



LA SIMULACIÓN COMPUTACIONAL Y LA PROBABILIDAD CONDICIONAL

July Celis y Gabriel Yáñez
Universidad Industrial de Santander (Colombia)
karinitacelis@yahoo.es, gyanez@uis.edu.co

En esta comunicación presentamos los avances de una investigación que pretende identificar los efectos de la simulación computacional en la comprensión y en la capacidad de resolver problemas de probabilidad condicional de estudiantes universitarios. Se propone la simulación como herramienta para resolver problemas de probabilidad condicional cuando no se conoce la estructura matemática que subyace a esos procesos. Además buscamos argumentos para garantizar que a través de la simulación computacional se puede favorecer la comprensión de procesos subyacentes a la probabilidad condicional, en particular, el razonamiento bayesiano.

PALABRAS CLAVE

Probabilidad condicional, enfoque frecuencial, simulación computacional, resolución de problemas.

INTRODUCCIÓN

Recordemos que la probabilidad condicional da cuenta del efecto que tiene la realización de un evento condicionante en la probabilidad de un evento condicionado. Su enseñanza y aprendizaje se ha realizado tradicionalmente adoptando casi que exclusivamente la expresión algebraica:

$$P(A|B) = \frac{P(A \text{ y } B)}{P(B)} \quad P(B) > 0$$

Esta expresión, envuelve en sí, dificultades cuando se requiere su uso para resolver problemas. No obstante que existen numerosas investigaciones en este campo que reportan las dificultades y las malas concepciones que los estudiantes se forman alrededor de este concepto (Contreras, 2011), no conocemos estudios que intenten explicar las razones por las cuales éstas se generan. Podría pensarse que una de las causas más influyentes sea la adopción de la medida de probabilidad desde el punto de vista clásico o laplaciano sin ninguna referencia a su interpretación frecuencial lo que conlleva un énfasis exagerado en la representación algebraica.

Si bien es cierto que desde hace un tiempo se ha generado entre la comunidad de estudiosos de la estadística educativa el interés por estudiar los efectos que podría tener el enfoque frecuencial de la probabilidad en la comprensión del concepto de probabilidad y de su modelación para resolver problemas prácticos (Serrano, 1996), también es cierto que en lo referente a la probabilidad condicional las investigaciones ya no son tan numerosas (Yáñez, 2003), razón por la cual vale la pena preguntarse



hasta qué tanto sería benéfica para generar en los estudiantes una mejor comprensión y competencia para resolver problemas que se modelen con probabilidad condicional.

Presentamos en este trabajo algunas reflexiones acerca de las bondades que a priori se pueden suponer posee el enfoque de simulación computacional en la comprensión de la probabilidad condicional.

MARCO DE REFERENCIA

A pesar de que la probabilidad condicional es tratada de manera implícita en algunos grados de escolaridad, por medio de uso de problemas que son resueltos a través de diferentes representaciones como tablas de doble entrada o diagramas de árbol, además de requerirse su razonamiento en la toma de algunas decisiones en la vida cotidiana, las investigaciones identifican la prevalencia de intuiciones incorrectas, sesgos de razonamiento y errores de comprensión y aplicación del concepto de probabilidad condicional.

Falk (1986) expone algunos problemas con la probabilidad condicional, uno es la llamada *falacia de la condicional transpuesta*, que se refiere a la confusión de $P(A|B)$ con $P(B|A)$, generado, tal vez, por la imprecisión del lenguaje natural. Cuando se escribe una probabilidad condicional usando la notación matemática, es claro cuál es el suceso condicionante y cuál el condicionado, pero en el lenguaje natural las cosas no siempre son tan claras..

Falk observó este error en problemas de contextos médicos, donde se confunde la probabilidad de que una persona tenga una enfermedad cuando el test diagnóstico le ha dado positivo, con la probabilidad de obtener un test diagnóstico positivo dado que tiene la enfermedad, y mostró cómo este error puede tener consecuencias importantes.

Falk además reporta la denominada falacia del eje temporal en donde el individuo no considera que la probabilidad de un evento que ocurre en la línea del tiempo, en un momento posterior, pueda influir en la probabilidad de ocurrencia de un evento que es anterior en el tiempo.

Konold (1989) describe un experimento de enseñanza con el uso del computador. Entre sus resultados muestra que algunos estudiantes cambian sus interpretaciones probabilísticas después de una etapa de instrucción, en tanto que otros no lo hacen y persisten en la defensa de ideas que se han obtenido antes de sus cursos de probabilidad, en situaciones de incertidumbre, en donde es común en su lenguaje el uso de términos probabilísticos como azar, independencia, probable, suerte.

Entre los resultados famosos de Konold (1991) está *el enfoque en el resultado aislado*, en el que el individuo solo se preocupa del próximo resultado sin referencia a los resultados previos o posteriores, en contravía total con el enfoque frecuencial de la probabilidad.

