UN ESTUDIO LONGITUDINAL DEL RAZONAMIENTO BAYESIANO CON ESTUDIANTES DE INGENIERÍA

Cindy Morgado y Gabriel Yáñez
Universidad de Santander (Colombia) - Universidad Industrial de Santander (Colombia)
nathalia.morgado@udes.edu.co, gyanez@uis.edu.co

Se presentan algunos resultados de una investigación realizada para conocer la transformación en el razonamiento bayesiano en estudiantes de ingeniería en un curso semestral de estadística. Para esto se diseñaron y aplicaron tres pruebas en momentos diferentes durante el semestre. Los resultados muestran que los estudiantes aplican las fórmulas aprendidas y utilizan el diagrama de árbol correctamente inmediatamente después de haber estudiado estos temas, pero poco después las fórmulas son utilizadas incorrectamente y el diagrama de árbol solo tiene la función de registrar la información dada por el problema y no la de producir información adicional. La concepción temporal de la probabilidad condicional permaneció en los estudiantes durante todo el tiempo de observación.

PALABRAS CLAVE

Probabilidad condicional, regla del producto**,** teorema de probabilidad total, razonamiento bayesiano, teorema de Bayes.

INTRODUCCIÓN

El razonamiento bayesiano hace referencia al cálculo de probabilidades condicionales inversas mediante el teorema de Bayes:

$$P(A_i \mid B) = \frac{P(B \mid A_i)P(A_i)}{\sum_{j=1}^{n} P(B \mid A_j)P(A_j)}$$

La ocurrencia del evento *B* está condicionada a la realización de algunos de los eventos A_j que conforman una partición del espacio muestral. La idea es que *B* ocurrió y se indaga por la probabilidad de que haya ocurrido A_i, $P(A_i \mid B)$ es la probabilidad a posteriori en tanto que $P(A_i)$ es la probabilidad a priori; los valores $P(B \mid A_i)$ son las verosimilitudes o probabilidad de que *B* ocurra cuando ha ocurrido A_i. Saber cómo utilizan este resultado los estudiantes para resolver problemas asociados con el razonamiento bayesiano motivó la realización de una investigación que pretendió en un semestre conocer la evolución de este razonamiento en estudiantes universitarios que cursaban su primer curso de probabilidad y estadística. En este mismo sentido, Morgado (2013) realizó una investigación con estudiantes universitarios que tomaban un curso de estadística para conocer la evolución que sufre el razonamiento bayesiano a lo largo de un semestre. La investigadora diseñó y aplicó tres cuestionarios en momentos diferentes durante el semestre. Los resultados mostraron que si bien los estudiantes responden mejor los problemas bayesianos inmediatamente después de haber estudiado la regla de Bayes y los temas afines, después de un tiempo esta capacidad se reduce a niveles iguales o inferiores a los mostrados en el primer cuestionario cuando no habían estudiado esos temas. Motivados por continuar este tipo de estudios realizamos una investigación con la misma estructura del trabajo de Morgado (2013) pero con cuestionarios mucho más reducidos y más homogéneos en términos de los problemas propuestos. Algunos de los resultados obtenidos se presentan en este trabajo.

MARCO DE REFERENCIA

Dado que se pretende conocer la forma como los estudiantes confrontan situaciones problemáticas relacionadas con la probabilidad condicional y el teorema de Bayes en diferentes momentos del semestre académico, así como identificar cuáles fueron las intuiciones, falacias o sesgos que permanecieron y desaparecieron como consecuencia del proceso de instrucción, se toma como marco de referencia la literatura en torno a la probabilidad condicional y, en particular, al teorema de Bayes. Uno de los primeros estudios fue el realizado por Kahneman y Tversky, (1972) quienes mostraron que los alumnos en su mayor parte no tienen en cuenta las probabilidades a priori en el cálculo de la probabilidad inversa, sesgo al que llamaron *olvido de la tasa base*. Además, se ha evidenciado que los estudiantes confunden el evento condicionante con el condicionado dando lugar a *la falacia de la condicional transpuesta,* es decir confunden $P(A|B)$ con la $P(B|A)$ (Falk, 1986). De otro lado, Gras y Totohasina (1995) destacan una *concepción temporal* que los estudiantes le adjudican a la probabilidad condicional en el sentido de que el evento condicionante siempre precede en el tiempo al evento condicionado. Ahora, seguramente el mayor problema al resolver situaciones problemáticas asociadas a la probabilidad condicional es la dificultad que existe para distinguir entre probabilidades condicionales y probabilidades conjuntas (Pollatsek *et al*., 1987), así como en la identificación correcta de los sucesos condicionante, condicionado y en la correcta partición del espacio muestral (Díaz y de la Fuente, 2006).

