se vacuna ni contrae la gripe común. Si un ciudadano se vacuna, ¿qué probabilidad tiene de contraer la gripe común?*

Mientras que en la primera versión el problema es percibido por el estudiante como un "problema de aritmética o de probabilidad" y a partir de esta percepción resuelve el problema, en la segunda esta percepción no tiene lugar pues el problema es un “problema de probabilidad” y lo que sí decide el resolutor es abordarlo como tal o bien traducirlo al sistema de signos de la aritmética y resolverlo en este campo.

4.1.3. *Variable lectura*

El término lectura se entiende aquí en el mismo sentido que le da el resolutor a la resolución de un problema dado y que aquí distinguimos como aritmética o algebraica. Lo tomamos prestado del mismo uso que se hace, por ejemplo, en Cerdán (2007) para los problemas de la familia de problemas aritmético-algebraicos. En efecto, como ya se ha dicho, los problemas ternarios de probabilidad condicional pueden considerarse como una subfamilia de aquéllos, aunque, en nuestro caso, implican cantidades y relaciones entre cantidades impregnadas de una fuerte carga conceptual. Así, sus resoluciones pueden acarrear lecturas analíticas teóricas tanto aritméticas como algebraicas a las que los resolutores pueden responder, indistintamente, con lecturas aritméticas o algebraicas (ver Figura 4).

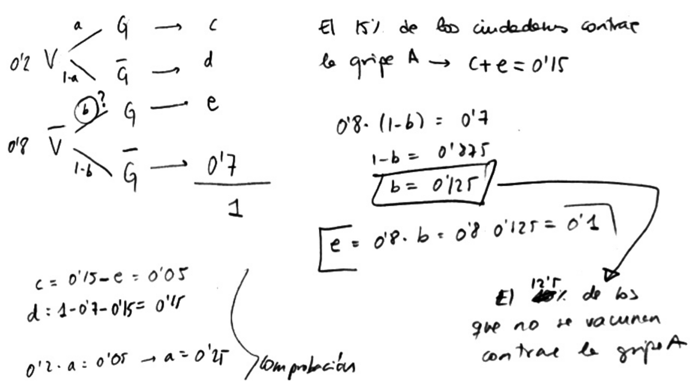


Figura 4. Resolución del problema 1 con lectura algebraica

En efecto, teóricamente hablando, si para resolver un problema solamente son requeridos los datos conocidos, se dice que tiene una lectura teórica aritmética, pero si, por el contrario, es necesario sobredimensionarlo con la incorporación de cantidades desconocidas, entonces tiene una lectura teórica algebraica. Pero un resolutor puede hacer una lectura aritmética o algebraica tanto de un problema que tiene una lectura teórica que es aritmética como de uno cuya lectura teórica es algebraica. Identificarlas y codificarlas, como aritméticas o algebraicas, puede ser determinante para el análisis del comportamiento de un resolutor a lo largo del cuestionario o en la comparación de resoluciones de un mismo problema con una lectura teórica dada. Además, el proceso de resolución del problema es dependiente del tipo de lectura que se realice. Los ejemplos de problemas mostrados hasta ahora tienen lecturas teóricas aritméticas. Compárelas el lector con la que haría del problema siguiente y compruebe que, a diferencia de aquéllas, ésta es algebraica.

> *Entre los ciudadanos que contraen la gripe común un tercio de ellos no se ha vacunado, mientras que de los ciudadanos que sí se han vacunado una cuarta parte ha contraído la enfermedad. Si entre los que no se vacunan el 12'5% contraen la gripe común, ¿qué probabilidad tiene un ciudadano cualquiera de contraer dicha enfermedad?*

La diferencia entre calificar la lectura de una resolución como algebraica o no está en la intención de usar las cantidades desconocidas como conocidas y operar con ellas. Una lectura puede ser aritmética aunque el resolutor use letras como incógnitas auxiliares con el único fin de obtener su valor en una expresión aritmética ternaria en la que dos de las tres cantidades son conocidas. En todo caso se trata de una variable dicotómica con dos posibles valores según sea la lectura aritmética o algebraica.

4.2. *El conjunto de las variables del proceso*

El conjunto de las variables del proceso de resolución lo constituyen las variables que llamamos organización de la información, cálculos intermedios y cálculo final.

4.2.1. *Variables de organización*

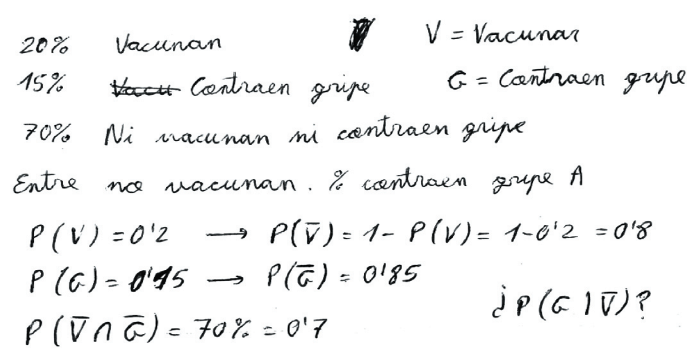
Esta variable está compuesta por otras variables como *lista*, *diagrama de árbol*, *tabla de contingencia* y *diagrama de Venn*. Un vector ordenado de cuatro componentes describirá la variable organización cuyos valores 0 o 1 indicarán, respectivamente, ausencia o presencia en la resolución de un medio de organización u otro.

