EL RIESGO COMO CONTEXTO EN LA ENSEÑANZA DE LA PROBABILIDAD CONDICIONAL

Carmen Batanero y Mª Magdalena Gea
Universidad de Granada. (España)
batanero@ugr.es, mmgea@ugr.es

Resumen

Aunque nuestra experiencia con la probabilidad condicional es amplia -debido a la necesidad de utilizar la información disponible para asignar probabilidades- y estudiamos la probabilidad condicional en varios niveles educativos, la investigación didáctica señala numerosos errores sobre este concepto. En esta conferencia paralela analizamos alguna de las causas de estos errores y proponemos algunos recursos para mejorar la interpretación de la probabilidad condicional y la toma de decisiones. Igualmente sugerimos que la toma de decisión en situaciones de riesgo, que son frecuentes en la vida diaria puede ser un contexto motivador para el estudio de la probabilidad condicional.

Palabras clave: probabilidad condicional, sesgos de razonamiento, situaciones de riesgo, recursos didácticos.

Abstract

Although our experience with conditional probability is wide -due to our need to use available information to assign probabilities- and we study conditional probability at different educational levels, didactic research points to numerous errors related to this concept. In this parallel lecture we analyse some of the reasons of these errors and propose some resources to improve the interpretation of conditional probability and decision making. We also suggest that decision making under risk situations, which are frequent in everyday life is a context that helps motivate students towards conditional probability study.

Key words: conditional probability, reasoning biases, risk situations, didactical resources.

Introducción

El motivo de centrarnos en este trabajo en las situaciones de riesgo es considerarlas un contexto motivador, por la atención que dichas situaciones reciben actualmente por parte de la prensa, políticos, responsables de salud o seguridad o gestores de empresas de seguros y financieros (Borovcnik, 2015). Además, fueron los contextos de riesgo, asociados a las apuestas en los juegos de azar, los que originaron el cálculo de probabilidades. Finalmente, la competencia para manejarse en estas situaciones se considera actualmente una parte importante de la cultura estadística (Gigerenzer, 2002). Por este motivo los estudiantes requieren estrategias elementales para modelizar las situaciones de riesgo y para verbalizar y comunicar dichas situaciones.

Ejemplos de situaciones de riesgo comprensibles para los estudiantes pueden encontrarse en multitud de contextos, como, por ejemplo, accidentes (domésticos, laborales, viajes), desastres naturales, efectos secundarios de tratamientos médicos u operaciones, epidemias, situaciones psicológicas (depresión, estrés, mobbing) o legales (juicios, contratos). Estos riesgos pueden ser también positivos, por ejemplo, riesgo de overbooking, popularidad excesiva, consecución de un crédito o un proyecto, o promoción a un puesto de responsabilidad. Otros ejemplos se describen en Batanero, Ortiz, Serrano y Albanese (2017).

Marco teórico

Slovic (2000) define el riesgo como un suceso incierto que puede producir un efecto negativo o positivo sobre una persona o grupo. Borovcnik (2015) diferencia entre riesgos personales (que afectan sólo a una persona) o sociales, que en casos como epidemias o desastres naturales se dan simultáneamente. Según dicho autor, la gestión del riesgo implica cualquier acción dirigida a eliminar o controlar impacto no deseado a un nivel aceptable. Gigerenzer (2003) indica que un riesgo lleva asociada una probabilidad que puede ser frecuencial (información de una larga serie de observaciones del suceso), subjetiva (grado de creencia personal), o basada en la propensión (tendencia de un fenómeno). En estos trabajos, así como en Till (2014), se destaca el papel de la probabilidad condicional como base de la comprensión de las situaciones de riesgo, por lo que será importante que el estudiante adquiera competencia con dicho concepto para manejarse en situaciones de riesgo.

En lo que sigue, en primer lugar resumimos las principales dificultades de comprensión de la probabilidad condicional y seguidamente describimos algunos recursos didácticos que pueden mejorar la comprensión de este concepto. Finalizamos con algunas implicaciones para la enseñanza de la probabilidad.

Dificultades con la probabilidad condicional

Numerosos autores, como, por ejemplo, Díaz y de la Fuente (2005) o Díaz, Batanero y Contreras (2010), muestran que algunas decisiones incorrectas en situaciones de riesgo son debidas a sesgos de razonamiento asociados a la probabilidad condicional, que resumimos a continuación. Ello es debido a las diferentes relaciones entre dos sucesos que pueden darse cuando se evalúa una probabilidad condicional y que se confunden (Tomlinson y Quinn, 1997).

