

Fases en la resolución de problemas de probabilidad condicional y variables de investigación

Huerta Palau, M. Pedro y Arnau Bresó, Joaquín

Universitat de València

Resumen

En este trabajo, basándonos en el modelo de fases de Pólya (1957) para la resolución de problemas de matemáticas y en la interpretación que hacen de ellas Puig y Cerdán (1988) para los problemas aritméticos escolares, mostramos las fases por las que transita un resolutor de problemas ternarios de probabilidad condicional (Huerta, 2009) de enunciado verbal. Por otra parte, basados en una interpretación de las variables de investigación en resolución de problemas de Kilpatrick (1975), citadas en Kulm (1979), mostramos, mediante ejemplos, cómo la identificación y observación con algún nivel de detalle de dichas fases permite definir una serie de variables dependientes para la investigación del comportamiento de un resolutor de problemas de probabilidad condicional. Los problemas que usamos de ejemplo están formulados atendiendo a un conjunto de variables independientes que pueden mostrarse influyentes en la actuación del resolutor, en particular la estructura de los datos, el formato de expresión de los mismos y el contexto en el que se formula el problema, y que les convierte en una tarea difícil para resolutores incluso con una formación alta en matemáticas (Arnau, 2012).

Palabras clave: Probabilidad condicional, problemas ternarios, resolución de problemas, fases en el proceso de resolución, variables del proceso y variables del producto.

1. Introducción

Desde hace algún tiempo, algunos trabajos proponen que la investigación en el terreno del razonamiento probabilístico se sitúe también en el campo de la resolución de problemas de matemáticas (Huerta, 2009), particularmente, el razonamiento implicado en la resolución de problemas de probabilidad condicional.

Vista la resolución de problemas de probabilidad condicional como un proceso, los resolutores, a lo largo del mismo, han de tratar con ella o bien como una cantidad conocida, por tanto informada en el problema, o bien como una cantidad desconocida, y por tanto preguntada. Así que, cuando se plantea un problema a un resolutor de lo que se trata es de enfrentarlo ante situaciones en las que la probabilidad condicionada no sea meramente la respuesta a un solicitado juicio subjetivo sobre un determinado suceso sino que sea como consecuencia del establecimiento de relaciones con otras probabilidades. Si es así, entonces los resolutores han de interpretar la probabilidad condicional y usarla, si es una cantidad conocida, o calcularla si es desconocida. Interesa entonces el significado de uso que los resolutores proporcionan a dicha probabilidad a lo largo del proceso y a su relación con el resto de probabilidades.

Por el hecho de considerar la resolución de problemas como un proceso, su observación puede hacerse, bien como dice Pólya (1957), al describirlo mediante un modelo de 4 fases, bien como dicen Puig y Cerdán (1988) quienes interpretando el modelo de Pólya amplían el número de fases permitiendo así describir el proceso de resolución con más detalle. Optamos por partir de éste último (lectura, comprensión, traducción, cálculo, solución y revisión-comprobación) ya que “la razón por optar por una subdivisión más menuda (de las fases) es que permite poner el énfasis en algunos aspectos del proceso, o

precisar mejor dónde predomina cada tipo de dificultad y, por tanto, hace posible organizar sugerencias para la instrucción en función de ello” (p. 25).

Por consiguiente, parece prometedor analizar la resolución de los problemas de probabilidad condicional con cierto detalle, del modo en que es propuesto en este trabajo, propuesta que es fruto de la experiencia en investigaciones anteriores en las que hemos actuado de este modo. No mostramos resultados de dichas investigaciones, aspectos que, no obstante, pueden encontrarse en las referencias citadas, sino que mostraremos cómo hemos observado las fases en resoluciones escritas y cómo de esta observación han surgido las variables que han servido para su posterior el análisis.

2. Objetivos

Dado que este no es un trabajo que reporte una investigación en particular, nos referiremos a los objetivos de la comunicación que presentamos y que podemos resumir en estos dos:

1. Mostrar cómo, a partir de una resolución escrita, pueden identificarse fases en el proceso de resolución de un problema de probabilidad condicional, y
2. Mostrar cómo la resolución de un problema puede analizarse con la consideración de un conjunto de variables dependientes observadas en cada una de las fases anteriores.