Aquello que el resolutor expresa de forma organizada como resultado de la lectura ordinaria del texto del problema proporciona al investigador una primera aproximación de la interpretación que hace de las cantidades en el problema.

Esta aproximación deberá confirmarse con el uso posterior de estas cantidades en las variables de cálculo. No obstante, ciertos errores conceptuales, falacias y sesgos asociados con la probabilidad y, concretamente, con la probabilidad condicionada, comienzan a manifestarse en esta fase inicial de la resolución del problema.

4.2.1.1. *Lista*

Decimos que las cantidades se organizan en una lista cuando se expresan como una enumeración de los datos conocidos y desconocidos en el problema, colocados unos a continuación de los otros (casi siempre en el mismo orden en el que aparecen en el texto del problema) y procedentes de una lectura ordinaria, aunque analítica, del enunciado.



20% Vacunan   V = Vacunar

15% ~~Vacu~~ Contraen gripe   G = Contraen gripe

70% Ni vacunan ni contraen gripe

Entre no vacunan. % contraen gripe A

$P(V) = 0'2 \longrightarrow P(\bar{V}) = 1 - P(V) = 1 - 0'2 = 0'8$

$P(G) = 0'15 \longrightarrow P(\bar{G}) = 0'85$

$P(\bar{V} \cap \bar{G}) = 70\% = 0'7$ ¿$P(G \mid \bar{V})$?

Figura 5. Elaboración de dos listas para el problema 1, en la misma resolución

En el ejemplo (Figura 5) la lista se ha producido tras una lectura analítica del enunciado, centrándose el resolutor en las cantidades. Estando ubicada la resolución del problema en el mundo de las probabilidades, la lista contiene, además, la traducción de los datos conocidos y desconocidos al sistema de signos de las probabilidades, expresando las cantidades en este sistema de signos.

Edo (2014) informa que puede producirse un error de cantidad cuando al expresarlas en una lista alguna de las componentes listadas no se corresponde con la cantidad expresada en el enunciado del problema. En el ejemplo siguiente, el error de cantidad se manifiesta en la cantidad desconocida pues el resolutor expresa la probabilidad condicionada preguntada como una probabilidad conjunta por la que el resolutor se interroga.

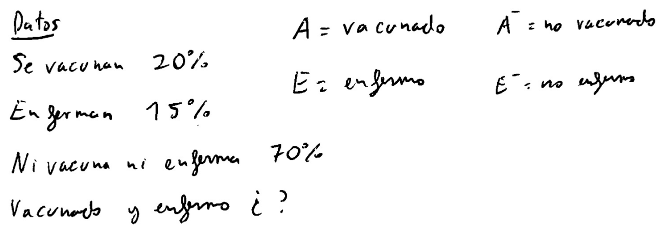


Figura 6. Elaboración de una lista para el problema 1 en la que puede observarse un error de cantidad para la cantidad desconocida (preguntada)

4.2.1.2. *Diagramas de árbol*

Junto con las tablas de contingencia, es habitual que este sistema de representación sea proporcionado por la enseñanza. Canónicamente se dice que las cantidades se expresan en un formato de diagrama de árbol cuando se describe una ramificación del espacio muestral en dos sucesos complementarios y, posteriormente, cada uno de éstos en otros dos sucesos, también complementarios, dando lugar a cuatro caminos que conducen a las cuatro intersecciones posibles. El proceso se considera compuesto de dos pruebas dependientes de tal manera que en la primera parte del diagrama se sitúan las probabilidades marginales y en la segunda las condicionadas. La regla del producto permite determinar las probabilidades conjuntas representadas por los cuatro caminos (véase, por ejemplo, Pluvinage, 2005).

Una organización de la información debería codificarse como árbol siempre que la representación de los datos mostrase, al menos, un análisis de posibilidades para los sucesos (con o sin las medidas de sus probabilidades) y las relaciones entre ellos, o bien una distribución de las cantidades numéricas (usualmente a partir del tamaño de la muestra o espacio muestral) aunque sea de manera incompleta o la representación no responda a la forma canónica. La siguiente representación se codificó como árbol (Figura 7):