Algunos de los sesgos de razonamiento relacionados con la probabilidad condicional ocurren al suponer que una relación condicional $P(A|B)$ implica siempre una relación causal, lo que no es cierto, en general (Falk, 1986). Por ejemplo, encontramos probabilidades condicionales para evaluar relaciones diagnósticas: cuando B es un síntoma asociado a una posible enfermedad A, si la persona tiene el síntoma B, la probabilidad de que la persona sufra la enfermedad $P(A|B)$ es mayor que la probabilidad de que la sufra otra persona cualquiera $P(A)$; pero un síntoma no es una causa, sino una consecuencia de la enfermedad.

Por otro lado, algunos estudiantes tienen problemas para calcular una probabilidad condicional $P(A|B)$ si la condición B sucede después de que haya ocurrido A, aunque en la vida diaria encontramos muchas situaciones que están condicionadas por un suceso posterior a las mismas. Por ejemplo, si una prueba de paternidad indica una alta probabilidad de que una persona sea el padre biológico de otra, la prueba genética en sí no es una causa de la supuesta paternidad sino un indicio de ella y además es posterior a la paternidad en sí misma. Otros ejemplos se dan en los juicios, donde una evidencia se encuentra después de que ocurre el delito, pero se usa como condición para emitir un veredicto sobre la posible culpabilidad del acusado. En general, siempre que aplicamos el teorema de Bayes invertimos el orden de ocurrencia en el tiempo de los sucesos condición y condicionado, y ello puede explicar la dificultad de resolver problemas que implican este teorema.

Es difícil también diferenciar las probabilidades condicionales $P(A|B)$ y $P(B|A)$, sobre todo cuando se enuncian en lenguaje ordinario (probabilidad de aprobar un examen si se ha estudiado o probabilidad de haber estudiado si se aprobó el examen). Esta confusión se conoce como *falacia de la condicional transpuesta* (Falk, 1986) y se da con frecuencia al interpretar el nivel de significación α en un contraste de hipótesis, donde α = P(Rechazar una hipótesis, siendo cierta) se confunde con P(Hipótesis cierta, cuando se ha rechazado). La falacia de la condicional transpuesta puede llevar al médico a recetar un tratamiento innecesario (a veces doloroso o agresivo) si se confunde la probabilidad de tener una enfermedad, dado un síntoma, con la probabilidad de tener el síntoma, dada la enfermedad.

Dificultades con la idea de independencia

El concepto de independencia está relacionado con la probabilidad condicional. Mientras que su definición es muy sencilla (A y B son independientes y solo si se cumple: $P(A \cap B) = P(A)xP(B)$), en la práctica no sabemos cuándo la debemos aplicar. Un ejemplo claro es la frecuencia con que las personas se abonan a un mismo número de lotería y juegan a él todos los años, aunque los números premiados anteriormente, no se retiran del bombo; por tanto, el hecho de que un número nunca haya resultado ganador no aumenta su probabilidad.

La aplicación de la inferencia estadística o la ley de los grandes números exigen que las repeticiones de un experimento sean independientes entre sí; pero en la práctica es difícil asegurar esta independencia: ¿Cómo sabemos, por ejemplo, que al lanzar sucesivamente un dado no vamos mejorando nuestra habilidad para conseguir el número que deseamos? ¿Podemos considerar a todos los votantes de un partido político o a todos los conductores de una cierta ciudad como el resultado de una serie de experimentos aleatorios idénticos e independientes? A pesar de la duda, cuando analizamos estadísticamente los resultados de las votaciones o se estudia la póliza a pagar en una compañía aseguradora, aplicamos la distribución binomial que asume la independencia de ensayos. Esto se explica porque los modelos matemáticos (en este caso la independencia) son simplificaciones de la realidad. Un modelo matemático no es nunca verdadero ni falso; lo utilizamos por su utilidad para describir una situación (Chaput, Girard y Henry, 2011).

Recursos didácticos para el estudio de la probabilidad condicional y la independencia

Hemos visto que la probabilidad condicional, importante en situaciones de riesgo, es un concepto difícil. En lo que sigue describimos algunos recursos, que podemos usar en el aula, para ayudar a los estudiantes a mejorar su razonamiento en probabilidad condicional. Dichos recursos están siendo experimentados en la enseñanza (véase, por ejemplo, Martignon y Kraus, 2009).