Por tanto, dependiendo de las variables independientes que se consideren ligadas a la formulación de los problemas, por ejemplo, el número de probabilidades condicionales conocidas y/o el contexto y el formato de expresión de los datos, la probabilidad condicionada puede estar presente en todas y cada una de esas fases del proceso de resolución, ya sea de un modo explícito o implícito, en su relación con otras cantidades conocidas del problema. En este sentido, puede examinarse con un mínimo detalle el uso de dichas cantidades, de ahí el interés por describir lo que pasa en cada una de las etapas o fases del proceso de resolución.

3. Elementos teóricos previos

Los problemas a los que nos referiremos en este trabajo son los que Cerdán y Huerta (2007) llamaron problemas ternarios de probabilidad condicional. Éstos se clasifican en 4 familias de problemas, atendiendo al número de probabilidades condicionadas conocidas que pueden considerarse en un enunciado de un problema, y en 20 subfamilias si a cada una de ellas se le hace participar el número de probabilidades marginales y conjuntas adicionales que, hasta un máximo de tres, completan su formulación (Huerta, 2009). La red de relaciones entre las probabilidades condicionales, las marginales y conjuntas necesaria para resolver el problema pueden dar una idea inicial sobre grado de complejidad del problema. Así, el problema que se usará para ejemplificar lo que decimos en este trabajo es ternario, formulado con dos probabilidades marginales y una conjunta, preguntándose al fin por una probabilidad condicional. Además está formulado en un contexto particular en el que las probabilidades conocidas se expresan por medio de cantidades expresadas en porcentajes. En nuestro trabajo, toda cantidad se ve como una terna (x, S, f) en el que x es un número, S una proposición que describe a x (o si se quiere, abusando del lenguaje, x mide a S) y f el formato mediante el cual se expresa el número.

La resolución de un problema puede verse como un proceso que comienza con la lectura de su enunciado y concluye con una respuesta a la pregunta formulada. Muchos autores han tratado de describir como es este proceso, siendo el llamado modelo de las 4 fases de Pólya (1954) uno de ellos y de los más conocidos. No obstante en este trabajo se considerará una visión más fina de este modelo, a partir del propuesto por Puig y Cerdán

(1988), que permite observar el proceso de resolución de problemas aritméticos con un mayor grado de detalle y describir e intuir en qué momento de la resolución de dichos problemas aparecen o pueden aparecer las dificultades en los resolutores. Este modelo presta atención a las fases que los autores llaman de lectura, comprensión, traducción, cálculo, solución y revisión-comprobación. Partimos pues de esta visión de las fases del proceso de resolución, pues vemos la resolución de los problemas ternarios de probabilidad condicional como un caso particular de la resolución de los problemas aritmético-algebraicos en algún momento del proceso, pero con la sustancial diferencia de que en aquéllos las cantidades tienen una fuerte componente conceptual, pues no en balde se refieren bien explícitamente a probabilidades marginales, conjuntas y condicionales bien implícitamente a ellas, siendo en este caso el resolutor quien las interpreta en un sentido probabilístico o no.

En Kulm (1979) puede verse una clasificación de las variables para la investigación en resolución de problemas de matemáticas en variables independientes y variables dependientes. De éstas segundas pueden considerarse lo que en su momento se llamó variables del proceso y del producto, que en brevedad describen lo que el resolutor hace hasta justo antes de dar una respuesta al problema y lo que proporciona como tal. En el trabajo citado también se muestra como estas variables dependientes están relacionadas con las fases de Pólya. Lo que mostramos aquí es como interpretamos esto mismo para una familia particular de problema de matemáticas, los problemas ternarios de probabilidad condicional.