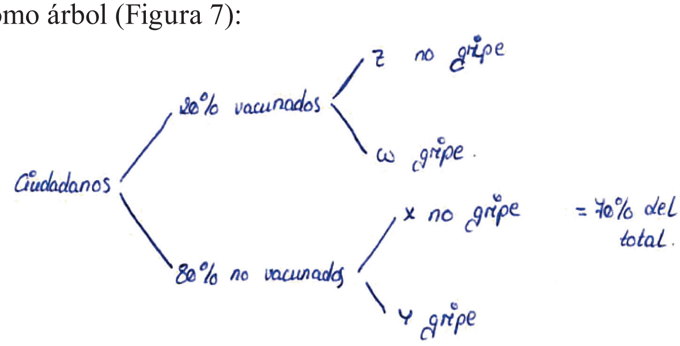


Figura 7. Construcción de un árbol para organizar la información del problema 1

Es preciso indicar que la lectura de las cantidades en un diagrama en árbol conlleva, muchas veces, la inclusión de las cantidades complementarias a las informadas explícitamente en el enunciado del problema, sean estas necesarias o no para la resolución posterior del problema.

La apariencia, por otra parte, no debería ser un factor determinante a la hora de decidir si la información se ha organizado en forma de árbol o no. Puede haber disposiciones en horizontal o en vertical, que usen (o no) líneas para conectar las diferentes pruebas compuestas, completas e incompletos, etc. La decisión de codificar una organización como árbol o no depende de que se cumplan total o parcialmente las características descritas más arriba. La siguiente organización (Figura 8) se codificó como diagrama de árbol en Amorós (2012):

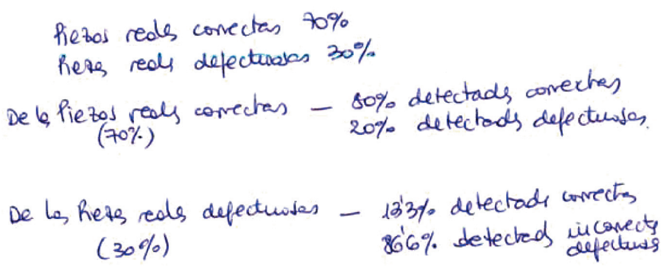


Figura 8. Representación de información en una resolución codificada como diagrama en árbol

4.2.1.3. *Tabla de contingencia*

Llamamos así a una tabla de doble entrada formada por tres filas y tres columnas en las que es posible representar hasta nueve cantidades: cuatro referidas a las cantidades marginales, cuatro a las intersecciones y una al tamaño de la muestra o espacio muestral. Estas tablas suelen usarse en los libros de texto escolares para la formulación de problemas de probabilidad. En estos casos, el papel de resolutor es distinto al del resolutor que usa las tablas de contingencia como una herramienta heurística para resolver el problema. En efecto, en este caso, el resolutor las usa para organizar la información de manera que le ayude a comprenderla y le proporcione además, un plan para la resolución (ver Figuras 2 y 3). En cambio, si esta información ya está organizada en el propio enunciado en formato de tabla, del resolutor se requiere comprensión de la información organizada de esta manera, lo que algunas investigaciones han mostrado que entraña dificultades y no mejora necesariamente la comprensión del problema (Contreras, Estrada, Díaz y Batanero, 2010).

Decimos que la información está organizada en una tabla de contingencia si el resolutor organiza la información en la forma habitual en la que se suele presentar como una tabla 2x2, ampliada o de cualquier otro modo, pero en la que aparezcan representados por etiquetas de cualquier índole los dos pares de sucesos y sus complementarios y, o bien las probabilidades que se enuncian explícitamente en el texto del problema o bien las derivadas de una lectura de las mismas en la propia tabla, como es el caso en el que la probabilidad conocida sea una probabilidad condicionada.

4.2.1.4. *Diagramas de Venn*

Habiendo casi desaparecido de los libros de texto escolares para la representación de conjuntos y operaciones entre conjuntos, todavía es posible encontrar diagramas de Venn en algunos libros de texto para la enseñanza de la probabilidad condicionada y el teorema de Bayes. Aunque su uso como herramienta para la representación de la información es más bien escaso, prácticamente nulo en la enseñanza secundaria, existen resolutores que prefieren representar la información referida a sucesos mediante los diagramas de Venn y sus probabilidades mediante las áreas de las zonas que delimitan dichos diagramas.

Diremos que en una resolución hay presencia de diagramas de Venn si existe una representación gráfica de los sucesos básicos y sus intersecciones mediante cualquier forma geométrica que encierre una superficie (generalmente ovoides, círculos o cuadriláteros, ver Figura 9). En algunas ocasiones se incluyen medidas para las áreas, representando entonces a las probabilidades de los sucesos.

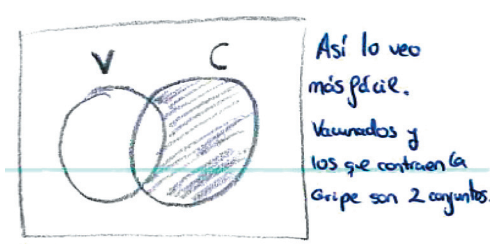


Figura 9. Inclusión de los diagramas de Venn en una resolución

Finalmente, es posible que una resolución contenga más de un medio de organizar la información del problema. En algunos casos, este hecho responde a la consideración de una manera de resolver el problema, un plan que pueden proporcionar las tablas de contingencia y los árboles pero no los diagramas de Venn ni las listas organizadas (Figura 10).