ASPECTOS METODOLÓGICOS

Se desarrolló un estudio a lo largo del tiempo con estudiantes universitarios de ingeniería que estaban realizando su primer curso de probabilidad y estadística con el fin de conocer la forma como los estudiantes confrontan situaciones problemáticas relacionadas con el teorema de Bayes. Se diseñaron tres pruebas con el fin de evaluar el razonamiento bayesiano de los estudiantes, aproximadamente se aplicó una prueba cada cinco semanas. Como la regla de Bayes incluye probabilidad condicional, probabilidad marginal, regla del producto y teorema de probabilidad total, las pruebas se elaboraron de tal forma que incluyeran ítems dirigidos a evaluar cada uno de estos conceptos. Cada prueba constó de tres ítems con el mismo nivel de dificultad. El primer ítem evalúa el razonamiento bayesiano en un contexto de canales. El segundo ítem evalúa el dominio del teorema de Bayes en un contexto de fabricación de máquinas. El último ítem es un problema de urna en un contexto de muestreo sin reposición y consta de cuatro incisos que evalúan la probabilidad condicional directa, probabilidad condicional inversa, la regla del producto para eventos dependientes y el teorema de probabilidad total.

La muestra es de 76 estudiantes de ingeniería industrial de la Universidad Industrial de Santander, Bucaramanga, Colombia, quienes estaban cursando la asignatura de estadística I en el quinto semestre de su carrera en tres grupos diferentes y tres profesores distintos. Las clases eran de tipo magistral: básicamente exponían los temas en el tablero, realizaban algunos ejemplos y resolvían algunos ejercicios de los textos guías (Montgomery y Runger, 2010 y Navidi, 2006). La primera prueba se aplicó en la segunda semana del curso; la segunda se aplicó una semana después de que los profesores enseñaron el teorema de Bayes; la tercera prueba se aplicó en la penúltima semana de clase. Cabe destacar que los profesores consideraron los resultados de estas pruebas para la nota final del curso.

DESARROLLO DE LA PROPUESTA

De los 76 estudiantes de ingeniería, 57 de ellos presentaron las tres pruebas. A continuación, en la Tabla 1 se muestran los porcentajes de éxito de cada ítem.

Contexto	Contenido	Primera Prueba	Segunda Prueba	Tercera Prueba
Canales	Teorema de Bayes	0%	9%	7%
Fabricación de artículos por dos máquinas	Teorema de Bayes	2%	75%	56%
Urna sin reposición	Condicional directa	61%	54%	53%
	Condicional Inversa	14%	2%	4%
	Regla del Producto	11%	40%	44%
	Teorema de Probabilidad Total	4%	11%	23%

Tabla 1. Porcentaje de respuestas correctas según el contenido de los ítems para las tres pruebas de los estudiantes

La Tabla 1 presenta varios hechos que vale la pena destacar:

- La enorme dificultad que representó a los estudiantes los problemas de canales, los porcentajes de éxito siempre fueron inferiores al 10% en las tres pruebas.
- Los altos porcentajes de éxito obtenidos en la segunda y tercera prueba en el ítem relacionado con las máquinas en contraposición con el bajísimo porcentaje de éxitos de la primera prueba. Sin embargo, se observa un descenso de éxitos en la tercera prueba respecto a la segunda.
- En el problema de urna, que se puede considerar libre de las influencias del contextos, se presentaron resultados que llaman la atención: las preguntas de probabilidad condicional fueron mejor respondidas en la primera prueba cuando los estudiantes escasamente conocían la forma clásica de calcular probabilidades, con altos porcentajes las de la probabilidad directa y con bajos porcentajes las de la probabilidad inversa. El ítem que indagaba por una probabilidad conjunta no dejó de aumentar a través del tiempo, cosa similar, aunque con porcentajes inferiores, ocurrió con el ítem que indagaba por una probabilidad marginal en la

segunda extracción. A continuación analizamos las respuestas de los estudiantes en cada uno de los contextos propuestos.

Ítem en un contexto de canales. Este ítem está dirigido a evaluar el teorema de Bayes. La mayoría de los estudiantes en la primera y tercera prueba respondieron ½ (Figura 1-a), es decir, asumieron la pregunta condicional como una conjunta: pase por el canal I y salga por R. En la segunda prueba, donde el canal era algo más complicado, se presentaron respuestas donde asumieron una probabilidad total de 1/2+1/4+1/8 (Figura 1-b) asumiendo que la pregunta indagaba por la probabilidad marginal de que la bola salga por R. En la tercera prueban algunos estudiantes realizaron el diagrama de árbol y nada más; otros estudiantes plantearon incorrectamente la fórmula de Bayes (Figura 1-c). Llama la atención las respuestas de algunos estudiantes que realizaron el diagrama de árbol y sin más respondieron ½ en la segunda y en la tercera prueba.