4. Instrumentos

El problema de probabilidad condicional que vamos a considerar como ejemplo en este trabajo es el primero de una serie de problemas ternarios de probabilidad condicional con los que se han elaborado diferentes cuestionarios. Este problema, además, se ha formulado en versiones diferentes, dependiendo de la muestra de resolutores a los que se le ha administrado y de los objetivos particulares de la investigación en la que estaba implicado. Fundamentalmente las versiones difieren en cuanto al formato de expresión de las cantidades conocidas y desconocidas. Aunque, para lo que nos interesa en este trabajo, este hecho no es influyente. Tal vez sí en cuanto a los resultados a los que se pudiera llegar con posterioridad.

Con el fin de hacer nuestra propuesta, usaremos un par de resoluciones escritas que tomadas como ejemplo permitirán observar con algún detalle el proceso de resolución. Estas resoluciones corresponden a dos estudiantes del máster de formación del profesorado de matemáticas de la Universitat de València y se han extraído de la investigación llevada a cabo por Arnau (2012). Estos estudiantes, en el momento de proponerles la resolución de los problemas, no habían recibido enseñanza sobre didáctica de las probabilidades aunque se les supuso suficientemente preparados en este campo de las matemáticas. En el contrato entre el investigador y el resolutor, entre otras cláusulas, se les solicitó que explicasen por qué hacen las cosas como las hacen y que una vez obtenida la solución de un problema interpretasen la respuesta obtenida. En el anexo encontraran dos resoluciones escritas completas del mismo problema, que se identifica como Problema 1.

5. Resultados. De las resoluciones escritas a las variables para el análisis.

Como hemos dicho, en el anexo ~~encontramos~~ pueden verse dos resoluciones del mismo problema por estudiantes distintos. En la primera (Anexo 1) es relativamente cómodo distinguir las diferentes etapas por las que transita un resolutor competente del problema, mientras que en la segunda (Anexo 2) la actuación no puede calificarse de competente, apareciendo errores en la resolución en algún momento o fase de la resolución.

No siempre en una resolución escrita son fácilmente identificables dichas fases. El enunciado del problema se presenta en la Figura 1.

Problema 1
 El 20% de los ciudadanos se vacunan para prevenir el contagio de 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?

Figura 1. Enunciado de un problema ternario de probabilidad condicional (Problema 1) con dos probabilidades marginales y una conjunta conocidas. Se pregunta por una probabilidad condicionada.

En ambos casos, lo que constituye la resolución del problema, y que va a ser objeto de análisis por el investigador, es todo aquello que el resolutor escribe justo debajo de su enunciado hasta que resalta de algún modo lo que constituye para él la respuesta a la pregunta planteada.

En primer lugar (consideremos la primera resolución en Anexo 1) si trazamos una línea horizontal imaginaria justo por debajo de la primera tabla 2x2, en la que el resolutor muestra los datos conocidos, todo lo que hace para leer y comprender el problema se ubica desde el enunciado hasta dicha línea horizontal. En la segunda resolución (Anexo 2) esta fase se ubica en el espacio que queda al trazar una línea vertical imaginaria que deja a la derecha un texto escrito y a la izquierda un gráfico. En esta fase los resolutores, seguramente después de una lectura ordinaria del texto del problema, pretenden hacer una lectura analítica del mismo, es decir, centrarse en los datos conocidos y desconocidos. Consecuencia de esta lectura analítica es que uno decide abordar el problema en el campo de las probabilidades (Figura 2), mientras que el otro pretende una resolución en el propio contexto del problema. En el primero de los casos se muestra una traducción de las cantidades conocidas al sistema de signos de las probabilidades, en el otro caso no. Ambos resolutores hacen una lectura analítica y organizada de las cantidades conocidas. El primero mediante una tabla de contingencia 2x2 (Figura 3) y el segundo mediante un diagrama de Venn (Anexo 2). La enseñanza, no obstante, proporciona otros dos medios organizados para realizar lecturas analíticas: la lista organizada y el diagrama de árbol, medios que no usan ninguna de los dos resolutores.