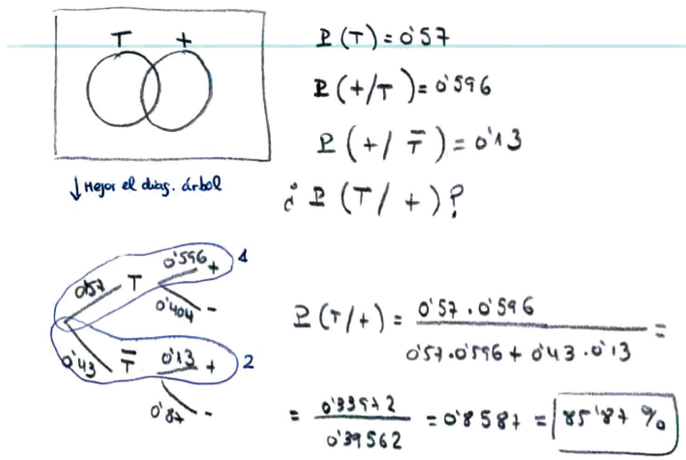


Figura 10. Resolución en la que es posible reconocer hasta 3 maneras de organizar la información de un problema: lista, diagrama en árbol y diagrama de Venn

4.2.2. *Cálculos intermedios*

Teóricamente hablando, el proceso de resolución del problema debería continuar con la determinación de cantidades intermedias, distintas de las conocidas y de la preguntada. Esta variable la descomponemos en otras dos, una que nos indica si hay presencia de cálculos intermedios y otra que nos indica si éstos se han producido con error. Evaluamos así si la respuesta es consecuencia de una serie de cálculos intermedios o no, no siendo extraño que ocurra esto último con resolutores jóvenes fuertemente influenciados por el contexto (Carles et al., 2009).

La variable cálculo se codifica dependiendo de la variable lectura del problema. Si la lectura es aritmética, decimos que en la resolución hay cálculos intermedios si aparece explícitamente el cálculo de al menos una cantidad intermedia, cálculo que está descrito mediante expresiones aritméticas expresadas de cualquier manera en la que se reconozca que la nueva cantidad producida se obtiene mediante su relación con otras dos cantidades ya conocidas. No reconocemos dicho cálculo intermedio cuando la cantidad obtenida sea el producto de una lectura del enunciado del problema, sin que ese cálculo se exprese explícitamente en la resolución, como podría ser el caso de las cantidades que son probabilidades complementarias de las enunciadas en el problema.

Si la lectura es algebraica, decimos que hay cálculos intermedios si las cantidades intermedias se expresan mediante expresiones algebraicas, producto de las relaciones entre cantidades conocidas, estando, necesariamente, alguna de ellas expresada de forma algebraica.

Decimos que hay error de cálculo, tanto si la lectura realizada es aritmética o algebraica, si alguna de las cantidades intermedias calculadas, necesarias para la obtención de la solución del problema, es errónea. Será errónea o bien porque es producto de alguna relación entre cantidades que es errónea o bien porque, aun siendo válida la relación que se establece, las cantidades que se relacionan son erróneas. Edo (2014) les llama, respectivamente, *error de relación* y *error de cantidad*.

4.2.3. *Cálculo final*

En la resolución de cualquier problema puede aparecer un cálculo específico para la probabilidad preguntada. Este hecho es muy acusado en aquellos problemas, como el problema 1, en los que se pregunta por una probabilidad condicionada. Así, parece apropiado explorar de qué manera los resolutores contestan a una pregunta de este tipo. Puede hacerse con la ayuda de un par de variables, la que informa de la presencia de dicho cálculo y aquella que informa de si éste se produce con o sin error.

El cálculo final es dependiente de la lectura realizada. Si la lectura es aritmética, se espera que el cálculo de la cantidad desconocida se realice por medio de expresiones aritméticas, mientras que si la lectura es algebraica entonces ese cálculo será consecuencia del establecimiento de una ecuación o sistema de ecuaciones cuya solución o soluciones conduzcan a la determinación de la cantidad preguntada, ya sea directamente o mediante alguna expresión aritmética posterior.

Decimos, entonces, que hay un cálculo final si en la resolución aparece un cálculo explícito o implícito para producir una cantidad final como respuesta a la pregunta (ver Figuras 2, 3), o bien si existen las comparaciones suficientes entre cantidades como para dar lugar a la ecuación o sistema de ecuaciones que resolvería el problema (Figura 4).

Si la lectura es aritmética, entonces el cálculo final que conduce a la respuesta puede ser realizado mediante una proporción (regla de tres), una relación multiplicativa o una relación aditiva e, incluso ser expresado de forma literal. Si la cantidad señalada como resultado final no procede de un último cálculo explícito, sino que ha sido elegido entre alguna de las varias cantidades intermedias disponibles, este cálculo no se considera como cálculo final, habiendo sido codificado en su momento como cálculo intermedio.