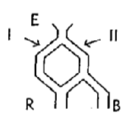
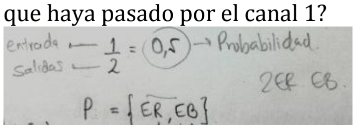
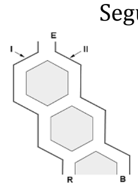
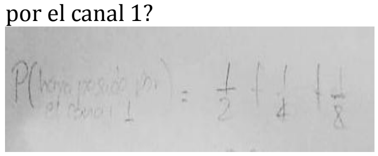
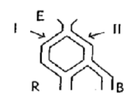
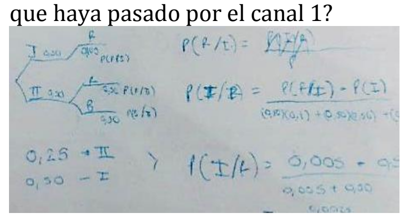
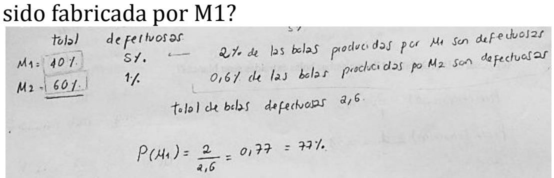
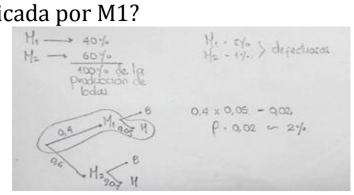
Primera Prueba	Segunda Prueba	Tercera Prueba
 Ítem 1. Una bola se suelta por la entrada E. Si sale por R, ¿Cuál es la probabilidad de que haya pasado por el canal 1?  a)	 Ítem 1. Una bola se suelta por la entrada E. Si sale por R, ¿Cuál es la probabilidad de que haya pasado por el canal 1?  b)	 Ítem 1. Una bola se suelta por la entrada E. Si sale por R, ¿Cuál es la probabilidad de que haya pasado por el canal 1?  c)

Figura 2. Respuestas de los estudiantes al ítem 1 en las tres pruebas

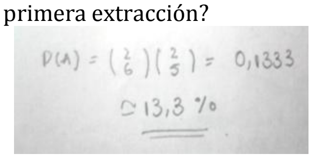
Ítem en un contexto de fabricación de artículos por dos máquinas. La relativa facilidad con que los estudiantes respondieron este ítem tal vez se pueda asociar a la total equivalencia que existe entre el enunciado del problema y el diagrama de árbol como lo explica Yáñez (2001). Un estudiante es especial respondió correctamente el ítem en la primera prueba sin hacer uso de la fórmula de Bayes, simplemente realizó regla de tres e interpretó la probabilidad como un cociente entre el 5% del 40% y la suma del 5% del 40% + 1% del 60% (Figura 2-a), este estudiante en la segunda prueba aplicó equivocadamente la fórmula de Bayes tomando los porcentajes incorrectos y en la tercera prueba volvió a responder incorrectamente, evidenciando esa confrontación entre una intuición aritmética esquematizada y una nueva modelación que no termina de aprehender. En la primera prueba algunos estudiantes asumieron una probabilidad conjunta ($P = 0.4 \times 0.05$), en la segunda prueba contestaron correctamente usando la fórmula de Bayes, pero en la tercera prueba vuelven a responder una probabilidad conjunta, lo que evidencia la dificultad que existe para distinguir entre probabilidades condicionales y probabilidades conjuntas (Pollatsek *et al.*, 1987). Además algunos estudiantes realizaron el diagrama de árbol correctamente pero no resolvieron el problema (Figura 2-b).

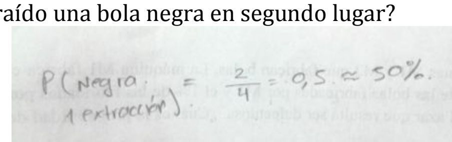
<table>
<tr>
<td>Ítem 2. Una fábrica dispone de dos máquinas M1 y M2 que fabrican bolas. La máquina M1 fabrica el 40 % de las bolas y la M2 el 60%. El 5% de las bolas fabricadas por M1 y el 1% de las fabricadas por M2 son defectuosas. Tomamos una bola al azar que resulta ser defectuosa. ¿Cuál es la probabilidad de que haya sido fabricada por M1?


a)</td>
<td>Ítem 2. Una fábrica dispone de dos máquinas M1 y M2 que fabrican bolas. La máquina M1 fabrica el 40 % de las bolas y la M2 el 60%. El 5% de las bolas fabricadas por M1 y el 1% de las fabricadas por M2 son defectuosas. Tomamos una bola al azar que resulta ser defectuosa. ¿Cuál es la probabilidad de que haya sido fabricada por M1?