En la lectura y análisis de Yáñez (2003) se encuentra que la simulación computacional de la probabilidad permite superar algunos de los sesgos o malas concepciones que los estudiantes poseen sobre las secuencias aleatorias o sobre el valor de las probabilidades en experimentos compuestos. Uno de estos sesgos es el *olvido de la*



tasa de base, que los estudiantes que participaron en su investigación dieron señales de poseerlo en el cuestionario diagnóstico pero que después en otras evaluaciones posteriores no se volvió a presentar. Yáñez (2003) lo atribuye al uso de la simulación o las tablas, que implican necesariamente la consideración de la tasa de base.

DESARROLLO DE LA PROPUESTA

Para ejemplificar el trabajo de simulación y visualizar las posibles ventajas que podría tener su implementación en el trabajo de clase asociado con la probabilidad condicional, se plantea la siguiente situación tomada de Falk (1979):

Una urna tiene en su interior dos bolas blancas y dos bolas negras. Se extraen dos bolas sin reemplazar la primera.

- 1) *¿Cuál es la probabilidad de que la segunda bola sea blanca, dado que la primera bola fue blanca?*
- 2) *¿Cuál es la probabilidad de que la primera bola sea blanca, dado que la segunda fue blanca?*

En notación matemática la primera pregunta pide calcular la probabilidad condicional $P(B_2|B_1)$ en tanto que en la segunda se indaga por la condicional inversa: $P(B_1|B_2)$. La solución a la primera pregunta es bien intuitiva pues habiendo sacado ya una bola blanca y teniendo en cuenta que no se sustituye, en la urna quedarán tres bolas, dos negras y una blanca, por lo tanto $P(B_2|B_1)=1/3$.

La segunda pregunta, en cambio, es mucho más exigente, su solución desde el punto de vista algebraico requiere de la regla de Bayes, que no es para nada fácil de comprender, no solo por su complejidad algebraica sino porque se opone a la concepción temporal de la probabilidad condicional que muchas personas asumen: la segunda extracción no puede condicionar al evento que sucede previamente: la primera extracción (Falk, 1986; Gras y Totohasina, 1995). Esta concepción lleva a las personas a asignar a esta probabilidad el valor $1/2$.

La solución algebraica es la siguiente:

$$P(B_1|B_2) = \frac{P(B_2|B_1)P(B_1)}{P(B_2)} = \frac{P(B_2|B_1)P(B_1)}{P(B_1)P(B_2|B_1) + P(N_1)P(B_2|N_1)} = \frac{1}{3}$$

Ahora veamos cómo se resuelve el mismo problema utilizando simulación. Utilizamos el paquete Fathom. Para ello basta definir los atributos que hacen referencia a las dos extracciones,

Primera: `Randompick("B","B","N","N")`

Segunda: `if(Primera = "N") { randomPick("B","B","N")
randomPick("B","N","N") }`

Esta programación da lugar a una tabla de resultados como la que se muestra en la Figura 1. Para estimar $P(B_2|B_1)$ lo que hay que tener claro es que en el enfoque frecuencial la probabilidad se aproxima por las frecuencias relativas asociadas a la realización del evento que se está estudiando. En este caso se trata de contar las veces que se obtiene blanca en la segunda extracción (B_2) cuando en la primera se obtuvo blanca (B_1), es decir, se trata de contar las veces en que estas dos cosas se dan



conjuntamente y calcular las frecuencias relativas respecto al número de veces que sucede el evento condicionante.

Ahora bien, lo que es más interesante es que para estimar $P(B_1|B_2)$ el razonamiento es exactamente el mismo: el numerador es el mismo de la primera pregunta, y como ahora el condicionante es extraer blanca en la segunda (B_2), el denominador es el número de veces que se obtuvo blanca en la segunda extracción. En esencia, desde el punto de vista frecuencial, los dos procedimientos son completamente semejantes, cosa muy distinta en el plano algebraico que exige la utilización del teorema de Bayes para resolver la segunda pregunta. Adicionalmente, obsérvese que este enfoque permite darse cuenta que el asunto del antes o después, asociado a la concepción temporal o la denominada falacia del eje temporal (Falk, 1986); no tiene sustento porque la probabilidad condicional es una frecuencia relativa asociada a las repeticiones favorables a los dos eventos en cuestión sin consideraciones temporales.

CONCLUSIONES

La simulación como estrategia de enseñanza de la probabilidad condicional goza, entre otras, de las siguientes propiedades:

Vivencia paso a paso del experimento lo que permite identificar la condicionalidad y, por consiguiente, la independencia o no entre eventos. La definición de la condicionalidad a través del comando *if* puede permitir con mayor claridad la identificación de los eventos condicionante y condicionado. Además del reconocimiento de las tasas base requeridas para realizar la programación inicial al asignar valores de probabilidad, al definir el espacio muestral y la distribución de probabilidad del problema estudiado.