Si la lectura es algebraica es suficiente con que en la resolución aparezcan tantas ecuaciones como incógnitas implicadas en la obtención de la cantidad desconocida para que entonces se codifique que hay un cálculo final en esa resolución con lectura algebraica.

Decimos que en el cálculo final hay presencia de error, con independencia de si la lectura es aritmética o algebraica, si o bien las relaciones utilizadas para ello no son adecuadas o bien, aun siéndolo, las cantidades empleadas no pueden relacionarse de manera que conduzcan a la respuesta del problema. Si la lectura del problema es algebraica, la no resolución de la ecuación o del sistema de ecuaciones pertinentes no la calificamos como error en el cálculo final.

4.3. *El conjunto de las variables del producto (resultado)*

El proceso de resolución culmina, razonablemente, con la respuesta a la pregunta formulada, si es que el resolutor proporciona alguna. Si este es el caso, entonces se espera que sea una cantidad expresada mediante las tres componentes que la definen.

4.3.1. *Resultado*

Hemos llamado así a la variable que pretende evaluar la capacidad del resolutor de proporcionar una respuesta a la pregunta formulada. No se pretende evaluar su idoneidad sino el hecho de que se proporcione alguna respuesta. El resolutor suele destacarla de algún modo en su resolución (ver Figuras 2, 3, 4 y 10).

4.3.2. *Número*

Con la variable número pretendemos, ahora sí, evaluar la tipología de respuesta proporcionada, fundamentalmente cuando la respuesta dada es numérica. Distinguimos entre número correcto y número incorrecto, por lo que se asume que un valor (1 o 0, respectivamente) en esta variable indica que el estudiante proporciona una respuesta numérica y que el número proporcionado es o bien correcto o bien incorrecto. Queda vacía si el estudiante no proporciona una respuesta, o bien, si la respuesta es no numérica. Un error de cálculo aritmético o algebraico conducente a una respuesta numérica errónea no debería ser evaluada como número incorrecto si las cantidades y las relaciones entre las cantidades conducentes a la respuesta numérica correcta fuesen las adecuadas para ello.

4.3.3. *Descripción correcta*

Mediante esta variable evaluamos la capacidad del resolutor de dar la respuesta del problema en términos de cantidades, acompañando a la respuesta numérica con una descripción, siendo ésta considera correcta (Figuras 2 y 4). Dado que hablamos de problemas verbales de probabilidad, este descriptor puede ser una expresión verbal que dé cuenta de lo que es medido por el número. Si el resolutor usa el lenguaje simbólico de la probabilidad, se considerará un descriptor correcto el hecho de traducir la expresión simbólica de nuevo al contexto en el que esté formulado el problema y no simplemente la expresión del resultado en dicho lenguaje simbólico.

4.3.4. *Descripción incorrecta*

En el caso en el que el resolutor sí describa el resultado numérico proporcionado, pero lo haga mediante una proposición incorrecta, entonces registramos este hecho mediante el valor 1 en esta variable. Si un resolutor solamente proporciona una respuesta numérica pero no la describe, entonces valoramos las dos variables con 0, indicando así ausencia de descripción del resultado numérico (en Figura 3 y 10).

En resumen, para el análisis de las resoluciones de los problemas de probabilidad condicional de enunciado verbal, este artículo propone tener en cuenta tres grandes variables que hemos llamado del enfoque, del proceso y del producto o resultado. En cualquiera de ellas puede observarse al resolutor en su relación con la probabilidad condicionada, bien cuando es una cantidad conocida en el problema, bien cuando es una cantidad desconocida. La actuación del resolutor durante el proceso dependerá de su correcta interpretación o no, lo que le permitirá establecer, o no, las relaciones pertinentes con otras cantidades: probabilidades marginales, probabilidades conjuntas y otras probabilidades condicionales. Esta actuación puede observarse con cierto detalle mediante el análisis de los resultados de un conjunto de, por lo menos, quince variables secundarias que explican las grandes variables anteriores: el enfoque con el que el resolutor aborda la resolución del problema, cómo lee, organiza e interpreta la información, y si esta interpretación es la esperada o hay presencia de errores de interpretación; si para dar respuesta al problema realiza los cálculos intermedios pertinentes y un adecuado cálculo final, sin la presencia de errores de relación o con ellos, que le permite obtener una respuesta a la probabilidad preguntada y si, finalmente, se responde o no a la pregunta del problema en términos de cantidades, es decir, mediante un número y una proposición que describa este número interpretado como una probabilidad. En este último caso, podemos evaluar la capacidad del resolutor de expresar probabilidades desconocidas y no solamente utilizarlas, como se hace al principio del proceso de resolución del problema.