Consideraremos el ejemplo siguiente: En una gran ciudad una de cada 1000 personas sufre narcolepsia. Si seleccionamos al azar una persona en esta ciudad, la probabilidad de sufrir narcolepsia será $P(N)$= 1/1000. Hacemos a la persona pasar una prueba que da positiva en 99 de cada 100 personas enfermas y también en 2 de cada 100 personas sanas, es decir $P(+|N) = 0{,}99$ y $P(+|S) = 0{,}02$. A la persona en concreto la prueba le dio positiva, ¿Cuál es la probabilidad de que la persona tenga narcolepsia $P(N|+)$?

Frecuencias absolutas y tablas de contingencia

Una forma de facilitar la resolución de problemas de probabilidad condicional es cambiar el formato en que se dan los datos del problema, pasando estos datos, cuando sea posible a frecuencias absolutas, en lugar de darlos en probabilidades o en porcentajes. Ello se debe, según sugieren Gigerenzer y Edwards (2003) a que nuestra mente está mejor equipada para razonar con frecuencias absolutas (que llama frecuencias naturales) que con probabilidades o porcentajes.

Un método de simplificar los cálculos para encontrar la probabilidad pedida es pensar qué ocurriría en una muestra de 100000 personas a las que pasamos la prueba. Supongamos que las personas no tienen síntomas (es decir, estamos en una situación de revisión masiva de la población o screening). En la Tabla 1 hemos realizado los cálculos; comenzamos calculando el número de enfermos (100) en la población, de los cuáles 99 obtendrán un resultado positivo y 1 un resultado negativo. Nos quedan 99900 personas sanas, de las cuales el 2% (1998) tendrán un resultado positivo; fácilmente completamos las celdas que faltan y hemos transformado todas las probabilidades en frecuencias.

El cálculo de la probabilidad $P(N|+)$ es ahora sencillo, pues podemos simplemente aplicar la regla de Laplace dividiendo el número de enfermos en la fila de personas con resultado positivo por el número total de resultados positivos: $\mathrm{P}(\mathrm{N}\,|+)=\frac{99}{2097}=0{,}0472$; un resultado que sorprende a muchos. La probabilidad de estar sano, aunque la prueba en un screening sea positiva, es muy alta; en el ejemplo, el complementario del anterior, es decir, 0,9527. La explicación es que, aunque la prueba sea muy fiable, son muchas las personas sanas que se someten a ella y, por tanto, hay muchos *falsos positivos*.

Tabla 1. *Número de personas sanas y enfermas en una muestra de 100000 personas según el resultado de una prueba (prevalencia= 0,001)*

	Test +	Test -	Total
Sufren narcolepsia	99	1	100
No sufren narcolepsia	1998	97902	99900
Total	2097	97903	100000

Diagramas en árbol y visualizaciones

La traducción de los datos de la Tabla 1 a un diagrama en árbol es aún más efectiva, pues nos ayuda a pensar en la secuencia temporal de los sucesos y separa claramente las dos subpoblaciones en estudio. En la Figura 1 trabajamos directamente con frecuencias absolutas y llegamos a la misma conclusión que con la tabla, pero de un modo más visual. En realidad, cualquiera de los dos métodos nos ha hecho transformar un problema típico de aplicación del teorema de Bayes en un problema de cálculo de probabilidades simples (regla de Laplace).

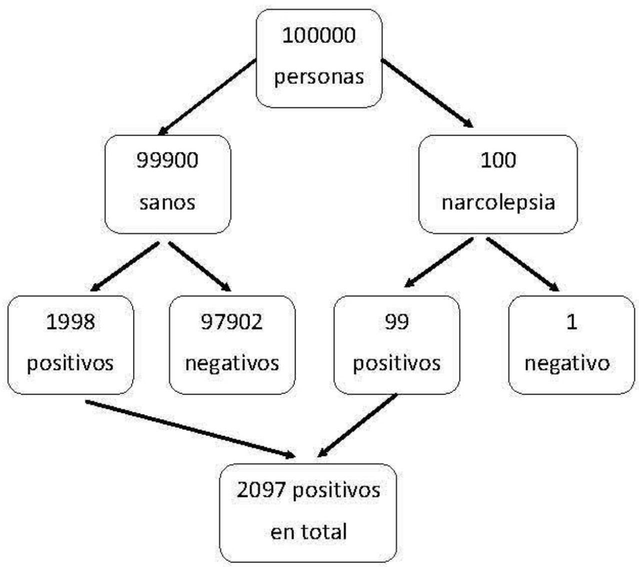

Figura 1. Visualización del problema mediante diagrama en árbol del ejemplo de narcolepsia