Este primer análisis de las resoluciones sugiere la posibilidad de considerar dos tipos de variables, la que denominamos variable del *enfoque* y la que denominamos *organización de la información*. La variable del enfoque, inicialmente, puede tomar los valores *probabilístico* o *aritmético*. La variable de organización de la información puede tomar los valores: lista, tabla, árbol y diagrama de Venn, como más habituales. En efecto, puede darse el caso de que un resolutor pueda necesitar de todos estos medios de organización para interpretar correctamente la información contenida en el enunciado de un problema.

| | |
|--------------------------------------|---------------------------------|
| $P(G)$ = probabilidad gripe | $P(V)$ = probabilidad vacunados |
| $P(\bar{G})$ = probabilidad no gripe | \bar{V} = no vacunados. |

Figura 2. Sistema de signos con el que el resolutor identifica las descripciones de las cantidades (básicas) implicadas en el enunciado del problema.

Con estos datos, voy a crear una tabla de contingencia para contener todas las probabilidades y obtener las incógnitas.

| | G | \bar{G} | |
|-----------|----|-----------|-----|
| V | | | 20 |
| \bar{V} | | 70 | |
| | 15 | | 100 |

Figura 3. Tabla de contingencia 2x2 que sirve en la resolución para identificar las cantidades conocidas y desconocidas y darles un significado concreto derivado de su posición en la tabla.

Si ahora trazamos otra línea imaginaria entre la anterior y lo que parece ser la respuesta dada a la pregunta del problema, vemos otra fase en el proceso de resolución que llamamos fase de cálculos. Ésta, a su vez, la podemos distinguir en otras dos: los cálculos intermedios y el cálculo final conducente a la respuesta del problema. En la primera resolución esta distinción es muy visible, pues la primera fase de cálculos intermedios (Figura 4) comprende lo que se escribe entre las tablas 2x2 (ambas incluidas, pues no deja de ser éste el plan que el resolutor ha considerado para la resolución del problema) y el cálculo final (Figura 5) que abarca desde que el resolutor se plantea la pregunta del problema y si ésta la puede encontrar al relacionar las cantidades conocidas de que dispone y cómo hacerlo. Esto mismo es apreciable en la segunda resolución, pues hasta que el resolutor responde a la pregunta del problema obtiene, mientras, las cantidades intermedias que considera necesarias para ello. En este caso, sin embargo, la lectura organizada del problema mediante un diagrama de Venn no parece proporcionar al resolutor un plan de resolución, sino una comprensión de las cantidades conocidas y sus relaciones.

Vemos también que la observación con algún detalle de esta fase de la resolución del problema nos ha proporcionado nuevas variables para su consideración. De una parte, la presencia o no de cálculos en la resolución mediante la consideración de lo que denominamos variable *cálculos intermedios* y *cálculo final*. Puede pensarse que todo resolutor proporcionará una respuesta al problema que será una cantidad calculada. Seguramente que el lector apreciará que este problema admite una respuesta subjetiva, independientemente de los datos proporcionados por el problema, asunto que ha sido ampliamente investigado por la psicología dando lugar a los conocidos sesgos en la probabilidad condicional (la relación causal si alguien se vacuna no enferma). Claro que los cálculos pueden estar afectados de error, por lo que a estas dos variables siempre se le acompaña las variables que identifican la presencia o no de errores en dicho cálculo.

Ahora paso a rellenar los datos incógnita simplemente por sumatorio. Pues la probabilidad total es el 100% o sea 1.

| | G | \bar{G} | |
|-----------|----|-----------|-----|
| V | 5 | 15 | 20 |
| \bar{V} | 10 | 70 | 80 |
| | 15 | 85 | 100 |

Figura 4. Tabla de contingencia 2x2 que sirve en la resolución para calcular las cantidades intermedias dándoles el significado concreto derivado de su posición en la tabla.

→ la pregunta es, si un ciudadano se vacuna (Partido, estamos en un caso de probabilidad condicionada en el que tenemos que aplicar la siguiente fórmula: $P(A/B) = \frac{P(A \cap B)}{P(B)}$

• lo que en este caso sea $P(V/G) = \frac{P(G \cap V)}{P(V)} = \frac{5}{20} = \frac{1}{4} = 0,25$

Figura 5. Fase de la resolución en el que se muestra un cálculo específico y final para responder a la pregunta formulada en el problema y la manera en la que se hace.