5. Uso de este esquema para el análisis de las dificultades de los problemas

Asociado con el estudio de las variables que afectan a la resolución de los problema, es tradicional el estudio de sus dificultades (Lesh & Zawojewski, 2007). Así, en Carles et al. (2009); Huerta, Cerdán, Lonjedo y Edo (2011); Amorós (2012) y Arnau (2012) se introduce la idea de dificultades de los problemas y una forma de medirlas mediante lo producido por los resolutores y analizado con este esquema. Si bien la idea de medir una dificultad por medio de la razón entre el número de fracasos y el número de participantes no es nueva, lo que sí aporta estos estudios es la tipología de dificultades que se miden, fruto del análisis pormenorizado del proceso de resolución mediante las variables que hemos presentado. Además, se estudia si las dificultades dependen de las variables independientes de la tarea consideradas en la formulación de los problemas: la estructura de los datos y el contexto (Carles et al., 2009) e, incluso, del formato de pregunta del problema (Amorós, 2012). Así, teniendo su origen en el trabajo de Cerdán (2007) y tomando, por ejemplo, en consideración las variables Abordado, Resultado, Número, Descripción correcta y Descripción incorrecta, ya descritas, se definen las dificultades: *Dificultad apreciada* (DA), *Dificultad del problema* (DP), *Dificultad de la solución del problema* (DSP) y *Dificultad de la descripción correcta* de la solución del problema (DDRESCP). Estas dificultades informan sobre los problemas, proporcionando así la posibilidad de considerarlas como nuevas variables independientes en futuras investigaciones cuyo objetivo sea observar el comportamiento de los resolutores dependiendo del grado de dificultad de los problemas.

La dificultad apreciada (DA) mide hasta qué punto el problema es apreciado como “difícil” de ser resuelto, pues se basa en la cantidad de resolutores que deciden no abordarlo o abandonarlo prematuramente, antes ni siquiera de intentar organizar la información del problema de alguna manera que nos permita decir si el resolutor ha realizado algún tipo de lectura del problema encaminada a su resolución. Medimos esta dificultad mediante la razón entre el número de resolutores que sí abordan el problema y el número total de resolutores a los que se les propuso su resolución expresándolo en porcentajes, como todas las demás dificultades. Dado que esta dificultad es calculada, deberíamos prestar atención a la confirmación de este hecho: si la dificultad apreciada de un problema es sentida realmente por el resolutor, por ejemplo mediante entrevistas o registros filmados o audio-grabados en los que el resolutor confirme las dificultad que para él tiene el problema y que no le permiten avanzar.

La dificultad del problema (DP), para aquellos que sí abordaron su resolución, mide hasta qué punto es difícil dar una respuesta, del tipo que sea, a la pregunta del problema. Si esto no ocurre es porque razonablemente el resolutor abandona la resolución durante el proceso, en cualquiera de las fases anteriores, dejando el problema sin respuesta. Esta dificultad se mide mediante la razón entre el número de resolutores que no dan respuesta al problema y aquellos que abordaron su resolución, expresándola en porcentajes.

Dada una respuesta, esta puede ser expresada mediante un número o no y éste, a su vez, puede ser un número correcto o no. La dificultad de dar una respuesta numérica correcta a un problema la medimos con la dificultad de la solución del problema (DSP). Esta dificultad puede medirse respecto de aquellos que abordaron el problema, respecto de los que dieron algún tipo de respuesta a la pregunta o a ambas. En todo caso, la medimos mediante la razón entre el número de resolutores que dan una respuesta numérica no correcta al problema y el número de resolutores que lo abordan. Finalmente, asociada a la dificultad de la solución numérica del problema está la dificultad de su descripción. Dado que la descripción de la respuesta, ya sea correcta o incorrecta, puede que no aparezca en la solución de un problema, puede definirse una dificultad para la descripción de la respuesta del problema (DDRESP) y entre las respuestas con descripción una dificultad de que la descripción dada sea la correcta (DDRESCP). Se miden mediante la razón entre el número de resolutores que dan una descripción y los que dan una respuesta al problema (DDRESP), o por la razón entre el número de descripciones correctas y el total de las descripciones dadas (DDRESCP).

Carles et al. (2009) informan de las dificultades elevadas de los problemas de nivel N_0 (problema 1, p. 7) para estudiantes de 15 a 16 años y de la influencia del contexto y de la estructura de datos sobre ellas. Huerta y Cerdán (2010), Huerta et al. (2011), Amorós (2012), Arnau (2012) y Edo (2014) muestran como estas dificultades elevadas se presentan también en estudiantes graduados en Matemáticas o Ingenierías, futuros profesores de matemáticas de la educación secundaria lo que sorprende dada la alta preparación en matemáticas de los estudiantes.

El estudio de las dificultades no acaba en las que se mencionan. Puede extenderse a las dificultades en la realización de cálculos intermedios y de cálculos finales con el fin de analizar en qué fases de la resolución de los problemas los resolutores encuentran nuevas dificultades y a qué son debidas.