Por otro lado, la visualización de los datos en un diagrama de barras apilado (Figura 2) nos permite introducir la idea de riesgo relativo, comparando dos riesgos:

- Riesgo de obtener una prueba negativa cuando se sufre de narcolepsia $P(-|N) = 0{,}01$;
- Riesgo de obtener una prueba negativa cuando se está sano. $P(-|S) = 0{,}98$;

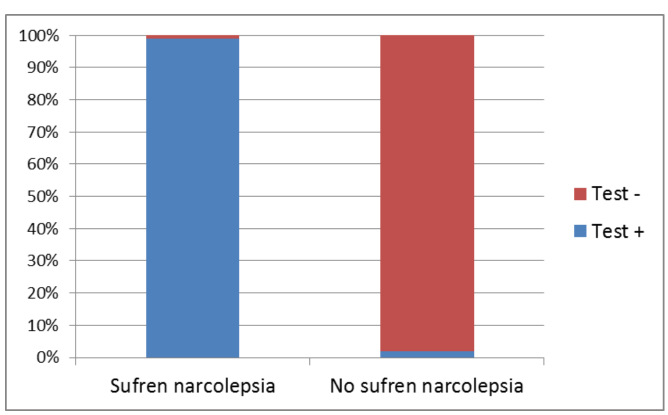

Figura 2. Visualización del problema mediante diagrama de barras apilado del ejemplo de narcolepsia

Estos riesgos pueden expresarse y compararse de diversas formas, que no tienen igual impacto en la persona que recibe la información; por ejemplo:

- $\frac{P(-|S)}{P(-|N)} = \frac{0,98}{0,01} = 98$; La razón entre el riesgo de un resultado negativo en las personas sanas respecto a las enfermas es 98 a 1;
- Por el contrario $\frac{P(+|S)}{P(+|N)} = \frac{0,02}{0,99} = 0,0202$: La razón entre el riesgo de un resultado positivo en las personas sanas respecto a las enfermas es 202 por cada 10000.
- Por cada 10000 resultados positivos en personas enfermas obtenemos 202 en personas sanas;
- Hay una reducción de 798 resultados positivos por cada 10000 en las personas sanas respecto a las enfermas.
- Se reduce el 7,98% de resultados positivos en las personas sanas, etc.

Estas actividades ayudan también a interpretar correctamente, en una situación de riesgo, las probabilidades pequeñas, que con frecuencia se desestiman, pero pueden ser importantes cuando un experimento (como es la prueba médica) se repite con frecuencia. Así, por ejemplo, supongamos que el riesgo de un accidente de tráfico tiene una probabilidad P(A)= 0,00002 (2 en 100000 viajes). Este riesgo es casi despreciable. Pero si consideramos una operación de tráfico de salida de vacaciones, cuando se producen, digamos un millón de desplazamientos, la probabilidad de que no se produzca ningún accidente sería:

$$P(\bar{A}) = (1 - P(A))^{1000000} = (0,99998)^{1000000} = 0,0000000020607$$

Por tanto, con casi toda seguridad (0,999999998) se producirá al menos un accidente, a pesar de la baja probabilidad en cada vez aislada que realizamos la actividad.

Simulaciones y programas de cálculo

Otra ayuda importante son los numerosos programas de cálculo y applets disponibles en Internet que nos permiten introducir una metodología exploratoria y activa en el aula, reduciendo los cálculos tediosos. Tienen un papel importante en los casos en que las soluciones de probabilidad chocan con nuestra intuición. Por ejemplo, se puede utilizar uno de los numerosos programas de cálculo del teorema de Bayes, como el disponible en http://www.ugr.es/~jsalinas/bayes.htm para analizar el problema de la narcolepsia, pudiendo incluso variar las probabilidades implicadas.

Es importante también recordar que en toda simulación se lleva a cabo una actividad de modelización (Chaput, Girard y Henry, 2011). Por ello el profesor debe poner atención al aprendizaje de los diferentes pasos en la modelización: Observación de la realidad; construcción del modelo matemático, trabajo con el modelo e interpretación de los resultados del modelo.