Lo que queda sin identificar todavía en la resolución escrita es lo que constituye la respuesta a la pregunta formulada por el problema, si es que el resolutor ha proporcionado alguna. El resolutor suele destacar ésta entre el conjunto de las cosas escritas y que constituyen la solución proporcionada (Figura 6). Otras veces ha de ser el investigador quien determine si ha habido respuesta a la pregunta formulada y qué constituye ésta. Como el problema pide una respuesta a una cantidad desconocida, se espera que el resolutor conteste en términos de cantidades, es decir, mediante las tres componentes de la cantidad: numérica, descripción y formatos adecuados. Responder en estos términos le permitirá interpretar la respuesta dada, si es el caso. Puede observarse (Figura 6) como en la primera resolución la respuesta es la cantidad (25, contraen la gripe si se vacunan, %), mientras que en la segunda (Anexo 2) es (0,05, vacunarse y coger gripe, probabilidades).

Si se vacuna tiene un 25% de contraer la gripe. 25%

Figura 6. El resolutor resalta lo que es la respuesta a la pregunta y la interpreta mediante una cantidad.

Así pues, de nuevo, prestar atención a la fase de solución del problema nos ha proporcionado la posibilidad de analizarla desde la consideración de nuevas variables. Así, mediante la variable que llamamos *respuesta* identificamos la presencia de aquellos resolutores que dan o no una respuesta a la pregunta del problema, sea ésta del tipo que sea: calculada o asignada. La manera en la que se expresa esta respuesta la analizamos mediante la variable que llamamos *número* que identifica respuestas numéricas y, a la vez, si éstas son correctas o incorrectas, y las variables *descripción correcta* o *descripción incorrecta* (del número) que determina si hay descripción del número y cómo la calificamos.

6. Conclusiones.

En este trabajo hemos mostrado como el proceso de resolución de problemas de probabilidad condicional es un proceso complejo que puede recorrer varias etapas. Algunos estudios han mostrado como estos problemas son una tarea difícil a la que se enfrentan muchos estudiantes en diferentes niveles educativos (Carles, Cerdán, Huerta, Lonjedo y Edo, 2009; Huerta, Cerdán, Lonjedo y Edo, 2011). Es por eso que reducir su análisis y conclusiones correspondientes basándose en el hecho de si el resolutor proporciona una respuesta correcta o no a la pregunta formulada en el problema, parece cuanto menos que descorazonador por la cantidad de información que se deja de lado y que puede dar luz sobre el comportamiento que conduce al resolutor a proporcionar dicha respuesta. Es, también, una tarea compleja para el investigador que se interesa por estudiar los procesos de resolución de los problemas. Cómo convertir la información en variables para su posterior análisis. Es lo que hemos hecho aquí, mostrando como cualquier estudio sobre la resolución de los problemas de probabilidad condicional implica a un conjunto de, cuanto menos, 14

variables relacionadas, dos describiendo el enfoque de la resolución, cuatro describiendo el proceso de lectura y comprensión del enunciado, otras cuatro describiendo la presencia o no de cálculos que conducen a la respuesta del problema y por fin otras cuatro que dan cuenta de la respuesta a la pregunta formulada y la forma en que es proporcionada. Estas las consideramos básicas y con cierto carácter generalista pues pueden tomarse en consideración independientemente de los sujetos, como hacen por ejemplo, Edo y Huerta (2010) con estudiantes de secundaria y Huerta, et al. (2011) y Amorós (2012) con estudiantes ya graduados, futuros profesores de secundaria. Pero no se agotan en ellas mismas, pues dependen también del interés del investigador en profundizar en alguna de ellas en particular, como hace Arnau (2012) quién investiga no solamente si hay presencia de cálculos en la resolución sino de qué tipo son estos y si están afectados de error.