6. Conclusiones

El comportamiento de un resolutor de problemas no se puede describir solamente por la respuesta que proporciona a la pregunta formulada, si es que proporciona alguna. Ésta depende, o es consecuencia, de todo un proceso anterior del que debemos ser conocedores antes de obtener conclusiones. Así, en lo que es particular a los problemas de probabilidad condicional, las dificultades y errores en su uso pueden aparecer en cualquier instante del proceso. Es por esto que disponer de una metodología de análisis del proceso completo de resolución de los problemas de probabilidad condicional puede facilitar el análisis de las dificultades de comprensión de este concepto o el de sus interpretaciones equivocadas mostradas por una amplia mayoría de resolutores independientemente de su formación inicial, como lo demuestran algunos de los trabajos empíricos mencionados con anterioridad y que son consecuencia de esta metodología de codificación y análisis del comportamiento de un resolutor real. Comportamiento que puede describirse por medio de tres grandes variables dependientes, que hemos llamado del enfoque, del proceso y del producto, y nueve secundarias que las explican: abordado, lectura, mundo, organización de la información, cálculos intermedios y final, resultado, número y descripción del número. El análisis del comportamiento es consecuencia de las posibles relaciones entre dichas variables y entre ellas y las tomadas como independientes. Es obvio que con estas variables no se agotan las posibilidades, pero pueden tomarse como suficientes para producir una buena descripción y análisis del comportamiento de los resolutores.

Por otro lado, con esta propuesta, se abren nuevas líneas de investigación en el terreno de la educación probabilística al considerar la investigación en resolución de problemas de probabilidad como un tema de investigación que no había sido tenido en cuenta hasta el momento (Huerta, 2014). Además, un mayor y mejor conocimiento sobre el comportamiento de los resolutores en la resolución de los problemas de probabilidad, describiendo competencias, errores, interpretaciones equivocadas y dificultades, puede permitir diseñar unidades de enseñanza, como hace Edo (2014), que mejoren la situación actual de dicha enseñanza que Lonjedo et al. (2012) califican de, al menos, como insatisfactoria.

Referencias bibliográficas

Amorós, R. (2012). *Un ejemplo de análisis de datos mediante la inferencia bayesiana en resolución de problemas de probabilidad condicionada*. Tesis de Maestría no publicada. Universitat de València. Valencia, España.

Arnau, J. (2012). *Un estudio exploratorio de la resolución de problemas de probabilidad condicional centrado en la fase de cálculo.* Tesis de Maestría no publicada. Universitat de València. Valencia, España.

Carles, M., Cerdán, F., Huerta, M. P., Lonjedo, Mª A. y 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 y J. Murillo (Eds.), *Investigación en Educación Matemática XIII* (pp. 173-185). Santander: SEIEM. Obtenido de http://www.seiem.es/publicaciones/actas.htm

Carles, M. & Huerta, M. P. (2007). Conditional probability problems and contexts. The diagnostic test context. In D. Pitta-Pantazi & G. Philippou (Eds.), *Proceedings of the Fifth Congress of the European Society for Research in Mathematics Education* (pp. 702-710). Retrieved from http://www.mathematik.uni-dortmund.de/~erme/CERME5b/WG5.pdf

Cerdán, F. (2007). *Estudios sobre la Familia de problemas Aritmético-Algebraicos*. Tesis Doctoral no publicada. Universitat de València. Valencia, España

Cerdán, F. y Huerta, M. P. (2007). Problemas ternarios de probabilidad condicional y grafos trinomiales. *Educación Matemática*, *19*(1), 27-61.

Contreras, J. M., Estrada, A., Díaz, C. y 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, y T.A. Sierra, (Eds.), *Investigación en Educación Matemática XIV* (pp.271-280). Lleida: SEIEM. Obtenido de http://www.seiem.es/publicaciones/actas.htm

Corter, J. E. & Zahner, D. (2007). Use of External Visual Representations in Probability Problem Solving. *Statistics Education Research Journal*, *6*(1), 22-50. Retrieved from http://www.stat.auckland.ac.nz/serj

Edo, P. (2014). *Estudios sobre los problemas ternarios de probabilidad condicional de nivel N_0 con estudiantes de secundaria (15-16 años).* Tesis Doctoral no publicada. Universitat de València. Valencia, España

Edo P., Huerta, M. P. y Cerdán, F. (2011). Análisis de las resoluciones de problemas de probabilidad condicional mediante grafos. Un ejemplo. En M. Marín, G. Fernández, L. J. Blanco & M. Paralea (Eds.) *Investigación en Educación Matemática XV* (pp. 337-350). Ciudad Real: SEIEM. Obtenido de http://www.seiem.es/publicaciones/actas.htm

Henry, M. (2005). Modélisation en Probabilités conditionnelles. En M. Henry (Ed.) *Autour de la modélisation en probabilités* (pp. 173-185). Presses universitaires de Franche - Comté-Université de Franche-Comté. Obtenu de http://www.univ-fcomte.fr/download/pufc/document/doc_en_ligne/ouvrages_en_ligne/autour_de_la_modelisation_des_probabilites.pdf