Conclusiones

Las reflexiones anteriores nos llevan a concluir la importancia de reforzar la enseñanza de la probabilidad condicional y conectarla con contextos que el estudiante experimenta en su vida cotidiana. Este puede comprender la importancia de saber interpretar correctamente un diagnóstico o tomar una decisión sobre un posible tratamiento; comprende que cuando se acude a una consulta médica, el diagnóstico depende de los diferentes síntomas, estilo de vida del paciente, edad y otros factores. Estudios sobre el posible efecto de riesgos como el fumar, tomar drogas o estar en sobrepeso sobre la salud se pueden utilizar para motivar el estudio de la probabilidad, a la vez que contribuir a la educación para la salud.

Otros contextos cotidianos engloban desde el pronóstico del tiempo, resultados de partidos o votaciones, calificaciones en exámenes, veredictos en juicios, o cambio climático. No tenemos excusa para seguir repitiendo únicamente problemas sobre juegos de azar o, lo que es peor, problemas sin ningún contexto, en que la finalidad es puramente adquirir una erudición matemática.

Como indica Borovcnik (2015), las situaciones de riesgo pueden enriquecer la enseñanza de la probabilidad, pero aún necesitamos encontrar caminos para introducirla en las aulas y ayudar a los estudiantes a mejorar su percepción del riesgo y la forma de afrontarlo.

Algunos investigadores, como Sánchez y Orta (2015), han realizado experiencias docentes utilizando estos contextos y evaluando el nivel de razonamiento de los estudiantes y su evolución en las experiencias. Sus resultados son alentadores y nos indican la posibilidad de utilizar estas situaciones para mejorar el razonamiento probabilístico de los estudiantes. Esperemos que otros profesores e investigadores contribuyan en el diseño de este tipo de tareas y su incorporación en el aula.

Agradecimientos: Proyecto EDU2013-41141-P y EDU2016-74848-P (AEI, FEDER) y Grupo FQM-126 (Junta de Andalucía).

Referencias bibliográficas

Batanero, C., Ortiz, J., Serrano, L. y Albanerse, V. (2017). Razonamiento sobre probabilidad condicional en situaciones de riesgo. *Suma, 83,* 73-80.

Borovcnik, M. (2015). Risk and decision making: The "logic" of probability. *The Mathematics Enthusiast, 12*(1), 113-139.

Chaput, B., Girard, J. C. y Henry, M. (2011). Frequentist approach: modelling and simulation in statistics and probability teaching. En C. Batanero, G. Burrill y C. Reading (Eds.), *Teaching Statistics in school mathematics-Challenges for teaching and teacher education* (pp. 85-95). New York: Springer.

Díaz, C. y de la Fuente, I. (2005). Razonamiento sobre probabilidad condicional e implicaciones para la enseñanza de la estadística. *Epsilon, 59*, 245-260.

Díaz, C., Batanero, C. y Contreras, J, M. (2010). Teaching independence and conditional probability. *Boletín de Estadística e Investigación Operativa, 26*(2), 149-162.

Falk, R. (1986) Misconceptions of statistical significance, *Journal of Structural Learning, 9(*1), 83-96.

Gigerenzer, G. (2002). *Calculated risks: how to know when numbers decieve you*. New York: Simon & Schuster.

Gigerenzer, G. (2003). *Reckoning with risk: learning to live with uncertainty*. Penguin UK.

Gigerenzer, G. y Edwards, A. (2003). Simple tools for understanding risks: from innumeracy to insight. *British Medical Journal*, *327*(741-744).

Martignon, L. y Krauss, S. (2009). Hands on activities with fourth-graders: a tool box of heuristics for decision making and reckoning with risk. *International Electronic Journal for Mathematics Education, 4*(3), 117–148.

Sanchez, E. y Orta, A. (2015). Levels of reasoning of middle school students about data dispersion in risk context. *The Mathematics Enthusiast, 12*(1), 275-289.

Slovic, P. E. (2000). *The perception of risk.* Londres: Earthscan Publications.

Till, C. (2014). Fostering risk literacy in elementary school. *Mathematics Education, 9*(2), 83-96.

Tomlinson, S. y Quinn, R. (1997). Understanding conditional probability. *Teaching Statistics, 19*(1), 2-7.