Referencias

- Amorós, R. (2012). *Un ejemplo de análisis de datos mediante la inferencia bayesiana en resolución de problemas de probabilidad condicionada*. Memoria de investigación. Master de Investigación en Didácticas Específicas. Universitat de València.
- Arnau, J. (2012). *Un estudio exploratorio de la resolución de problemas de probabilidad condicional centrado en la fase de cálculo*. Memoria de investigación. Master de Investigación en Didácticas Específicas. Universitat de València.
- Cerdán, F.; Huerta, M. P. (2007). Problemas ternarios de probabilidad condicional y grafos trinomiales. *Educación Matemática*, 19(1), 27-62.
- Edo, P. & Huerta, M. P. (2010). Estudios sobre los problemas ternarios de probabilidad condicional de nivel N_0 . Comunicación presentada en el grupo de Probabilidad y Estadística. Lleida: SEIEM. Edición en CD.
- Huerta, M. P., Cerdán, F., Lonjedo, M. A., y Edo, P. (2011). Assessing difficulties of conditional probability problems. En M. Pytlak, T. Rowland, y E. Swoboda (Eds.), *Proceedings of the Seventh Congress of the European Society for Research in Mathematics Education*, pp. 807-817. University of Rzeszów, Poland.
- Huerta, M. P. (2009). On conditional probability problem solving research – Structures and context. En M. Borovcnik y R. Kapadia (2009), Special issue on “Research and Developments in Probability Education”. *International Electronic Journal of Mathematics Education*, 4(3), 163-194.
- Kulm, G. (1979). The classification of problem-solving research variables. En G. A. Golding y C. E. McClintock (Eds.), *Task Variables in Mathematical Problem Solving* (p. 1-12). ERIC.
- Polya, G. (1957). *How to solve it*. Princeton N Y: Princeton University Press. (Traducción castellana: *Cómo plantear y resolver problemas*. México: Trillas, 1965)
- Puig, L. y Cerdán, F. (1988). *Problemas aritméticos escolares*. Madrid: Síntesis.

Anexo I

$P(G)$ = probabilidad Gripe
 $P(\bar{G})$ = probabilidad no Gripe
 $P(V)$ = probabilidad vacunado
 $P(\bar{V})$ = probabilidad no vacunado

Con estos datos, voy a crear una tabla de contingencia para contener todas las probabilidades y obtener las incógnitas.

| | | | |
|-----------|----|-----------|-----|
| | G | \bar{G} | |
| V | | | 20 |
| \bar{V} | | 70 | |
| | 15 | | 100 |

Ahora poco a poco voy a rellenar los datos incógnita completamente por sumatorio. Pasa la probabilidad total es el 100% o sea 1.

| | | | |
|-----------|----|-----------|-----|
| | G | \bar{G} | |
| V | 5 | 15 | 20 |
| \bar{V} | 10 | 70 | 80 |
| | 15 | 85 | 100 |

→ La pregunta es, si un ciudadano se vacuna (Piense, estamos en un caso de probabilidad condicionada) en el que hecho que aplica la siguiente fórmula: $P(A/B) = \frac{P(A \cap B)}{P(B)}$

lo que en este caso sea $P(V/G) = \frac{P(G \cap V)}{P(V)} = \frac{5}{20} = \frac{1}{4} = 0,25$

Si se vacuna tiene un 25% de contraer la gripe. 25%

Anexo II

CIUDADANOS VACUNADOS = V
 CIUDADANOS CONTRA GRIPE = G
 CIUDADANOS

TOTAL CIUDADANOS = 100

$100 - 70 = 30$ CIUDADANOS SE HAN VACUNADO O HAN CEGIDO GRIPE

LA SUMA DE LOS VACUNADOS Y LOS QUE HAN CEGIDO GRIPE DEBE SER POR TANTO 80, CON LO QUE:

$20 + 15 = 35$
 $80 - 35 = 45$ SON LOS CIUDADANOS QUE SE HAN VACUNADO Y HAN CEGIDO GRIPE

PROBABILIDAD DE VACUNARSE Y CEGER GRIPE = $\frac{35}{100} = 0,35$