Al tener los dos atributos, haciendo referencia a las columnas del formato matricial que nos ofrece el software, asociados a la primera y segunda extracción, con los valores de las diferentes opciones posibles en diferentes extracciones, producto de repeticiones del experimento, muy seguramente este permitirá conjeturar, analizar y finalmente ayudará a combatir *el sesgo en el resultado aislado* de Konold (1991) en que el individuo se interesa por un resultado específico, sin acabar de comprender que la probabilidad solo se refleja en los resultados de muchas repeticiones del experimento (ley de los grandes números).

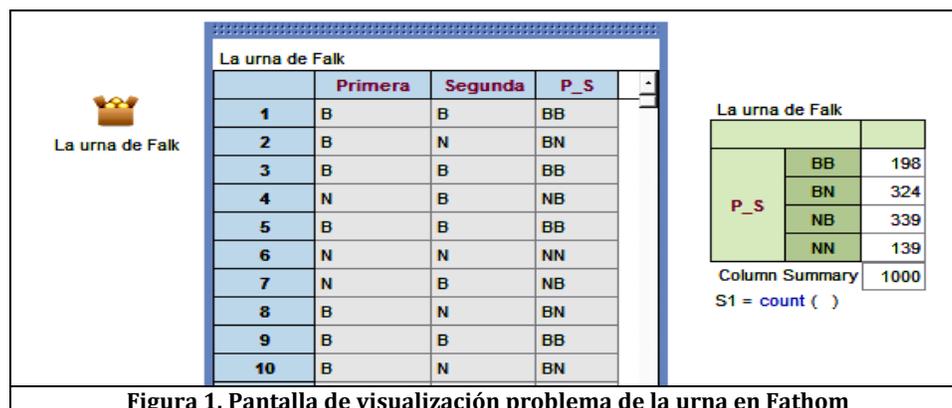


Figura 1. Pantalla de visualización problema de la urna en Fathom



El software ofrece diferentes herramientas que facilitan la solución de un problema, además de que permite diferentes formas y el uso de diferentes representaciones para resolverlo, incentivando la coordinación entre ellas, y permitiendo al mismo tiempo dar cuenta del nivel de comprensión del individuo (Duval, 1999).

En el caso del problema de urnas de Falk (1979), por ejemplo, se observa que la simulación identifica claramente que se trata de un problema de dos dimensiones y que, por consiguiente, las probabilidades conjuntas, mejor dicho, las frecuencias relativas conjuntas, se constituyen en la primera prelación. Con el espacio muestral bivariado construido, no hay lugar para pensar que un evento sucede antes o después y que lo único importante a la hora de calcular probabilidades condicionales es conocer quién juega el papel de evento condicionante y quién de condicionado. Precisamente, en los experimentos diacrónicos, que es donde se presenta la concepción temporal, son los que más fácilmente permiten identificar estos eventos. En otras palabras, en el proceso mismo de la simulación se reconoce la equivalencia que existe entre la versión diacrónica (una bola después de la otra, sin sustitución) y la versión sincrónica (extraer simultáneamente las dos bolas).

REFERENCIAS

- Contreras, J.M. (2011). *Evaluación de conocimientos y recursos didácticos en la formación de profesores sobre probabilidad condicional*. Tesis de doctorado. Granada: Universidad de Granada.
- Duval, R., (1999), *Semiosis y pensamiento humano. Registros semióticos y aprendizajes intelectuales*. Traducción del Instituto de Educación y Pedagogía, en la Universidad del Valle, del original *Sémiosis et Pensée Humaine. Registres sémiotiques et apprentissages intellectuels*. France: Peter Lang.
- Falk, R. (1986). Conditional probabilities: Insights and difficulties. *Proceedings Second International Conference on the Teaching of Statistics* (pp. 992-997). Quebec, Canada: University of Victory.
- Falk, R. (1979). Revision of probabilities and the time axis. *Proceedings of the Third International Conference for the Psychology of Mathematics Education* (pp. 64-66). England: Warwick.
- Gras, R. y Totohasina, A. (1995). Chronologie et causalité, conceptions sources d'obstacles épistémologiques à la notion de probabilité conditionnelle. *Recherche en Didactique des Mathématiques*, 15, (1), 49-55.
- Konold, C. (1991). Understanding students' beliefs about probability. En E. Von Glasersfeld (Ed.), *Radical Constructivism in Mathematics Education* (pp. 139-156). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction*, 6, 59-98.
- Serrano, L. (1996). Significados institucionales y personales de objetos matemáticos ligados a la aproximación frecuencial de la enseñanza de la probabilidad. Tesis de doctorado. Granada: Departamento de Didáctica de la Matemática, Universidad de Granada.
- Yáñez, G. (2003). Estudios sobre el papel de la simulación computacional en la comprensión de las secuencias aleatorias, la probabilidad y la probabilidad condicional. Tesis de doctorado. México: CINVESTAV - IPN.