Hoffrage, U., Gigerenzer, G., Krauss, S. & Martignon, L. (2002). Representation facilities reasoning: what natural frequencies are and what they are not. *Cognition, 84*(3), 343-352. doi: 10.1016/S0010-0277(02)00050-1

Huerta, M. P. (2009). On Conditional Probability Problem Solving Research − Structures and Context. *International Electronic Journal of Mathematics Education*, *4*(3), 163-194. Retrieved from http://www.iejme.com/

Huerta, M. P. (2014). Researching conditional probability problem solving. In E. J. Chernoff & B. Sriraman (Eds.), *Probabilistic Thinking: Presenting Plural Perspectives* (pp. 613-639). Dordrecht, The Netherlands: Springer. doi: 10.1007/978-94-007-7155-0_33

Huerta, M. P. y Arnau, J. (2013). Fases en la resolución de problemas de probabilidad condicional y variables de investigación. En J. M. Contreras, G. R. Cañadas, M. M. Gea y P. Arteaga (Eds.), *Actas de las I Jornadas Virtuales en Didáctica de la Estadística, Probabilidad y Combinatoria* (pp. 327-334). Granada, Departamento de Didáctica de la Matemática de la Universidad de Granada. Obtenido de : http://jvdiesproyco.es/index.php/actas.

Huerta, M. P. y Cerdán, F. (2010). El cálculo de probabilidades en la formación del profesorado de matemáticas de secundaria. En M. M. Moreno, A. Estrada, J. Carrillo y T. A. Sierra (Eds.), *Investigación en Educación Matemática XIV* (pp. 353-364). Lleida: SEIEM. Obtenido de http://www.seiem.es/publicaciones/actas.htm

Huerta, M. P., Cerdán, F., Lonjedo, Mª. A. & Edo, P. (2011). Assessing difficulties of conditional probability problems. In M. Pytlak; T. Rowland & E. Swoboda (Eds*.), Proceedings of the Seventh Congress of the European Society for Research in Mathematics Education* (pp. 807-817). University of Rzeszów, Poland. Retrieved from http://www.cerme7.univ.rzeszow.pl/WG/5/CERME_Huerta-Cerdan-Lonjedo-Edo.pdf

Huerta, M. P. & Lonjedo, M. A. (2007). The same problem in three presentation formats: Different percentages of success and thinking processes. In D. Pitta-Pantazi & G. Philippou (Eds.), *Proceedings of the Fifth Congress of the European Society for Research in Mathematics Education* (pp. 732-741). Retrieved from http://www.uv.es/lonjedo/documentos/cerme5huertalonjedo.pdf

Jones, G. A., Langrall, C. W. & Mooney, E. S. (2007). Research in Probability (Responding to Classroom Realities). In F. Lester (Ed.), *Second Handbook of Research on mathematic teaching and learning* (pp. 909-956). NCTM: Information Age Publishing.

Jones, G. A. & Thornton, C. A. (2005). An overview of research into the teaching and learning of probability. In G. Jones (Ed.), *Exploring Probability in School (Challenges for Teaching and Learning)* (pp. 65-92). New York, USA: Springer.

Kulm, G. (1979). The classification of Problem-Solving Research Variables. In G. A. Golding & C. E. McClintock (Ed.), *Task Variables in Mathematical Problem Solving*, (pp. 1-22). Retrieved from http://files.eric.ed.gov/fulltext/ED178366.pdf

Lesh, R. & Zawojewski, J. S. (2007). Problem solving and modeling. In F. K. Lester, Jr. (Ed.). *The Second Handbook of Research on Mathematics Teaching and Learning* (pp. 763-804). Charlotte, NC, USA: Information Age Publishing.

Lonjedo, Mª 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.

Pluvinage, F. (2005). Árboles de transiciones etiquetadas en cálculo de probabilidades. *Revista Latinoamericana de Investigación en Matemática Educativa*, *8*(1), 91-99.

Puig, L. y Cerdán, F. (1988). *Problemas aritméticos escolares*. Madrid, España: Síntesis.

Tversky, A. & Kahneman, D. (1982). Evidential impact of base rates. In D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 153-160). Cambridge, USA: Cambridge Academic Press.

Watson, J. M. & Kelly, B. A. (2007). The development of conditional probability reasoning. *International Journal of Mathematical Education in Science and Technology*, *38*(2), 213-235. doi 10.1080/00207390601002880

Zahner, D. & Corter, J. E. (2010). The process of Probability Problem Solving: Use of External Visual Representations. *Mathematical Thinking and Learning, 12*, 177-204. doi: 10.108/10986061003654240

Autores

M. Pedro Huerta. Universitat de València, España. manuel.p.huerta@uv.es

Patricia I. Edo. IES Cova Santa, Segorbe, España. paegual@gmail.com

Rubén Amorós. Universitat de València, España. ruben.amoros.salvador@gmail.com

Joaquín Arnau. Universitat de València, España. joarbre@alumni.uv.es