c)</td>
</tr>
<tr>
<td colspan="2">Figura 2. Respuestas de los estudiantes al ítem 2 en la primera y segunda prueba</td>
</tr>
</table>

Ítem en un contexto de urna sin reposición. Respecto al inciso que evaluaba la condicional directa, en la segunda y tercera prueba algunos estudiantes la asumieron como una probabilidad conjunta (Figura 3-a). En el inciso de la condicional inversa se evidenció la concepción temporal de la probabilidad condicional (Gras y Totohasina, 1995) cuando muchos estudiantes respondieron en todas las pruebas con ½ (Figura 3-b). En el problema que indaga por la probabilidad de obtener una bola blanca en la segunda extracción, típica aplicación del teorema de probabilidad total, se observa que los estudiantes que no responden acertadamente utilizan la estrategia del 'depende' ya reportada por Yáñez (2003) y que refleja la concepción del enfoque del resultado aislado (Konold, 1991).

<table>
<tr>
<td>Ítem 3. Una urna contiene dos bolas blancas y dos bolas negras. Extraemos a ciegas dos bolas de la urna, una después, sin devolver la primera a la urna. ¿Cuál es la probabilidad de extraer una bola blanca en la segunda extracción si se sabe que se extrajo una bola negra en la primera extracción?


a)</td>
<td>Ítem 3. Una urna contiene dos bolas blancas, dos bolas negras y dos rojas. Extraemos a ciegas dos bolas de la urna, una después, sin devolver la primera a la urna. ¿Cuál es la probabilidad de haber extraído una bola roja en primer lugar, sabiendo que hemos extraído una bola negra en segundo lugar?


c)</td>
</tr>
<tr>
<td colspan="2">Figura 3. Respuestas de los estudiantes al ítem 3 en la primera y segunda prueba</td>
</tr>
</table>

CONCLUSIONES

Además de corroborar los hallazgos ya reportados por Morgado (2013), los resultados de esta investigación ponen de relieve la dificultad que existe cuando se pretende modificar intuiciones primarias por otras utilizando metodologías de enseñanza centradas fundamentalmente en los aspectos algebraicos de los modelos sin siquiera intentar enlazar esas intuiciones primarias, de carácter aritmético en el caso de la probabilidad condicional, con las expresiones probabilísticas de corte netamente algebraico. Para destacar también el enorme reto que representó para los estudiantes el problema de los canales que prácticamente no fueron capaces de resolverlo. Este problema donde la información no se explicita sino que está contenida en el gráfico que se presenta y que podría pensarse debería reflejar un claro razonamiento bayesiano en el sentido de que una información adicional modifica una probabilidad inicial, evidenció claras muestras de un proceso de enseñanza deficiente en estos temas.

La investigación también comprobó la dificultad que existe en la superación de los sesgos que ya son clásicos de encontrar en todas las investigaciones que se realizan con la probabilidad condicional: la confusión entre condicionales y conjuntas y la concepción temporal de la probabilidad condicional.

REFERENCIAS

Díaz, C. y de la Fuente, I. (2006). Dificultades en la resolución de problemas que involucran el teorema de Bayes. Un estudio exploratorio en estudiantes españoles de psicología. *Educación Matemática*, *18*, 75-94.

Falk, R. (1986). Conditional probabilities: insights and difficulties. En R. Davidson y J. Swift (Eds.), *Proceedings of the Second International Conference on Teaching Statistics* (pp. 292-297). Canadá: University of Victoria.

Gras, R. y Totohasina, A. (1995). Chronologie et causalité, conceptions sources d'obstacles épistémologiques à la notion de probabilité conditionnelle. *Recherches en Didactique des Mathématiques, 15* (1), 49-95.

Kahneman y Tversky (1972) Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3*, 430-454

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.

Montgomery, D.C. y Runger, G.C. (2010). *Probabilidad y Estadística aplicadas a la ingeniería.* Limusa Wiley

Morgado, C. (2013). Un estudio longitudinal del razonamiento bayesiano en estudiantes universitarios. Tesis de maestría. Bucaramanga, Colombia: Universidad Industrial de Santander.

Navidi, W. (2006). *Estadística para ingenieros y científicos.* México: McGraw-Hill.

Pollatsek, A., Well, A.D., Konold, C. y Hardiman, P. (1987). Understanding conditional probabilities. *Organization, Behavior and Human Decision Processes, 40*, 255-269.

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.

Yáñez, G. (2001). El álgebra, las tablas y los árboles en problemas de probabilidad condicional. En P. Gómez y L. Rico (Eds.), *Iniciación a la investigación en didáctica de la matemática. Homenaje al profesor Mauricio Castro* (pp. 355-375). Granada: Universidad de Granada.